Robust Question Answering

Gracinda Carvalho

Tese de Doutoramento em Informática

UNIVERSIDADE ABERTA

July 2011

Robust Question Answering

Gracinda Carvalho

Tese de Doutoramento em Informática

Orientadores: Prof. Doutor Vitor Jorge Ramos Rocio

Prof. Doutor David Manuel Martins de Matos

UNIVERSIDADE ABERTA

July 2011

Agradecimentos

A primeira palavra de agradecimento dirige-se aos orientadores por me terem dado o privilégio de contar com o seu apoio e colaboração, além dos seus conhecimentos e competência. Um muito obrigada também pela paciência com que sempre me acompanharam ao longo deste percurso. A ambos tenho ainda a agradecer o terem-me dado a conhecer e aberto as portas desta área de conhecimento e dos respectivos centros de investigação.

Um agradecimento é devido à minha instituição de origem, a Universidade Aberta, por me ter proporcionado os meios necessários à realização deste trabalho. Saliento em especial o "capital humano" da instituição que me permitiu sempre sentir acompanhada na realização deste projecto: agradeço assim as múltiplas palavras de incentivo dirigidas por colegas dos três corpos da Universidade: docente, não docente e discente. Saliento os meus colegas de secção, que demonstram inequivocamente a sua solidariedade ao assegurar a carga lectiva acrescida durante a minha dispensa, mas que além disso dispuseram do seu tempo e conhecimentos na sessão de avaliação intermédia decorrida em Outubro de 2009, e me privilegiaram com os seus comentários: tenho assim que agradecer ao colega Luís Cavique, organizador da iniciativa, e aos colegas José Coelho, Paulo Shirley, Jaime Remédios e Carlos Castilho.

Desde o meu ingresso na Universidade Aberta, o percurso até este plano de trabalhos foi longo, fruto da distância em anos da última prova académica prestada, o mestrado, em 1997. Houve neste trajecto alguns pontos que tiveram especial relevância: a frequência da escola de verão internacional em lógica, línguas e informação ESLII 2005, a participação no fórum de avaliação internacional QA@CLEF 2008 e a possibilidade da escrita da tese ser feita em Inglês por forma a poder contar com membros do júri estrangeiros. Nesse processo foi determinante a colaboração dos Directores do DCET à data, nomeadamente, o Prof. Alexandre Cerveira, a Prof.ª Filomena Amador, e o Prof. Fernando Costa a quem muito agradeço o terem conduzido, da melhor forma, estes processos. Agradeço ainda pelos mesmos motivos, e já na recta final, ao actual Director, Prof. Adérito Marcos. Também os coordenadores de secção tiveram um papel determinante de incentivo, mais ou menos activo consoante as características de cada um

e as exigências das diversas fases do projecto. Assim dirijo os meus agradecimentos ao Filipe Figuiredo, Paulo Shirley, José Coelho, Luís Cavique e Henrique São Mamede.

Uma palavra muito especial de agradecimento à Rosário Ramos pela sua disponibilidade e colaboração com comentários minuciosos e certeiros relativos ao tratamento estatístico dos dados. Agradeço ainda a colaboração do Jorge Santos nesta matéria. A boa colaboração com os colegas dos serviços da informática foi também fundamental ao longo do tempo, salientando sem desprimor dos restantes, João Lima e Paulo Castelo Branco.

Um agradecimento especial ao Prof. Vargas e à Prof.ª Isabel Trancoso, que em Junho de 2006 colaboraram na fase inicial de definição deste projecto e permitiram que a tese fosse esta e não outra, e ainda bem... Aos meus "avôs científicos": Prof. Nuno Mamede pelo apoio e por no Mestrado me ter feito ver que talvez não fosse má ideia continuar para um Doutoramento, ou seja ter colocado a ideia na minha mente; Doutor Gabriel Pereira Lopes pelo seu apoio, e pelas suas "pistas".

Aos organizadores do CLEF e Linguateca agradeço muito especialmente o empenho e profissionalismo, aos primeiros na organização das tarefas de avaliação e aos segundos em tornar o Português uma possibilidade a nível do CLEF, através da obtenção e disponibilização de recursos.

Os meus agradecimentos à FCT, L2F /INESC-ID Lisboa e CITI/FCT-UNL pelos recursos colocados ao meu dispor para participação nos eventos acima referidos. Agradeço o ambiente de aprendizagem proporcionado pelo L2F/INESC-ID, salientando a útil colaboração prestada por: Hugo Meinedo, Fernando Batista, Luísa Coheur, Ana Mendes e Ricardo Ribeiro.

Por último,"last but not least", à minha família e aos meus amigos por terem sido todos 100% colaborativos e pelo constante apoio e incentivo. Sem pretender transformar a tese num documento demasiado enfadonho desde tão baixo número de página, aqui vão: Mãe, Nhã, Faninha, Luísa, Amílcar, Zé Pedro, Maria João, Ju, Bernardo, Marcos, Bábá, João, Ana, Calucha, Cristina, Maria Augusta, Quim, Isabel, Mónica, Cristina, Luís Santos, Luís Miguel, Silvério, Paulo Rogério, Luís Bernado, Filipa...

Lisboa, 20 de Julho de 2011

Gracinda Carvalho

Dedicada a Ana de Sousa

(Tia Aninhas)

Abstract

A Question Answering (QA) system should provide a short and precise answer to a question in natural language, by searching a large knowledge base consisting of natural language text. The sources of the knowledge base are widely available, for written natural language text is a preferential form of human communication. The information ranges from the more traditional edited texts, for example encyclopaedias or newspaper articles, to text obtained by modern automatic processes, as automatic speech recognizers.

The work described in the present document focuses on the Portuguese language and open domain question answering, meaning that neither the questions nor the texts are restricted to a specific area, and it aims to address both types of written text. Since information retrieval is essential for a QA system, a careful analysis of the current state-of-the-art in information retrieval and question answering components is conducted.

A complete, efficient and robust question answering system is developed in this thesis, consisting of new modules for information retrieval and question answering, that is competitive with current QA systems. The system was evaluated at the Portuguese monolingual task of QA@CLEF 2008 and achieved the 3rd place in 6 Portuguese participants and 5th place among the 21 participants of 11 languages.

The system was also tested in Question Answering over Speech Transcripts (QAST), but outside the official evaluation QAST of QA@CLEF, since Portuguese was not among the available languages for this task. For that reason, an entire test environment consisting of a corpus of transcribed broadcast news and a matching question set was built in the scope of this work, so that experiments could be made. The system proved to be robust in the presence of automatically transcribed data, with results in line with the best reported at QAST.

Resumo

Um sistema automático de pergunta resposta tem como objectivo dar uma resposta curta e precisa a uma pergunta formulada em língua natural, pesquisando uma base de conhecimento constituída por texto em língua natural. As fontes deste tipo de conhecimento são numerosas, dado que o texto escrito constitui uma forma preferencial de comunicação humana. A informação varia desde o tradicional texto editado, como é o caso das enciclopédias e dos artigos de jornal, até texto obtido através de modernos processos automáticos, como os reconhecedores automáticos de fala.

O trabalho descrito no presente documento centra-se na língua Portuguesa e em sistemas de pergunta resposta de domínio aberto, o que significa que nem a pergunta nem a colecção de textos se restringem a uma área específica. Ambas as formas de texto escrito referidas no parágrafo anterior são consideradas.

Dado que a recuperação de informação é essencial num sistema de pergunta resposta, as técnicas mais actuais utilizadas nestas duas áreas neste tipo de sistema são objecto de um estudo aprofundado, tanto no que diz respeito aos seus aspectos mais práticos, como às suas motivações teóricas. Uma vez que um sistema nunca pode ser simples demais, desde que cumpra as especificações e produza resultados de elevada qualidade, é feita uma análise de custo benefício das técnicas passíveis de serem utilizadas, dando preferência a soluções simples.

O principal objectivo do presente trabalho é assim estudar e desenvolver componentes inovadores para recuperação de informação e pergunta resposta, e a construção de um sistema de pergunta resposta completo, eficiente e robusto, capaz de competir com os sistemas mais avançados existentes actualmente.

Uma opção importante tomada foi a utilização da língua Portuguesa, uma língua falada por um vasto número de pessoas, o que constitui um requisito importante para um sistema de pergunta resposta, quer pela existência de um volume importante de texto escrito disponível nesta língua, quer pelo número de possíveis utilizadores de uma aplicação específica para o

Português. Há no entanto que ter em conta a existência de menor número de recursos linguísticos para a língua Portuguesa, especialmente se comparada com a língua Inglesa, que é correntemente a "língua franca" da investigação científica. É precisamente este o motivo do presente documento estar escrito na língua Inglesa: permitir a participação nos trabalhos e a validação de resultados internacionalmente, sendo este facto totalmente compatível com a focalização do estudo e dos trabalhos na língua Portuguesa, alargando inclusivamente a sua divulgação para públicos não falantes da mesma.

Na abordagem para a realização deste trabalho esta opção foi tomada em conjunto com uma outra que foi explorar as potencialidades da Wikipedia como recurso de base de QA, e que se revelou de extrema utilidade em várias vertentes do trabalho desenvolvido. As características da Wikipedia que se consideraram mais relevantes foram o facto da informação estar disponível gratuitamente, e de resultar do esforço conjunto de um elevado número de utilizadores, o que viabiliza o desenvolvimento de aplicações para as quais seja útil conhecimento enciclopédico e conhecimento de natureza ontológica. Ambas as vertentes foram utilizadas de forma inovadora no presente sistema.

Apresenta-se neste trabalho o sistema de pergunta resposta, que foi desenvolvido de raiz, e que provou estar ao nível dos melhores sistemas de pergunta resposta, dado que foi submetido a avaliação em 2008 no Fórum de Avaliação Internacional CLEF (Cross Language Evaluation Fórum) e se classificou em terceiro lugar entre os seis participantes concorrentes na categoria de sistemas de pergunta resposta em Português, onde era o único sistema a participar pela primeira vez. A taxa de primeiras respostas correctas foi de 32,5%. Este resultado permitiu obter o 5º lugar entre os 21 sistemas participantes nas 11 línguas disponíveis, sendo de referir o elevado nível dos sistemas concorrentes para o Português, dado que nos três primeiros lugares se classificaram dois sistemas para o Português, com o sistema da companhia Portuguesa Priberam ocupando a primeira posição com uma taxa de primeiras respostas correctas de 63,5% e o sistema da Universidade de Évora classificado em terceiro lugar, com uma taxa de primeiras respostas correctas de 46,5%.

Os melhoramentos introduzidos após a análise dos resultados obtidos, que foi feita considerando quer as respostas do próprio sistema, quer as respostas produzidas pelos restantes sistemas, resultaram num considerável aumento da taxa de primeiras respostas correctas, para 50,5%, o que se seria correspondente a um segundo lugar nos resultados para o Português.

O sistema desenvolvido é eficiente na indexação e resposta a perguntas, levando, na sua versão melhorada, apenas 4 horas para indexar toda a colecção de textos utilizada na tarefa do QA@CLEF 2008, e cerca de dois minutos a responder às 200 perguntas da tarefa, o que corresponde a uma média de 0,6 segundos para responder a uma pergunta. De referir que nenhum participante divulgou dados sobre a eficiência do sistema. Apenas se encontraram publicados dados de eficiência para um sistema que não participou na avaliação, que reporta valores médios de resposta por pergunta de 22 segundos.

O sistema foi ainda testado num caso de estudo envolvendo perguntas efectuadas sobre o conteúdo de peças faladas. A base de textos que é pesquisada neste caso, consiste nos textos obtidos de forma automática a partir do reconhecimento automático da fala. Dado que a tarefa do Fórum de Avaliação CLEF para sistemas automáticos a responder a perguntas sobre transcrições automáticas (QAST - Question Answering over Speech Transcripts) não incluía a língua Portuguesa, os dados tiveram que ser todos recolhidos e organizados tendo sido criado um recurso que permite fazer testes de sistemas para o Português. Este recurso tem como base um corpo constituído pelos Telejornais da Rádio Televisão Portuguesa, RTP, nas suas edições da noite dos canais 1 e 2, correspondente aos meses de Junho a Setembro de 2008. Este corpo consiste em cerca de 180 horas de duração, transcritas automaticamente e enriquecidas com colocação automática de pontuação. Foi feito um conjunto de 100 perguntas, baseadas em transcrições manuais, e que foi utilizado para testar o sistema. O sistema demonstrou ser robusto, pois mesmo na presença de texto com palavras incorrectamente reconhecidas, ou pontuação colocada fora dos locais correctos, o sistema obteve 30% de taxa de primeiras respostas correctas, 42% de taxa de respostas correctas nas três primeiras respostas, e uma taxa de 60% de localização do excerto onde se encontra a resposta correcta. Este último valor tem uma aplicação interessante de localização de um determinado tema num conjunto de diversas horas de vídeo, através de uma pergunta formulada em língua natural. Os resultados obtidos estão ao nível dos melhores reportados nas avaliações QAST do QA@CLEF.

Dado que o principal objectivo traçado para o presente projecto de doutoramento, era estudar e desenvolver componentes inovadores de recuperação de informação e pergunta resposta que conduzissem à construção de um sistema de pergunta resposta para o Português, completo eficiente e robusto, e com resultados ao nível dos melhores sistemas, considera-se que o objectivo foi plenamente atingido.

Relativamente ao uso do Português como língua de trabalho, confirma-se o facto de que os resultados obtidos para esta língua na área de sistemas de pergunta resposta estão ao melhor nível dos sistemas actuais para outras línguas, provando-se ser possível ultrapassar o problema de escassez de recursos. Os resultados validam também o conceito da existência de corpus onde coexistem textos com origem em distintas variantes de Português, nomeadamente Europeia e Brasileira, mas não só. No que diz respeito a língua falada, os resultados obtidos no caso de estudo indicam uma necessidade de tratamento específico para estas duas diferentes variantes do Português.

Palavras Chave
Keywords

Palavras Chave

Sistemas de Pergunta Resposta

Recuperação de Informação

Extracção de Informação

Keywords

Question Answering

Information Retrieval

Information Extraction

Contents

List of Figures

List of Tables

List of Algorithms

Notation

K	Total number of terms in the collection
K_j	Document length of term d_j in number of words
Kb_j	Document length of term d_j in bytes
M	Number of distinct terms in the collection
N	Number of documents in the collection
Q	Number of questions in the question set
R	Set of Relevant documents in the collection for a query q, According to our conventions the notation $\mathbf{R}$ should be used since we are talking about a set, however to simplify the expressions we use the representation R instead. Another simplification we use for readability sake is to omit the query: R_q would be more correct.
Z	question size in number of words
$\bar{R}$	Set of non Relevant documents in the collection for a query q, According to our conventions the notation $\bar{\mathbf{R}}$ should be used since we are talking about a set, however to simplify the expressions we use the representation $\bar{R}$ instead. Another simplification we use for readability sake is to omit the query: $\bar{R}_q$ would be more correct.
$\bullet$	Inner product between vectors
$\mathbf{D}$	Set $\{d_1,...,d_j,...,d_N\}$ of documents in the collection, or Document Collection. Each document is assigned a unique *document number* in the collection, from 1 to N
$\mathbf{Q}$	Question set
$\mathbf{T}$	Set $\{t_1,...,t_i,...,t_M\}$ of distinct terms in the collection. Each term is assigned a unique *term number* in the collection, from 1 to M
$avdlb$	**Av**erage **d**ocument **l**ength of the collection in **b**ytes
cf_i	Collection term frequency of term t_i, i.e. the number of occurrences of term t_i in the entire collection
cs	Collection size in words, i.e.total number of words in the entire collection
d_j	A generic document in the collection The associated index subscript is usually j

df_i Document frequency of term t_i in the document collection,
i.e. the number of documents in the collection that have one or more occurrences of term t_i

i The index subscript usually associated to a generic term t in the collection

j The index subscript usually associated to a generic document d in the collection

k The index subscript usually associated to a generic term v_{jk} in document d_j in the collection

q A generic query in an Information Retrieval system

qtf_i Frequency of term t_i in query q

$rank(d_j, q)$ Ranking or scoring of document d_j with respect to query q. Different IR ranking models calculate this value according to different formulas.

t_i A generic term in the collection
The associated index subscript is usually i

tf_{ij} Term frequency of term i in document j,
i.e. the number of occurrences of term i in document j

v_{jk} The term in position k in document d_j

w_{ij} Weight of term i in document j
This concept differs slightly depending on the retrieval model, as explained in Section 4.2

I Introductory Concepts and Approach

Introduction to Part I

Chapter 1 describes the problem and motivation, and summarizes the thesis objectives and approach. The chapter continues with the identification of the main contributions of the thesis and the related publications. Next, key definitions, terminology and style conventions are given, and the chapter ends with the thesis outline.

In Chapter 2 a survey on QA is presented. It includes an overview of historical QA systems, followed by an explanation on how they are evaluated in international initiatives, including the task descriptions, as well as evaluation metrics, question types, and text collection characteristics. Next, the approach and options taken by Portuguese question answering systems are described. The chapter ends with the research directions to follow and a short description of the components of the system developed in the scope of this thesis presented, with the main differences to other QA systems identified.

1 Introduction

1.1 Description of the Problem and Motivation

The purpose of a Question Answering (QA) system is to provide an answer, in a short and precise way, to a question in Natural Language. Answers are produced by searching a knowledge base that usually consists of Natural Language text. The usefulness of this type of system is to find the exact information in large volumes of text data. With the wider availability of this type of resources, whether it is in the form of newspaper collections, or texts obtained through ASR (Automatic Speech Recognition), or encyclopaedic resources, or the blogosphere, or even the World Wide Web, there is an increasing interest in this type of system.

1.2 Thesis Objectives and Approach

The techniques employed by current state-of-the-art Information Retrieval (IR) and Question Answering (QA) systems are investigated and subject to a careful inspection of practical aspects, as well as of their theoretical motivations. Since we believe that a system can never be too simple, as long as it complies with the specifications and produces good results, we make a careful analysis of the cost/benefit of the techniques liable to be employed, valuing simpler solutions.

The main goal of this work is to study and develop innovative components of IR and QA, to build a complete, efficient and robust QA system, that can compete with the current state-of-the-art QA systems. The name of the system developed is IdSay, a short name for ”I would say” or ”I dare say”.

The key concepts taken into account in the development of the present work are:

- efficient implementation;
- robust implementation;

- validation of the results through evaluation with peer systems;
- explore Wikipedia as a resource for QA;
- use developed QA system in speech transcripts.

If the system is not efficient, it will not be useful for a real question answering application, since users expect fast answers. We are only interested in fast components of IR and QA, that are less likely to give enough time for the user to consider the decision of waiting for the answer or giving up.

A robust system, that performs well not only under the most favourable conditions but also under unusual circumstances, has greater chances of being valuable in more situations, so we will direct the development towards the robustness of the components. In our case, the most favourable circumstances refer to well formed text, as opposed to text that contains incorrections such as that obtained through automatic methods.

It is extremely important to validate the system in an international forum, not only to be able to compare it with peer systems, and to share the evaluation effort with others, but mainly to have the results certified by an international organization, that prevents misleading analysis and conclusions on the performance of the system. Therefore, whenever possible, we use this option.

An important option taken is to use the Portuguese Language, a widely spoken language, which makes it both eligible in terms of the extensive text data present for instance in the web and also the usefulness for a large number of potential users. Despite this fact, language resources are still less abundant, especially if we make a comparison to English, the current "standard language" for research, but also to some other languages. However facts have proven that it is possible to achieve state-of-the-art results using Portuguese, compensating the possible lack of specific resources.

Wikipedia combines two characteristics that we consider interesting, one is the fact that it is freely available for public use, and the other is the fact that it results from a collaborative effort from millions of people around the globe, bringing it the benefits of diversity and volume. Both these features contribute for the creation of quality working material. We intend to make use of Wikipedia for improving the system efficiency.

Finally we want to test the robustness of the system using it on data obtained from Automatic Speech Recognition (ASR) applications. Because this data is less well formed than written text, due to the word error rates (WER) of the recognisers, a good performance in this scenario validates the robustness of the system. It is an important application of QA since more data from ASR is becoming available, along with the corresponding need for search mechanisms to cope with it.

1.3 Contributions

The major contribution of the present work is the QA system developed to satisfy the aims of the thesis, but we highlight the most relevant contributions in terms of generality and usability by other researchers:

- A statistically validated study was conducted regarding pre-processing options for an IR system working in the QA context for Portuguese. We found out that converting text to lowercase and removing punctuation marks increase retrieval efficiency but there is no statistical evidence of improvements derived from the use of stop lists, lemmatization or stemming, for the experiments conducted. These results were surprising, since most of these techniques are assumed to be relevant, but the tests made point out the necessity to validate its application for a given IR system and application (Chapter 3).

- An efficient search mechanism for large document collections to be used for QA. The data structure for storing documents uses one number per word, instead of strings. This data structure, in the text collection used, requires an average of 4,28 bytes/word while the string version would require 10,43 bytes/word. With this data structure, instead of string manipulation one integer comparison is done to compare words. This leads to improvements in both space and time (Chapter 5). The component is based on the Boolean Retrieval Model as the result of the study conducted (Chapter 4), and the search is done considering separately words and entities(groups of contiguous words) identified from the question (Chapter 6).

- A strategy to remove the most frequent keywords from the query, only if no satisfactory results have been produced. This can be seen as a dynamic application of stop lists, but

instead of blindly removing stop words at indexing time, the words are selectively removed, if needed, based on the question being processed (Chapter 6).

- A method of Results Quadrants is introduced, that summarizes the information related to a system when compared to other systems performing the same task, in this case answering questions. The Results Quadrant for a system allows the identification of such characteristics as its degree of innovation or coverage of easy questions, in perspective with peer systems (Chapter 7).

- A Question Answering over Speech Transcripts corpus for Portuguese: This corpus consists of video recordings of the evening editions of the Broadcasting News from the two channels of the Portuguese public television network, Rádio Televisão Portuguesa, RTP, from around 3 months in the summer of 2008 along with a set of 100 questions. The data contains over 180 hours of audio with the corresponding automatic transcripts, enriched with punctuation marks and topic detection, obtained by the SSNT system developed at INESC-ID. Approximately 60 excerpts were transcribed manually, and a set of 100 questions was made based on these transcripts (Chapter 9).

1.4 Publications

The work developed originated, so far, the following publications:

- Gracinda Carvalho, David Martins de Matos & Vitor Rocio (2007). Document Retrieval for Question Answering: A Quantitative Evaluation of Text Preprocessing. In *Proceedings of ACM first Ph.D. Workshop, PIKM 2007, in the 16th ACM Conference on Information and Knowledge Management, CIKM 2007, Lisbon, Portugal, November 5-10, 2007*, pp. 125-130, ACM. ISBN: 978-1-59593-832-9 DOI: http://dx.doi.org/10.1145/1316874.1316894;

- Gracinda Carvalho, David Martins de Matos & Vitor Rocio (2008). IdSay: A Question Answering system for Portuguese powered by Wikipedia. In *Propor 2008 Special Demonstration Session Promoted by Microsoft Language Department Center: Applications of Portuguese Speech and Language Technologies, September 10, 2008, Curia, Portugal.* http://download.microsoft.com/download/E/6/3/E63EEEBE-BEF8-4607-B381-BA2C5F6F9AE6/IdSayAQuestionAnsweringsystem.pdf;

- Gracinda Carvalho, David Martins de Matos & Vitor Rocio (2008). IdSay: Question Answering for Portuguese. In *Working Notes of the 9th Workshop of the Cross-Language Evaluation Forum CLEF 2008 Workshop of the European Conference on Research and Advanced Technology for Digital Libraries - ECDL 2008, Aarhus, Denmark, September 17-19, 2008.* http://www.clef-campaign.org/2008/working_notes/carvalho-paperCLEF2008.pdf;

- Gracinda Carvalho, David Martins de Matos & Vitor Rocio (2009). IdSay: Question Answering for Portuguese. In *Evaluating Systems for Multilingual and Multimodal Information Access: 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers, LNCS Series Volume 5706*, pp. 345-352, Springer-Verlag, Berlin, Heidelberg. ISBN: 978-3-642-04446-5. DOI: http://dx.doi.org/10.1007/978-3-642-04447-2_40;

- Gracinda Carvalho, David Martins de Matos & Vitor Rocio (2010). Improving IdSay: a characterization of strengths and weaknesses in Question Answering systems for Portuguese. In *Proceedings of the 9th International Conference on Computational Processing of the Portuguese Language, PROPOR 2010, Porto Alegre, RS, Brazil, April 27-30, 2010, LNCS Series Volume 6001*, pp. 1-10, Springer-Verlag, Berlin, Heidelberg. ISBN 978-3-642-12319-1. DOI: http://dx.doi.org/10.1007/978-3-642-12320-7_1.

1.5 Definitions, Terminology and Style Conventions

Words and **terms** are used interchangeably throughout the thesis. **Tokens** is used in the thesis in the early stage of pre-processing before Information Retrieval is done, to define the smallest unit that is considered in the system (and form the basic data structure of the system). In the case of our system, for Portuguese, our tokens correspond to words. Therefore, it has the same meaning of words and terms, but we restrict token usage to pre-processing.

We use the expression "number of words" (for instance in a collection or document) to express the number of all occurring words, even though some of them may appear more than once. If we only count each distinct word once, we call it "number of distinct words", M. Sometimes in the literature it is common to find token as associated to the total number of words, so the number of tokens would be our number of words, but we do not use this convention.

The concept of Multi Word Expressions (MWE) is used in the literature to include Named Entities (NE) and proper nouns, as well as numbers, dates or collocations. This corresponds roughly to our notion of entity, which is the name used throughout this document to identify a group of contiguous words that have a specific meaning together. The correct order of the words is a requirement for an entity.

The work described in this thesis is done for the Portuguese language, so the document collection, questions and corresponding answers are all in Portuguese, unless otherwise mentioned. Since the thesis is written in English, whenever we use the Portuguese language to exemplify some point, we use a sans-serif font immediately followed by the English translation in brackets. To give an example, Este é um texto escrito em Português [This is a text written in Portuguese].

1.6 Thesis Outline

The thesis is divided in four parts:

- Part I - **Introductory Concepts and Approach**: besides the present chapter this part includes, in Chapter 2, a survey on question answering systems for Portuguese and the research directions followed.
- Part II - **Information Retrieval and Question Answering** starts with Chapter 3 in which an analysis of the different pre-processing techniques is done, followed by Chapter 4 assigned to study the information retrieval models, for using in question answering, and the part ends with Chapter 5 that contains the description of the proposed IR system, IdSearch, the information retrieval for IdSay.
- Part III - **IdSay** presents at first Chapter 6 with all IdSay components used in the search for an answer, followed by Chapter 7 that presents the evaluation results of IdSay at QA@CLEF 2008 and its analysis and comparison with other systems results. The part ends with Chapter 8 that describes the improvements added to IdSay after the evaluation, reporting the corresponding increase in results.
- Part IV - **Case Study and Conclusions** is composed by two chapters, Chapter 9 with a case study of IdSay in QA over speech transcripts, validating the robustness of IdSay, and the second, Chapter 10, with the contributions and future work.

- Part V corresponds to the **Appendices**. Although the thesis text is self-contained, the appendices have additional information on the experiments made, including the full tests results.

The printed version of the thesis consists of the the first four parts and Bibliography, and includes in the back cover a CD with the full version of the thesis containing the appendixes, in pdf format.

2 Question Answering: Survey and Research Directions

2.1 Introduction

Question Answering (QA) systems have the aim of providing a short and precise answer to an information request stated in Natural Language (NL). The knowledge base usually consists of a large amount of NL text which is searched using Information Retrieval (IR) techniques. The usefulness of these systems is to find the exact information in large volumes of text data.

However QA has been an area of interest for researchers since the mid 50s. In the present chapter we will make a brief survey of question answering, following an historical time line from the early days to the present days, focusing also on the relationship between QA and IR.

We proceed to look in more detail at the tracks dedicated to QA of the above mentioned international evaluation initiatives and the very active research in this field they fostered.

We survey the QA systems for the Portuguese language, and conclude the chapter with research directions.

2.2 Overview and Historical Perspective

The first Question Answer systems, followed an approach of Natural Language Understanding and date from the end of the 50s and beginning of 60s. They were usually dedicated to very specific domains, following the view that understanding natural language requires the knowledge, as thorough as possible, of the context they deal with.

The concept of Question Answering was used in a broad sense to refer to systems that communicated with the user or dealt with natural language (English) statements, in a similar way to the human question and answer pattern. That was the understanding of (Simmons 1965), which made a survey of sixteen systems that fall in this category.

One of these is the BASEBALL system (Green et al. 1961), that was projected by Fredrick C. Frick, Oliver G. Selfridge, and Gerald P. Dineen and implemented by Bert F. Green, Jr., Alice K. Wolf, Carol Chomsky, and Kenneth Laughery at MIT. This system answered questions related to the baseball games of the American League, and was able to give information on the date and location of the matches, as well as teams and results, for a one year season.

Example questions are:

- "Who did the Red Sox lose to on July 5?"
- "Who beat the Yankees on July 4?"
- "Did any team play at least once in each park in each month?"

It used ad-hoc natural language techniques that for instance made the correspondence to "who" in the system dictionary as "Team=?".

Another system was the Student system (Bobrow 1964) that was developed at MIT as Dan Bobrow PhD project. The system was intended to solve elementary algebra problems stated in natural language.

Example question is

- "The sum of two numbers is 96, and one of the number is 16 larger than the the other number. Find the two numbers."

That produced the resulting output: "One of the numbers is 56"; "The other number is 40".

Techniques used:

- The program converted the natural language formulation of the problem into a set of algebraic equations by breaking it into simple patterns and searching for words or phrases that could be replaced by arithmetic expressions or variables.
- It was able to deal with connected discourse and not simply with isolated sentences.
- It used general knowledge information base to assist in the solution of algebra problems

- Could expand its knowledge base by interacting with the user and eliciting further information from him.

We can see that the concept of QA at the early years was closely connected to the Natural Language processing side of it. However another important area to the field is Information Retrieval. It has been present since the early years, but its influence was not fully integrated. In both his surveys, Simmons referred to document retrieval systems, with Salton work and the SMART information retrieval system is mentioned in the category of "Miscellaneous" in (Simmons 1970).

The research on the specific challenges of using IR techniques within QA systems has been the purpose of two workshops IR4QA, the first taking place in 2004 as part of ACM SIGIR 2004, and the second in 2008, held as part of ACL COLING 2008.

In recent years there has been very active research in this field, that led to the creation of tracks dedicated to QA in several international evaluation initiatives that, although presenting several variations, generally focus on the task of extracting answers from large open domain text collections. Such is the case of the Text REtrieval Conference (TREC) that had a QA track from 1999 to 2007 and the Text Analysis Conference (TAC) in 2008, for the English language. Another case is the bi-annual NTCIR Workshop that has been running a QA track, Question Answering Challenge (QAC), since its third edition in 2002, and whose main focus is on Asian languages. The Cross-Language Evaluation Forum CLEF, an initiative co-sponsored by the European Commission is running a QA track, QA@CLEF, since 2003, including the Portuguese language since 2004. Although this forum concentrates its attention on Cross-Language issues, it also has available mono-lingual tasks.

Another impulse for QA in the present days is the challenge that was undertaken by IBM to produce a QA system, Watson[1] (Ferrucci et al. 2010), to compete in a TV Quiz Show, Jeopardy, much in the style of the Deep Blue challenge of the turn of the century, that put man and machine face to face in a Chess competition. The challenge took place in the 14th, 15th and 16th of February 2011, where the winners of the show confronted the machine, with the final result being favourable to the latter.

[1]Named after the founder of IBM Thomas J. Watson.

2.3 Evaluation of QA systems

QA has been an area of interest for researchers, particularly over the last few years. As mentioned, this interest is in part motivated by the international QA evaluation forums, namely the Text REtrieval Conference (TREC), NTCIR and more recently the Cross-Language Evaluation Forum (CLEF) through QA@CLEF, running from 2003 and that since 2004 includes the Portuguese language.

In these evaluations, a collection of written documents is provided, as well as a set of questions, which are to be answered by the participating systems. Each answer is assessed manually, and, based on that assessment, an overall score is attributed to each participating system. Despite being dedicated to the common task of QA, each of the evaluations has its specificities. We dedicate the following sections to making an analysis of these initiatives.

We will centre our attention in TREC and CLEF. The bi-annual NTCIR Workshop, organised by NII from Japan, that has been running a QA track, Question Answering Challenge (QAC), since its third edition in 2002. We can also mention an interesting feature QAC has, of defining an interface for information to be exchanged between modules, that allows the influence of each module in the overall performance of the system to be determined.

2.3.1 TREC

The Text REtrieval Conference[2] (TREC)(Voorhees & Tice 2000), organised by the National Institute of Standard and Technology[3] (NIST) from the United States, has been conducting a track for QA dedicated to the English language. It was called TREC QA and it run from 1999 to 2007. This task was transferred to the Text Analysis Conference (TAC), also sponsored by NIST, in 2008.

The TREC QA had 9 editions and it was one of the most popular tracks (Voorhees 2007) with 20 participants in 1999 to 36 participants in 2001, with a minimum of 28 participant systems in 2000, 2004 and 2007.

Since an exhaustive study of the 9 editions is not possible in the scope of this work, we will describe in more detail the rules of the first edition, to give an idea on the nature of the task.

[2] http://trec.nist.gov/
[3] http://www.nist.gov/

The first edition of the evaluation, in 1999, consisted in assembling a collection of text documents, in English, from several different sources, namely: Foreign Broadcast Information Service, Los Angeles Times, Financial Times, Congressional Records, and Federal Register. The documents were formatted with SGML tags. A set of questions were produced based on the collection. The participants should answers these questions with passages produced from the document collection. Two different sizes for passages were allowed, short (up to 50 bytes) and long (up to 250 bytes). The question were fact based, and some examples of questions are:

- What year was the Magna Carta signed?
- How far is Yaroslavl from Moscow?
- Who was President Cleveland's wife?

The participants had the option to use their own retrieval system, or to use the documents provided from the organization, that were retrieved by the Zprise system.

Each system may give up to five answers per question, and each answer must be accompanied by the identification of the text it comes from. The five answers must be ordered using a rank from 1 to 5, with 1 being attributed to the most likely answer, and so on until 5.

The answers are then evaluated manually by assessors, who must provide a binary value to each answer: if an answer is correct the value 1 is attributed, and 0 otherwise. The answers were evaluated by three assessors and in case of differences in judgements an adjudicator reviews the assessment, and a final score of 0 or 1 is attributed for the answer.

The evaluation metric used is the Mean Reciprocal Rank, MRR, which is the mean of the reciprocal of the rank of the first answer that is correct for each question.

Its calculation is based on the Boolean function that represents the final judgement of the assessors as wrong (0) or correct (1). The definition of that function, which we will name F_{ij} for the case of answer j to question i, is:

$$F_{ij} = \begin{cases} 1 & \text{if answer } j \text{ to question } i \text{ is considered correct} \\ 0 & \text{if answer } j \text{ to question } i \text{ is considered wrong} \end{cases} \tag{2.1}$$

If we consider a set of questions, from 1 to Q, and if for each question up to N_a answers may be provided, we can define function J_i as:

$$J_i = \begin{cases} \frac{1}{\min_{j=1\ldots N_a}\{j:F_{ij}=1\}} & \text{if there is } j = 1\ldots N_a \text{ for which } F_{ij} = 1 \\ 0 & \text{if } F_{ij} = 0 \text{ for all } j = 1\ldots N_a \end{cases} \tag{2.2}$$

The definition of MRR would be:

$$\text{MRR} = \frac{1}{Q}\sum_{i=1}^{Q} J_i \tag{2.3}$$

In the first edition of TREC QA, the values are Q=200 and N_a=5.

Another metric that was later introduced was CWS (Confidence Weighted Score). It was used for the first time at TREC QA 2002 (Voorhees 2002), in which a file was provided by the participants with just one answer per question, with the answers ordered not by question number, but in order of decreasing confidence in the answer.

A formal definition for CWS would be:

$$\text{CWS} = \frac{1}{Q}\sum_{i=1}^{Q} \frac{\sum_{j=1}^{i} F_{j1}}{i} \tag{2.4}$$

It is important to keep in mind that index i follows the answers in order from most confident response to least confidence response, and that there is exactly one answer per question (hence the index 1 for answer number in F_{j1}).

2.3.2 CLEF

The Cross-Language Evaluation Forum[4] (CLEF) has dedicated a track to QA, the QA@CLEF track (Magnini et al. 2006), that occurred from 2003 to 2008, and that since 2004 includes the Portuguese language (Santos & Rocha 2005).

[4] http://www.clef-campaign.org/

2.3.2.1 QA@CLEF Main Task Description

QA@CLEF started in 2003 (Magnini et al. 2003; Magnini et al. 2004) with the aim of providing an evaluation environment in which QA systems for European languages (besides of English that was the language of TREC QA) could be tested. It is an initiative sponsored by European Union at the beginning under program DELOS Network of Excellence for Digital Libraries[5]. Given the language diversity of Europe, the Forum centred its attention not only in establishing a multilingual framework in which systems using European languages could be evaluated, but also stimulated the cross-lingual aspects of the tasks offered. According to this philosophy, tasks could be monolingual, that is the question language (source language of the information request) is the same as the language of the collection (target language) that is going to be searched in order to find an answer. The cross-lingual task are the ones in which the source language is different from the target language, thus involving some automatic translation mechanism. In the first edition of the track, three languages were offered as source and target (Dutch, Italian and Spanish), with the cross-lingual tasks being offered with five source languages (Dutch, French, German, Italian and Spanish) for an English target corpus.

There has also been national initiatives for example for the French language, in the form of EQueR-EVALDA (Campagne d'évaluation en question-réponse) that took place in 2004 (Ayache et al. 2006), sponsored by the French Ministry of Research within the scope of the EVALDA project and the Technolangue action (Interminesterial action to produce an infrastructure for the production and diffusion of French language resources). In this edition 7 research groups submitted their systems, with results reported as similar of those obtained at QA@CLEF 2004 (Ayache et al. 2005). This work, namely the collection built, was integrated in QA@CLEF in 2004.

Collections were being built for several other languages, as the case of Portuguese, and in 2004 (Magnini et al. 2004) there were already seven languages as target (Dutch, English, French, German, Italian, Portuguese and Spanish) to which two more could be used as source (Bulgarian and Finnish).

The information on the task rules and question categories and distribution are not language specific. The text collections were built so that they contain information regarding the same years, for instance the newspapers are all for the years of 1994 and 1995[6] so that questions could

[5] http://www.delos.info

[6] Except for languages where that was not possible, as the case of Bulgarian which was added later to the

be translated and used for different languages. This allowed system results to be compared even if systems were participating in tasks for different languages.

The monolingual task of QA@CLEF consists in finding answers to 200 questions within a text collection, all in the same language.

Systems should be able to answer three categories of questions, namely factoids, definitions and closed list questions, some of which may include a temporal restriction.

The type of the questions can be very diverse, ranging from questions about persons, organizations or locations, for instance. Neither the question category, nor its type or if it is temporally restricted, are explicitly given with the question, therefore it is up to the system to find out that information.

The questions can be stand-alone questions, or organized in clusters of up to four questions about a common topic, with the topic being determined by the first question/answer pair. Questions 2 to 4 of the cluster can contain co-references to the cluster's topic. The clusters are identified by a number that is given to the system.

It is possible that a question has no answer in the data collection, in which case the answer should be NIL.

Each question, except NIL questions, must be supported with an excerpt from a document in the collection, therefore an answer consists of the exact answer (with no more information than strictly needed), an identifier from a document in the text collection, and an excerpt of up to 700 characters from that document that supports the answer [7].

The system may produce up to three answers for each question, and two complete answer files can be submitted for evaluation.

Each participant receives a file with the questions, and has one week to return one or two answers files in an XML format provided by the organization.

collection.

[7] In fact, up to three separate excerpts from the same document can be provided, as long as the limit in characters is not exceeded.

Table 2.1: Summary of Participations at QA@CLEF with Top 5 Systems in Monolingual Tasks.

Year	2003	2004	2005	2006	2007	2008
# Participants	8	18	24	30	21	21
Total Runs	17	48	67	77	37	51
Monolingual Runs	6	20	43	42	23	31
Cross lingual Runs	11	28	24	35	14	20
1º	-	UAMS NL-NL 91	PRIB PT-PT 129	PRIB PT-PT 134 ES-ES 105	SYN FR-FR 108	PRIB PT-PT 127 ES-ES 86
2º	-	FUHA DE-DE 67	SYN FR-FR 128	SYN FR-FR 129	PRIB PT-PT 101 ES-ES 89	SYN FR-FR 113
3º	-	ALIV ES-ES 65	GRON NL-NL 99	INAOE ES-ES 102	UE PT-PT 84	UE PT-PT 93
4º	-	UE PT-PT 57	UAMS NL-NL 88	ULIA FR-FR 88	INAOE ES-ES 69	DKFI DE-DE 74
5º	-	IRST IT-IT 56	DKFI DE-DE 87	VEIN & DKFI ES-ES DE-DE 80	DKFI DE-DE 60	UAb PT-PT 65

Participants:
ALIV - University of Alicante, Spain
DKFI - Deutsche Forschungszentrum für Künstliche Intelligenz (German Research Center for Artificial Intelligence), Saarbrücken, Germany
FUHA - Fern University, Hagen, Germany
GRON - University of Groningen, Netherlands
INAOE - Instituto Nacional de Astrofísica, Óptica y Electrónica, México
IRST - ITC (Instituto Trentino di Cultura), IRST Institute for Scientific and Technological Research, Trento, Italy
PRIB - Priberam, Portugal
SYN - Synapse Développement, France
UAb - Universidade Aberta, Portugal
UAMS - University of Amsterdam, Netherlands
UE - University of Évora, Portugal
ULIA - Laboratoire d'Informatique d'Avignon, France
VEIN - VEng (Vanguard Engineering) + INAOE, Puebla, México
Languages:
DE - German **ES** - Spanish **FR** - French **IT** - Italian **NL** - Dutch **PT** - Portuguese
(Results expressed in number of correct first answers in 200)

In Table 2.1 we show a summarized information on the QA@CLEF campaigns from 2003 to 2008, with the top 5 systems in monolingual tasks. This information is not given for 2003 since most of the groups participated in cross-lingual tasks. In 2006 DKFI and VEIN occupied ex aequo the fifth position, with a score of 80 correct first answers. To be noted that one to three Portuguese systems are among the best five in each year.

QA systems for Portuguese at QA@CLEF are worth a careful analysis because they produced state of the art results (Forner et al. 2009), with the system from the Portuguese company Priberam (Amaral et al. 2009) being the best system overall in the 2005, 2006 and 2008 campaigns (2nd in 2007). At QA@CLEF 2008, out of the 21 systems participating, there were two systems for Portuguese among the three best systems: Priberam (1st) and Senso (Saias & Quaresma 2008a) (3rd). Our system, IdSay, was 5th overall.

2.3.2.2 Portuguese Text Collection

In 2004 the document collection for Portuguese was created, consisting of the newspaper articles from Portuguese newspaper "Público" to which the news articles from the same period from the Brazilian newspaper "Folha de São Paulo" were added in 2005. The edition of the Portuguese Wikipedia was added to the collection in the 2007 edition, with a frozen version of the HTML edition of the Portuguese Wikipedia from November 2006.

This data collection was used for the 2007 and 2008 editions of the evaluation campaign, and it is the reference data collection we consider for the tests described in the present work.

Some statistics on the sizes of the collection are presented in Table 2.2[8].

2.3.2.3 Question Categories and Distribution

In the first two editions of QA@CLEF all questions were factoids[9] or definitions.

From 2006 onwards another category, list questions category, was introduced. In the 2006 edition the lists were open lists, similar to the ones used at TREC QA, in the sense that the number of possible answers were not necessarily known beforehand, and the answers could occur

[8]Some Wikipedia files are not content files.

[9]We use in the text the term "factoid", because it is widely used in the literature. However we would prefer the term "questions about facts", or "factual questions", because of the meaning of "factoid" as "an invented fact usually taken as true".

Table 2.2: Collection Size

Data	Collection Value
Size of Público	341 MB (358 493 681 bytes)
Files of Público	726
Size of Folha	222 MB (233 722 868 bytes)
Files of Folha	730
Size of Wikipedia	7.13 GB (7 662 009 520 bytes)
Files of Wikipedia	602 002
Size of Collection	7.68 GB (8 254 226 069 bytes)
Files of Collection	603 458
Documents	414 895
Distinct Words	942 990
Total Words	170 290 141

separately. In the 2007 and 2008 editions the type of list considered was close lists, in the sense that the number of answers was known beforehand, though it may or not be stated in the question, and occurrences should be located together in the same document.

In terms of the Portuguese variants European and Brazilian, the questions, although made by the Portuguese Researcher, Paulo Rocha from Linguateca, took into account Brazilian Portuguese vocabulary and spelling, when the Brazilian newspaper articles were added to the collection. As mentioned in the previous section, in 2007 a frozen version of the Portuguese edition of Wikipedia was added to the data collection. This contributed to enrich the diversity of the texts, with the presence of both Brazilian and European Portuguese (and possibly other sources) coexisting together, as well as different writing styles.

Regarding the type of question and if it is temporally restricted or not, the information about the number and distribution of questions per question category and per year is presented in Table 2.3. The years are from 2004 to 2007, and the questions of 2008 are presented in Appendix C and will be treated in more detail in Chapter 7.

We will give some examples of questions of each type:

- Factoid - Person

  Quem escreveu "Ulisses"? [Who wrote "Ulysses"?]

- Factoid - Date/Time

  Quando é que abriu o Sony Center na Kemperplatz de Berlim? [When did the Sony Center

Table 2.3: Portuguese Question distribution per year, type and category

Category	Type	2004	2005	2006	2007	Total
F	Person	42	46	31	21	140
	Date/Time	15	15	19	19	68
	Location	41	35	25	31	132
	Organization	14	23	21	21	79
	Measure	23	18	21	16	78
	Count	0	0	0	21	21
	Object	8	0	0	5	13
	Manner	4	0	0	0	4
	Other	21	21	28	26	96
D	Person	18	27	9	9	63
	Organization	14	15	7	6	42
	Object	0	0	7	6	13
	Other	0	0	23	9	32
L	-	0	0	9	10	19
Total		200	200	200	200	800
Temporal Restriction		0	23	21	20	64

open in Kemperplatz in Berlin?]

- Factoid - Location

  Em que cidade fica a mesquita de Al Aqsa? [In what city is the Al Aqsa mosque located?]

- Factoid - Organization

  A que partido pertence Bill Clinton? [What party does Bill Clinton belong to?]

- Factoid - Measure

  Quantos milhões de imigrantes ilegais há na União Europeia? [How many illegal immigrants are there in the European Union?]

- Factoid - Count

  Quantos focos tem uma elipse? [How many focus points does an ellipsis have?]

- Factoid - Object

  Que espada usavam as legiões romanas? [What sword did the roman legions use?]

- Factoid - Manner

  Como morreu Jimi Hendrix? [How did Jimi Hendrix die?]

- Factoid - Other

  Quero o nome de um vinho húngaro. [I want the name of an hungarian wine.]

- Definition - Person

  Quem é Nelson Mandela? [Who is Nelson Mandela?]

- Definition - Organization

  O que é a FIFA? [What is FIFA?]

- Definition - Object

  O que é Quinoa? [What is Quinoa?]

- Definition - Other

  O que é o Acordo de Dayton? [What is the Dayton Agreement?]

- List - Open List (2006)

  Quais são os sintomas da Doença de Parkinson? [What are the symptoms of the Parkinson's Disease?]

- List - Closed List (2007)

  Quais são os signos do Zodíaco? [What are the Zodiacal signs?]

- Temporally Restricted Question

  Qual a divisa austríaca antes de 2002? [What was the austrian currency before 2002?]

2.3.2.4 Evaluation Metrics

The answers returned by the systems are evaluated manually by human assessors, in a process similar to that of TREC QA which we have described. The metrics used to rank the systems are also based on the Boolean function that takes into account the judgement of the assessors.

In the case of QA@CLEF several assessment values can be attributed to answers: in case they are correct the value 'R'(Right) is attributed. In case of incorrect answers the possible values are: 'X' (ineXact) when the support of the answer contains the answer, but the answer presented is inexact due to extraction problems, 'U' (Unsupported) in the case of a correct answer but which is not supported by the accompanying passage and 'W'(Wrong) in the case of a wrong answer. These values are sometimes even further detailed by the assessor, as for instance in the

case of 2005 in which the wrong answers were further classified in three breakdown values[10] and in the 2007 and 2008 edition, where the X assessment was break into two values X^+ and X^- that indicated respectively that the system had extracted more words or less words from the support than the exact answer.

The Function F_{ij} for the case of answer j to question i, defined for TREC in 2.1, now becomes:

$$F_{ij} = \begin{cases} 1 & \text{if answer } j \text{ to question } i \text{ is assessed as 'R'} \\ 0 & \text{if answer } j \text{ to question } i \text{ is assessed as 'X','U'or'W'} \end{cases} \tag{2.5}$$

The main evaluation metric used in QA@CLEF 2008 is accuracy over the first answer, which is the average of first answers that where judged to be correct. A more formal definition of accuracy over the first answer can be given as:

$$\text{Accuracy over the first answer} = \frac{1}{Q}\sum_{i=1}^{Q} F_{1j} \tag{2.6}$$

Another metric used is MRR or the mean of the reciprocal of the rank of the first answer that is correct for each question. As defined for TREC in the expression 2.3 using the auxiliary function defined in 2.2. In QA@CLEF, the values are Q=200 and N_a=3.

Another common measure used for Question Answering systems is the accuracy over all answers. If we define R_i, for question i, as:

$$R_i = \begin{cases} 1 & \text{if there is } j{=}1 \ldots N_a \text{ for which } F_{ij}{=}1 \\ 0 & \text{if } F_{ij}{=}0 \text{ for all } j{=}1 \ldots N_a \end{cases} \tag{2.7}$$

We can define the accuracy over all answers in the following way:

$$\text{Accuracy over all answers} = \frac{1}{Q}\sum_{i=1}^{Q} R_i \tag{2.8}$$

[10]NIL answers are assessed as"null", "rubbish" is used for answers that did not risk being considered the correct answers, and "dangerous" in the case a user could be misled and take the wrong answers as correct.

This metric takes into account the correct answers returned for each question regardless of the fact that they where returned in the first position or not. Correct answers that are given earlier in the answer list have a higher contribution to the score. The value of MRR is higher than accuracy over the first answer (assuming $N_a > 1$) and lower that the accuracy over all answers, in a way that takes into account the relative position of the first correct answer in the list of returned answers.

Another metric that was provided for the systems was CWS (Confidence Weighted Score) as defined in expression 2.4 . It takes into account exactly one answer per question (the first one, or top ranked one for each question), with the answers ordered not by question number, but in order of decreasing confidence in the answer. This measure was not calculated for systems that did not provide a value for the score of the answer.

2.3.2.5 Other Tasks of QA@CLEF

The main task is not the only task offered at QA@CLEF. Other tasks are the Answer Validation Exercise, AVE, that gives pairs of questions and answers for the systems to determine if the answers are valid for the given question. It is an important feature for a QA system to have, since it may help automatic validation of answers. The task has similarities with the Recognizing Textual Entailment task, RTE, at the TAC Conference, that can be defined as follows: "given a corpus and a set of "candidate" sentences retrieved by Lucene from that corpus, RTE systems are required to identify all the sentences from among the candidate sentences that entail a given Hypothesis."[11].

Another task of QA@CLEF, is the Question Answering over Speech Transcripts, QAST, that is of particular interest to the current work since our aim is to investigate and develop techniques that can be applied to data obtained from automatic speech recognizer, ASR systems. This kind of text data differs from written data because it can have errors that usually written texts do not have, ranging from morphological and syntactic errors to semantic ones[12]. This difference is natural since we are comparing text of human origin, and additionally written text usually undergoes editing processes, be it the formal editorial process of a newspaper or the less formal

[11] Quoting the main page of RTE-7, at http://www.nist.gov/tac/2011/RTE/

[12] We are considering written text of human origin, excluding text obtained by automatic means, as the case of scanning hand written text.

collaborative editing of Wikipedia, that make errors even less prone to happen. QAST will be described with more detail in Chapter 9.

2.4 Question Answering Systems for Portuguese

As far as participation at the Portuguese monolingual task at QA@CLEF is concerned, it has been increasing since the first edition, in which there were two participants, University of Évora with the Senso system and Linguateca with the Esfinge system.

We present in Table 2.4 the results of the systems that participated in the five years that the Portuguese monolingual task of QA@CLEF was available. The results are presented in the form of the number of questions of the 200 question set that each system answered correctly in the first answer, according to the overview papers for each year (Magnini et al. 2004; Vallin et al. 2005; Magnini et al. 2006; Giampiccolo et al. 2007; Giampiccolo et al. 2008)

Table 2.4: Results Overview of Portuguese Monolingual Task at QA@CLEF (Correct First Answers in 200).

System	2004	2005	2006	2007	2008
Senso	57	50	-	84	93
Esfinge	30	46	50	16	47
Priberam	-	129	134	101	127
Raposa	-	-	26	40	29
QA@L2F	-	-	-	26	40

In the following sub-sections we identify the most significant features of these systems, as published in the QA@CLEF literature.

2.4.1 University of Évora - Senso

The University of Évora was one of the two systems that participated in the first edition of the Portuguese monolingual task QA@CLEF in 2004 (Quaresma et al. 2004). The approach followed was to use deep linguistic analysis to process the documents, and to store the semantic information derived from them in a knowledge base. This is a preparatory phase, and upon its completion the system is able to answer questions. The process of answering questions is to find the DRS (Discourse Representation Structures) representation corresponding to the question,

and then to find the answer using an inference mechanism.

The authors report that treating the information of all documents together would be too complex, so the inference process is restricted to 50 documents. Therefore for each question the inference process is applied to the 50 documents most likely to contain the answer, as provided by an IR system used in parallel for that purpose.

The preliminary phase related to the creation of the knowledge base has the following steps:

1. Linguistic processing of the text based on the parser PALAVRAS (Bick 2000), obtaining a file with the syntactic representation of each document.
2. Semantic interpretation of the syntactic structure to first order logic form, in the form of DRS (Discourse Representation Structures).
3. Semantic/Pragmatic interpretation of the sentences rewritten in DRS format. First an ontology is created combining the DRS sentences of the text with a pre-existing global ontological base. The ontology creation uses the OWL (Ontology Web Language) format. It is obtained using a logic programming framework, ISCO, developed at University of Évora, that: "allows the integration of Prolog-like inference mechanisms with classes and inheritance, and constraint solving algorithms". The knowledge base of facts is then derived from the text using the GNU Prolog Finite Domain (FD) constraint solver. This step involves the integration of ontological information from other external sources.

The authors illustrate the process trough the following example: O gato do João comeu o rato do Manuel [João's cat ate Manuel's mouse]. After step 2. the DRS obtained, with 4 referents (the cat, João, the mouse and Manuel) and the corresponding relations would be:

DRS([def-A-m-s, def-B-m-s, def-C-m-s, def-D-m-s], [cat(A), rel(of,A,B), name(B,'João'), eat(A,C), mouse(C), rel(of,C,D), name(D,'Manuel')]).

After the incorporation of the ontological information the DRS would become:

DRS([def-A-m-s, def-B-m-s,def-C-m-s, def-D-m-s], [cat(A), owns(B,A), person(B), name(B,'João'), eats(A,C), mouse(C), owns(D,C), person(D), name(D,'Manuel')]).

When a question is processed, it undergoes the same process. The example question Quem comeu o rato do Manuel? [Who ate Manuel's mouse?] would result in the following representations:

DRS([who-A-X-Y, def-B-m-s, def-C-m-s], [eat(A,B), mouse(B), rel(de,B,C), name(C,'Manuel')])

and

DRS([who-A-X-Y, def-B-m-s, def-C-m-s], [eat(A,B), mouse(B), owns(C,B), person(C), name(C,'Manuel')]).

The system would then try to infer the answer.

The IR part of the system indexes the texts in the collection removing stop words (the list is not provided) and using lemmatization through the POLARIS system (Lopes et al. 1994). The system used is an adaptation for Portuguese of a system from the SINO system, developed originally by the Australasian Legal Information Institute, AustLII.

Besides the process of obtaining the DRS for the question, the question is also processed to generate a query that is input to the IR module. Three queries are forwarded to SINO, who returns a list of documents ranked according to several criteria such as results from the more specific queries are ranked higher, and number of word hits and word hits in the title of the document. The top 50 documents are subject to the inference process.

The three queries input to the system specify different levels of specificity, through the use of the Boolean operators 'AND' and 'OR'. The first question of the QA@CLEF 2004 campaign, Em que cidade se encontra a prisão de San Vittore? [In what city is the San Vittore prison?] is used by the authors to exemplify these queries:

- cidade AND encontrar AND prisão AND (San AND Vittore)
- cidade AND (encontrar OR prisão OR (San AND Vittore))
- cidade OR encontrar OR prisão OR (San AND Vittore)

The results of the system were 47 correct first answers. The main difficulties encountered, besides the complexity of treating a large number of documents, was to extract the semantic information for the texts, because, even considering a single document the texts had a very large number of concepts involved and it was very difficult to extract the relations between them. Another problem reported was that the information retrieval module, that was a critical factor for the correct answer to be found, failed in identifying the documents that contained the answer.

Other problems that were mentioned was how to solve the semantic ambiguity, and the problem that the same entity having two different forms that are identified as being the same, or when the same referent being incorrectly unified for distinct entities.

These queries are presented in order of decreasing specificity, but no indication is given on the criteria on how the queries are constructed, for instance the fact that the words San Vittore must appear together, except that they are constructed based on the DRS for the question.

In the second participation of University of Évora in QA@CLEF (Quaresma & Rodrigues 2005), in 2005, the system abandons the use of an IR system to select the top 50 documents to treat for each question. Instead of that, the documents are processed (in a preliminary step) and all the information derived from them is stored in a common database.

Apart from that, the process is fundamentally the same as the previous year's. Evolution in specific aspects of the system is reported, like:

- the algorithm takes into account the syntactic category of answers that can be part of the answer;
- it verifies if words appearing in the question also appear in the answer, and tries to avoid that;
- a special treatment was introduced for dates and locations, with access to databases of dates and locations;
- there is a comparison between all answers found for a question, with a scoring based on frequency being introduced.

It is still unclear the origin of the information included in the ontology constructed and used. No performance data is explicitly indicated for the system.

In this edition the text collection was no longer solely the news articles from "Público", but it also includes the news articles from "Folha de São Paulo". It is reported that there was not enough time to process the added Brazilian news articles.

The results obtained were 50 correct answers.

The main problems encountered are the incomplete ontological information, with a future intention of automatic ontology creation being expressed, the incorrect syntactical analysis, the

problem of choosing the best interpretation for the information returned from the parser, and the lack of a synonyms mechanism.

The next participation of University of Évora in QA@CLEF is in 2007, and with the name Senso being used for the system (Saias & Quaresma 2007; Saias & Quaresma 2008b). The system resumes its original architecture of a parallel usage of an IR system and the inference engine procedure. However the usage of the IR system has now a more prominent role in the system, as it includes new mechanisms of searching for answers "autonomously". This process is described as "ad hoc" module, and the answers produced by this module are later joined with those produced by the logic solver (inference mechanism used in the last edition) in what is called the solver module. The reason indicated for using again an IR module is the volume of data to be processed. In fact in 2007 the addition of the Portuguese Wikipedia to the text collection, represented an increase of over half a million files to process. The IR system now used is Lucene[13]. The pre-processing options are not mentioned.

The ontological information (proprietary Senso ontology) was expanded and is now used for search terms expansion and for the verification of concepts using the *isA/specialization* relations. The ontology contains about 3500 concepts and the relations covered include *isA*, *usedFor*, *locatedAt*, *capableOf* and *madeOf*.

The results were 84 correct answers, with 111 questions being answered as NIL, of which 12 were correct. This represents a considerable improvement of previous results. As for problems identified there is the fact that the documents returned by the IR system do not include the documents containing the correct answer, and they report the intention of analysing the best ways to construct the query to Lucene. The problem of semantic analysis of complex sentences is still mentioned, along with limitations in the ontology. Another difficulty is the lack of precision in extracting semantic information, with the intention of improving the desambiguity at this level being referred. The manual revision and extension of the ontology is mentioned as an intended improvement.

To close our analysis on the Senso system we look at the system version that participated at QA@CLEF 2008 (Saias & Quaresma 2008a). In this edition the system maintains the architecture of the previous edition, with the answers being produced in the solver module by two alternative procedures, the logic solver and the "ad hoc" solver based on the selection of the texts

[13] http://lucene.apache.org/

using the Lucene IR system followed by a rule based answer extraction. There was an improvement in the mechanism to build the query from the question, but details are not specified. The pre-processing phase of the documents was also changed, with an original stemming procedure being introduced, but no details are given on its implementation. In the "ad hoc" module, a new way to answer definition questions was implemented through a pattern based approach used for Bulgarian for the QA@CLEF 2006 edition (Tanev 2006). Another improvement to this year's system was introduction of a validation of the list of answers produced by the solver module. A "a quick Web search to measure the answer value popularity with respect to the question" is mentioned, but no further details are presented.

The results of the system improved to 91 correct answers with most of them reported as being produced by the "ad hoc" module. The number of NIL questions was drastically reduced from 111 in the previous year to 21. As future improvements it is mentioned that the rules for the extraction of answers need to be adjusted.

2.4.2 Esfinge [Sphinx]

The Esfinge system (Costa 2004) was the other system that participated at the first edition of QA@CLEF for Portuguese. It is developed at the Linguateca/SINTEF in Norway, and it has participated in all editions of QA@CLEF.

The starting point for the system was the architecture of Eric Brill (Brill 2003), and it consists of four modules:

- Question reformulation
- n-grams harvesting
- n-grams filtering
- n-grams composition

An n-gram is described as a sequence of one or more words that occur together in the answers extracted. The most frequent ones are selected as answers. Esfinge considers n-grams from 1 to 3 words, using the Ngram Statistics Package (NSP) (Banerjee & Pedersen 2003).

The question reformulation applies a set of patterns. It consists of matching the question with a regular expression (the question pattern), and if the process is successful, the module produces an answer pattern and a score. The answer pattern will be used to look for passages in the document collection. For example, the question pattern O que X Y? [What X Y?] transforms to answer pattern "Y X". This question pattern matches the question O que é a MTV? [What is the MTV?] generating the answer patter a MTV é [the MTV is]. The patterns and score were obtained manually, and the author states that this list could be improved at a later stage. For cases not covered by the more specific question patterns, a pattern that consists in all words in the question by any order is also considered as a last resort.

The n-grams are extracted from the first 100 documents returned that match the answer pattern. All n-grams are considered possible answers, and are scored according to the following formula:

$$n-gram\ frequency\ \times\ score\ of\ the\ answer\ pattern\ \times\ n-gram\ length.$$

For returning documents for a given answer pattern, Esfinge uses the IMS Corpus Workbench (Christ 1994), a software package designed to process large text corpora (100 million words and more). The Corpus Workbench (CWB) has been developed at the Institut für Maschinelle Sprachverarbeitung (IMS) from the University of Stuttgart since 1993, and Esfinge uses version 2.2 from 1999. The results of a query to Google[14] search engine were also used. The passage extraction for Google was already solved, since Google returns snippets, but for the document collection several strategies were tried: 50 contiguous words; documents divided by sentences; document divided in sets of three sentences. For indexing the document collection, Esfinge uses stop lists, discarding the 22 most frequent words in the collection CETEM/Público (Santos & rocha 2001).

After obtaining the answers, they are filtered by type. Examples are given, but not the complete set of options: for instance, a "when?" question requires an answer that is a date, and all other answers could be discarded; a "how many?" question requires a number answer, and other types of answers can be discarded as well. Two filters were used: one to discard answers that are contained in the question, and the other that makes use of the morphologic analyser

[14] http://www.google.com/

jspell (Simões & Almeida 2001) to determine possible part of speech tags for the answers to help make a selection valuing name phrases. However a misinterpretation in the output of the package is reported, which probably resulted in incorrections.

In the last module, n-grams composition, the answers are arranged in list form, if the question formulation requires a list answer, returning the first N answers.

Esfinge submitted two runs, the first run based solely on the document collection and the second run using the Google interface to extract the answer, that is later checked using the document collection. This second run proves to be the best one, resulting in 30 correct answers, when for the first run only 21 questions were answered correctly.

In the second participation of Esfinge at QA@CLEF (Costa 2005), the main objectives were to correct mistakes learned from the previous year participation, as well as to make an attempt of cross-language QA with English as the source language and Portuguese as target language.

The document collection is stored in the IMS Corpus Workbench but the option is to segment each document in sets of three sentences, since in the previous year it performed slight better than the other options. It now uses stemming (Orengo & Huyck 2001) to increase the possible matches in the document collection. The stop list now contains some words not wanted in the answer patterns, that are thus discarded from the document collection. Examples of these words are chama [is called], *fica* [is located] and *país* [country].

The Web is still searched for answers in one of the submitted runs, with answers being searched in the data collection afterwards. To improve the web results, a set of list patterns is used to exclude several web sites, for instance blog or humour sites.

The best 200 n-grams are considered possible answers, and named entity recognition (NER) is now used. The system used is SIEMÊS (Sarmento 2006d), that was developed together with the gazetteer REPENTINO (Sarmento 2006c) and had obtained good results at the HAREM NER evaluation for Portuguese (Santos et al. 2006). SIEMÊS detects and classifies named entities in a wide range of categories, and Esfinge uses a sub-set of these categories, namely Human, Country, Settlement (includes cities, villages, etc), Geographical Locations (locations with no political entailment, like for example Africa), Date and Quantity. A list of question patterns/answer types was manually built, and if the question pattern is matched, the corresponding answer type is checked by the NER system. Each n-gram classified by SIEMÊS as one

of the expected answer types gets an increase in ranking.

The answers are then filtered, first discarding words contained in the question. Second, a list of undesired answers is discarded. This list was built by manually analysing past answers, and have entries like: pessoas [people], nova [new], lugar [place], *grandes* [big] and exemplo [example]. Third, the morphologic analyser jspell is used, if no type of answer was possible to determine or in case the NER system failed to identify the answers as the expected type. The system discards all answers for which the first and last word is not one of the following categories: adjective; common noun; number or proper noun. Finally, a filter is used to verify if the answer can be supported by a document in the collection, for the case in which the web search was used.

The answer composition have also an improved procedure. If an answer passes the previous filters, the answer is not yet returned if other possible answers contain that answer. In that case, if one of those answers pass the filters, this answer is discarded and selected the answer that contains this one.

The final answer is the answer with the highest score.

These changes result in an improvement of accuracy at first answer, with the version of the system that uses the web first still getting slightly better results. The number of correct first answers in the 200 set is 46 using the web, and 43 using only the data collection.

As far as cross-language is concerned, the question in English is given to Altavista/Babel fish for translation for Portuguese, with the process being the same as for the monolingual operation. The results obtained in this situation are lower (24 correct answers).

The third participation of Esfinge (Costa 2006b; Costa 2006c) maintains the global algorithm with a slight change, it now checks the document collection, before the web, to allow NIL answers to be returned more quickly. It started to use the parser Palavras (Bick 2000) to obtain the number of answers expected in list answers.

The NER system SIEMÊS is still used to refine answer patterns for some type of questions, to which the use of the database of co-occurrences BACO (Sarmento 2006a) was added to adjust the scoring of answers, increasing the scoring in the less frequent co-occurrences. This process turns out not to improve the results in a significant way. The results of this different test situations were subject to analysis in (Costa & Sarmento 2006).

The results of this year improve to 50 with web and 46 without web, with the gap between

the two situations widening slightly, favouring the use of the web.

The fourth participation of Esfinge (Cabral et al. 2007) coincided with two major changes: the introduction of Wikipedia in the text collection and the cluster questions, liked through anaphoric references. The paper describes the anaphora system developed, that is based on the PALAVRAS parser, that increases its role in Esfinge. Also described is the use of a MySQL database to store the Wikipedia, instead of IMS Workbench, since the increase in size of data seemed to indicate a dangerous increase in processing time. The authors manage to index words in MySQL with 3 or more characters, with words with 2 and 1 character being ignored in retrieval. The global strategy showed some evolution, with answers being searched for by the order in the Web then in the Newspaper collection and then in Wikipedia.

The overall result of this year dropped to 16. This is explained by the existence of bugs later detected in the system, as well as an increase in the complexity of the questions.

In its 5^{th} participation at QA@CLEF (Costa 2008) Esfinge introduced answer retrieval patterns, to allow the extraction of specific answers from the documents. In previous years the treatment of questions heavily relied on patterns and the match of the question text with a question pattern resulted in the selection of an answer pattern. This answer pattern was to be used to identify the sentences from the document collection where an answer was liable to be found. This year answer retrieval patterns started to be used by Esfinge for the purpose of answer extraction. Only if no answers are generated by answer retrieval patterns, is the NER (SIEMÊS) used to extract answers of specific types, and if still no answers are found, then the n-grams technique is used, with the selection as answers of the most frequent n-grams starting and ending with an adjective, common noun, number or proper noun, as determined by jspell.

Also, in document retrieval, if the answer pattern does not result in any answers, a pattern generated by PALAVRAS is used, and if there are still no answers, the same pattern without the verbs is tried.

The introduction of answer retrieval patterns improved results but only for definition questions. The results for correct first answers was 47.

As for general comments for Esfinge, we can say that it is a system very close to our research intents, namely the aim in producing a QA system with simple techniques, that the authors make available to other researchers to use.

The system however has one problem that we want to avoid in our proposal, that is the existence of many manual lists. Looking in all the papers, we can count up to six lists in the final system in 2008:

1. question answer/answer pattern (for each match of a question with the question pattern, search in documents for the answer pattern)
2. list of patterns of undesired sites (if a site has one of the undesired patterns, ignore it)
3. stop word list with words undesired in document match (these words are not indexed because they could disturb passage extraction)
4. question answer/answer type (to verify if answers are of a given type, using NER/SIEMÊS)
5. undesired answers (any answer in this list is discarded)
6. answer retrieval pattern list (to extract answers from the passages)

Although the use of patterns are an acceptable mean of dealing with language specific characteristics, excluding a set of sites, setting as stop list undesired words in document match, and excluding undesired answers, are not lists that can be obtained and maintained easily, so we do not concur with its usage.

The components used are most of them available for researchers, such as n-grams statistics package, jspell, PALAVRAS, SIEMÊS, BACO, IMS Workbench, MySQL, and several tools from Linguateca. This might not be compatible however with the efficiency that we aim at, as we would risk becoming "hostages" of the possibilities offered by a tool.

2.4.3 Priberam

The system from Priberam participated at QA@CLEF for the first time in the 2005 edition (Amaral et al. 2005; Amaral et al. 2006). Priberam is a Portuguese company that specializes in linguistic tools for Portuguese. The system is based on the company NLP workbench and on a proprietary Information Retrieval system. This IR system is based on the participation of Priberam in the project TRUST (Text Retrieval Using Semantic Technologies) that is described in (Amaral et al. 2004) and also in the company's system for Legal Information System, LegiX[15].

[15] http://www.legix.pt

The architecture of the system consists of five modules: Indexing process, Question analysis, Document retrieval, Sentence retrieval e Answer extraction, and is considered by the authors to be "fairly standard" or "a standard approach".

The system works based on three types of patterns specified using a proprietary language to support patterns, that supports more options that regular expressions. The three patterns are:

- QP - Question Patterns - used to categorize questions;
- AP - Answer Patterns - used in the indexing process to indicate for each sentence the potential answer type it may contain (more that one answer type can be attributed to a sentence);
- QAP - Question Answer Patterns - more detailed answer patterns, that are "activated" after the categorization of the question.

The text base is processed in an off-line "a priori" (Indexing process phase) using AP's. This corresponds to the predictive annotation phase described in (Prager et al. 2000; Prager 2006). QP's are attributed to the question and QAP's are only used afterwards, in Answer Extraction. Both QP's and QAP's have scores associated to them, that will be used to score candidate answers. These scores are heuristically adjusted, reflecting information such as the existence of optional terms is rewarded and terms occurring at a longer distance in terms of words that the maximum admissible are penalized.

Regarding the question classification the taxonomy created in project trust is used, that consists of 86 question categories, including: <DENOMINATION>, <DATE OF EVENT>, <BIRTH DATE>, <DATE OF DEATH>, <LOCATION>, <TOWN NAME>, <COUNTRY>, <FUNCTION>, <AIM>, <CAUSE>, <CONSEQUENCE> and <CONDITION>. The option for using a supervised learning method as in (Ferrés et al. 2004) was discarded due to the unavailability of training data, especially given the large number of categories used. The questions are classified using QP's and for the 2005 it is allowed to attribute more than one category per question.

The question Quem é Jorge Sampaio?[Who is Jorge Sampaio?] that enquires on the function of Jorge Sampaio, the former President of the Portuguese Republic (current in the data collection), is used to exemplify the QP, QAP and AP patterns, that are presented in Figure 2.1.

```
// Example of a question answer block encoding QPs and QAPs:

Question (FUNCTION)
     :  Word(quem) Distance(0,3) Root(ser) AnyCat(Nprop, ENT) = 15
        // e.g.  ''Quem é Jorge Sampaio?''
     :  Word(que) QuestIdent(FUNCTION_N) Distance(0,3) QuestIdent(FUNCTION_V) = 15
        // e.g.  ''Que cargo desempenha Jorge Sampaio?''
Answer
     :  Pivot & AnyCat (Nprop, ENT) Root(ser) {Definition With Ergonym?} = 20
        // e.g.  ''Jorge Sampaio é o {Presidente da República}...''
     :  {NounPhrase With Ergonym?} AnyCat (Trav, Vg) Pivot & AnyCat (Nprop, ENT) = 15
        // e.g.  ''O {presidente da República}, Jorge Sampaio...''
     ;

// Example of an answer block encoding APs:

Answer (FUNCTION)
     :  QuestIdent(FUNCTION_N) = 10
     :  Ergonym = 10
     ;
```

Figure 2.1: Pattern examples for Priberam's QA System

The "with" keyword means the second term has to be included as part of the first term, in the case of the second QAP where the string "current President of the Republic" contains "President of the Republic", which makes both strings acceptable answers.

The indexing phase as mentioned is done by an adapted version of a proprietary IR system to index semantic information, ontology domains, and questions categories among other QA specific information. The indexation is done at document level, but for each word besides the document identification, the references of the sentences that contain it within the document are also stored (for performance reasons). The documents are subject to a thorough analysis: before being indexed the text must be divided in sentences and for each sentence information is collected on the ontological and terminological domain identification and the question categories it may possibly answer to (via AP's). Then there is a treatment at word level, with several information being flagged in the text for instance stop words, named entities (NE), numbers, dates and fixed expressions. Each word is represented by a triple [lemma, head of derivation, POS].

When a question is presented to the system, its text is processed in a similar way than the documents, so that the documents that provide the best matches can be selected. The information that is selected from the question for search purposes is called the pivots. They are extracted after the question text is morphologically disambiguated and lemmatization is

performed. The question is categorized, producing the following results: one or more question categories are selected for the question, QAP's are activated for extraction purpose, and a score for each QP pattern that matched the question text. The information about the pivots include the lemma of the pivot, its head of derivation its POS and synonyms. The ontological and terminological domains for the question are also determined.

In the document retrieval module, the information collected from the question described above is used to search the text collection. A score is attributed to words and based on that documents are also scored. The 30 top ranked documents, with sentences in which pivot words occur marked, are forwarded to the sentence retrieval module. The score at word level is a combination of the:

- POS score, with NE's scored higher than common nouns, that are in turn scored higher than adverbs, with verbs being attributed a lower score,
- inverted lexical frequency, *ilf*, corresponding to the logarithm of the inverted relative frequency of the word in the collection, and
- inverted document frequency, *idf*

The score of a document involves a component that is a function of the score of the words and two components that reflect the influence of the question category(ies) and ontological domain(s) in common between the document and the question. For a given word the contribution to the document score is given through a the multiplication of the word score by a weight factor. The presence of the following words related to a pivot in a documents are considered, in decreasing order of influence:

- lemma;
- head of derivation;
- synonym

The contribution of each synonym is further quantified by a value that reflects the semantic proximity between the existing synonym and the original pivot lemma.

The sentences that contain the pivots for the selected documents are analysed by the passage retrieval module. Although the system has a parameter that allows a neighbouring fixed number of phrases to be analysed before and after the phrase the parameter is used as 0 for the current edition.

The external resource used by the system comprise a lexicon, a thesaurus that is part of the Priberam proofing tools for Portuguese, FLiP, *Ferramentas para a Língua Portuguesa*[16] and an ontology that is part of the Priberam as described in (Amaral et al. 2004). The lexicon was based on a partnership with Porto Editora, and has information, for each lexical unit, on its part of speech (POS) function, sense definitions, semantic features and ontological and terminological domains. The thesaurus has synonyms that are used for query expansion in the document retrieval stage. The ontology, a multilingual resource that was based on the work of the French company Synapse Développement for its search engine *Chercheur*, that was extended in the project TRUST (Text Retrieval Using Semantic Technologies) to other languages besides of French, including the Portuguese language that was treated in the Portuguese module of TRUST, that was developed by Priberam (Amaral et al. 2004). Other languages were English, Italian and Polish.

In second participation (Cassan et al. 2006) of Priberam at QA@CLEF the system architecture was kept due to the results obtained in the previous edition, in which Priberam obtained the best results among all participants. However the following main improvements have been introduced:

- Treatment of temporal restrictions;
- Validation of the extracted answers;
- Cross-language operation.

Other minor improvements were also made, such as:

- A priority system to increase the assertiveness of the question categorization module, that no longer attributes more than a question category per question, and

[16] http://www.flip.pt

- the use of non exact matching techniques based on the Leveshtein distance for proper nouns was introduced.

In terms of the treatment of temporal restriction, it includes both the temporal expressions in the text collection as well as the meta-data, i.e. the date of the newspaper the news article comes from. The meta-data was indexed to allow a boolean search to be done returning all documents between two dates. Regarding the temporal references in the text they can occur as absolute dates, incomplete absolute dates or periods. New functions were added to the system to make the comparison of dates, translation of time units and the validation if a specific date is within a certain given time interval. For a question with temporal restriction, dates are converted to a normalized format. To specific dates, a few days are added and subtracted so that the documents of the resulting period can be searched (using the meta-data). Additionally, documents that contain the original data are searched for, but incomplete dates are also considered, but are attributed lower scores. The documents obtained with both processes are joined in a set of resulting documents.

Concerning the validation of answers, it had been identified as a process that was responsible for many failures in the system, therefore a mechanism was developed to verify the sentence the answer originated from. The validation consists in the following steps: first the sentence must match (at least partially) all proper nouns and named entities in the question, and second it must contain a number of nominal and verbal pivot from the question. The matching makes use of lemmas, heads of derivation or synonyms.

The introduction of cross-language operations was done for Spanish. The systems was prepared to participate at the monolingual Spanish task, as well as both options with Portuguese and Spanish as source and target languages. The system design was kept for Spanish, including the 86 types used for question identification. The authors mention that adapting the system for Spanish did not require a lot of changes, emphasising the fact that the system is self-contained, i.e. does not depend on external software or the use of the web for the translation. The main work was related to the language resources: lexicon, thesaurus, ontology and QA patterns.

The lexicon for Spanish is reported to have been acquired, and the process of adaptation to the format used by the system took 3 months and was done by a team of four people. Some words were added, mainly proper nouns, such as toponyms and anthroponyms, to be used in the recognition of named entities.

Spanish was added to the ontology (160 000 words and expressions through their conceptual domains organised in a four level tree with 3 387 terminal nodes) in a joint effort with Synapse Développement.

Many pattern for Portuguese were also used for Spanish due to the similarity of the two languages (both Romance Languages), however there was the need to translate and revise information such as groups of semantically related words and question identifiers and contextual rules, such as the ones for morphological disambiguation and for NER.

Like in Portuguese, Spanish morphological disambiguation is done in two stages: first contextual rules are used and afterwards the remaining ambiguities are suppressed using a statistical POS tagger based on a second-order hidden Markov model (Thede & Harper 1999). For training, a corpus previously disambiguated with SVMTool (Giménez & Màrquez 2004) was used.

For this edition no thesaurus for Spanish was used.

The translation is done directly using the ontology. For Portuguese and Spanish there is a direct translation, whereas for other language pairs for which a direct translation is not possible, the translation is done using English as the intermediate language. The translation obtained from the ontology is then refined by statistical information derived from the Europarl parallel corpus (Koehn 2005). The alignment of the sentences in the corpus gives the number of times that two concepts are associated to one another, and that value is used as the score of that translation pair. An example is the Spanish word *"hijo"* that belongs to the ontological domain família/linhagem [family/lineage]. It has two possible translations to Portuguese filho [son] and criança [child], but the word filho [son] is selected because it has a higher score. The Europarl corpus was also used to enhance the translation database, with word pairs extracted based on co-occurrence in the aligned text, that were extracted using likelihood and Chi-squared criteria.

The cross-language operation is done by the analysis of the question in the source language, and after the pivots are extracted they are translated to the target language using the ontology. In case of several possibilities the one with highest score is used, and the remaining ones are used as synonyms.

After that process the QAP's are activated: while in monolingual operation there is a one to one correspondence between a QP and a QAP, in the cross-lingual operation QP's are in the source language and QAP's are in the target language, therefore it is possible that several QAp's

become active for a question.

The results obtained were very good, with Priberam achieving the first place in the competition overall, both for the Portuguese monolingual task (134 questions answered correctly in the first answer) and the Spanish monolingual task (105 correct first answers). In the cross-language tasks the results obtained were not so high, but they were also among the best.

Despite the high results achieved some points were identified for improvements, concerning the ontology and the hyperonymy, hyponymy, antonymy and synonymy relations that need to be improved for Portuguese and introduced in Spanish, the improvement of the lexical information with senses for Spanish, as well as the introduction of a Thesaurus for this Language. The necessity of a better translation of entities is demonstrated by the example of the movie name Guerra das Estrelas [Star Wars] that fails to be translated to *La guerra de las galaxias* in Spanish.

The third participation (Amaral et al. 2007) introduced syntactic processing in the system, and due to the changes in the evaluation rules, the treatment of several questions related to one topic (in clusters of up to four questions) had to be addressed, as well as the addition of the Wikipedia to the corpus.

The syntactic processing occurs at the level of the question, with QP patterns enhanced with syntactic information. The pivots are now identified according to the syntactic structure for the question, that identifies the main constituent of the question and secondary ones. For instance in the question De que estado brasileiro foi governador Adhemar de Barros? [Of which brasilian state was Adhemar de Barros governor?] the main constituent is "Adhemar de Barros".

The syntactic information was also used to improve QAP's, that now include information about each pivot's syntactic specifications, which the system tries to match with the answer pivots. The parser implemented was based in the algorithm of (Earley 1970), and it proved useful in a number of situations, for example the sentence "Jorge Sampaio (presidente de Portugal) deslocou-se em visita de estado à República Popular da China" [Jorge Sampaio (president of Portugal) went on a state visit to the People's Republic of China] can be used to answer the definition question Quem é Jorge Sampaio? [Who is Jorge Sampaio?] using the text in brackets, however this text can be ignored, and a different path can be explored to answer a question like Que país visitou Jorge Sampaio? [Which country did Jorge Sampaio visit?]. The recursive nature of the parser developed is also useful to extract closed list answers, a new feature introduced in this year evaluation rules. Since the authors opted for storing the Wikipedia text with the links,

to be able to discard them when looking for answers is also an important feature.

Regarding the Wikipedia pages, as just mentioned, the links present in a Wikipedia page are kept: they are converted form the format `[[<Article title>|<Link text>]]` to the format `<Link text>(<Article title>)` and index it as natural language text. This procedure is indicated by the authors as being responsible for the success in answering questions with acronyms. On the other hand the title of the Wikipedia page is added to the page content for sentences without a subject or a pronoun indicating an anaphoric reference, and then the phrase is parsed again. It is pointed out though that other referents may exist that are disregarded. This procedure led to answers being assessed as unsupported in the Spanish monolingual task. Only Wikipedia content pages were indexed, and the content of information contained on tables or boxes is not considered.

Other improvements were made especially to the Spanish resources of the system, with the lexical information being extended and a thesaurus added as a new resource. However, no reference on the origin of such resources is mentioned.

The results for this year participation dropped in relation to the ones obtained in the previous year, following the general trend, with participants commenting on the difficulty of the new rules. Nevertheless Priberam maintained its participation at an excellent level, with second best results in monolingual tasks overall, and best results for Portuguese and Spanish.

The fourth participation (Amaral et al. 2008; Amaral et al. 2009) aimed mainly at stabilizing the system and recover the performance levels obtained previously, before the last edition, with that aim being attained since Priberam returned to the leading position in overall results with 127 correct answers in the monolingual task. Although no specific data on performance was ever published, in the 2008 participation it is said that the response time was reduced to half.

2.4.4 Raposa [Fox]

The RAPOSA (FOX) system participated at QA@CLEF for the first time in 2006 (Sarmento 2006b).

A study on QA systems and components has been made by the authors of Esfinge and Raposa (Costa & Sarmento 2006), so it is natural that both systems use some common resources

as basis.

RAPOSA uses MySQL to store the document collection, storing snippets as each sentence in raw text. It has a Question Parser to process the question, a Snippet Extractor to extract relevant sentences, a Candidate Generator to extract answers from the sentences, and a Answer Selector to select an answer to return.

The Question Parser uses the NER system SIEMÊS to extract entities, and then uses a set of rules to identify question type and its elements.

The Snippet Extractor uses MySQL to extract relevant sentences, but only words with 4 or more letters are considered. A simple stemmer was also used, that replace the last 2 or 3 letters of a word by an asterisk, resulting in extracting many unrelated sentences.

Candidate Extractor uses SIEMÊS to extract answers, as well as a set of rules to extract answers depending on the question type, for instance for a question in the form Quem é X? [Who is X] the system looks for patterns like "... <job title> X ..." or "... X, <job title> ...". These patterns are not fully specified.

Answer Selector gives higher importance to the number of supporting snippets, making use of redundancy.

In 2007 RAPOSA (Sarmento & Oliveira 2007) improved pattern rules (72 rules, not specified) in Question Parser and in Candidate Extractor, resulting in a considerable increment in the number of correct answers. This confirms the high importance of pattern rules.

In 2008 RAPOSA (Sarmento et al. 2008) makes query expansion of verbs using a statistically made verb thesaurus, keeping the same stemmer. This allowed RAPOSA to answer 4 more answers than without verb query expansion. The global results decrease, and the reason that was identified was that pattern rules were not adequate for that year questions.

The RAPOSA uses an interesting philosophy and system architecture, but the use of MySQL as an IR system seems to imposes a limitation on the possible results of the system.

2.4.5 QA@L2F

This system participated for the first time at QA@CLEF 2007 (Mendes et al. 2007) and is based on linguistic analysis of the text collection (and the question). It was developed at L^2F INESC-

ID using the research group natural language processing chain, that consists in the morphological analyser and spell checker/corrector Palavroso (Medeiros 1995), the Morphossyntactic Ambiguity ResolVer module Marv (Ribeiro et al. 2003), Rudrico (an improved version of PAsMo (Paulo et al. 2002)), which splits tokens and recognizes simple and compound terms identifying them as single tokens, but also, and the Xerox Incremental Parser, XIP[17] based on the work described in (Aït-Mokhtar et al. 2001; Roux 1999) that returns the input organized in chunks and connected by dependency relations.The system also makes use of NER systems based in the NLP chain.

The philosophy of the system is to treat the document collection in an off-line process and to extract information to be stored in a data-base format called relation-concepts, in which the information is stored together with the supporting text snippet, to be retrieved later. However there are several reasons why the relevant information may not be stored this way, for instance the fact that the document collection was not integrally processed, or the important information was not detected, or was detected but was not converted to structured format.

In this case different strategies are used for instance keeping a database with the raw text and one with the named entities identified to be used to help find relevant snippets. Wikipedia is stored in WikiXML format in another MySQL database.

As far as answer extraction is concerned, the system first attempts to find a linguistic pattern match in the relation-concepts database. Second, linguistic reordering is used, that tries to answer with a set of patterns rules, that work better on definition questions. Third it uses named entities recognition, and look in the raw texts database for snippets with the named entity and auxiliary words in the query, identified by the NLP, and return the most frequent entity in the snippets. As a last resource the system makes a query with all the words to look for text snippets from the raw database, and the best snippets go through the NLP chain returning the most frequent concept match.

At its second participation in QA@CLEF (Coheur et al. 2008) the system maintained its architecture, with a new module for co-reference resolution being added to it.

Even with incomplete information on the structure of the relation-concepts database, or how the snippets are obtained from the raw text, we can say that the reliance upon linguistic tools is

[17] http://open.xerox.com/Services/XIPParser

not the best option for our aims, specifically to produce a robust system, capable of dealing with incorrectly formulated sentences. On the other hand, the efficiency we are trying to achieve is intended to be able to process text and find the information on-the-fly, at question time. That is because we believe that the unstructured nature of natural language text, especially if we are talking about open domain, is too rich to be kept in a structured way. The use of a database for retrieval can also present a problem in terms of efficiency.

2.4.6 Summary and Other Portuguese QA Systems

Table 2.5 presents a summary of the main characteristics of the systems described.

There are other two systems that participated at QA@CLEF in tasks involving Portuguese. We will analyse each of them briefly.

The USP (the Brazilian University of São Paulo) participated in QA@CLEF 2006 with the system (GistSumm) (Filho et al. 2006). It is not a QA system, it is a summarization system, with high precision in identifying the gist of texts (Pardo 2002; Pardo et al. 2003). The results obtained were very low, indicating that summarization techniques by themselves are not enough for a QA task.

The LCC Power Answer participated at QA@CLEF 2006 (Bowden et al. 2006) and 2007 (Bowden et al. 2007; Bowden et al. 2008), with cross-language version between several language pairs: English as source and French, Portuguese and Spanish as targets.

LCC (Language Computer Corporation), now Lymba Corporation[18], was a Texas based spin off company from SMU (Southern Methodist University), with a strong research area on QA which had as starting point a QA system developed as a PhD project. The systems from this company have consistently been among the top 5 systems of the TREC QA.

We present in Table 2.6 the most significant results for cross-lingual tasks of QA@CLEF, in the form of the number of questions of the 200 question set that each system answered correctly in the first answer, according to the overview papers (Magnini et al. 2004; Vallin et al. 2005; Magnini et al. 2006; Giampiccolo et al. 2007; Giampiccolo et al. 2008)

[18] http://lymba.com/

Table 2.5: Summary of Characteristics of Systems for Portuguese at QA@CLEF 2008

System	Collection Pre-Analysis, Storage and Retrieval	Question Treatment	Answers Extraction and Scoring	Knowledge used besides the collections	Additional tools and techniques used
Priberam	Proprietary IR system; Stop Words and Lemmatization; Syntactical Processing; Indexing includes lemmas, heads of derivation, synonyms, ontological domains, and question categories.	Question Classification (around 84 categories); Syntactic Processing.	Manually obtained patterns; Scoring through patterns; Redundancy.	Proprietary Ontology; Thesaurus for query expansion; Lexicon for Lemmatization.	Proprietary parser; Morphological disambiguation; Named Entity Recognition.
Senso	Information Retrieval: system Lucene; Stop Words and Proprietary Stemmer.	Multi Strategy: Semantic Representation; Extraction of query words for ad hoc search.	Multi Strategy: Top Documents from IR are treated in two different ways: Semantic Representation and inference mode PROLOG style; Search documents via manually obtained patterns. Answer Validation using web search.	Proprietary Ontology including Thesaurus for query expansion	Parser Palavras; Answer Validation using web search.
Esfinge	Documents stored as sentences: IMS Workbench for News Articles; MySQL for Wikipedia.	Question transformed into patterns in two ways: String processing; And using Parsing.	Three strategies: 1st Manually constructed patterns; 2nd NER SIEMÊS; 3rd N-grams model. Score: Answer frequency; Words of question in support; Candidate answer length.	Database of Word Co-Occurrences BACO.	Parser Palavras; Named Entity Recognizer SIEMÊS; Morphological Analyser JSpell; N-gram filtering Ngram Statistics Package (NSP); Search answers in web using Yahoo search API; Filters to exclude undesirable answers.
QA@L2F	MySQL News Articles; Heavy linguistic processing to build two data bases from the collection: Relation-Concept database; Named Entities database; and Raw text database. Wikipedia: Raw text data base.	Questions analysed by type and a script per type builds query for database.	Patterns; Named Entities Recognition.		L2F morphology processing chain (POS tagging, Tokenization, Disambiguation: Palavroso, Rudrico, Marv) XIP Parser (syntactic analysis). NER based in the above NLP chain.
Raposa	MySQL; stores the collection as sentences or paragraphs.	Rule based classification of question with special incidence in people questions.	NER; Context Rules or if there aren't any, Semantic Label/ Redundancy;	Database of Word Co-Occurrences BACO; Verb Thesaurus automatically built from an external large collection.	Morphological Analyser JSpell; Named Entity Recognizer SIEMÊS.

Table 2.6: Best Results of Cross-lingual Tasks at QA@CLEF (Correct First Answers in 200).

Year	Participant	# correct first answers	source language	target language
2006	Synapse	94	PT	FR
2006	Synapse	86	EN	FR
2006	Priberam	72	PT	ES
2006	Priberam	67	ES	PT
2007	LCC	81	EN	FR
2007	LCC	56	EN	PT

As a comment to these results we can say that they do not reach the level of results for the monolingual tasks but they are higher than the average results of systems for monolingual tasks. It can be seen that there is a correlation between the monolingual results, since these systems are top performer for both monolingual tasks and cross-lingual ones. These participants are all companies dedicated to the study of language and text search tools. Another detail to be pointed out is the fact that the best results involve the Portuguese language, at pair with English, French and Spanish.

Other QA Systems for Portuguese that did not participate at QA@CLEF, are XisQuê and a QA system developed by Eckhard Bick.

XisQuê (Branco et al. 2008a; Branco et al. 2008b) is a QA system with an working web page[19], that searches the web to provide answers, using the results of established search engines in the web namely Ask[20], Google, MSN Live[21] and Yahoo![22]. The system makes use of shallow linguistic processing tools, including part-of-speech annotation, morphological analysis, lemmatization or named entity recognition for the tasks of question processing, answer type detection and keyword extraction. The keywords are then submitted to the commercial search engine, and after a number of procedures to strip HTML related information, the text from the web pages returned is available for further treatment related to the extraction of answers. This step involves the segmentation of the text to sentences, its ranking based on the number of occurrences of the search keywords, followed by the extraction phase that starts with the linguistic annotation of the sentences, followed by the annotation of named entities with their

[19] http://xisque.di.fc.ul.pt
[20] http://www.ask.com/
[21] Currently this service is provided by Bing http://www.bing.com/
[22] http://search.yahoo.com/

semantic types[23] by the NER system, and finally the application of patterns for extraction of the answer. The system was not present in a formal evaluation initiative, but was informally evaluated using a set of 60 questions picked randomly from the "Trivial Pursuit" game, with results reported as similar to those obtained by Esfinge. The efficiency of the system is a concern of the authors, with a reported average time of 22 seconds per question, of which 14 seconds are used by the interface to the third part IR component. This values, according to the authors, outperform Esfinge, that takes in average 91 seconds per question.

The QA system from Bick (Bick 2003a) was presented in PROPOR 2003. It relies upon the pre-existing general constraint grammar parser PALAVRAS (Bick 2000), to which a name entity recognizer module has been added (Bick 2003b). The system was developed as a prototype to investigate if rule based linguistic analysis could provide a sound basis for a QA system, and the conclusions seem to indicate so, even though the system was never present at a formal evaluation such as QA@CLEF.

2.5 Research Directions

In the present section we describe the research directions followed to fulfil the objectives defined for the present thesis, in Section 1.2, to study and develop new components of IR and QA, to build a complete, efficient and robust QA system, that can compete with the current state-of-art QA systems.

We decided that we would build a new QA system because the motivation for their existence is there, in their three main characteristics, that are:

- Interface with the user - input: user information request in natural language;
- a large amount of text to be searched in which the information is liable to be found;
- Interface with the user - output: a list of ranked answers.

That practical forms of interfacing with the system described for a QA system constitute added value to the traditional system users nowadays for searching information namely the

[23]The semantic types used closely match those of QACLEF, viz. person, organization, location, number, measure or time.

search engines that are based in IR systems that differ from QA precisely in the HCI parts stated above, namely the input is a set of words, a query instead of a question, and the output is a ranked list of documents instead of a list the full text of documents that must be inspected by the user for the information request to be completed. The added value required to accept the different input text and for the output to be produced with a higher convenience for the user can only be achieved through the addition of more modules to the system than the traditional search component.

The form of interface with the system in natural language in the statement of the information need is a natural way of doing it that has the advantage of being an interface that the user has been using all life long, therefore fully understood. In addition it allows a precise need to be stated, that requires a precise answer (or a ranked list of answers) hence the practical side of the output too, that allows a considerable less effort than having to read a full text document in search of the answer. Despite the interest of QA systems, and the strong research effort in the filed, the current state-of-the-art of the area has many limitations that make it worthwhile investing in new QA systems to try new approaches and better results. The main questions that still have be addressed and improved for the systems to be widely used by the general public can be summarized as:

- the **coverage of questions**;
- the **coverage of text data sources obtained automatically**;
- the **reliability** of the answers;
- the **speed** of system in producing the answers.

Our objectives for the current thesis address these items cover mainly the last item, as we intended to produce answers rapidly (speed-efficiency) and we are targeting different types of data sources that may contain errors (coverage of text data sources obtained automatically-robustness) and we intend to be competitive with current state-of-the-art results which means that we expect to be able to participate in the international evaluation initiatives with good results.

These objectives point in the direction of developing a QA system, with the architecture fully designed for the task it is meant to address in mind, since a good integration of components

is a key factor for efficiency, without having to convert information between formats used by general purpose pre-existing components. Also if the we use pre-existing components, we loose the flexibility to change the components design to address our needs in terms of efficiency and may also be forced to use functionality that we do not really need, but are unable to turn-off. This is the classical view of an engineering problem, and that is associated to QA systems in (Prager 2006), and that is connected with good results and to the realization that sometimes a state-of-the-art method or component when tested on its intended function, does not result in an improvement in overall performance, when integrated into a QA system.

Bearing these considerations in mind, our approach will be to develop a new QA system, which we called IdSay. It is an open domain Question Answering system for Portuguese that was developed from scratch, with the objective of optimizing resources, so that response time could be short. Since we wanted it to be applicable to different sources of data we developed it to answer to open domain question as defined for QA@CLEF and submitted to the 2008 evaluation campaign a first version of the system, fulfilling the need to validate it among the peer systems described in the previous section.

The version of the system that participated at QA@CLEF was the core version of the system, and it is based in Information Retrieval techniques.

This approach seemed a natural one since if one bears in mind the fact that the basis of the system is a large text collection that has to be searched. It is the basis of an Information Retrieval component, namely in the indispensable item of our present daily life: the search engine, with a different interface. It is intended to process efficiently large amounts of data and it is independent of the content. Therefore closely matches two of our requirements.

We proceed to describe the initial version of our system, starting by a short description of each module of the system, followed by the information indexing, reflecting the research motivations and directions we were faced with, and how we solved them. We indicate the chapter were each topic will be further discussed.

The core version of IdSay uses little linguistic knowledge, and is as close as possible to simple keyword search. The only external information that we use besides the text collections is lexical information for Portuguese (Alves 2002).

Our option for using little linguistic knowledge is to maintain the system less dependent of

a specific language and it rules is to be able to use the system when each of this conditions is not met. Also the language processing tools take additional time, and some times are very time consuming, so the process is slowed down.

This points to an architecture for IdSay based on the modules presented in Figure 2.2.

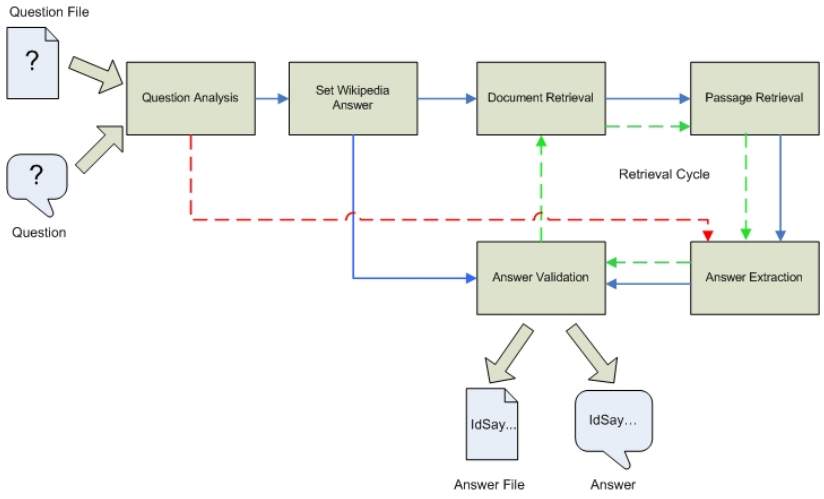


Figure 2.2: IdSay system architecture

IdSay accepts either a question written by the user (manual interface), or a set of questions in an XML file (automatic interface). In the manual interface the user can select if he wants to search for the information in the entire data collection, or in one or both the newspaper collections or just Wikipedia. It can also specify the maximum number of answers allowed, and other system related parameters, for example if lemmatization or stemming should be considered, or the detail of information to be shown by the system. The system allows for various statistics of the data to be consulted. In the manual interface, the system is not prepared to treat co-reference between questions.

Each question is analysed in the question analysis module to determine the question type and other variables to be used in the answer extraction and validation modules. The question analysis also determines a search string with the information of which words and entities to use in the document retrieval module to produce a list of documents that match both. This list of documents is then processed by the passage retrieval module, responsible for the search of passages from the documents that contain the search string, and with length up to a given limit. The passages are then sent to the answer extraction module, where short segments of text

(candidate answers) are produced, that are then passed on to the answer validation module . This module validates answers and returns the most relevant ones. If in one of the steps no data is produced, the search string is revised and the loop starts again, in a process we identify as retrieval cycle.

This is a classic QA system architecture, but contrary to most QA systems, we do not store in the IR passages, but documents, and we will extract passages in answering time. This option allow us not to fix the passage size to a one sentence, or three sentences as Esfinge, allowing to be returned passages short and long, that all have the keywords that we need.

In the pre-processing options from Portuguese QA systems we know that are systems using stop words, lemmatization, and stemming, but normally not all information is available, neither we can infer any conclusion on what is best to do. We decided to conduct a study the pre-processing options in Chapter 3, to base our decisions on what to use based on that study.

The information storage and search options for the QA systems in the previous section are the following: proprietary; MySQL; Lucene; IMS Workbench.

The use of a database as search system was an option here, this however could compromise one of the goals of the thesis, to build an efficient QA system. A query would result in an heavy process in all documents stored as strings in the database, that we cannot optimize, and the systems that used MySQL were faced to the fact it not index words with less than 4 letters.

Using an existing information retrieval system such as Lucene was also an option. Our option for building an information retrieval system is based on the study conducted in Chapter 4 in which we decide upon the retrieval model that better suits the needs of our system. Also the tests conducted in Chapter 3 related which to pre-processing options to perform on the text data were conducted using Lucene, and it was not fast enough considering our expectations for speed. These reasons led us to build a new system, IdSearch, that is fully described in Chapter 5, and that integrates with the other QA system components, and we have the option to optimize the data structure that could speed up all QA components.

IdSay's architecture is based on indexing techniques that were developed from scratch for the system. However, these techniques are general purpose IR and are not specific for Question Answering alone. The IR engine was also built with cross-language usage in mind, so we tried to develop it modularly, with the language specific information clearly separated from other generic

components. For this purpose we analyse the input text data in successive levels, building an index file for each layer, as depicted in Figure 2.3.

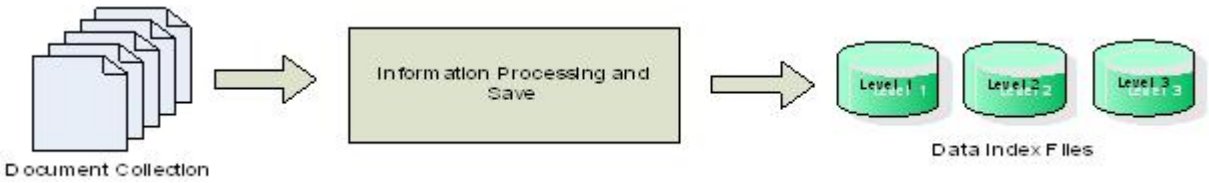


Figure 2.3: Information Indexing in IdSay

The index files for the text collection occupy 1.15 GB of disk space, and took about 4 hours to build. The load time is around 1 minute, and the time to process 200 questions is less than 1 minute, i.e. more than 3 questions are answered per second[24].

In level 1 the documents are kept as close to the original text as possible, apart from the compression techniques used. It includes also tokenization and the minimal pre-processing to allow efficient retrieval, namely separation of words with spaces and lowercase conversion.

In level 2, according to the results of our previous work (Carvalho et al. 2007) in which lemmatization and stemming were compared we opted for doing only lemmatization[25]. We will cover this topic in depth in Chapter 4. We do not remove stop words from the texts, but words maybe removed at a later stage, during the retrieval cycle.

Since this process is used to increase the retrieval efficiency in finding relevant documents, and having in mind the fact that we may need the original information (or as close as what we can have, that is to say level 1 index) we store more information in the level 2, but we also keep the information of level 1. That prevents us from having to go back to the original document, which helps us improve the performance in terms of response time.

Finally, level 3 corresponds to making equivalence classes based on related words at a linguistic level, and therefore it is one of the levels that is more markedly language-specific.

In this level, which we call the entity level, we find all sequences of words that occur often in the text collections, and if they occur more than a given threshold, we consider them an entity

[24]The tests were made using a machine with an AMD Athlon 64 processor (2.21 GHz), with 4GB of RAM, running Windows XP.

[25]Both options are available; when we say we use lemmatization, we are talking about the system setup for QA@CLEF.

whether it corresponds to a meaningful entity like the name of an organization, or to a common string of words. For the time being, we rely on our ranking mechanism to eliminate the second kind of entities, but we may do some further work in this area in the future. Are also considered an entity, the titles of Wikipedia pages.

Next we present a more detailed description of IdSay modules:

- Question Analysis

  This module processes the question and produces a search string. It needs to identify the category of the question (within the CLEF standard categories factoid, definition, and closed list question) and the type of answer expected. This analysis depends on finding specific patterns for Portuguese, which are normally used to formulate each kind of question. For instance, if we have a question in the form "O que é X?" [What is X?] or "Quem é X?" [Who is X?], we conclude that the category of the question is definition .

  If we are unable to determine the question category and type, we treat it as a generic question, for which we have a default procedure.

  Besides the category and type of the question, which will be used at a later stage to get answers of the correct type, this module also identifies the appropriate information that should be used to guide the search for documents related to the question, which we call reference entities, and in the example definition question "Quem é X?" [Who is X?] it would be X. These entities are also used for co-reference resolution in the case of clusters of questions.

  The question is searched for reference entities in the following manner: in a first stage, we rely on the indications of the user. Therefore if there are words capitalized, or words enclosed in single or double quotation marks or guillemots, we assume them to be entities . In a second stage, the text of the question is searched looking for entities that we have found in the text collections, and these are registered to be used in the search.

- Document Retrieval

  This module takes the words from the search string produced previously and generates a list of documents that contains all of the words in the search string, and also all the

entities.

This is done efficiently, since we have the documents indexed by words and by entities. The document list is built through the intersection of both the list of documents that contain each single word and the list of documents that contain the entities present in the question.

If the list of documents is empty, the process is repeated, removing the most frequent word from the search string. This process is identified in Fig. 1 as retrieval cycle, and intends to increase the possibility of finding the correct answer, turning down the words with higher frequency of occurrence and therefore with lower discriminative power.

The only exception for this rule is when we are looking for the definition of a concept that has a page in Wikipedia. In this case, the answer of IdSay corresponds to the first sentence in the page.

- Passage Retrieval

The aim of this module is, given a list of documents, to produce all the passages from the documents where the words we are looking for occur. The passage length should not exceed a given limit (currently, 60 words).

This is done in the following way: each document is searched once for the words of the question. Each time a word is found, its position in the document is registered. After storing this information for a newly found word, we check if all the words already have a position registered. If that is the case we check the total length of the passage by subtracting to the current position the minimum of the set of positions of all words. If the length does not exceed the limit we add this passage to the passage list.

Each passage is then adjusted, adding words to the beginning and to the end, in such a way that the passage corresponds, as much as possible, to one or more full sentences. For this purpose two punctuation marks sets are defined: the terminators which include for instance full stop (.), question mark (?) and exclamation mark (!) and the separators that include for instance the comma (,), semi-colon (;) and some words as 'e' [and] 'ou'[or].

The adjustment is made until the nearest terminator is found, both before the passage and after the passage. If a terminator in not found within a distance in words that allows the

passage length limit to be respected, then the system searches for the nearest separator, and if it also exceeds the length limit then the passage of the maximum length is considered, even if it breaks sentences in the middle.

Each document in the list is searched, with the corresponding passages (which can be zero for a given document) being added to a global passage list common to all documents.

- Answer Extraction

The input for this component of the system are passages, and from them we intend to extract answers, that is to say, we intend to eliminate the unnecessary portion of the passage, retaining no more information than what is absolutely needed to answer the question.

We analyse each passage searching for entities. Each entity found is considered a candidate answer. If no entities are found, the low frequency words are considered candidate answers.

A word is considered low frequency if its frequency is less than the double of the frequency of the least frequent word in the passage. All frequencies considered are the absolute frequencies in all the text collection.

If the question category is D (definition) the candidate answers considered are the phrases immediately before and after the word or entity for which the definition is being sought.

The extraction phase takes into account the category and type of the question if they were identified in the question analysis phase, in a way that prevents us from extracting answers that are not related to what we are looking for. For instance, if the answer type is date, time or numeric (count or measure) then the system searches the passage looking for numeric values followed by the corresponding units if they occur in the passage.

- Answer Validation

This module validates if the answers produced are the best ones given the other answers and their supporting passages. It receives the answers with the respective passages from where the answers were extracted, and orders the answers by the scoring based on frequency.

The answers that have the same supporting document, even if the passages are slightly different, are merged if they are close in the document. For instance, if the passage "... Ms.

X is the prime-minister of country Y. . . " supports two different answers, "prime-minister" and "Y", a new answer is produced joining both answers from the passage: "prime-minister of country Y".

Afterwards, the list of answers is filtered, by removing the answers with common words, with lower score. For instance, using the same example as above, if the answer list contains, in a lower position, an answer "minister", it will be removed since is contained in "prime-minister of country Y".

Finally, the support used for each question is the smallest passage associated with each answer.

II

Information Retrieval and Question Answering

Introduction to Part II

In Chapter 3 a classification of pre-processing techniques for IR in the context of QA is proposed, and a tests are conduced to allow statistically-validated conclusion to be taken regarding the advantage of using each technique. Several statistical tests are used: sign test; Wilcoxon signed rank test; T student test, and also less commonly tests: Bootstrap; NcNemar; Friedman.

In Chapter 4 the information retrieval models are analysed in some detail, and the usage of retrieval models for QA is surveyed. The chapter ends with a discussion of our choice for retrieval model, and the justification for our option to build a information retrieval from scratch.

Chapter 5 describes the proposed information retrieval system IdSearch, its data structures and algorithms. We propose a new data structure for document data in the QA context, and a new algorithm for calculating entities by frequency. It is shown the theoretical advantage for both proposals in terms of time and space complexity, when compared with alternatives.

3 Pre-Processing

3.1 Introduction

It is important to clarify what we call pre-processing. When using this concept we mean the thin layer that precedes the use of a component of a system that is prepared to treat different types of data for different purposes. The aim of that layer is to get the best results out of the module, regarding the specific type of data one wants to process.

If the architecture of a system is such that it uses a text classifier module prior to using an IR module, we do not consider the text classifier as pre-processing but another module of the system. The text classifier will probably need its own layer of text pre-processing. The architecture for such system is shown in Figure 3.1.

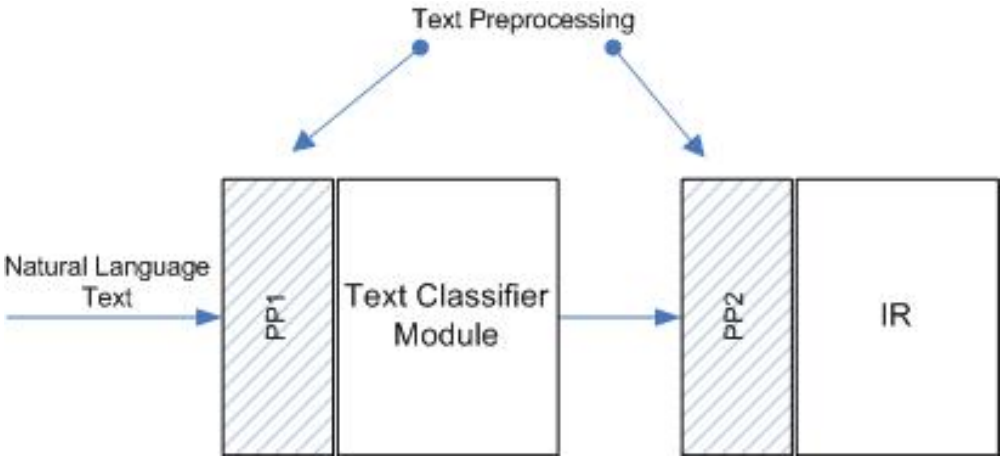


Figure 3.1: Architecture of a hypothetical system using text Pre-Processing

It would depend on system design if the two pre-processing layers would be the same or slightly different, or even if the pre-processing of the text classifier (PP1) would make it unnecessary for the IR to have its own layer (PP2). In the case of the present work, the data type is natural language text and the component we consider is the IR.

An IR system is usually prepared to treat any kind of information, regardless of its nature. Its internal organization requires the information to be organized into units that will probably occur many times in the data. These units are commonly called tokens.

If the nature of the data is known in advance, pre-processing is a way to introduce meaning to the data in the IR, so that retrieval will be easier.

Pre-processing the input data can also have the goal of saving space and processing time, so the full data after pre-processing becomes a logical representation of that data with the most representative tokens of data chosen. However, the advance in the speed of computer components and the compression techniques, and despite the very large amounts of data available for processing, we concentrate mainly in efficacy rather than efficiency.

There are mainly two approaches to text pre-processing: one is normalization and the other is the removal (or "cleaning") of elements from the original text, that we believe contain "little" information and/or introduce "noise" in the retrieval.

The process of normalization can be interpreted in terms of defining equivalence classes between different representations, and the use of one of the representations for all the occurrences of that class.

Both of these techniques can be done at a graphical level or at a conceptual level. Generally, the graphical tasks are performed first because they are useful for the conceptual tasks.

The classification we propose for pre-processing methodologies is depicted in Figure 3.2. The different lines of action we just described are on the top part of the figure, while the bottom shaded area corresponds to the particular implementation of these approaches that we consider in our work, and that we will now describe in more detail.

Starting by the left-hand side, we find the graphical tasks. These are the first pre-processing tasks and they consist of pre-processing the full text, doing case folding (we chose lower case) and removing punctuation marks, as there is little information we can derive from them isolatedly. The result is a collection of words written in lower case. They will be used as tokens to IR because our knowledge base is natural language text, and words are the core concept of a language. We have thus performed tokenization.

We now have one representation for each word, even if it occurs in different formats in the text. For instance, the words that refer to Lisbon: Lisboa, lisboa and Lisboa? are all converted

<table>
<tr><td>Removal of Elements</td><td colspan="3">Normalization</td><td>Removal of Elements</td></tr>
<tr><td>Graphical</td><td>Graphical</td><td colspan="2">Conceptual</td><td>Conceptual</td></tr>
<tr><td>Graphical Signs (Punctuation Marks)</td><td>Case Folding</td><td>Stemming</td><td>Lemmatization</td><td>Words (Stop Words)</td></tr>
</table>

Figure 3.2: Classification of Pre-Processing techniques

to lisboa.

The conceptual normalization is implemented by means of two different techniques: Stemming and Lemmatization. Both techniques aim at aggregating words that are related morphologically.

Stemming is a technique that tries to find the base (or stem) of the word by removing its affixes. It then replaces the word by its stem, thus combining the words that come from the same base. That is the case for instance of the words amável, amigo and amor ("kind", "friend" and "love", respectively) that all have for basis "am" which comes from Latin and means union and friendliness. The stem of a word, as shown in the above example, does not need to be a word itself; generally it is just one syllable long.

There are algorithmic approaches to perform stemming, based on rules related to the affixes of the language. These rules do not always produce correct results. For instance in the case of the words proteger and tecto ("to protect" and "roof") the common origin is "teg" (Latin for "cover"), but it is a case in which, through time, the letter "g" turned into a "c".

Lemmatization is a technique in which a valid word of the language (the lemma) is used as a representative for all the lexical variations that may apply. It is the headword that appears in a dictionary definition, as in the case in which andar (to walk) subsumes words as andando or andei ["walking" or "walked"].

These techniques are expected to produce good results for highly inflectional languages since they use the same representation for words with similar meaning. For instance, the sentence "A Maria vai ao Algarve" ["Mary goes to the Algarve"], might be converted to "a maria ir ao

Algarve"["Mary to go to the Algarve"] since the word vai (third person, singular, future of the verb "to go") is converted into its lemma ir (infinitive of the verb "to go").

The removal of elements from the text at a conceptual level consists of removing a set of words, called stop words, that have little information per se (like conjunctions and articles). For instance, the sentence "O João comeu a sopa" ["John ate soup"], might be converted to "joão comeu sopa" if the stop list include the words "o" and "a".

The definitions that we have made so far include some concepts that are ambiguous in the area of text processing. Although these techniques are part of almost all QA systems that use IR, questions like "What is the relevance of including a specific stop word in the stop list?" or "Should I use lemmatization?" are rarely addressed and even less quantified as far as its impact on IR performance for QA usage is concerned. In this chapter we help clarifying this sort of questions, with statistical support.

We are aware that pre-processing takes out some information that was present in the full text, since we believe that, apart from involuntary mistakes like spelling mistakes or mistyped words, everything in the text has its function, that we are losing should we remove or change it. However, we are looking for a balance as far as IR is concerned, which means we may have to make some removals or changes to enhance IR performance. In any case, these changes are temporary, since we keep the identification of the texts from the retrieval phase, and use the full text again to do further processing after the retrieval phase. In this way the loss of information that the pre-processing might have introduced, will not affect the end goal of the QA system.

In the literature, the difference between Stemming and Morphological Query Expansion for English QA was analysed in (Bilotti et al. 2004), with results favouring the use of morphological query expansion, with the use of stemming being reported as producing in a lower recall. Flores et al. (Flores et al. 2010) conducted a study using several stemming techniques for Portuguese to investigate how they affected the performance of an IR system, and concluded that accurate stemming is not critical in information retrieval, allowing for the simplification of the algorithms used. Stop word lists are frequently different from application to application, but authors rarely discriminate what they use. Comparisons of different lemmatization and stemming techniques can be found, for instance, in (Flores et al. 2010; Kanis & Skorkovská 2010), but systematized, statistically-validated evaluation of different pre-processing configurations in QA systems has not been addressed in the literature, as far as we know, and that is the subject of the present

chapter.

3.2 Test Design

3.2.1 Performance Measures

IR system performance is generally evaluated in terms of two standard measures, namely, precision and recall, since the Cranfield Tests (Cleverdon 1991) in early 1960's. These measures are based on a binary classification of relevance of documents to the query and precision is the number of relevant documents retrieved in proportion to the total number of documents retrieved by the system, while recall gives the number of relevant documents retrieved in proportion to the total number of relevant documents to the query. We will use specific measures defined for evaluating IR performance for QA usage: coverage and redundancy. If we consider the first n documents of the hits list, coverage indicates the probability of having a relevant document among those n documents, whether redundancy indicates the number of relevant documents in those n documents (Roberts & Gaizauskas 2004).

These measures are preferable in QA because in QA the IR system is used to find a number (n) of documents that may contain the answer. Those documents must be processed to check if they contain the answer and in the positive case, build the answer. This methodology fails if the IR system does not return any relevant document in the first n documents. So we are interested in knowing if a relevant document is among those n first hits, and that is what coverage represents, the probability that a relevant document is processed.

If we take into account the following definitions:

- Considering a set of questions, $\mathbf{Q}$, with cardinality Q, a document collection $\mathbf{D}$, let $A_{D,q}$ represent the set of documents that contain a valid answer for a given q belonging to $\mathbf{Q}$.
- For an IR system S, let $R^S_{D,q,n}$ be the n top-ranked documents retrieved by the system for question q.

The coverage measure can be expressed as follows:

$$coverage^S(\mathbf{Q}, \mathbf{D}, n) \equiv \frac{|\{q \in \mathbf{Q} | R^S_{D,q,n} \cap A_{D,q} \neq \emptyset\}|}{Q}. \tag{3.1}$$

As mentioned earlier, the coverage measure for an IR system represents the percentage of questions for which at least one document containing an answer is retrieved among the list of n top-ranked documents. We use this measure because it captures well the importance of the document retrieval component of a QA system, and how in a pipelined architecture, it is dependent on having at least a document containing an answer to be processed by the subsequent modules. Another reason to use this measure in our work is that it does not require complete classification of the data collection, unlike the traditional IR measures of precision and redundancy, and the other measure defined in (Roberts & Gaizauskas 2004), redundancy of a IR system in the QA context, that represents the average number, per question, of documents within the top n ranks retrieved which contain a correct answer. If a measures requires complete information, in the case of question answering that is quite hard thing to obtain, since it is not only necessary to make a binary decision if the document is relevant or not for a topic, but to look inside the document considering a specific question and search for a valid answer in all the document, that is adequately supported according to the rules of the evaluation. This is a problem that affects not only the languages with less resources, as Portuguese, but all languages. Since the amount of information available is consistently growing, the problem is going to be tackled with, otherwise the gap between the test collections and the reality will widen. This problem has been addressed, in the context of IR, by Buckley and Voorhees (Buckley & Voorhees 2004), suggesting precisely as way to overcome it, the introduction of new performance measures.

3.2.2 Tests

We conducted a series of experiments to test the different text pre-processing techniques described in Section 3.1.

In all the experiments, the same pre-processing used for the text collection is applied to the question, and the result is used to query the IR. We then search the hit list returned by the IR for the reference of one of the documents that contains the answer.

We conducted nine tests covering different pre-processing options. Figure 3.3 presents the techniques used in each test. It is an extension of Figure 3.2, where a line in light grey was added with specific implementations of the concepts of the dark grey line. A line was also added for each test, marking the technique(s) used and the number that appears on the bottom left-hand side of the cell indicates the order in which the different techniques were applied (1 - first to 3 -

third).

		Removal of Elements		Normalization			Removal of Elements		
		Graphical		Graphical	Conceptual		Conceptual		
		Graphical Signs (Punctuation Marks)		Case Folding	Stemming	Lemmatization	Words (Stop Words)		
		maintain hyphen	remove hyphen	lowercase	Porter Stemmer	POLLUX	SL1	SL2	SL3
Phase 1	Test0								
	Test1			✓ 1					
	Test2	✓ 2		✓ 1					
	Test3		✓ 2	✓ 1					
Phase 2	Test4		✓ 2	✓ 1			✓3		
	Test5		✓ 2	✓ 1				✓3	
	Test6		✓ 2	✓ 1					✓3
Phase 3	Test7		✓ 2	✓ 1		✓3			
	Test8		✓ 2	✓ 1	✓3				

Figure 3.3: Summary of Tests

The tests were divided in three phases that we describe in the following subsections.

3.2.2.1 Phase 1 - Basic Pre-processing

In this phase, the techniques that work at graphical level are tested. Test0 corresponds to the full text, without any kind of processing, to establish a baseline to compare when introducing pre-processing. We proceed to Test1 in which only case folding was done (turning all letters to lower case). The tests related to the removal of punctuation marks were divided into two different situations: Test2, where the hyphen was the only punctuation mark that was kept, and Test3, where the hyphen was removed along with the rest of the punctuation marks.

We gave special attention to the treatment of the hyphen for two reasons:

1. The use of the hyphen in composite words like co-orientador [co-advisor].

2. The use of the hyphen in Portuguese in the enclitic pronouns like in "Ela disse-me que ..." ["She told me that ..."] and in the mesoclitic pronouns like "Ela dir-me-ia se ..." ["She would tell me if ..."].

We also treat unknown characters as word delimiters. For instance information like e-mail addresses or URLs are split up.

The parametrization of this phase that conducts to best results will be used in subsequent phases. As will be shown in the next section, it corresponds to that of Test3.

3.2.2.2 Phase 2 - Stop Lists

We have used several instances of stop lists for Portuguese. One, SL1, is composed by the 100 most frequent words in the corpus, and is published by Linguateca. It is presented in Figure 3.4.

a	à	agora	ainda	ano	anos	ao	aos	apenas	as
às	até	bem	cento	com	como	contos	contra	da	das
de	depois	dia	do	dois	dos	durante	e	é	em
entre	era	esta	está	estado	este	fazer	foi	foram	governo
grande	há	hoje	isso	já	lisboa	mais	mas	mesmo	mil
milhões	muito	na	nacional	não	nas	no	nos	num	numa
o	onde	ontem	os	ou	outros	país	para	parte	pela
pelo	pode	por	porque	portugal	presidente	público	quando	que	quem
são	se	segundo	sem	ser	seu	seus	só	sobre	sua
também	tem	ter	todos	três	tudo	um	uma	vai	vez

Figure 3.4: Stop List SL1

SL1 has many words specific to the corpus, i.e. commonly found in the newspaper context.

Examples of this kind of words are:

- Lisboa - Lisbon
- nacional - national
- país - country
- Portugal

- presidente - president
- Público - (the name of the newspaper).

Another list of Stop Words that we used, SL2 (Figure 3.5), is published by the University of Neuchâtel and is the Portuguese version of the procedure described in (Fox 1989). This list is composed of 356 words.

a	à	adeus	agora	aí	ainda	além	algo	algumas	alguns
ali	ano	anos	antes	ao	aos	apenas	apoio	após	aquela
aquelas	aquele	aqueles	aqui	aquilo	área	as	às	assim	até
atrás	através	baixo	bastante	bem	bom	breve	cá	cada	catorze
cedo	cento	certamente	certeza	cima	cinco	coisa	com	como	conselho
contra	custa	da	dá	dão	daquela	daquele	dar	das	de
debaixo	demais	dentro	depois	desde	dessa	desse	desta	deste	deve
deverá	dez	dezanove	dezasseis	dezassete	dezoito	dia	diante	diz	dizem
dizer	do	dois	dos	doze	duas	dúvida	e	é	ela
elas	ele	eles	em	embora	entre	era	és	essa	essas
esse	esses	esta	está	estar	estas	estás	estava	este	estes
esteve	estive	estivemos	estiveram	estiveste	estivestes	estou	eu	exemplo	faço
falta	favor	faz	fazeis	fazem	fazemos	fazer	fazes	fez	fim
final	foi	fomos	for	foram	forma	foste	fostes	fui	geral
grande	grandes	grupo	há	hoje	horas	isso	isto	já	lá
lado	local	logo	longe	lugar	maior	maioria	mais	mal	mas
máximo	me	meio	menor	menos	mês	meses	meu	meus	mil
minha	minhas	momento	muito	muitos	na	nada	não	naquela	naquele
nas	nem	nenhuma	nessa	nesse	nesta	neste	nível	no	noite
nome	nos	nós	nossa	nossas	nosso	nossos	nova	nove	novo
novos	num	numa	número	nunca	o	obra	obrigada	obrigado	oitava
oitavo	oito	onde	ontem	onze	os	ou	outra	outras	outro
outros	para	parece	parte	partir	pela	pelas	pelo	pelos	perto
pode	pôde	podem	poder	põe	põem	ponto	pontos	por	porque
porquê	posição	possível	possivelmente	posso	pouca	pouco	primeira	primeiro	próprio
próximo	puderam	qual	quando	quanto	quarta	quarto	quatro	que	quê
quem	quer	quero	questão	quinta	quinto	quinze	relação	sabe	são
se	segunda	segundo	sei	seis	sem	sempre	ser	seria	sete
sétima	sétimo	seu	seus	sexta	sexto	sim	sistema	sob	sobre
sois	somos	sou	sua	suas	tal	talvez	também	tanto	tão
tarde	te	tem	têm	temos	tendes	tenho	tens	ter	terceira
terceiro	teu	teus	teve	tive	tivemos	tiveram	tiveste	tivestes	toda
todas	todo	todos	trabalho	três	treze	tu	tua	tuas	tudo
um	uma	umas	uns	vai	vais	vão	vários	vem	vêm
vens	ver	vez	vezes	viagem	vindo	vinte	você	vocês	vos
vós	vossa	vossas	vosso	vossos	zero				

Figure 3.5: Stop List SL2

List SL2 contains almost all the word from SL1 (apart from some of the examples above), and SL3 is a subset of both lists SL1 and SL2.

We have automatically built a stop list, Stop list SL3, that consists of the words that are in at least 75% of the documents of the collection. This list contains 22 words, and is shown if Figure 3.6, where the word is followed by the percentage of documents it which it occurs. The idea behind this list is that a word that belongs to practically all documents, does not contribute to make a distinction between them, so they belong to the class of "little" information.

de	99%	a	98%	o	98%	e	96%	da	95%
que	94%	do	94%	em	92%	os	88%	um	88%
se	88%	no	87%	para	87%	com	86%	na	86%
uma	85%	por	83%	dos	81%	as	79%	ao	78%
à	76%	não	76%						

Figure 3.6: Stop List SL3

3.2.2.3 Phase 3 - Stemming and Lemmatization

Stemming consists of automatically shortening the word down to its stem, based on a set of rules, while Lemmatization replaces a word by its linguistic lemma (also a word), and therefore requires linguistic knowledge. In the literature sometimes Lemmatization is named Dictionary based Stemming.

The lexical knowledge came from the POLLUX system (POrtuguese Lexical Largely Usable and eXtensible) (Alves 2002). This database has a table with 925,275 Portuguese lexical items, including inflected ones. Based on this information, a text file with the words and their lemma was built. This file is loaded into memory to be consulted in run time. If a word does not belong to the list, it is maintained; otherwise it is replaced by its lemma. POLLUX was a re-engineering project of the POLARIS system (Lopes et al. 1994).

The stemming algorithm follows Martin Porter's approach. The implementation of the Neuchâtel University was used. This approach consists of successive steps of word reductions like removal of suffixes, normalization of gender and removal of accented characters.

3.2.3 Working Environment

In our experiments, we focus on domain-independent QA for Portuguese.

The text collection used is made available by Linguateca, and the texts belong to the knowledge base of the Question Answering task of the Cross-Language Evaluation Forum (QA@CLEF) for Portuguese. This collection consists of news articles from the Portuguese daily newspaper Público, from Lisbon, for the years of 1994 and 1995. The edition of a given day is divided into news articles, to which a unique identification is assigned. In our case, a document for the IR system corresponds to a news article. The total number of documents is 106,821. The questions used are from the year 2004 evaluation campaign and they total 180. We use questions from this year because they are the only ones that have the information about the relevant documents, which allows automatic calculation of the coverage measure.

In our experiments we use CLucene, the C++ version of the open source IR API of Apache Lucene. This IR system is commonly used in QA systems, with satisfactory performances (Tellex et al. 2003).

The index size for the experiments described in Figure 3.3 is presented in Figure 3.7. In the second column, the number of terms are distinct terms in the collection. By removing punctuation marks (Test2 to Test8) the number of distinct terms drop to less than half. This can be explained by the fact that punctuation marks appear together with words, without space, which is accounted as a new distinct word. In terms of disk space, the saving obtained by the different techniques used, were quite modest when compared with the reduction in the number of distinct terms.

Index ID	Number of Terms	Disk Space (MB)
Indice0	1,473,716	508
Indice1	1,353,817	504
Indice2	622,783	478
Indice3	538,116	477
Indice4	538,039	363
Indice5	537,821	432
Indice6	538,101	401
Indice7	457,117	461
Indice8	454,287	430

Figure 3.7: Index Sizes

3.3 Results

The results for the coverage measure are presented in Figure 3.8. The different columns indicate several values for the cut-off of the hit list, so the search for documents that answered the question would be limited to the documents until that rank.

The values of 10, 20, 50, 100 and 1,000 were used, and, naturally coverage increases for higher cut-off values.

		Cutoff value				
		10	20	50	100	1000
Phase 1	Test0	8,9%	12,8%	21,1%	26,7%	45,0%
	Test1	31,1%	35,0%	44,4%	50,0%	69,4%
	Test2	36,7%	46,1%	54,4%	61,7%	76,7%
	Test3	38,9%	47,8%	55,6%	63,3%	79,4%
Phase 2	Test4	41,1%	50,0%	58,3%	64,4%	80,6%
	Test5	40,0%	50,0%	57,2%	66,7%	80,6%
	Test6	42,2%	47,2%	56,7%	63,9%	80,0%
Phase 3	Test7	37,2%	43,3%	54,4%	62,2%	81,7%
	Test8	38,3%	44,4%	53,3%	61,7%	79,4%

Figure 3.8: Coverage for the different Pre-Processing tests

Appendix A contains more detailed information, namely the 180 questions used and the details per question of the results presented in Figure 3.8.

The information regarding which documents contain an answer to the question is limited to only one document reference in 98% of the cases. We have manually processed a number of questions, and we have found numerous other documents that contain the correct answer, and they usually score higher than the ones indicated. We believe that this is the main reason why the coverage of our IR system is not better. We intend to improve this information (for instance by searching for the answers instead of the questions) so that the information is more comprehensive in terms of references of documents where answers can be found. It will also allow us to calculate the redundancy of the system, which will be useful to determine at what rank on the hit list the cut off must be done. We also intend to increase the number of questions used.

3.3.1 Statistical analysis

Given the results of Figure 3.8, we decided to make a statistical analysis to determine which results were sufficiently supported by the experimental evidence to allow a sound decision to be made when determining which pre-processing options to choose, or on the contrary which results could have been due to mere chance.

That is the aim of confirmatorial statistics, and conclusions about statistical significance of observed data can be made by means of Statistical Hypothesis Testing or through the Calculation of Confidence Intervals.

In our case we are trying to access which pre-processing options applied to a reference text document collection lead to better results when used by an IR system for a reference set of questions. That is a case of matched data in which different methods of pre-processing (the experimental tests T0 to T8) are being applied to the same testing environment (text collection and set of questions). The matched data in this case are the results per question obtained by the different pre-processing methods. Our statistical tests are based in the comparison of the results of coverage for each of the corresponding 180 questions in our question set that are obtained for different pre-processing methods, i.e. the experimental tests T0 to T8.

Now that we established our working context, we will give some more detail on the statistical analysis we will do: the basic assumption of statistical tests is to consider a null hypothesis according to which the results of the different methods being compared were obtained by chance, which is the same as to say that the methods are the same, or indistinguishable in our experimental setup. The alternative hypothesis is that the methods are not the same, or at least one differs from the others and we will try to understand from the data if we have evidence that allows the rejection of the null hypothesis, keeping the alternative hypothesis.

The criterium that will allow us to make that decision is the **p-value**, which is the probability that reflects the measure of evidence against the null hypothesis, and is calculated by statistical methods. A small p-value correspond to strong evidence against the null hypothesis and for a test we define the value with which the p-value is to be compared to decide the rejection of the null hypothesis. That value, α, is the significance level, and it is the maximum p-value for which the null hypothesis is rejected, and the alternative hypothesis is accepted. We will use in our tests the value 0,05, that is often used. We are considering a bi-caudal or two tailed p-value,

since the alternative hypothesis says the methods being compared are different, as opposed to assume and test a better or worse situation, that lead to one sided p-values.

In the case of the calculation of a Confidence Interval (CI), we are able to make the decision whether methods differ or not, but in this case based on numeric values that reflect information on the data being analysed. The confidence interval is a range of values, again calculated by statistical methods, in which the value of the parameter being investigated lies, with a probability called confidence probability or confidence level, represented as (1 -α). A confidence level of 95% is usally selected. If we have an interval built at the 95% confidence value, this means that in 95% of the cases the value of the variable of interest should be within the range of the confidence interval.

Conclusions can be taken about statistical significance by means of the CI, if we define the "no effect" or "zero effect" of the statistical function being used. For instance if we the test statistic is "equal means" the "no effect" value in this case is 0. The results are statistically significant if the confidence interval does not include the "no effect" value. If this decision is taken based on a CI of 95%, it corresponds to the situation of hypothesis testing with significance level of 0.05 with the null hypothesis being rejected for a p-value less than 0.05.

In our analysis we present results of statistical tests for the p-value calculated at a significance level of 0.05 and values for the limits of the confidence interval at a 95% confidence level, with the value "-" indicating either a situation that does not apply or a situation in which the value was not calculated.

We have planned our experimental tests in three phases: Phase 1 that corresponds to basic pre-processing, Phase 2 in which the use of three stop lists are being investigated, and Phase 3 where we investigate the usefulness of Lemmatization and Stemming. These phases are performed sequentially, as shown in Figure 3.8, and we will divide the statistical analysis by Phase. We will centre our analysis in three cut-off values, namely 10, 100 and 1000.

We will now give some details on the statistical tests we chose for our analysis. We are evaluating the performance of an IR system, so we base our decisions on previous work by researchers in this area. However there are some extra facts that we take into consideration, one being the fact that when we are comparing more than two methods we prefer to use a statistical test prepared for that situation rather than perform a lot of pairwise comparisons, because we can account all the information at the same time and have a more accurate measure

of the real p-value than the multiple pairwise comparisons. The other fact is the specificity of our performance measure, coverage, that, as explained before is a binary value per each of the 180 questions.

Given these considerations, and starting by the specific ones, for the statistical test for multiple comparisons, we had several options such as ANOVA or the Friedman test. We decided for the non-parametric option, since it has fewer assumptions, and used the Friedman test. For matched pairs in IR, the tests that we have seen more commonly used in the literature are the Sign test, the Wicoxon Signed Rank test and the Student's T-test. In 1968 Salton (Salton & Lesk 1968) reported the incorporation in SMART of two separate testing procedures, T-test and Sign test, explaining the normality requirements of the data and independence among search requests by the t-test, and emphasising the less demanding requirements of the sign test, especially the fact that it does not require the data to follow a normal distribution. A few years later the Wilcoxon Signed Rank test was added to SMART. These three tests have become over the years references in the area to this day (Croft et al. 2010; Büttcher et al. 2010) so we used them to test our data. The Wicoxon Signed Rank test is not particularly design for our type of data, since we have binary data, and the magnitude of differences will be equal, but we can still perform the test under these conditions. Surveys on statistical tests applied to IR can be found, for instance, in (Hull 1993; Smucker et al. 2007).

We added to these three tests the McNemar test, in which our binary data will be compacted as a 2x2 matrix (Box et al. 2005).

Another approach to decide upon the statistical confidence we used, was the bootstrap method (Efron & Tibshirani 1993; Davison & Hinkley 1997) for the calculation of the confidence intervals. It is a distribution free resampling method that allows the statistical measure to be chosen. The idea behind bootstrapping is to draw a large number of re samples from a single sample. The resampling is done with replacement and the sample size of the re sample is the same as the original sample size. The idea is to calculate the statistic for each re sample, and the bootstrap distribution based on the re samples. Then the principle is to assume the bootstrap distribution of the statistic based on the re samples represents the distribution of the statistic, as would be obtained by many samples. Under certain regularity conditions the variance of the estimator obtained as described is inversely proportional to $R - 1$, with R being the number of re samples, or re sample factor, which allows the precision of the estimate to be increased by

increasing the number of re samples. We used a non parametric bootstrap with the mean of differences statistic to generate confidence intervals.

All values presented were obtained by freely available implementations in the R language[1], with the corresponding functions presented are in Table 3.1. We present in the same table the abbreviated name (column STest) that will be used in the subsequent tables of results for each statistical test.

Table 3.1: Statistical Functions used in R

Statistical Test	STest	R function
McNemar	McN	mcnemar.test
Sign Test	sign	binom.test
Wilcoxon Test	wilcoxon	wilcox.test
Student T-test	t-test	t.test
Bootstrap Resampling	BS	boot
Bootstrap Confidence Interval		boot.ci
Friedman	-	friedman.test

For the bootstrap tests we used the boot package (Canty 2002), that is the R implementation of the package originally written as an S-Plus library that was released in conjunction with Davison and Hinkley (Davison & Hinkley 1997). We used the resampling function boot, using the random number generator of Mersenne-Twister. In our tests, we used a resampling factor of 9999, with the samples are of size 180, resampling with replacement from the original data, and for each sample the chosen statistic measure, the mean of differences statistic in our case, is calculated. We present the results for two different types of confidence intervals for bootstrap of function boot.ci, normal, and BCa, that are identified in the results tables as BS1 and BS2, respectively.

Phase 1 - The statistical tests we will do start in Phase 1, and since it has four tests (T0-T3) we started by a performing a statistical test for multiple comparisons, the Friedman test applied only to phase 1 data. The results are presented in Table 3.2, and the first line corresponds to the Friedman test for experimental tests T0-T3, and the second for tests T1-T3. The very low p-value in the first case is to be expected since values for T0 are extremely different from all the others, and the second line of the table tells us that it is worth looking for differences

[1] http://www.R-project.org/

between the remaining tests T1 to T3.

Table 3.2: Friedman Test for the Experimental Tests of Phase 1

	Cut-off **10**	Cut-off **100**	Cut-off **1000**
	p-value	p-value	p-value
All Tests: T0 to T3	$\ll$ 2.2e-16	$\ll$ 2.2e-16	$\ll$ 2.2e-16
Three Tests: T1 to T3	8.326e-04	6.503e-08	1.278e-08

We will proceed then to pair wise comparisons, investigating the following questions:

- Experimental Tests Pair T1-T0 - Is it better to do lowercase conversion or to maintain the original text?
- Experimental Tests Pair T2-T1 - Is it better to remove all punctuation marks **except the hyphen** after lowercase conversion, or leave the original punctuation?
- Experimental Tests Pair T3-T1 - Is it better to remove all punctuation marks after lowercase conversion or leave the original punctuation?
- Experimental Tests Pair T3-T2 - After lowercase conversion is it better to remove all punctuation marks or leave the hyphen?

The results are presented in Table 3.3, and will be discussed later in this chapter.

Phase 2 - Again we start by doing the Friedman test for the tests in this phase, T4-T5-T6 that correspond to different stop lists.

The results, presented in Table 3.4, indicate that there is no statistical evidence in different results by the use of any of the three stop-lists tested.

We will investigate if there is any difference in using stop-lists at all, when compared with the situation of experimental test T3. The statistical tests used are the same than the pairwise tests done in phase 1, with the results being presented in Table 3.5

Phase 3 - In phase three we investigate the benefits of using Lemmatization or Stemming after lowercase conversion and removal of punctuation marks, which correspond respectively to Experimental Tests Pair T7 - T3 and Experimental Tests Pair T8 - T3, whose results are presented in Table 3.6.

Table 3.3: Statistical Analysis for Experimental Tests Phase 1

STest	Cut-off **10**			Cut-off **100**			Cut-off **1000**		
	p-value	CI min	CI max	p-value	CI min	CI max	p-value	CI min	CI max
Pair T1 - T0									
McN	6.737e-10	-	-	5.919e-10	-	-	2.141e-10	-	-
sign	1.955e-11	0.8743	0.9994	3.075e-11	0.8516	0.9947	8.363e-12	0.8574	0.9949
wilcoxon	7.05e-10	-	-	6.16e-10	-	-	2.225e-10	-	-
t-test	7.251e-11	0.1590	0.2855	6.163e-11	0.1671	0.2995	1.701e-11	0.1773	0.3115
BS1	-	0.1602	0.2835	-	0.1670	0.2996	-	0.1768	0.3111
BS2	-	0.1556	0.2778	-	0.1667	0.2944	-	0.1778	0.3056
Pair T2 - T1									
McN	0.01842	-	-	2.669e-05	-	-	0.001616	-	-
sign	0.03088	0.5236	0.9359	1.943e-05	0.7397	0.9902	0.00235	0.6355	0.9854
wilcoxon	0.01969	-	-	2.856e-05	-	-	0.001772	-	-
t-test	0.018	0.009640	0.1015	1.78e-05	0.06446	0.1689	0.00145	0.02816	0.1162
BS1	-	0.0096	0.1012	-	0.0643	0.1689	-	0.0287	0.1159
BS2	-	0.0056	0.1000	-	0.0611	0.1667	-	0.0278	0.1167
Pair T3 - T1									
McN	0.002838	-	-	1.177e-05	-	-	2.386e-4	-	-
sign	0.004344	0.5972	0.9481	8.43e-06	0.7347	0.9789	2.772e-4	0.6763	0.9734
wilcoxon	0.003014	-	-	1.242e-05	-	-	2.543e-4	-	-
t-test	0.002611	0.02751	0.1280	7.218e-06	0.07642	0.1902	1.911e-4	0.0482	0.1517
BS1	-	0.0279	0.1276	-	0.0776	0.1895	-	0.0488	0.1514
BS2	-	0.0278	0.1222	-	0.0778	0.1889	-	0.0500	0.1500
Pair T3 - T2									
McN	0.0455	-	-	0.2568	-	-	0.09558	-	-
sign	0.125	0.3976	1	0.4531	0.2904	0.9633	0.1797	0.3999	0.9718
wilcoxon	0.07186	-	-	0.2986	-	-	0.1096	-	-
t-test	0.04519	0.0004811	0.04396	0.258	-0.01232	0.04565	0.09568	-0.004946	0.06050
BS1	-	0.0007	0.0440	-	-0.0116	0.0455	-	-0.0046	0.0601
BS2	-	0.0056	0.0444	-	-0.0154	0.0444	-	-0.0056	0.0611

3.3.2 Discussion

Taking into account the results of the statistical tests performed, we will now discuss the main conclusions of each phase. The tests were done for 3 cut-off values but results are generally the same across cut-offs, so we will draw general conclusions, referring to specific cut-offs only when discrepancies occur.

Phase 1 - Test pair T1-T0: the results clearly indicate that all tests reject the null hypothesis with a confidence level of 95%. This means that converting all letters to lowercase is indeed good for improving the document retrieval efficiency. Test pair T2-T1: in this test we wanted to check the difference between doing only lowercase conversion or additionally remove all punctuation marks from the text, keeping the hyphen. Again, all tests done reject the null hypothesis with a confidence level of 95%. The p-value of the sign test is close to 0.05, the

Table 3.4: Friedman Test for the Experimental Tests of Phase 2

	Cut-off **10**	Cut-off **100**	Cut-off **1000**
	p-value	p-value	p-value
All Tests: T4 to T6	0.4724	0.1738	0.7788

Table 3.5: Statistical Analysis for Experimental Tests Phase 2

STest	Cut-off **10**			Cut-off **100**			Cut-off **1000**		
	p-value	CI min	CI max	p-value	CI min	CI max	p-value	CI min	CI max
Pair T4 - T3									
McN	0.2850	-	-	0.4795	-	-	0.3173	-	-
sign	0.4240	0.3514	0.8724	0.7266	0.2449	0.9148	0.625	0.1941	0.9937
wilcoxon	0.3014	-	-	0.5297	-	-	0.4237	-	-
t-test	0.2863	-0.01878	0.06322	0.481	-0.01994	0.04216	0.3187	-0.0108	0.0330
BS1	-	-0.0179	0.0631	-	-0.0196	0.0417	-	-0.0103	0.0330
BS2	-	-0.0222	0.0611	-	-0.0222	0.0389	-	-0.0111	0.0278
Pair T5 - T3									
McN	0.593	-	-	0.08326	-	-	0.3173	-	-
sign	0.7905	0.2886	0.8234	0.146	0.4281	0.9451	0.625	0.1941	0.9937
wilcoxon	0.6179	-	-	0.09148	-	-	0.4237	-	-
t-test	0.5944	-0.02999	0.05221	0.08326	-0.004430	0.07110	0.3187	-0.0108	0.0330
BS1	-	-0.0297	0.0517	-	-0.0043	0.0709	-	-0.0106	0.0332
BS2	-	-0.0333	0.0444	-	-0.0056	0.0667	-	-0.0111	0.0278
Pair T6 - T3									
McN	0.01431	-	-	0.7055	-	-	0.5637	-	-
sign	0.03125	0.5407	1	1	0.1841	0.9010	1	0.0942	0.9915
wilcoxon	0.01966	-	-	0.7768	-	-	0.7728	-	-
t-test	0.01389	0.006858	0.05981	0.7066	-0.02352	0.03462	0.5652	-0.0134	0.0245
BS1	-	0.0069	0.0597	-	-0.0234	0.0342	-	-0.0132	0.0244
BS2	-	0.0111	0.0611	-	-0.0278	0.0278	-	-0.0167	0.0222

limit to reject the null hypothesis, as opposed to the other tests. That is understandable, since the signed test is the weakest as it only takes into account the number of successes or failures, discarding the ties. A note to the p-values for cut-off 100, that register smaller values than the other two cases, but all in agreement. We conclude that removing all punctuation marks execept the hyphen helps to improve document retrieval. Test pair T3-T1: in this test we check if it is better to remove all punctuation marks, including the hyphen after lowercase conversion or leave the original punctuation. All tests reject the null hypothesis, with a confidence level of 95%. We conclude that is better to remove all punctuation marks including the hyphen. Test pair T3-T2: in this test we check if, when removing punctuation marks after lowercase conversion, it is better to keep the hyphen or not. It turns out that the tests could not reject the null hypothesis, except for the cut-off 10, and for a marginal difference for the McNemar test, t-test and bootstrap tests.

Table 3.6: Statistical Analysis for Experimental Tests Phase 3

STest	Cut-off **10**			Cut-off **100**			Cut-off **1000**		
	p-value	CI min	CI max	p-value	CI min	CI max	p-value	CI min	CI max
Pair T7 - T3									
McN	0.2568	-	-	0.5271	-	-	0.1573	-	-
sign	0.4531	0.03669	0.7095	0.7539	0.1216	0.7376	0.2891	0.3491	0.9681
wilcoxon	0.2986	-	-	0.5653	-	-	0.1817	-	-
t-test	0.258	-0.04565	0.01232	0.5286	-0.04584	0.02361	0.1579	-0.0086	0.0531
BS1	-	-0.0455	0.0117	-	-0.0454	0.0232	-	-0.0087	0.0527
BS2	-	-0.0556	0.0056	-	-0.0556	0.0167	-	-0.0111	0.0500
Pair T8 - T3									
McN	0.7055	-	-	0.3173	-	-	1	-	-
sign	1	0.09899	0.8159	0.5078	0.07485	0.7007	1	0.1870	0.8129
wilcoxon	0.7768	-	-	0.3506	-	-	1	-	-
t-test	0.7066	-0.03463	0.02352	0.3187	-0.04956	0.01622	1	-0.0347	0.0347
BS1	-	-0.0349	0.0234	-	-0.0492	0.0162	-	-0.0345	0.0338
BS2	-	-0.0444	0.0167	-	-0.0556	0.0111	-	-0.0389	0.0278

These later cases favour the use of T3 configuration. It is not a fully conclusive test, except for C@10, but we opt for the situation of test T3 removing all punctuation marks, which is also the simpler solution, applying to the hyphen the same treatment as to other punctuation marks. This was the solution adopted as the result of phase 1, and it will serve as the basis of comparison for the other phases.

Phase 2 - In this phase we checked if the use of stop word lists makes a difference for document retrieval, when compared with the situation of experimental test T3. For the cases of stop lists SL1 and SL2, corresponding to experimental test pairs T4-T3 and T5-T3 respectively, none of the tests done could reject the null hypothesis with a confidence level of 95%, so with this data we cannot confirm that any of the the stop word lists SL1 and SL2 are good to use. In the case of stop list SL3, the one we created based on the text collection, we have a slightly different situation since, for cut-off 10 the null hypothesis is rejected, while in the other two cut-offs the tests did not find enough evidence to reject the null hypothesis. We conclude that the increase of C@10 from 38,9% to 42,2% when the words of SL3 list are removed from the text, is statistically significant. For the other cut-off values tested we can not confirm statistically the benefit in retrieval efficiency using stop list SL3.

Phase 3 - In this phase we investigated the influence of the use of lemmatization (test pair T3-T7) and stemming (test pair T3-T8) in retrieval efficacy. The statistical test results all have the same result: the null hypothesis cannot be rejected at a confidence level of 95%, which

means that we cannot confirm an advantage in the use of either lemmatization or stemming for document retrieval.

As general conclusions we can say that it is worth doing lowercase conversion and removal of punctuation marks as pre-processing to increase retrieval efficacy. The case of the hyphen was investigated, but no evidence was found that it should be treated differently from the other punctuation marks.

The results regarding stop word removal were surprising, since the advantage of using stop words is often taken for granted, but especially for high cut-off values, this advantage cannot be easily verified. The same happened as far as lemmatization and stemming is concerned.

In terms of statistical significance testing, in the tests conducted during the present work, at the confidence level of 95%, the results of all statistical tests are the same, with the single exception of test pair T3-T2, in which the sign test could not confirm the other tests. This confirms what we already knew about the lower power of these tests: this resides in the fact that they take into account less information: in the case of the sign test only success or failure is considered, and the ties are removed which in general cases are less information than what the data contains. In the case of our tests the coverage measure is a binary value, so the sample size is the information that is lost in this context, but in the case of ordinal data with more than two levels or continuous data, information is also lost that way. As for the McNemar test usage to test a paired experiment of binary data we have a similar effect, but it is a particular application of the test that does not perform very well. In this case we conclude that, for a sample number of 180 the effect of using binary data should not be the main criteria when choosing the statistical test to apply. Our tests also point in the direction that it is reasonable to assume the normality of the data for a large sample size as ours, which is reflected in the high level of agreement of the t-test and bootstrap. A surprising fact is the good performance of the Wilcoxon signed rank test in a situation of binary data, considering the fact that this test was designed to absorb the absolute differences in magnitude of the data. As a final conclusion the use of more powerful tests as bootstrap, did not allow in this case to give support to conclusions that were not already possible with the classic matched pairs t-test.

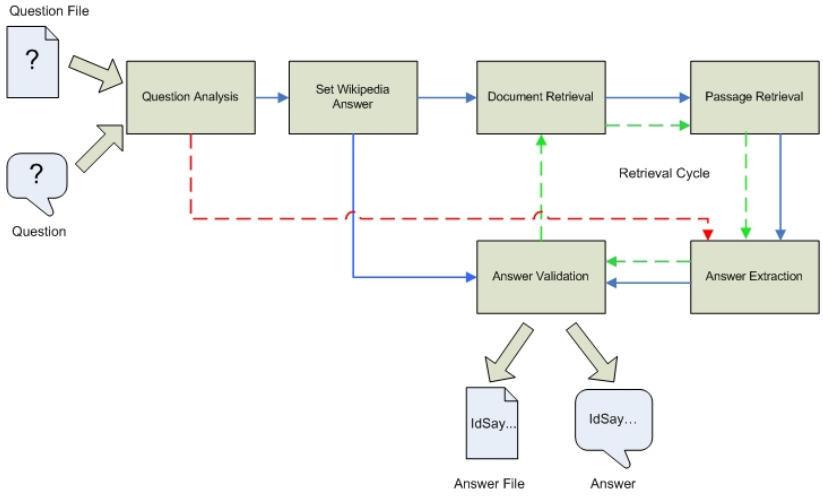


Figure 3.9: IdSay system architecture

3.4 Design options of IdSay

The results of the analysis led to the architecture of IdSay that is presented in Fig. 3.9, and that is going to be explained in that light.

In the question analysis module, the user question is processed, and if there is not a direct Wikipedia answer, a query string is built from the question. The question and query string are divided into single words and entities, with the latter being a set of more than one word that is treated as a whole, keeping word order. The query string is then processed by the document retrieval module, that will return all the documents that contain those words. After collecting all documents, the algorithm proceeds to the phase of passage retrieval, extracting all passages in the returned documents, with all query words in the passage, as long the passage does not exceed a maximum length parameter. With all the passages, the next phase is to extract answers, and we use different methods depending on the question. Finally, the most frequent answers are returned.

The concept of cut-off is not used as in a traditional IR system: while it is not reasonable to expect that a human user (or a system with heavy processing needs) will be able to process a very long list of documents, we can deal with a large number of documents, hence we dropped the traditional concept of cut-off value.

We use a multi layered organisation to keep the information, with the first level being the

original data, and level one data after pre-processing and so on. Since we have to care about efficiency, we seldom use the original level except to present the information to the user, in the answer or supporting passage. It would be a waste of time having to retrieve the original text when looking for passages, therefore we opted for not removing punctuation since we are again dealing with human produced information (meaningful text), so the punctuation marks are important information, particularly when passage extraction is concerned to determine sentence boundaries. We decided to keep punctuation marks, but in a normalised format according to which all punctuation marks are treated as words, and separated by a space from the surrounding words even if it was not the case originally.

We do not use a list of stop words as we could not prove its influence the document retrieval. However the only list that showed statistical evidence of being efficient in our tests, was SL3, the one we built based on the most frequent words in the data collection. We evolved along this line, and created the feedback loop in the system, that we called retrieval cycle, through which we try again leaving out the most document frequent word in the collection that is also in the query string, if we are not satisfied with the results of a cycle. This way we take into account the word frequency of the collection, and also of the query string, so we can drop words that would not be dropped if we used a fixed stop list built beforehand.

As we explained our system does not use the traditional cut-off concept. For this reason it is more adequate to rely on the data provided for a high cut-off value in the previous tests. That was one of the reasons why we decided to use lemmatization in our system. On the other hand, lemmatization, or dictionary based stemming, as it is called by some researchers, often incorporates the information collected by experts, which should be used, in the cases it is available.

4 Information Retrieval and Question Answering: Theoretical Models

4.1 Introduction

Open domain QA systems seek to give a concise answer to a question, addressed in natural language that is not restricted to any specific field. The knowledge base of a QA system is usually a large collection of documents, also in natural language.

Considering the size of the information involved, many QA systems use IR modules in their architecture, because of their techniques to process and store the information in a way that enables a query over a large amount of data to be retrieved in a reasonably short time.

IR systems process and store large quantities of unstructured information, that does not need to obey a rigid format (usually text) in an efficient manner, so that it is able to quickly return the information that is relevant to a given request.

Information is input into the IR system through the document concept. A document is a block of text that will be returned as a whole, by the IR system, as a match to a query to the system. The returned documents of the IR, called hits, are usually ordered by a scoring function that tries to determine the relevance of the document to the query. The decision about the granularity of the documents is up to the user of the system. For instance, if one wants to feed the novel "War and Peace" to an IR system to find out details about the action, one can either consider each chapter a document, each paragraph a document, or each sentence a document, depending on the level of detail of the analysis to be made.

The main difference between an IR system and a QA system is that while the former returns to the user the documents that are more likely to be of interest to the query, the latter aims at producing a succinct answer extracted from the document(s), not the list of documents.

In a QA system, the IR component is generally used to filter out documents that have nothing to do with the question, retaining only the documents that are related, for further

processing. It is therefore of fundamental importance that among the documents retrieved by the IR is the one (or several ones) that contains the answer.

Given the typical architecture of a QA system, which has an IR component at the beginning of a processing chain of several modules, a poorly performing IR module places the first limitation on the overall performance of a QA system. Therefore we believe it is of the utmost importance to understand clearly the inner workings of the IR component, theoretically speaking. This fact has been often highlighted in the QA literature, however the choices made are not clearly explained, or in most cases not even specified.

In the present chapter we begin with an overview of what the QA literature has been saying on the subject of Information Retrieval. Afterwards we make a theoretical analysis of retrieval models so that we decide on which one to choose with a stronger basis. We proceed in previous chapter a study on an area that precedes the IR module, but is crucial to the success of the retrieval process, which are the pre-processing options. In the next chapter we describe the inner parts of our IR component, IdSearch, describing both its data structures and functionalities. We then describe our entity related processes, since they play an important part in he retrieval process. The description of our components are made in the light of its computational efficiency, as we identified earlier as a goal to our overall system.

4.2 Choosing a Retrieval Model

We consider the information retrieval problem as a situation in which there is a (usually large) collection of text organized in documents, and a user that has a specific information need that he wants (hopes) to satisfy by accessing the right documents within the collection.

An information retrieval system is a computer application that stores the collection of documents and has the user information need as its input, producing as output a set of documents relevant to the input query.

The information retrieval has three main processes: the representation of the documents, the representation of the query and the process that compares the two of them to decide if they match (exact match retrieval) or to what extent do they match (ranked retrieval).

The theoretical models of information retrieval provide the mathematical foundations for these processes, allowing them to be defined formally in a consistent manner.

In the next subsections we will cover briefly the four classical IR models. These approaches are:

- Boolean Model;
- Vector Space Model;
- Probabilistic Model;
- Language Model.

4.2.1 Boolean Model

The Boolean model is based on set theory and Boolean algebra. The weight of each word in a document is a Boolean value: 1 if it is present and 0 if it is absent. The query can be formulated as a Boolean expression using the operators *and*, *or* and *not* to connect the words. As an example, for a conjunctive query, a document is returned if all the terms in the query are present in the document. The output of the system can be a set of documents, but no ranking is associated to them.

Advantages from the IR point of view: The semantics of the query are precisely defined and easy to understand.

Disadvantages from the IR point of view: The list of returned documents is not ranked

Even though the Boolean Model is firmly grounded mathematically in the set theory, it is one of the most criticized models of retrieval because it does not provide a ranking of documents.

For the information retrieval problem, ranking is considered of the utmost importance because if the user is to be presented with a list of documents let's say for instance 1000, it is important that the most relevant ones are present in front of the list, for he will not probably want to go over all the 1000 documents, but only a more reasonable smaller number.

The models we will cover in the next sections are all ranking models.

4.2.2 Vector Space Model

The vector space model has its foundations in concepts of algebra and geometry.

It considers documents and queries represented as vectors in a M-dimensional space, where M is the total number of different terms in the collection (assuming full text indexing). In this model, a non binary weight $w_{ij} \geq 0$ is assigned to each word t_i in document d_j. This weight is the value of the coordinate t_i of the vector $\vec{d_j}$, that represents d_j, so document d_j would be represented in the M-dimensional space by vector

$$\vec{d_j} = (w_{1j}, w_{2j}, \ldots, w_{Mj}). \tag{4.1}$$

The same can be done in respect to the query q, that becomes represented as vector

$$\vec{q} = (w_{1q}, w_{2q}, \ldots, w_{Mq}). \tag{4.2}$$

M is defined as the number of distinct words in the collection, however the query can introduce new words that are not in the collection. For these definitions of the vector space approach to be fully correct, M should be the number of distinct words in the collection and in the query. In practice a query can add new dimensions to the hyper space, if it contains words not present in the collection. However such words will not be considered in the ranking of documents, as they do not appear in any document.

To establish a ranking between the documents, a similarity measure between the document and the query is defined, and for that query, documents are ranked in decreasing order of similarity.

The similarity function is usually given by the inner product of the two vectors

$$sim(d_j, q) = \vec{d_j} \bullet \vec{q} \tag{4.3}$$

A normalized version of the similarity measure that is also commonly used is the cosine of the angle between the document vector and the query vector (which yields a value between 0 and 1 since all weights are positive). The similarity function of the vector space model can thus be written as:

$$sim(d_j, q) = \frac{\vec{d_j} \bullet \vec{q}}{\| \vec{d_j} \| \cdot \| \vec{q} \|} = \tag{4.4}$$

$$= \frac{\sum_{i=1}^{M} w_{ij} \cdot w_{iq}}{\sqrt{\sum_{i=1}^{M} w_{ij}^2 \cdot \sum_{i=1}^{M} w_{iq}^2}} \tag{4.5}$$

The sum is done only for words that are present both in the document and in the query, because otherwise either w_{ij} or w_{iq} or both would be 0, as defined above.

In the vector space model a score is attributed to each document according to equations 4.3 or 4.5. Documents are returned to the user ranked in decreasing order of similarity to the query. A document can be retrieved even if it matches the query only partially.

With the above definitions the model is complete in theoretical terms, however to be used in practice a very important question remains to be answered: "What should be used as weights to obtain the best results?".

One of the most effective techniques proved to be what become known as *tf-idf* weights. Although the model is a similarity model, in the sense that it bases its ranking on how similar a document is to the query, the reasoning for the introduction of *tf-idf* weights is of a discriminating nature, based on the notion of clustering. Let us suppose that we want to make a cluster containing all the documents that are similar to the query, leaving outside all the other documents. For that we need to identify the characteristics that the documents within the cluster have in common, and value their contribution to the weights, while in regard to the features that make documents to be left outside the cluster, they must have a contribution that results in lower weights. The former characteristics provide an *intra-cluster similarity*, while the latter determine *inter-cluster dissimilarity*. *Tf-idf* weights are a result of the balance between these two contributions:

- the *tf* part stands for *term frequency* of a word in a document, i.e. the number of occurrences of the term in that document, and plays the role of intra-cluster similarity: it is assumed that the more times a query word appears in the document the more similar to the query that document must be;
- the *idf* part stands for *inverse document frequency* of a word in the document collection, i.e. the inverse of the number of documents in the collection in which the term occurs, and plays the role of inter-cluster similarity: it is assumed that the more documents in

the collection that contain that word, the less meaningful it must be in determining the document similarity to the query (hence the inverse in its contribution to the weight);

This reasoning can also be interpreted in terms of the two measures that are usually employed to evaluate an information retrieval system, based on the relevance of a document to a query, namely *precision* that is the proportion of retrieved documents that are relevant and *recall* that is the proportion of relevant documents that are retrieved. The cluster of documents similar to the query would be the set of relevant documents to the query, and the best performance that an information retrieval could achieve would be 1 both for precision and recall, meaning that all relevant documents to the query and only those would be retrieved by the system.

Another factor that was later added to the *tf-idf* weighting scheme:

- the *normalization factor* that was especially useful for collections that were not uniform in terms of document length. The normalization factor would act as a corrector for the fact that longer documents would be prone to get higher scores, because they were more likely to have higher term frequencies. This idea was behind the use of the cosine similarity measure of 4.5.

The model was created by Gerard Salton in the 1970's (Salton et al. 1975), but has been subject to a lot of investigation especially in the decades of 1970's and 80's. Also the TREC evaluations, that were created in the 1990's triggered a lot of research and new results in the field.

The SMART Information Retrieval system has been developed based on this theoretical model. It's first implementation was in an IBM mainframe, but later versions were written to UNIX.

It was widely used in the research of the term weighting investigation. Extensive investigation and tests over many decades introduced new schemes to value the contribution of the three weighting factors, and they evolved from the raw frequencies counts to more elaborate functions of these quantities.

A notation, which came to be known as the SMART notation, was created to identify different *tf-idf* weighting schemes, and it consists of two triples identifying the three weighting factors in the following order:

- *term frequency*

  A function, $f(tf_{ij})$, where tf_{ij} stands for *term frequency* of a word i in a document, i.e. the number of occurrences of word i in that document;

- *inverse document frequency*

  A function, $g(\frac{1}{df_i})$, where df_i stands for *document frequency* of a word i in the document collection, i.e. the number of documents in the collection that have one or more occurrences of word i;

- *normalizing effect*

A complete scheme would be identified by

(ddd,qqq)

in which ddd would be the factors used for the document, and qqq the factors used for the query, because different criteria could be best suited for documents and the query.

The similarity measure would be the inner product between vectors 4.3.

A summary of tests conducted with several term weighting strategies can be found in (Salton & Buckley 1988). As a result of these tests, the following configuration is suggested for best performance for documents that consist of natural language texts:

Weights for the document j:

$$w_{ij} = \frac{tf_{ij} \cdot \log(\frac{N}{df_i})}{\sqrt{\sum_{i=1}^{M} \left(tf_{ij} \cdot \log\left(\frac{N}{df_i}\right)\right)^2}} \tag{4.6}$$

Weights for the query:

$$w_{iq} = \left(0.5 + \frac{0.5 \cdot tf_{ij}}{\max_q tf_{ij}}\right) \cdot \log\left(\frac{N}{df_i}\right) \tag{4.7}$$

These expressions correspond to a raw term frequency for the document, and an augmented normalized term frequency for the query, i.e. *tf* normalized by maximum *tf* in the vector (the query vector in this case, hence the $\max_q tf_{ij}$) and with a further normalization that makes it lie between 0.5 and 1. The inverse document frequency for both the query and document is the

logarithmic function of the inverse of the proportion of total documents that contain the term: this function gives weight of 0 to a term present in all documents ($\frac{N}{df_i} = 1$) and gives a higher weight to words that exist only in a few documents, the bigger the document collection, the bigger the weight. The normalizing factor is cosine normalization for the document and none for the query.

For an historical perspective, the original SMART system used a configuration with raw term frequencies and no inverse document frequency for both the document and the query, and cosine normalization for the document only, and a unit weight for words without *idf* or *scaling* would correspond to a scoring given by the number of words matching between the document and the query.

These were the results by the end of the 1980's. In the 1990's tests motivated in large part by TREC, new results were introduced, namely:

- a pivoted document length normalization (Singhal et al. 1996)

  As mentioned earlier, document length normalization intends to give different length documents fair chances of retrieval. That was the motivation for the use of the cosine similarity function of 4.5. However later studies using the TREC collection showed that this method tended to retrieve short documents with a higher probability than their probability of relevance, whereas for long documents the probability of retrieval was lower than their likelihood of retrieval[1]. This means that the correcting effect of the normalization was not the same across all document lengths and tended to penalize too strongly longer documents.

  Figure 4.1, which was taken from (Singhal et al. 1996), depicts this effect on the left hand side. The point where the curves intersect is called the *pivot*. The idea was to compensate this effect by rotating the value of the normalizing function around the pivot, as indicated in the figure of the right hand side. In practice it would mean to use as new normalization function the line with slope s that has the same value of the old normalization function at the pivot p:

[1] At this time configuration lnc.lnt was used for SMART.

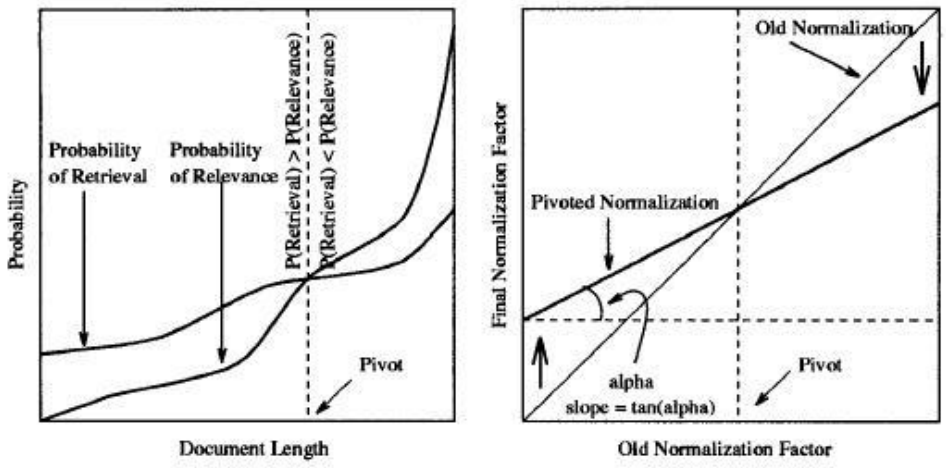


Figure 4.1: Pivoted Normalization - Graphical Interpretation

$$pivoted\ normalization = (1.0 - s) \times p + s \times old\ normalization \tag{4.8}$$

If this equation is divided $(1.0 - s) \times p$ it can be written in the form:

$$pivoted\ normalization = 1 + c \times old\ normalization \tag{4.9}$$

where c is a constant that has the value

$$c = \frac{s}{(1.0 - s) \times p} \tag{4.10}$$

It corresponds to removing one parameter by fixing one of the values, and has no loss in generality. Usually the pivot is fixed at the *average old normalization factor* and the final normalization factor becomes:

$$pivoted\ normalization = (1.0 - s) + s \times \frac{old\ normalization}{average\ old\ normalization} \tag{4.11}$$

This normalization compensation is independent of the original normalization function used, although results were based on cosine normalization usage. However alternatives to the cosine normalization were proposed based on the fact that for extremely long docu-

ments, the cosine function tended to favour those documents in retrieval, and with this compensation that tendency is aggravated even further.

Two alternative functions were introduced:

- *pivoted unique normalization*
  based on the unique (or distinct, in our nomenclature) number of terms in a document, and
- *pivoted byte size normalization*
  based on the number of bytes of a document,

both yielding better results than the cosine function in the experiments conducted.

- a double logarithmic function to calculate the term frequency (Singhal et al. 1998)
- byte length normalization
- a different *idf* function
- The query vector weights became simply the raw frequency of query terms.

A combination of all these factors has proved to achieve high quality results (Singhal 2001). The ranking expression for a document is the following:

$$rank(d_j, q) = \sum_{t_i \in d_j, q} \frac{1 + \log(1 + \log(tf_{ij}))}{(1 - s) + s\frac{Kb_j}{avdlb}} \cdot qtf_i \cdot \log \frac{N + 1}{df_i} \tag{4.12}$$

calculated for all terms that belong both to the document and the query, with

N Number of documents in the collection

tf_ij frequency of the term in the document d_j

qtf_i frequency of the term t_i in the query

df_i document frequency of the term t_i

Kb_j document length of document d_j, in bytes

$avdlb$ average document length of the collection in bytes, and

s slope is the parameter originated by the pivoted normalization (typically 0.2).

4.2.3 Probabilistic Model

There are several models for IR based on Probabilistic approaches: in this section we consider the Binary Independence Model (BIM), a precursor within the approach that is known as Probabilistic Relevance Framework, and in the next section we consider the language model application to IR, another probabilistically oriented approach. An extensive discussion on the Probabilistic Relevance Framework (PRF) is given in (Robertson & Zaragoza 2009).

In the Binary Independence Model the documents are represented as vectors in the words' dimensions (in the same sense as in the Vector Space Model) and are binary (in the same sense as in he Boolean Model):

- $w_{ij} = 0$ if term t_i is not present in document d_j
- $w_{ij} = 1$ if term t_i is present in document d_j

It is assumed that there is independence between terms.

The Probabilistic Model is based in Probabilities theory, and it has as central concept the set of relevant documents to a given query.

The way in which documents are ranked is based in the probabilistic ranking principle that defines that for a user query, documents should be returned in order of decreasing probability of relevance to that query. The model can thus be seen as a discriminative model, in which the relevant documents get separated from the non-relevant. That can be achieved if the ranking function is defined as the following ratio:

$$rank(d_j, q) = \frac{P(d_j \ relevant \ to \ q)}{P(d_j \ not \ relevant \ to \ q)} \tag{4.13}$$

This is the Odds of a document d_j being relevant to query q, and produces the same ordering of documents as using just the numerator, and simplifies the expressions.

If we consider R the set of relevant documents for a query q and $\bar{R}$ the complement of R, i.e. the set of non relevant documents to query q, we can write:

$P(d_j \ relevant \ to \ q) = P(R|d_j)$,

and

$P(d_j \text{ non relevant to } q) = P(\bar{R}|d_j)$,

which means the probabilities of a random document d_j belonging to the set of relevant documents R and to the set of non relevant documents $\bar{R}$, respectively. We omit the query q from the conditional probability to the notation becomes simpler.

The ranking function is then:

$$rank(d_j, q) = \frac{P(R|d_j)}{P(\bar{R}|d_j)} \tag{4.14}$$

Using Bayes Rule it becomes:

$$rank(d_j, q) = \frac{P(d_j|R) \cdot P(R)}{P(d_j|\bar{R}) \cdot P(\bar{R})} \tag{4.15}$$

$P(d_j|R)$ indicates the probability of a random document taken out of the set of relevant documents R being document d_j, and $P(d_j|\bar{R})$ has the same meaning for the set of non relevant documents $\bar{R}$. Since $P(R)/P(\bar{R})$ is constant across documents for a given query, the expression can be further simplified to:

$$rank(d_j, q) = \frac{P(d_j|R)}{P(d_j|\bar{R})} \tag{4.16}$$

If independence between terms in a document is assumed (Naive Bayes assumption), then the conditional probability of a given document being d_j is given by the multiplication of the probability of its coordinates (i.e. weights w_{ij}) for all dimensions, so the rank can be calculates as:

$$rank(d_j, q) \simeq \frac{P(d_j|R)}{P(d_j|\bar{R})} = \prod_{i=1}^{M} \frac{P(w_{ij}|R)}{P(w_{ij}|\bar{R})} \tag{4.17}$$

If we take into consideration that the weights are binary, and separate the case $w_{ij} = 1$ (i.e. term t_i present in d_j) and $w_{ij} = 0$ (i.e. term t_i not present in d_j), we can write:

$$rank(d_j, q) \simeq \prod_{i=1\ :\ w_{ij}=1}^{M} \frac{P(t_i|R)}{P(t_i|\bar{R})} \cdot \prod_{i=1\ :\ w_{ij}=0}^{M} \frac{P(\bar{t_i}|R)}{P(\bar{t_i}|\bar{R})} \tag{4.18}$$

$P(t_i|R)$ is the probability that term t_i is present in a document randomly selected from R and $P(t_i|\bar{R})$ is the probability that term t_i is present in a document randomly selected from $\bar{R}$. $P(\bar{t_i}|R)$ and $P(\bar{t_i}|\bar{R})$ have the corresponding meaning, but for term t_i being absent from the randomly selected document.

If we assume that terms not occurring in the query are equally distributed in relevant and non relevant documents, then we can consider only terms that occur in the query:

$$rank(d_j, q) \simeq \prod_{i=1\ :\ w_{ij}=1, w_{iq}=1}^{M} \frac{P(t_i|R)}{P(t_i|\bar{R})} \cdot \prod_{i=1\ :\ w_{ij}=0, w_{iq}=1}^{M} \frac{P(\bar{t_i}|R)}{P(\bar{t_i}|\bar{R})} \tag{4.19}$$

If we take into account the fact that

$P(t_i|R) + P(\bar{t_i}|R) = 1$ and $P(t_i|\bar{R}) + P(\bar{t_i}|\bar{R}) = 1$,

we can rewrite the expression in terms of the probability of occurrence of t_i:

$$rank(d_j, q) \simeq \prod_{i=1\ :\ w_{ij}=1, w_{iq}=1}^{M} \frac{P(t_i|R)}{P(t_i|\bar{R})} \cdot \prod_{i=1\ :\ w_{ij}=0, w_{iq}=1}^{M} \frac{1 - P(t_i|R)}{1 - P(t_i|\bar{R})} \tag{4.20}$$

If we include in the right product the query terms that occur in the document (and divide by them in the left product so that the overall result is unchanged) the expression becomes:

$$rank(d_j, q) \simeq \prod_{i=1\ :\ w_{ij}=1, w_{iq}=1}^{M} \frac{P(t_i|R)}{P(t_i|\bar{R})} \cdot \frac{1 - P(t_i|\bar{R})}{1 - P(t_i|R)} \cdot \prod_{i=1\ :\ w_{iq}=1}^{M} \frac{1 - P(t_i|R)}{1 - P(t_i|\bar{R})} \tag{4.21}$$

The rightmost product is constant for a given query, so we can consider only the first product:

$$rank(d_j, q) \simeq \prod_{i=1\ :\ w_{ij}=1, w_{iq}=1}^{M} \frac{P(t_i|R)}{P(t_i|\bar{R})} \cdot \frac{1 - P(t_i|\bar{R})}{1 - P(t_i|R)} \tag{4.22}$$

If we take logarithms of the expression we get a new ranking function, proportional to the previous one, and that can also be used for ranking:

$$rank(d_j, q) \simeq \sum_{i=1\ :\ w_{ij}=1, w_{iq}=1}^{M} log\ \frac{P(t_i|R)}{P(t_i|\bar{R})} \cdot \frac{1 - P(t_i|\bar{R})}{1 - P(t_i|R)} = \sum_{t_i \in d_j, q} log\ \frac{P(t_i|R)}{P(t_i|\bar{R})} \cdot \frac{1 - P(t_i|\bar{R})}{1 - P(t_i|R)} \quad (4.23)$$

The same expression can be written in a more condensed way, if we define

$p_i = P(t_i|R)$

as the probability of a term t_i appearing in a document relevant to the query, and

$u_i = P(t_i|\bar{R})$

as the probability of a term t_i appearing in a non relevant document to the query. Then 4.23 becomes:

$$rank(d_j, q) \simeq \sum_{t_i \in d_j, q} log\ \left(\frac{p_i}{1 - p_i}\right) \Big/ \left(\frac{u_i}{1 - u_i}\right) \quad (4.24)$$

Since

$$\frac{p_i}{1 - p_i} \quad (4.25)$$

is the odds of a term t_i appearing in a relevant document, and

$$\frac{u_i}{1 - u_i} \quad (4.26)$$

is the odds of a term t_i appearing in a non relevant document, we can see that the ranking function for document d_j is the log odds ratio, calculated for terms that belong simultaneously to the document and the query.

Having coming this far, we are now in a position in which the model is defined (making several assumptions), but we need to make estimates for p_i and u_i to use the model in practice, since in general we do not know what the set of relevant documents R is.

A common way to make an initial estimation is:

(a) assuming that the relevant documents are a lot less than the total documents, N, then

the distribution of terms in the non relevant documents can be approximated by the distribution in the entire collection. If the number of documents in the collection that contain term t_i is df_i, then:

$u_i^{(0)} = df_i/N$

(b) assuming that each term has equal probability of appearing in a relevant document, typically a 0.5 probability, then an approximation is,

$p_i^{(0)} = 0.5$

Using these initial estimates and the ranking formula 4.24 we can obtain a ranked list of documents. This set of documents is ordered by decreasing order of relevance to the query, and can be used to improve the estimation of probabilities, and therefore produce a better ranked list. This process can be repeated iteratively in a relevance feedback loop until results are satisfactory.

Next we describe a methodology to improve the estimation of the probabilities p_i and u_i called pseudo relevance feedback. Pseudo relevance feedback is used here in the sense that the process is fully automated and the feedback is not provided by the user, however user intervention can be considered if it is available (the user analyses a sample of documents taken from the initially produced ranked list, and manually determines their relevance to the query).

The pseudo relevance feedback process, after the initial ranking of documents, let us consider a subset of these documents, typically a fixed number of documents taken from the top of the initial ranked list (supposedly containing the most relevant documents to the query). Regarding this sample $\mathbb{S}$ of relevant documents, we define:

N_S - Number of documents in the sample

n_{Si} - Number of documents in the sample that contain term t_i

R_S - Number of relevant documents to the query in the sample

r_{Si} - Number of relevant documents to the query in the sample that contain term t_i

We can make an improved estimation in the following way:

(a) assuming the distribution of terms in the relevant documents retrieved so far is a good estimation for p_i:

$$p_i = \frac{r_{Si}}{R_S} \tag{4.27}$$

(b) assuming that the distribution of terms in the documents belonging to the sample but not relevant to the query is a good estimation for u_i,

$$u_i = \frac{n_{Si} - r_{Si}}{N_S - R_S} \tag{4.28}$$

To avoid problems due to the small size of the sample (for instance the case of $R_S = 1$ and $r_{Si} = 0$), the following adjustment or smoothing is used for equations 4.27 and 4.28, which become respectively:

$$p_i = \frac{r_{Si} + 0.5}{R_S + 1} \tag{4.29}$$

$$u_i = \frac{n_{Si} - r_{Si} + 0.5}{N_S - R_S + 1} \tag{4.30}$$

Replacing these last expressions,4.29 and 4.30, in 4.25 and 4.26 we obtain the following:

$$\frac{p_i}{1 - p_i} = \frac{r_{Si} + 0.5}{R_S + 1} \cdot \frac{R_S + 1}{R_S - r_{Si} + 0.5} = \frac{r_{Si} + 0.5}{R_S - r_{Si} + 0.5} \tag{4.31}$$

$$\frac{u_i}{1 - u_i} = \frac{n_{Si} - r_{Si} + 0.5}{N_S - R_S + 1} \cdot \frac{N_S - R_S + 1}{N_S - R_S - n_{Si} - r_{Si} + 0.5} = \frac{n_{Si} - r_{Si} + 0.5}{N_S - R_S - n_{Si} - r_{Si} + 0.5} \tag{4.32}$$

And finally the ranking function of 4.24 becomes:

$$rank(d_j, q) \simeq \sum_{t_i \in d_j, q} log \frac{(r_{Si} + 0.5)(N_S - R_S - n_{Si} - r_{Si} + 0.5)}{(R_S - r_{Si} + 0.5)(n_{Si} - r_{Si} + 0.5)} \tag{4.33}$$

This ranking scheme was established in (Robertson & Spärck Jones 1976) and thus are

represented as w_{ij}^{RSJ}. The work on he formula continued with the introduction of a smoothing function and the incorporation of the document length in the model, and as a result BM (Best Match) formulas were tried, for instance BM11 and BM15 (Robertson & Walker 1994). During the test of TREC-3 with the Okapi system, BM11 and BM15 were combined into a single function that became known as BM25 (Robertson et al. 1994), a four parameter model. With continuing experiments BM25 formula is given in(Robertson & Zaragoza 2009) as 4.34.

$$rank(d_j, q) \simeq \sum_{t_i \in d_j, q} \frac{tf_{ij}}{k_1((1-b) + b\frac{dl}{avdl}) + tf_{ij}} w_{ij}^{RSJ} \tag{4.34}$$

In 4.34 the constants take the following values $1 \leq k_1 \leq 2$, typically 1.2 and $0 \leq b \leq 1$, typically 0.75. The formula is known as w_{ij}^{BM25} and it considers the influence of the original or expanded query terms according to the factor 4.35, with constant k_3 taking values between 1 and 1000.

$$\frac{(k_3 + 1)qtf}{k_3 + qtf} \tag{4.35}$$

4.2.4 Language Model

The language model approach (LM approach) to Information Retrieval is based on probabilistic language modelling. A language model is a probability distribution over sequences of words, that is, a function that associates a probability to a sequence of words from a text. The idea of building a model of language was originally used in the 1980's in the automatic speech recognition area, where the model is derived from samples of spoken text, but will be used to generalize beyond the sample data, to deal with new words and sequences of words that are liable to occur (a language model can work as a recognizer or as a generator of strings). There are many applications that use of the concept of Language Model with slightly different functions (namely spelling correctors or machine translators) and that is the case of Information Retrieval.

Language modelling is quite a broad field and there are some decisions to be made particularly as far as how to determine the probabilities of the sequences of words and how to measure the probability of the query for the derived model.

The simplest model is to consider independence between word which means that the con-

ditional probability - this model is called unigram model. If we take into consideration the influence of the previous word to generate the next one we will have a bi-gram model, and so on. Tri-gram models are commonly used in ASR applications.

In Information retrieval applications it is common to use unigram models, and one of the reasons for that is that unlike the case of ASR applications in which we want to guess the next word as accurately as possible, in IR it is considered acceptable to model word occurrences at the document level without regard to word order. Even though no order is considered to model, the probability of a query can be obtained, however the queries with the same words with different word orders will have the same probability. Another argument for the use of unigram models is that since the model is built based on a single document that is considered as a representative of "the related documents class" and since documents can be short in length, it usually chosen the simplest model, i.e. the unigram model, because losses from data sparseness tend to out weight gains from the use of higher level models.

A language model can be built based on the occurrence of a word given the occurrence of earlier words. Using the Chain Rule

$$P(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} P(x_i | x_{i-1}, \cdots, x_1) \tag{4.36}$$

If we estimate the probability of words as occurring independently of other words we have the simplest form of language model, in which no conditional probabilities are considered:

$$P(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} P(x_i) \tag{4.37}$$

A framework for using a language model approach for IR was presented in (Ponte & Croft 1998). The estimation of the probability of the query being generated by the language model derived for a document d_j, which we will represent as LM_j, used was the maximum likelihood estimator (MLE), and for a unigram model it is:

$$P(q|LM_j) = \prod_{t_i \in q} \hat{P}_{mle}(t_i|LM_j) \tag{4.38}$$

The maximum likelihood estimate of the probability of term t_i given the distribution of term

for document d_j is:

$$\hat{P}_{mle}(t_i|LM_j) = \frac{tf_{ij}}{K_j} \tag{4.39}$$

where tf_{ij} is the term frequency of term t_i in document d_j, and K_j is the document length of document d_j in words.

This approach has the advantage that document statistics are an integral part of the model, with the normalization effect of dividing by the word length included, while in other models these statistics are introduced heuristically.

However using this LM approach suffers from two main problems:

- One of them happens when a document does not contain one of the words of the query. In this case the document is assigned a probability, and a ranking, of 0, and it is never selected for that query. This results in a strictly conjunctive semantics for the model. This effect is known as the *zero probabilities* problem;
- The other problem is caused by the language model to be built based on a relatively small amount of data (the length of a document), which may result in a poor estimation of word probabilities. Above we have considered the case in which query words are not part of the document, and now we are worried about words that occur only a few times in the document, for instance one time *hapax legomena*. This effect is known as the *sparse data* problem.

The answer for these problems is usually introduced through smoothing, for which there are multiple possibilities. Smoothing works in a way to give some probability mass to unseen or rarely seen words in a document.

Up until now the document statistics were introduced as part of the model, but to answer these two problems several adjustments are made on an heuristic basis, and to introduce collection statistics as well as document statistics as a component of term weighting. This makes the model become closer to a term weighting model (as those described in the previous two subsections). That is for instance the case of estimating the probability of a query word not present in the document for

$$\frac{cf_i}{cs}$$

Where

- cf_i is the collection term frequency of term i, and

- cs is the collection size in words, i.e.total number of words in the entire collection

as in (Ponte & Croft 1998), but which is highly sensitive for example for the case of full text indexing without stop words removal.

4.2.5 The use of *idf*

The *idf* concept has played such an influential role in the IR field (Harman 2005) that it is worth looking at the several ways it has been considered in the ranking of documents for retrieval. It was proposed by Sparck-Jones in the early 1970s (Spärck Jones 1972), and over the years there has been several formulas for *idf*. One of them, which we call Function 1, corresponds to formula 4.40.

$$\log\left(\frac{N}{df}\right) \tag{4.40}$$

Formula 4.40 corresponds to the collection frequency component identified as f in the term weighting scheme of (Salton & Buckley 1988).

The graphical representation of function 4.40 is presented in Figure 4.2.

The plot emphasizes the fact that the weight of a word decreases with the increasing number of documents of the collection in which it is present, and in the limit the weight of a word that exists in all documents is null. However it extremely high for words that exist seldom in the collection. If a word belongs to a query or a lexicon and it is not in the document collection (df=0) it is not possible to calculate its effect.

Another *idf* function, Function 2, is suggested in (Robertson & Spärck Jones 1976) that corresponds to the collection frequency component identified as p, for "probabilistic inverse collection frequency factor" in the term weighting scheme of (Salton & Buckley 1988).

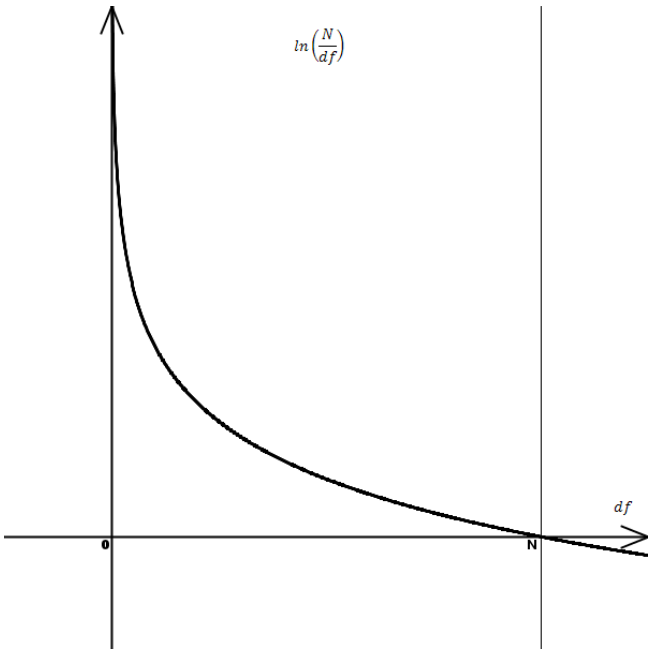


Figure 4.2: Graphical Interpretation of IDF Function 1

$$\log\left(\frac{N-df}{df}\right) \tag{4.41}$$

Function 4.41 is presented in Figure 4.3.

This plot has a considerably different shape from the one in Function 1, and it has a root at (N/2), which means that the weights of a term that belong to less than half of the documents in the collection is positive, while the weight of a document that belongs to more than half of the documents of the collection is negative. Its asymptotic behaviour near 0 and near N can however present problems for terms that belong to the query but do not belong to the document collection (as in Function 1), but as well as for terms that belong to all documents in the collection.

The function used in the probabilistic model, Function 2, also appears in a slightly changed format in (Robertson & Spärck Jones 1976), that of Function 3 (4.42), being used in BM25 (Robertson & Zaragoza 2009).

$$\log\left(\frac{N-df+0.5}{df+0.5}\right) \tag{4.42}$$

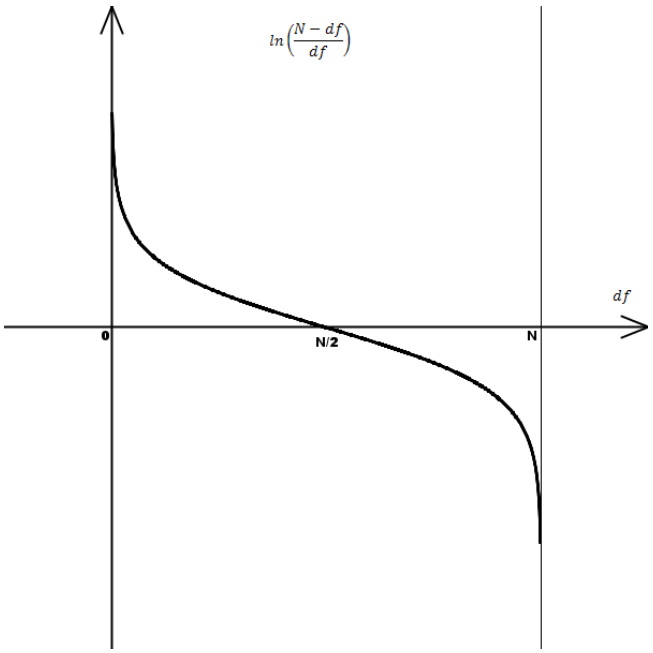


Figure 4.3: Graphical Interpretation of IDF Function 2

Function 3, represented in Figure 4.4, also has a root in N/2, meaning that the *idf* factor still weights documents with a positive or negative value according to the fact that they belong to less of the half of the documents in the collection or to more of half, respectively. The problems described with the asymptotic behaviour of Function 2 no longer apply in this case, since the asymptotes are now shifted by a factor of 0.5 to the left of the origin of coordinates (rare words) or to the right of the total number of documents in the collection (very common words in the documents of the collection).

The *idf* Function 1 was also subject to evolution, with a new formula, and in (Singhal et al. 1998) and (Singhal 2001) a new function which we will name Function 4 (4.42) was used.

$$\log\left(\frac{N+1}{df}\right) \tag{4.43}$$

The plot of Function 4 (Figure 4.5) shows a different value for the root, N+1, so the weight of a term, even if belongs to all documents in the collection will never be zero.

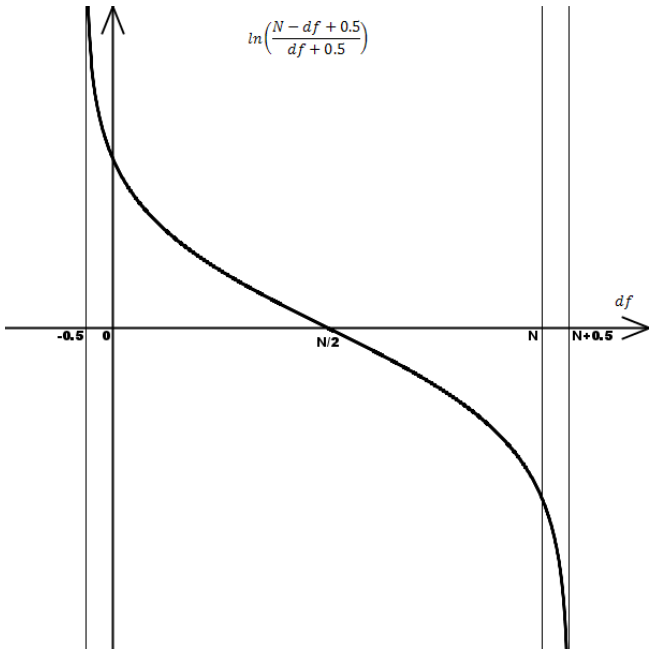


Figure 4.4: Graphical Interpretation of IDF Function 3

4.3 IR Models in QA Systems

Although a very large number of QA systems report the use of an information retrieval component, the characterization of such component in terms of the theoretical model and related options it uses is not usually specified in the literature.

There are exceptions to this rule, starting with the top performing system of TREC QA 1999 (TREC-8) and 2000 (TREC-9) from the Southern Methodist University in Dallas, that uses Boolean search, with no ranking of documents produced, with the corresponding favour on recall rather than precision. In their participation at TREC-QA 1999 they report (Moldovan et al. 1999) the use a modified version of the Zprise IR system made available from NIST[2]. For the 2000 TREC QA edition, they reported on the use of the SMART search engine (Paşca & Harabagiu 2001) without getting into further details on the parameters used.

The AT&T system (Abney et al. 2000) participating at TREC, also based in the SMART

[2] An IR system based on the vector space model, that returned the references for the 1000 top ranked documents for each question. Provided by the NIST, D. Dimmick. Guide to Z39.50/PRISE 2.0: Its Installation, Use, & Modification. http://www.itl.nist.gov/iaui/894.02/works/papers/zp2/zp2.html,2000, but no longer available at this location.

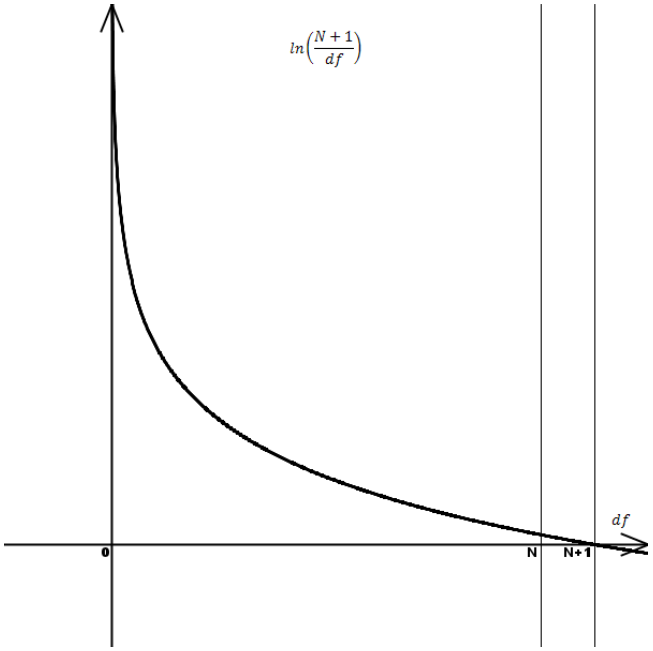


Figure 4.5: Graphical Interpretation of IDF Function 4

IR system that explicitly reports the usage of a scoring mechanism for passages based on *idf* weights.

In the work (Tellex et al. 2003) that motivated our choice of IR system for the tests of the previous chapter, Lucene, compares algorithms for passage retrieval as reported by several participating systems in TREC QA. The passages are obtained from documents obtained by two different IR systems NIST PREISE and Lucene. The two systems are described as follows: "PRISE is representative of the state of the art in information retrieval, incorporating many modern advances in query and term weighting, i.e., bm25" and "(. . .)Lucene, a freely available open-source IR engine. Lucene supports a boolean query language, although it performs ranked retrieval using a standard tf.idf model.". The results obtained were thus described: "An immediate conclusion from our study is that in terms of passage retrieval, the performance obtained using the Lucene document retriever is comparable to the performance obtained using the PRISE document retriever. In fact, passage retrieval algorithms using Lucene actually achieve a higher MRR on average. We found this result surprising because in terms of pure document retrieval, boolean query models have been consistently outperformed by more modern approaches. Yet, for the passage retrieval task, the effectiveness of these different document retrievers is quite sim-

ilar. This result confirms the intuition of many in the question answering community: boolean queries can supply a reasonable set of documents for down-stream components in a question answering system."

The same conclusion was reached by (Bilotti & Nyberg 2008) who tested the effect in the overall performance of a QA system of replaced the text retrieval module by a high-precision module based on the Indri search engine[3] with no a significant improvement in accuracy.

The work of (Saggion et al. 2004) compared several ranking models (Okapi, Lucene and Prise) with boolean retrieval strategies to conclude that while the IR performance of the ranking models were higher, the boolean strategies employed had advantages when considering the QA system as a whole.

4.4 Our choice for model

The theoretical model is a very important underlying aspect of a retrieval system, but it is not the only factor to take into account.

The information system organizes and processes the information in a way so that the results are produced in an efficient way. It has the underlying theoretical model at its base, but this model does not specify the implementation details. For that the basic computer science techniques should be taken into consideration. The efficiency of the system is a result of the careful application of these techniques.

Another way to evaluate the results of an information retrieval system is through the "quality" of the output produced. We have introduced the notion of the user expecting the system to retrieve the "right" or relevant documents. Although relevance is an ambiguous concept that may vary from user to user, and even for the same user it may vary with time, it important to establish some compromising definition of relevance that allows the creation of test document collections to be used in the evaluation and comparison between information retrieval systems, from an end results point of view. It is the case of the international evaluation initiatives such as TREC, CLEF and NTCIR.

The results of these evaluation initiatives are very important on the one hand to define

[3] http://www.lemurproject.org

which models work best in practice and, on the other hand, they are an incentive to improve the techniques employed. The several models define a basis for the system, but they all leave way to heuristic methods to be tested (either because of approximations that are necessary to introduce in the theoretical models, or because of adaptations that are required for practical reasons) so the end results are usually taken into consideration.

Because of the different nature of the tasks of Information Retrieval and Question Answering we will not use the end system results as the main criterium for our decision.

We will use the Boolean Model in a conjunctive way to retrieve all documents that match the query words. In the QA literature it is often mentioned the fact that the retrieval stage is employed to reduce the number of documents in which to look for answers (sometimes because the "downstream" modules are too expensive in terms of computational time to allow the treatment of many documents). We aim at building light modules that allow the treatment of many documents, and still be efficient in terms of system response time, starting with the retrieval component itself. We put however a limit in the number of documents to be processed in the retrieval component for the case of very general questions, that we usually set to 10 000. This limit is not reached often.

With this choice we give more emphasis to the traditional evaluation strategies of attempting to estimate and be efficient in the costs of computational time and space employed.

We summarize our options for the retrieval component, emphasizing the reasons that led us to consider other factors than simply the evaluation results in terms of IR evaluation Forums (evaluation in terms of relevance of the document to the query):

- we have criteria for relevance in QA that are different from IR (QA relevance is more "exact" in the sense that the the document either contains an answer to the question or not, while being undeniably relevant to the question topic)
- ranking methods allow documents that contain only one word of the query to be considered relevant, for instance it is only necessary to have a non zero coordinate in one dimension of the vector space model for the cosine to be non zero; likewise, a similar effect is obtained through smoothing in the Language Model approach. That is a consequence of the difference in criteria between IR and QA: it is more probable for concepts in a loose query to be independent than in a question in Natural Language in which there are dependences

between the individual words. The dependences between query words can be stronger in some cases than in others (in NLP and IR) and even non existing (in IR). Taking this factor into account, a Language Model approach to retrieval would seem more appropriate to the QA task than the Vector Space Model and Probabilistic model, particularly when smoothing is not done or done in a way that it has a stronger effect on longer queries than shorter ones.

- we are aiming at a very high recall (get **all** the documents that contain all documents in the query) instead of using a cut-off value to use only a few of the documents that scored higher according to some criteria. So even if these criteria were very good, they could not improve our results.

- documents are not ranked for document retrieval purposes (all are considered and treated), but a rank is attributed to answers considering results from the several modules in the processing chain.

Just as a final remark on the subject of IR for QA: it would be nice to have an information retrieval system that always produced very good results for QA, meaning that it would return at the top of a ranked list only the few documents that actually contained the answers to the question. Therefore the other modules could make a deeper analysis of the documents to extract the answers. However that ideal IR system is in fact the QA system (all of it), that returns the ranked answers instead of documents, taking into consideration additional criteria in the other modules.

To summarize we will develop an information retrieval component using the Boolean Model, with efficiency in mind. The reason we found it necessary to develop such component from scratch was precisely the efficiency reason, the fact that IR (general purpose) modules based on the Boolean model are not easy to find because they are considered old-fashioned. It is possible to use ranking IR systems as Boolean systems, but they are not very efficient since part of the information they store and use in those models are useless in this case. Moreover we wanted to add the functionality that we describe in the next paragraph, that can also be implemented in existing IR modules, but not so straightforwardly as we wanted.

As a way to overcome the problem of separating words that are part of a single concept in a question, we identify a series of such concepts, which we call entities, but that could also be

identified as MWE (multi word expressions) or collocations or NE (named entities) or several other names depending on the different fields of application. We define an entity as a sequence of one or more words that co-occur in that specific order and has a specific meaning together, different than that of the independent words that constitute it. For the time being we just introduce this concept to say that we use it as part of our retrieval component. We build a separate index for entities, and if we are able to identify an entity in the question we will look for it as an entity and not as separate words, allowing us, for instance, to look for information about Nova Yorque [New York] in documents where the two words co-occur together in that order, leaving out documents that have both words, but occurring separately. We will come to this subject in Chapter 5, and also in Chapter 8 where semantic relations between words and entities are introduced.

5 IdSearch Data Structures and Algorithms

5.1 Introduction

In Chapter 3 we considered the options related to how to obtain the tokens from the text collection. For that we had to analyse the content of what was being indexed (Portuguese text in our case) and make the appropriate decisions.

In the present Chapter, we take the tokens that are the output of the pre-processing phase, and make decisions on how to store them so that the access to them can be effective and efficient. Therefore, our objective in this Chapter is to define the data structures and algorithms that allow the search to be done efficiently.

In other words we are building an IR system, which we name IdSearch. Its function is to return the set of documents that contain a set of terms. These terms correspond to the ones in the query, and each one must appear at least once in each of the documents returned.

From this chapter onwards, and according to our conventions explained in Chapter 1, we use a different name for tokens: we use terms preferably to words, because they include punctuation and numeric values that are not words, strictly speaking.

We divide the data structure in levels, as already introduced in Chapter 2, with each level built upon the data of the previous levels.

In the next section we describe the original document level (level 1 or L1), where the document data is stored as close the original as possible, followed by a section describing the word root level (level 2 or L2) where we store information on inverted indexes and replacing the words by its root. This way we are able to return all documents containing occurrences of words in the same morphological class. As already mentioned in Chapter 3 we use Lemmatization to do this normalization.

Afterwards we describe the entities level (level 3 or L3) that stores data as sequence of words,

that have meaning together, the entities, so that search can be done for these words together, with the order preserved. The Chapter ends with conclusions, in section 5.

5.2 Original Document Level - L1

The aim of this level, henceforth called L1, is to store the document collection as close as possible to the original text, i.e. subject only to the graphical pre-processing described in Chapter 3.

The first step of the pre-processing is to turn all the document text into lowercase, including the letters with accents. Afterwards we remove all URLs since we do not consider them neither in questions nor in answers. The URLs appear often in our text collection, especially in Wikipedia. That is done in a passage through the document in which the URLs to be removed are identified as strings starting with "http" and ending with the space characters or a quotation mark. These space characters are also used in the next step, which is to split the documents into tokens. A token is a sequence of non-space letters separated by a space character. Space characters are presented in Table 5.1.

Table 5.1: Token Separator Characters

Name	Code (Windows-1252)
Space (SP)	32
Tab (TAB)	9
Line feed (LF)	10
Carriage return (CR)	13
Non breaking space (NBSP)	160

A string ending in a punctuation mark is split into two tokens, one with the word and the other with the punctuation mark. In the case of punctuation marks or numbers between letters, the string is split into different tokens, each containing only letters, punctuation marks or numbers. An example of this later case is the string "sdf!sfde1" that originates the four following tokens: "sdf" "!" "sfde" "1". The process avoids the problem of keeping punctuation marks identified in Chapter 3, without the need to remove them.

Next comes a pre-processing step specific for numeric values. This process was reformulated for the improved version of IdSay, therefore it is described entirely in detail in Section 8.4, in the pre-processing area. In this version we just split the digits in groups of three, for reasons

explained later in this chapter.

The next step, this one exclusively done in the improved version of IdSay, deals with treatment, identification and normalization of abbreviations and acronyms, and is described in Section 8.5.

Now that we have identified the tokens, we must decide what to do with them. Each token is from now on called a term in the collection.

We introduce now an example of a small text collection in order to illustrate the data structures and algorithms in the rest of the Chapter. The collection, which we will call Poem Text Collection, is composed of nine verses from nine poems from well known Portuguese and Brazilian poets. We give more information in Appendix B. The original version of the texts is shown in Table 5.2, and Table 5.3 shows them after the pre processing step.

Table 5.2: Poem text collection: Original Documents

Doc #	Document
1	Eles não sabem que o sonho
2	Assim devera eu ser se não fora não querer.
3	Eu te amo porque te amo,
4	Ai que prazer não cumprir um dever
5	Eu amo o longe e a miragem
6	É um não querer mais que bem querer;
7	Eu não sou eu nem sou o outro,
8	Se as coisas são inatingíveis... ora! Não é motivo para não querê-las...
9	Eu te peço perdão por te amar de repente

Table 5.3: Poem text collection: Documents after pre-processing

Doc #	Document
1	eles não sabem que o sonho
2	assim devera eu ser se não fora não querer .
3	eu te amo porque te amo ,
4	ai que prazer não cumprir um dever
5	eu amo o longe e a miragem
6	é um não querer mais que bem querer ;
7	eu não sou eu nem sou o outro ,
8	se as coisas são inatingíveis ... ora ! não é motivo para não querê - las ...
9	eu te peço perdão por te amar de repente

Up until now we have documents as strings, one string for each document. These strings came from the different collection files, whose different formats have been converted to text, as described in Chapter 3.

From now on we will store the documents differently. The way we chose to store the documents is the following: each term is assigned a sequential number, the *term number*, and each occurrence of the term in the document collection is replaced from its character version by the corresponding *term number*. That is, instead of storing text strings we store numbers, *term numbers*.

All documents are processed and in each document the terms are looked into in order of appearance: if a term has not yet occurred in the collection, a new *term number* is assigned to it, otherwise its assigned *term number* is used. In the end of this processing, the documents will become a sequence of *term numbers* instead of a string.

For this process we need a table to make the correspondence between a *term string* and a *term number*. For our example text collection this correspondence is presented in Table 5.4. Note that *term numbers* are assigned sequential to *term strings* by the order of occurrence in the text collection. Table 5.5 presents the example text collection with documents represented as a sequence of *term numbers*.

Table 5.4: Poem text collection: Correspondence Term Number - Term String

Number	String	Number	String	Number	String	Number	String
1	eles	14	.	27	miragem	40	ora
2	não	15	te	28	é	41	!
3	sabem	16	amo	29	mais	42	motivo
4	que	17	porque	30	bem	43	para
5	o	18	,	31	;	44	querê
6	sonho	19	ai	32	sou	45	-
7	assim	20	prazer	33	nem	46	las
8	devera	21	cumprir	34	outro	47	peço
9	eu	22	um	35	as	48	perdão
10	ser	23	dever	36	coisas	49	por
11	se	24	longe	37	são	50	amar
12	fora	25	e	38	inatingíveis	51	de
13	querer	26	a	39	...	52	repente

In order to store the information in this way from the original text we need to have a process

Table 5.5: Poem text collection: Documents as sequences of term numbers

Doc #	Document
1	1 2 3 4 5 6
2	7 8 9 10 11 2 12 2 13 14
3	9 15 16 17 15 16 18
4	19 4 20 2 21 22 23
5	9 16 5 24 25 26 27
6	28 22 2 13 29 4 30 13 31
7	9 2 32 9 33 32 5 34 31
8	11 35 36 37 38 39 40 41 2 28 42 43 2 44 45 46 39
9	9 15 47 48 49 15 50 51 52

to do it in linear time since we are dealing with a large amount of text data[1], and we need to process all term strings to convert them to the corresponding *term numbers*.

For this process we use a dictionary based method: each time we process a term we check if it is already in the dictionary (in the example text collection Table 5.4): if it is, we store the corresponding *term number*, otherwise we add it to the dictionary, attributing to it a new *term number*.

Both dictionary operations:

- look up a term in the dictionary, and
- add a new term to the dictionary,

must be done in constant time since at least the first operation will be used for all terms, and the second for all distinct terms in the collection.

For implementing the dictionary we used hash tables, as these structures have been shown to perform better than other data structures (Zobel et al. 2001). The hash tables we use have fixed size and allow collisions, this decisions also supported by the same article. We used a hashing function from Robert Jenkins[2].

The hash table is an intermediate structure to speed the process of finding if a term string has already been allocated a term number. This operation would have a linear time complexity

[1] Data not as collection of datum but used in the singular, in a more general sense.
[2] Available as public domain at `http://burtleburtle.net/bob/c/lookup3.c`.

if the *Term String - Term Number table* is not ordered, and would have logarithmic complexity otherwise using a binary search. With the hash table we have constant time complexity for this operation. The size of the hash table we use for the data collection is 821 647, that is a prime number of the same magnitude of the number of distinct words in the collection 975 723. The result of the hash function for the first word in our example, eles, is 215 520 526. The remainder of the integer division of the result of the hash function by the hash size is the *hash value*, and for the word eles is 249 012. The *hash value* is the index of the hash table where the set of *term numbers* with the same *hash value* will be stored. To check if a term is in the dictionary, we need to check only the terms in this set.

For the example text collection we choose a small value for the hash table size, 11, to illustrate the collisions in the hash table. We present in Table 5.6 the hash values for this table size, and the corresponding hash table can be found in Table 5.7.

Let us exemplify how the dictionary works for the word amar[love]: If we want to check if the word amar is already in the dictionary, first we calculate the *hash value* that in our case is 0 (Table 5.6). In the hash table (Table 5.7), we will have to check if any of the term numbers in entry 0 correspond to the term string amar by using Table 5.4. In this case there are 7 *term numbers* in entry 0: 7 15 19 22 23 43 50 that correspond to the words assim te ai um dever para amar. In this case amar is in the list and its term number is 50. If the word was not in the list, we would add a new term number for it.

The procedure for both dictionary operations is presented in Algorithm 5.1, that returns a *term number* for a *term string*. If the term is not in the hash table, we can add it in lines 8- 11, updating the hash table with the new term.

The maximum number of collisions in our example (third column in Table 5.7) is 9, that by chance is exactly the same that exists in our data collection. It can happen that there are unused entries in the hash table. That does not happen in our example. However in our data collection the number of empty slots in the hash table is 250 642. This is not a problem since the hash table has 821 647 slots, and with 975 723 distinct term strings, the average of terms in non-empty slots in the hash table is 1.7 terms. So, in average we just check 1.7 strings, and in the worst case we check 9 strings, that prove to be appropriate for this problem.

For the hash table to be efficient, in the pre-processing phase we had the concern not to let the number of terms in the dictionary grow unnecessarily large. This could happen due to the

Algorithm 5.1 Word: calculate the *term number* for an arbitrary *term string*

Parameters:
Input: *term string* as ts
Output: *term number* as tn
Using:
hash table size as $htsize$
hash function as $Hash$
hash table as $htable$
hash value as $hvalue$

$htable(hvalue)$ represents the set of *term* for which the $Hash$ function returns $hvalue$.
t is an iterator for the above set.
$string(t)$ is the *term string* associated with *term number* t.

1: $hvalue \leftarrow Hash(ts) \bmod htsize$
2: **for all** $t \in htable(hvalue)$ **do**
3: **if** $ts = string(t)$ **then**
4: $tn \leftarrow t$
5: **return**
6: **end if**
7: **end for**
8: $M \leftarrow M + 1$
9: $tn \leftarrow M$
10: $string(tn) \leftarrow ts$
11: $htable(hvalue) \leftarrow htable(hvalue) \cup \{tn\}$

possible existence in the collection of a large number of codes[3] and specific numeric values (each arbitrarily large). The way to minimize this problem was to separate letters from numbers, already mentioned in the pre processing, and grouping numeric values in words of three digits[4].

We store in a file, L1 file, in binary mode, the information of these three data structures:

- Table of correspondence Term Number - Term String (Table 5.4 of our example, $string(t)$ in Algorithm 5.1)
- Hash Table for the correspondence between a term string and a term number (Table 5.7 of our example, $htable(hvalue)$ in Algorithm 5.1)
- Original data collection (documents as sequences of term numbers) (Table 5.5 of our example)

This choice of data structures was done in order to save space and time.

[3]For instance the list of asteroid names.
[4]See Section 8.4 for more information.

Table 5.6: Poem text collection: Correspondence Term String - Hash Value

String	Hash	String	Hash	String	Hash	String	Hash
eles	1	.	7	miragem	10	ora	3
não	7	te	0	é	3	!	9
sabem	5	amo	5	mais	5	motivo	5
que	8	porque	3	bem	8	para	0
o	1	,	7	;	2	querê	6
sonho	3	ai	0	sou	4	-	6
assim	0	prazer	7	nem	8	las	8
devera	1	cumprir	10	outro	9	peço	10
eu	2	um	0	as	1	perdão	3
ser	10	dever	0	coisas	3	por	3
se	9	longe	1	são	4	amar	0
fora	3	e	9	inatingíveis	8	de	2
querer	6	a	3	...	7	repente	1

In terms of space each word will be stored using a number (4 bytes) instead of the wide-character [5] representation of the word that would required 2 bytes per character, plus the separator characters, one per word. That way 4 bytes would be enough just to represent a separator character and a one letter word. Considering that the mean number of characters per word for a Portuguese text is bigger than one, that clearly represents a saving in storage space.

The sizes of the example text collection are presented in Table 5.8. The total characters in the table take into account the end of string character or a separator character, therefore each word requires one more character to be stored in memory.

From the table we can quantify the difference in terms of space of our solution compared to the string based storage of the documents. We will start by calculating the required bytes to store our data structure:

- Table of correspondence Term Number - Term String

  Size: the 52 distinct words have 194 characters, each with an end of string character, so (194+52)=246, with two bytes per character will take 246*2 = 492 bytes.

- Hash Table for the correspondence between a term string and a term number

  Size: the 52 distinct words as term numbers stored in 4 bytes each: 52*4 = 208 bytes.

[5] Visual Studio compiler standard that supports conversion function to and from UTF-8.

Table 5.7: Poem text collection: Hash Table

Hash value	Term numbers	Count
0	7 15 19 22 23 43 50	7
1	1 5 8 24 35 52	6
2	9 31 51	3
3	6 12 17 26 28 36 40 48 49	9
4	32 37	2
5	3 16 29 42	4
6	13 44 45	3
7	2 14 18 20 39	5
8	4 30 33 38 46	5
9	11 25 34 41	4
10	10 21 27 47	4

Table 5.8: Poem text collection: Collection Size

Data	Value
Documents N	9
Distinct Words M	52
Total Characters in Distinct Words	246
Total Words K	79
Total Characters	331

- Original data collection (documents as sequences of term numbers)

  Size: the 79 words in the example collection require 79*4 = 316 bytes.

The total size of the L1 structures is 1 016 bytes for our example text collection. The string version has a total of 331 characters that would require 331*2=662 bytes. In this case there is not a saving in space since the collection is very small. We will proceed with a similar analysis of the CLEF data collection. The sizes of the collection and of the L1 index file are presented in Table 5.9.

In this case the sizes needed are:

- Table of correspondence Term Number - Term String

  Size: the 942 990 distinct words have 9 166 861 characters, with two bytes per character will take 9 166 861*2 = 18 333 722 bytes.

- Hash Table for the correspondence between a term string and a term number

Table 5.9: Collection Size and L1 Index Size

Data	Value
Size of Público	358 493 681 bytes
Size of Folha	233 722 868 bytes
Size of Wikipedia	7 662 009 520 bytes
Size of Collection	8 254 226 069 bytes
Documents	414 895
Distinct Words	942 990
Total Characters in Distinct Words	9 166 861
Total Words	170 216 572
Total Characters	887 928 331
L1 Index Size	727 976 380 bytes

Size: the 942 990 distinct words as term numbers stored in 4 bytes each: 942 990*4 = 3 771 960 bytes.

- Original data collection (documents as sequences of term numbers)
  Size: the 170 216 572 words in the example collection require 170 216 572*4 = 680 866 288 bytes.

The total bytes required are 702 971 970. These are the values that correspond to the example test collection, that we constructed manually. In practice we must take into account some additional information related to implementation details: In the first structure we must add the space of the pointer to the string, one (4 bytes) per distinct term number, that yields 3 771 960 bytes; for the second structure we must add for each entry in the hash table (hash size 821 647) a pointer to the array of term numbers and an integer which is the number of entries in the slot, which makes a total of 6 573 176 bytes, and in the third structure, the documents, we store the total number of words of the document and a pointer to the document which takes 3 319 160 bytes. The space for the L1 files becomes 716 636 266 bytes. There is a small difference between this value and the L1 index size in disk (last line of Table 5.9) due to minor implementation and format saving details.

The string version has a total of 887 928 331 characters that would require 887 928 331 * 2 = 1 775 856 662 bytes. In this case our data structures use just 40.1% of space required by the string version. The average number of bytes per word, total of characters times 2 divided by the total number of words, for the string version is 10.43 bytes, while for our data structures it

is the L1 index size including all structures divided by the total number of words is 4.28 bytes. Since the minimum number of bytes per word is 4, to store the integer value of a term number, the auxiliary data structures require only 0.28 bytes per word, saving 6.15 bytes per word in average. We can notice that the space required to store the terms of the collection in our format is approximately 94% of the total number of bytes of the L1 index. To have a reference, the original unprocessed text requires 48.49 bytes per word representing in this case a very large saving in space due to the pre processing operations. The space used by L1 index is 8.8% of the space required by the original collection.

In practice this gain is achieved by the use of a fixed value of 4 bytes per word, and also due to the fact that while in the original text words are separated by spaces (or a character with the corresponding function) in our approach we do not store these spaces. We store the documents as sets of numbers, and between each number a space is implicit, and only when we want to present the text for a user to read, do we write it explicitly. Contrary to spaces, punctuation marks are very important for text understanding and processing, therefore they are also given *term numbers* and are stored with the document.

In terms of time we have given special attention to efficiency, and we will present the time complexity of the algorithms used and developed, in the sections where we describe them.

As an example of the time gain we expect, we can consider the operation of finding a word in a text, that occurs very often in text processing algorithms. With the numeric representation we use, we are able to check if a given text position contains a word we are searching for with a single integer comparison instead of comparing two strings.

We summarise the process in Algorithm 5.2, that apply the level settings 1, transforming a string in a sequence of numbers. It has three main parts: converting to lowercase, removing URLs, and converting the string to a sequence of integers. For the first part can be used normally a standard function, and the second one, we removed all tokens starting with “http:”. The third part uses the Algorithm 5.1, the function $Word$, to convert each token strings to numbers.

Since all operations inside the cycle have constant time complexity, Algorithm 5.2 have complexity of $O(K)$, with K being the total number of terms in the strings.

Keeping punctuation marks separated from the words and storing words as numbers is a novelty in QA systems, and allow us to maintain efficacy in passage extraction, where punctua-

Algorithm 5.2 Apply L1 Settings

Parameters:
Input: *string*
Output: *ints*
Using:
$LowerCase(string)$ returns the string in lowercase
$Left(token, k)$ returns the first k letter in $token$
$Word(token)$ corresponds to Algorithm 5.1

```
1: ints ← ()
2: string ← LowerCase(string)
3: for all token ∈ string do
4:    if Left(token, 5) ≠ "http :" then
5:       ints ← ints + Word(token)
6:    end if
7: end for
```

tion marks are needed, and to have a gain in both, time and space complexity, when comparing with the string version.

5.3 Word Root Level - L2

In this level, L2, the goal is to store the information to support a fast retrieval of the documents that contain a set of words. We use in this level a conceptual normalization technique: we group the class of words that have the same root. This decision is based on the study conducted in Chapter 3, and aim mainly to respond to the high degree of verbal flexion of the Portuguese language. This linguistic based equivalence classes have the root as the representative of the class. The root itself can be found using two different methods Stemming (rule based affix stripping) or Lemmatization (dictionary based replacement using a Lexicon for Portuguese with lemma information). These alternatives correspond to parameters for level2 in our implementation, however in our work and in the rest of the text, we use lemmatization. Although we build the normalization classes based on roots of words in level2, we keep the original form of the words unchanged in level1.

We have considered how to represent and store the documents, now it is time to consider, given a term (or more) how to access the documents that contain it, or that contain all of them, since we are considering Boolean conjunctive queries, as a result of the study conducted in Chapter 4.

The simplest way to do that would be to search all documents sequentially, looking for the occurrences of terms in them, in a process similar to grep in UNIX. However for large collections of documents that is not feasible because it takes too long, for instance in our text collection it would mean go trough 414 895 documents with 170 216 572 terms.

For large collections the option is to organise the information beforehand, in a way that will facilitate the search process. For *ad hoc* queries, i.e. arbitrary queries, inverted indexes are usually used. We quote two of the most prominent books in this area: "Indeed, this inverted index structure is essentially without rivals the most efficient structure for supporting *ad hoc* text search." (Manning et al. 2008) and "In most applications inverted files[6] offer better performance than signature files and bitmaps , in terms of both the size of the index and the speed of query handling" (Witten et al. 1999) . Inverted indexes are used for static collections (like ours) and also for semi-static collections, for which the indexing occurs with a convenient frequency, for instance daily or weakly, such as the case of search engines.

The inverted index is composed of the vocabulary and the occurrences of the terms in the documents, in our case for each *term number* we have a document list of *document numbers*.

It is worth taking a look at the inverted index name: if we consider the computer science terminology an index allows an element of an array to be accessed (in the position indicated by the index). Therefore the inverse function of an index is to obtain the position of a given element in the array (its index). In this context, we can understand the name "inverted index" if we consider we have the element (word) and we want to get its position. If we however think about the meaning of the index in a book and bibliology, it means a list of terms, possible including just a selection of the most important of them, with an indication of the place where they can be found in the text, then the name inverted index may sound less intuitive, even redundant (Manning et al. 2008). There is also the question of granularity: an inverted index with the information we described above (term number - document list) is sometimes called in the literature an inverted list or even an inverted file. To summarize we use the term inverted index, but alternative names for the same concept in the literature are: inverted file, inverted list and postings file[7].

Different granularities other than document granularity can also be considered, for instance

[6] Inverted file as synonym of inverted index, but we will detail this terminological issue shortly in this section.
[7] This later name refers to the list of occurrences of terms as postings.

word granularity can be used. In that case one could keep the information of the number of the document the word occurs in, plus the number of the word within the document for all occurrences of the word in the document. Alternatively one can store all occurrences of words in the collection and not even store the document information, or even go further into character level. In our case however, at this stage we are only interested in retrieving documents, and we will process them further in the other modules of our system, as described in the next chapter.

The inverted index in L2 will have for each *term number* the list of *document numbers* in which the term occurs. Since we use conjunctive Boolean queries and do not need to attribute a score to the documents, we also do not need to store additional information, namely the frequency of the term within the document. The list of documents is sorted by *document number* and without duplicates, to facilitate the merging of lists at retrieval time, as will be explained in 6.3.

We will introduce the following definitions to facilitate the specification and analysis of the algorithms used:

- Set of documents in the collection $\mathbb{D} = \{d_1,...,d_j,...,d_N\}$
- Set of distinct terms in the collection $\mathbb{T} = \{t_1,...,t_i,...,t_M\}$
- K total number of terms in the collection
- K_j is the number of words of document d_j
- Term in position k in document d_j is v_{jk}
- Document d_j in the collection is a sequence of words $d_j = (v_{j1},...,v_{jk},...,v_{jK_j})$

We add to these definitions, the function *Root*, whose result for each term is its root, or lemma, according to the lexicon used and the procedures described later in this section.

The steps of the algorithm for level2 are now detailed further: after clearing the document lists, we start by building the inverted index for all terms in the collection, in a first phase without considering lemmatization. To build the document lists we go through all documents d_j and all words v_{jk} of the document and for each occurrence of v_{jk} we update the collection frequency cf_i of the corresponding term number t_i and add the document number to the document list of the

term. We do not order document lists at this step, and allow duplicates in them. However we have a criterium to stop adding further documents to a document list and that is when the total number of occurrences of t_i, cf_i, exceeds 10% of all the documents in the collection ($N/10$). In that case we clear the document list for term t_i,which means we make the approximation of considering that the term belongs to all documents of the collection. The threshold is an empirical value based on the values we already need to calculate, so that we can save the time of further calculations.

After having treated the entire collection, we go trough all terms and order the respective document lists and remove duplicates. This will allow us to perform set operations in the document list and process them more efficiently.

We present in Table 5.10 the inverted index for the example text collection, after the first phase of the index construction, that is the procedures we had described so far.

Table 5.10: Poem text collection: Inverted Index Phase 1

Term #	Doc List	Term #	Doc List	Term #	Doc List
1	1	19	4	37	8
2	1 2 4 6 7 8	20	4	38	8
3	1	21	4	39	8
4	1 4 6	22	4 6	40	8
5	1 5 7	23	4	41	8
6	1	24	5	42	8
7	2	25	5	43	8
8	2	26	5	44	8
9	2 3 5 7 9	27	5	45	8
10	2	28	6 8	46	8
11	2 8	29	6	47	9
12	2	30	6	48	9
13	2 6	31	6 7	49	9
14	2	32	7	50	9
15	3 9	33	7	51	9
16	3 5	34	7	52	9
17	3	35	8		
18	3	36	8		

We can see that term number 2, não [no], is the one that appears in more documents, 6 documents, followed by term number 9, eu [I], that appears in 5 documents.

To proceed in the inverted index construction the next step is to consider word roots.

The main objective of considering roots in the retrieval is to be able to locate documents that contain different linguistically related forms of the term expressed in the query. We build equivalence classes of words that have the same root, and whenever one of the forms occurs in a query, we will look for documents containing the root. Therefore we will retrieve documents that contain all related forms with the same root.

This is a process that is specific for Portuguese, and that is detailed later in this section.

In Table 5.11 we present the lemmas for the terms in the example text collection.

Table 5.11: Poem text collection: Correspondence Term Number (#) - Root Term Number (R#)

#	String	R#	#	String	R#	#	String	R#	#	String	R#
1	eles	53	14	.	14	27	miragem	27	40	ora	40
2	não	2	15	te	15	28	é	10	41	!	14
3	sabem	54	16	amo	16	29	mais	56	42	motivo	42
4	que	4	17	porque	17	30	bem	30	43	para	43
5	o	5	18	,	18	31	;	18	44	querê	44
6	sonho	6	19	ai	19	32	sou	10	45	-	18
7	assim	7	20	prazer	20	33	nem	33	46	las	46
8	devera	23	21	cumprir	21	34	outro	34	47	peço	59
9	eu	9	22	um	55	35	as	26	48	perdão	48
10	ser	10	23	dever	23	36	coisas	57	49	por	49
11	se	11	24	longe	24	37	são	37	50	amar	50
12	fora	12	25	e	18	38	inatingíveis	58	51	de	18
13	querer	13	26	a	26	39	...	14	52	repente	52
53	ele	53	55	1	55	57	coisa	57	59	pedir	59
54	saber	54	56	mal	56	58	inatingível	58			

Table 5.12 contains the final (taking lemmatization into consideration) inverted index for the example text collection.

The document list of the inverted index, *docs*, can be expressed as the following function:

$docs : \mathbb{T} \rightarrow 2^{\mathbb{D}}$

$docs(t_i) = \{d_j \in \mathbb{D} : \exists_k \; v_{jk} = t_i\}$

All the algorithms will be shown in detail, but without specific implementation options that would blur the clarity of the algorithm. In algorithms we use sequences and operator + to join sequences, when order is important, and sets and operator $\cup$ to join sets, when order is not

Table 5.12: Poem text collection: Inverted Index

Term #	Doc List	Term #	Doc List	Term #	Doc List
1		21	4	41	
2	1 2 4 6 7 8	22		42	8
3		23	2 4	43	8
4	1 4 6	24	5	44	8
5	1 5 7	25		45	
6	1	26	5 8	46	8
7	2	27	5	47	
8		28		48	9
9	2 3 5 7 9	29		49	9
10	2 6 7 8	30	6	50	9
11	2 8	31		51	
12	2	32		52	9
13	2 6	33	7	53	1
14	2 8	34	7	54	1
15	3 9	35		55	4 6
16	3 5	36		56	6
17	3	37	8	57	8
18	3 5 6 7 8 9	38		58	8
19	4	39		59	9
20	4	40	8		

important and its elements can be sorted to allow to find if a element belongs to the set in logarithmic time with binary search, and to allow to join and intersect two sets in linear time.

Algorithm 5.3 describes the construction of function *docs*, the inverted index.

It is a one time pass at all the terms, whose complexity is O(K), with K being the total number of terms. Note that inverted indexes are vectors that must be sorted to allow implementing linear time complexity intersections of sets. Sorting the inverted indexes only in the end, allow the operation in line 6 to be done in constant time.

5.3.1 Roots for the Portuguese Language

We determine the root of each word via lemmatization, using a lexicon for (European) Portuguese with lemma information. This lexical knowledge base is named POLLUX (POrtuguese Lexical Largely Usable and eXtensible) (Alves 2002). The database has a table with 925,275 Portuguese lexical items, including inflected ones.

Algorithm 5.3 Build the inverted index

Parameters:
Input: Document Collection
Output: *docs* updated
Using:
Function root as *Root*

1: **for** $i = 1$ to M **do**
2: $docs(t_i) \leftarrow \{\}$
3: **end for**
4: **for** $j = 1$ to N **do**
5: **for** $k = 1$ to K_j **do**
6: $docs(v_{jk}) \leftarrow docs(v_{jk}) \cup \{d_j\}$
7: **end for**
8: **end for**
9: **for** $i = 1$ to M **do**
10: **if** $t_i \neq Root(i)$ **then**
11: $docs(Root(t_i)) \leftarrow docs(Root(t_i)) \cup docs(t_i)$
12: $docs(t_i) \leftarrow \{\}$
13: **end if**
14: **end for**

The first version of Lemmatization have some shortcomings. We can indicate incomplete lexical information and the fact that words obtained by derivations that imply a change in morphological class are not considered: for example: “democracia” (“democracy”), the noun, will not be related to “democrático” (“democratic”), the adjective. Since we do not do any morpho-syntactic analysis that allows us to have a notion on the morphological class, whenever a word form has different lemmas we opt for leaving the original word, because we have no basis for deciding which lemma we should consider and we prefer to leave the original word instead of making a blind guess. One example of this situation is “fez” (noun - “the hat from the north of Africa and Turkey”) whose lemma is the word itself, and “fez” (verb, 3rd person singular past - “did”), whose lemma is “fazer” (“to do”). This type of situation, however is not frequent in Portuguese. This method is refined in the improved version (see 8.8.

Based on this information, a text file with the words and their lemma is build and loaded to memory. This list of terms is ordered alphabetically to speed up the search process.

It can be verified that the lexicon lists lemmas using the plural/singular criterion and female/male criterion have the same root.

The lexicon we use is for the European Portuguese, and as far as we known there is no

freely available corresponding information for Brazilian Portuguese. Our data collection has both variants of Portuguese, which means the written forms of the same terms are sometimes different. In the texts coming from newspaper articles the two variants could easily be separated: Público - European Portuguese, and Folha de São Paulo - Brazilian Portuguese, however with the Wikipedia the separation would be very hard, since there is a single version of Wikipedia in Portuguese.

We believe the processing of Portuguese text mixing both variants is not a problem to automatic processing, since the high degree of similarity of the two languages, at least as far as question answering is concerned. This opinion is shared by other researchers: "The CLEF evaluation showed that Brazilian Portuguese was not a relevant problem for a system that only used a European Portuguese lexicon. There were not many questions with exclusive Brazilian spelling or Brazilian terms, and the system was able to retrieve correct answers from Brazilian target documents." (Amaral et al. 2006)[8] The advantage of having a bigger volume of information this way is also pointed out in the literature: "This year, in addition to the European Portuguese text collection (Público), the organization also provided a Brazilian Portuguese collection (Folha de São Paulo). This new collection improved the performance of Esfinge, since one of the problems encountered last year was precisely that the document collection only had texts written in European Portuguese, but some of the answers discovered by the system were written in the Brazilian variety and were therefore difficult to support in a European Portuguese collection" (Costa 2006a).

Despite being the same language, there are some differences in the written forms of some words, and we added some rules so that the root of a word would be stored used the simplest form (with less letters), which correspond to the Brazilian writing. To unify the same words in the European variant and the Brazilian variant we need to replace the following five groups of 2 letters by the second letter, in the root. The original text of the documents in the collection (level1) remains unchanged. These rules are presented in Table 5.13, that is implemented in function $PTBRTransform$ in Algorithm 5.4.

In practice these rules are not universal, in the sense that not all words containing a group

[8]Nevertheless Priberam claims to have the European and Brazilian variants covered in the following ways: "FLiP includes a grammar checker, a spell checker, a thesaurus, a hyphenator for both European and Brazilian Portuguese, bilingual dictionaries and a verb conjugator" and "Besides the European and Brazilian Portuguese grammars, SintaGest is currently being tested with Polish".

Table 5.13: Rules to Uniformize Roots from PT-BR

2 letter group	1 letter replacement	Example
cc	c	accionista → acionista
cç	ç	acção → ação
ct	t	actor → ator
pç	ç	excepção → exceção
pt	t	óptimo → ótimo

of these two letters in European Portuguese writing are written with just the second letter in Brazilian Portuguese writing. However the application of these rules in these cases is not a problem because generally these rules do not map words to other meaningful words in Portuguese, so at most the problem in these cases is that we use an internal representation that does not correspond to a Portuguese word. Some counter examples to the rules are words that despite having one of these groups of two letters are written in the same way in European and Brazilian Portuguese, for instance opção [option] and aptidão [aptitude].

Another source of words with different orthography between the European and Brazilian variants of Portuguese is accentuation, generally caused by differences in pronunciation. We consider useful to use the word without accents as root for the differently accented versions of the word (which is also a solution for the common mistake in written texts that is "to forget to use accents"), but we have not implemented this. Some examples of such situations are:

- the use of the acute accent in the European Portuguese word is replaced for a circumflex accent in the Brazilian Portuguese version of the same word, as in económica/econômica [economical] and ténis/tênis [tenis],
- the use of the diaeresis over the letter u in the Brazilian Portuguese writing to indicate that this u is not mute, while the European Portuguese writing does not include such convention, as occurs in consequência/conseqüência [consequence] and frequentador/freqüentador [regular (visitor)],
- the use of the acute accent in the Brazilian Portuguese writing in words not accented in European Portuguese writing, such as Europeia/Européia [European] and estreia/estéia [première].

The Algorithm 5.4 apply level settings 2 to a string, returning its root. The final version uses a lemmatizer, and then apply a normalization between words in Portuguese and Brazilian format. The lemmatizer have function uses a dictionary with all words and its roots, and if the word is in the dictionary return its root, otherwise return the word unchanged. The PTBRTransform implements the replacement described in Table 5.13.

Algorithm 5.4 Apply L2 Settings

Parameters:
Input: *string*
Output: *string*
Using:
1: $string \leftarrow Lemmatizer(string)$
2: $string \leftarrow PTBRTransform(string)$

5.4 Entity Level - L3

This level supports the use of entities, defined as a set of words with specific meaning when used together. The sources for entities are the ones occurring with high frequency in the document collection, and also the ones identified from Wikipedia. These last entities correspond to the name of each Wikipedia page that has content. Like words, we need to index the entities and to have a way to check if a set of words is an entity or not.

To find the entities by frequency, we have a parameter $entityMinimalFreq$ that is the minimal frequency that a sequence of words must have to be considered an entity. The Algorithm 5.5 basically goes through all document in the collection and find all entities, updating the data structures. First it finds all pairs of words and select the ones with frequency greater than $entityMinimalFreq$, and mark them as entities. Then, based on the entities of size N-1, it builds the frequency of entities of size N, and select the ones with frequency bigger than $entityMinimalFreq$, and mark them as entities. The process stops when there are no entities of size N.

Explaining in more detail, the first 4 lines setup the local variables used. The variable *now* is the *n*umber *o*f *w*ords considered, that start in 2 and will grow until no more entities can be found, E is the entity count, that start at zero, and $lastLevel$ is an index for the first entity in the last level, which is meaningful when 3 or more words are considered, otherwise there is no last level. The variables nw and nwf structures are updated by function $ResetCountStructures$

to empty sets, corresponding to one empty set for each distinct term in the text collection. The variable nw stores the words next to each term, and nwf will store the frequency. When 3 or more words are considered, these structures will have a set for each entity in the previous level, and are reset in line 26. The cycle starts by processing all documents, and making a copy with the L3 applied: $docI$.

We will process all sequences of words in $docI$ of size now, and if we are checking pairs of words, we must certify that the first word is neither a number nor a one letter word, as we do not want an entity to start with either of them. If it is ok, $UpdateWordFreq$ will update the data structures nw and nwf for the pair of words considered. In the case of sequences with more than 2 words, first it is verified if an entity of $now - 1$ words exists in the preceding words in the document, by calling the function $Entity$, and in the positive case, the $UpdateWordFreq$ is also called to update the data structures nw and nwf for the entity and the word considered. If an entity does not exist in the preceding words with one word less, then the sequence of words considered cannot be an entity, since a sub-entity does not have enough frequency, much less the entity that contains it.

After processing all documents, we call $StoreEntities$ to store the sequences with more than $entityMinimalFreq$, and discard the other entities. The variable $lastLevel$ is updated with the index of the last entity stored in the current level, that will be the last level in the next cycle. Before the next cycle, the nw and nwf structures are updated by calling $ResetCountStructures$, setting a size equal to the number of entities created in this cycle, that will be required in the next cycle.

After the main cycle, the algorithm ends by deleting the entities marked for deletion, and update the hash table of the entities. The structure of the entity hash table is used in function $Entity$, that is called in line 14. In order for it to be possible to use the entities from the previous level, the hash table is updated in each cycle, in line 25, function $UpdateEntityHashTable$. This information is not valid after the entities are deleted, since the entity number will change, so we just reset the hash table structure and add all entities not deleted in line 33.

The time complexity of Algorithm 5.5 is $O(K.M.Z)$, with K as the number of terms in the text collection, M the number of distinct terms, and Z the largest entity size in the text collection. Since function $UpdateWordFreq$ used in lines 11 and 16 has a time complexity of $O(M)$, and is inside two cycles, in lines 6 and 8, that go through all terms in the text collection,

Algorithm 5.5 Entity Calculation by Frequency

Parameters:
Input: Document Collection
Output: *docs* updated
Using:
Set of sets nextWord as nw
Set of sets nextWordFreq as nwf
number of words as now
Entity count as E

1: $E, lastLevel \leftarrow 0$
2: $toDelete \leftarrow \{\}$
3: $now \leftarrow 2$
4: $nw, nwf \leftarrow ResetCountStructures(nw, nwf, M)$
5: **while** $\#nw > 0$ **do**
6: **for** $j = 1$ to N **do**
7: $docI \leftarrow ApplyL3Settings(d_j)$
8: **for** $k = now \rightarrow \#docI$ **do**
9: **if** $now = 2$ **then**
10: **if** $length(string(docI(k-1))) \neq 1$ **and not** $IsNumber(string(docI(k-1)))$ **then**
11: $nw, nwf \leftarrow UpdateWordFreq(nw, nwf, docI(k-1), docI(k))$
12: **end if**
13: **else**
14: $entityNumber \leftarrow Entity(docI, k - now + 1, k - 1)$
15: **if** $entityNumber \geq 0$ **then**
16: $nw, nwf \leftarrow UpdateWordFreq(nw, nwf, entityNumber - lastLevel, docI(k))$
17: **end if**
18: **end if**
19: **end for**
20: **end for**
21: $nw, nwf, toDelete \leftarrow StoreEntities(nw, nwf, now, lastLevel, toDelete)$
22: **if** $now > 2$ **then**
23: $lastLevel \leftarrow lastLevel + \#nw$
24: **end if**
25: $UpdateEntityHashTable(lastLevel + 1, E)$
26: $nw, nwf \leftarrow ResetCountStructures(nw, nwf, E - lastLevel)$
27: $now \leftarrow now + 1$
28: **end while**
29: $entity, E \leftarrow DeleteEntities(toDelete)$
30: **for** $i = 0 \rightarrow \#ehtable$ **do**
31: $ehtable(i) \leftarrow \{\}$
32: **end for**
33: $UpdateEntityHashTable(1, E)$

the complexity becomes $O(K.M)$, but all this is done inside a cycle in line 5 that will iterate from 2 until the largest sized entity is found. If we define Z as the size in number of words of the largest entity, the time complexity became $O(K.M.Z)$.

The space complexity of Algorithm 5.5 is the size of the local data structures, that are nw and nwf. Considering reasonable values for $minimalEntityFreq$, the largest size of this data structures will occur in the first cycle, with the word pairs. In the worst case those sets will have one set for each distinct term, at most with all distinct terms, making a space complexity of $O(M^2)$. If an n-dimensional counter (n-dimensional array) is used, the space complexity would be $O(M^Z)$, that could make it impossible to calculate the frequencies of sequences of words greater than 3 or 4. Since our algorithm re-uses the space, in line 26 reset count structure, a static counter is not needed.

The Algorithm 5.6 resets the nw and nwf count structures, with $size$ empty sets. It will be called for sequences of 2 words with the size of M, and for the other levels, with the size equal to the number of entities in the last level.

Algorithm 5.6 Reset Count Structures

Parameters:
Input: $nw, nwf, size$
Output: nw, nwf
Using:
Set of sets nextWord as nw
Set of sets nextWordFreq as nwf

1: $\#nw \leftarrow size$
2: $\#nwf \leftarrow size$
3: **for** $i = 1 \rightarrow size$ **do**
4: $\quad nw(i) \leftarrow ()$
5: $\quad nwf(i) \leftarrow ()$
6: **end for**

The Algorithm 5.7 updates the data structures next word, and next word frequency. If $word$ was already found after $entity$, then the frequency is updated, otherwise the new word is inserted in set of $entity$. In the first level, the $entity$ are just the existing words. Since any word can be inserted in the selected set, and exists M different words, the complexity of finding the word is $O(M)$, and so is the time complexity of the algorithm. If the sets are sorted, the time complexity of function $Find$ drops to $O(logM)$, but inserting a new term requires $O(M)$, keeping the complexity of the algorithm in $O(M)$. We prefer the first option, treating the sets

as sequences.

Algorithm 5.7 Update Word Frequency

Parameters:
Input: $nw, nwf, entity, word$
Output: nw, nwf
Using:

1: **if** $0 \leq entity < \#nw$ **and** $length(string(word)) > 1$ **and not** $IsNumber(string(word))$ **then**
2: $i \leftarrow Find(nw[entity], word)$
3: **if** $i < 0$ **then**
4: $nw(entity) \leftarrow nw[entity] + (word)$
5: $nwf(entity) \leftarrow nwf[entity] + (1)$
6: **else**
7: $nwf(entity)(i) \leftarrow nwf(entity)(i) + 1$
8: **end if**
9: **end if**

The Algorithm 5.8 goes through all sequences of words considered in the current cycle, and if the frequency is greater than $entityMinimalFreq$, create an entity for that sequence. The frequency of the sequence is also stored, and updated the frequency of sub-sequences of the current one, that must have is total frequency subtracted. If this subtraction does not take place, a sequence of words (A, B) with frequency 150, would kept as a valid entity even if the sequence (A, B, C) has a frequency of 120. This means that sequence (A, B) as a frequency of 150, but 120 exist due the sequence (A, B, C) exists, and not by the (A, B) itself. We subtract the 120 to the 150 of (A, B), and if the remaining value is less than $entityMinimalFreq$, the entity (A, B) is inserted in the set of entities marked to be deleted. For simplicity the algorithm does not have other sub-sequences, but the frequency of a sequence (A, B, C) is subtracted not only on (A, B) but also on (B, C).

The Algorithm 5.9 update the entity hash table starting at entity number $from$ and ending at entity number *to*. This algorithm assume the *ehtable* structured reset to empty sets, and can be used to update the entities in each level, and in the end, after removing the entities marked to delete, is called to update all the remaining entities.

The Algorithm 5.10 just return the entity number of a set of words, like the function $Word$ for words, but without inserting a new entity if it not exits. The entities are inserted in the hash table in Algorithm 5.9.

The Algorithm 5.11 convert *doc* in *docI* with the words replaced by its roots and the high

Algorithm 5.8 Store Entities

Parameters:
Input: $nw, nwf, now, lastLevel, toDelete$
Output: $nw, nwf, toDelete$
Using:
Entity count as E
Entity number i as $entity(i)$
Frequency of entity number i as $entityFreq(i)$

1: **for** $i = 0 \rightarrow \#nw$ **do**
2: **for** $j = 0 \rightarrow \#nw(i)$ **do**
3: **if** $nwf(i)(j) \geq entityMinimalFreq$ **then**
4: $E \leftarrow E + 1$
5: **if** $now = 2$ **then**
6: $entity(E) \leftarrow (i, nw(i)(j))$
7: **else**
8: $entity(E) \leftarrow entity(i + lastLevel) + nw(i)(j)$
9: **end if**
10: $entityFreq(E) \leftarrow nwf(i)(j)$
11: **if** $now > 2$ **then**
12: $entityFreq(i + lastLevel) \leftarrow entityFreq(i + lastLevel) - entityFreq(E)$
13: **if** $entityFreq(i + lastLevel) < entityMinimalFreq$ **then**
14: $toDelete \leftarrow toDelete \cup \{i + lastLevel\}$
15: **end if**
16: **end if**
17: **end if**
18: **end for**
19: **end for**

Algorithm 5.9 Update Entity Hash Table

Parameters:
Input: $from, to$
Output:
Using:
hash table size as $htsize$
hash function as $Hash$
entity hash table as $ehtable$
hash value as $hvalue$

1: **for** $i = from \rightarrow to$ **do**
2: $hvalue \leftarrow Hash(entity(i)) \bmod htsize$
3: $ehtable(hvalue) \leftarrow ehtable(hvalue) \cup \{i\}$
4: **end for**

Algorithm 5.10 Entity

Parameters:
Input: *entity* as *ent*
Output: *entity number* as *id*
Using:
hash table size as *htsize*
hash function as $Hash$
entity hash table as *ehtable*
hash value as *hvalue*

1: $hvalue \leftarrow Hash(ent) \bmod htsize$
2: **for all** $t \in ehtable(hvalue)$ **do**
3: **if** $ent = entity(t)$ **then**
4: $id \leftarrow t$
5: **return**
6: **end if**
7: **end for**
8: $id \leftarrow -1$

freq words removed. This level is ideal for matching sentences, since it uses the word root information. Note that function *Root* does not call *ApplyL2Settings*, since that operation is done only once for each word, and its root index stored for each word, reducing function *Root* to just a memory access.

Algorithm 5.11 Apply L3 Settings

Parameters:
Input: *doc*
Output: *docI*
Using:
Function root as *Root*

1: $docI \leftarrow ()$
2: **for** $i = 0 \rightarrow \#doc$ **do**
3: **if** $doc[i] \geq 0$ **and not** HighFreq(Root(doc[i])) **then**
4: $docI \leftarrow docI + Root(doc[i])$
5: **end if**
6: **end for**

The inverted index for entities is done like words in level 2, to allow the use of entities with the same flexibility as words.

The use of Wikipedia to extract entities and the algorithm presented to calculate entities by frequency are new and allow us to achieve a good performance. The entity inverted index will be used later in the document retrieval component, and will be described in Chapter 6.

5.5 Conclusions

In this chapter we have described the data structures used on the IR system we developed, IdSearch.

In level 1 it stores document collection text as lists of term numbers, instead of strings, using less space, with the conversion taking linear time complexity with the number of tokens. Since term numbers are stored, testing if a word is in a sentence will require less operations, as one integer comparison is enough to test if two terms are equal. The data structure requires an average of 4.28 bytes/word, when the string version requires an average of 10.43 bytes/word, for the CLEF data collection. For the best of our knowledge this data structure is new in QA context. Punctuation marks are kept separated from words, allowing them to be used in passage segmentation without deterioration in the document retrieval performance.

In level 2 with the inverted index, classes of words (root) and language uniformization, we provide a data structure capable of supporting fast boolean queries. This level also has linear time complexity with the number of terms. It is however much faster than level 1, since the passage through all terms means dealing numbers, since terms are already converted to numbers, contrary to what happens in level 1, in which there is a pass through all term strings.

In level 3 we present a new method to obtain entities by frequency, and we index the entities to allow its use in boolean queries. The process to check if a sequence of terms is an entity takes constant time complexity, allowing to use entities with the same efficiency as we using words. The task of automatically obtaining entities by frequency is the slowest in the indexation process, and takes $O(K.M.Z)$, where Z is the length of the largest entity in the collection measured in number of terms, M the number of distinct terms, and K the total number of terms. This can be considered a good time complexity, considering the amount of information involved. In terms of space complexity, it requires $O(M^2)$, avoiding the allocation of a counter for all possible sequences of words, that would need $O(M^Z)$. This allows us to calculate the frequency entities of any number of terms without memory problems. Entities are also extracted from Wikipedia resulting in a high quality entity set, improving the performance of several components: document retrieval, passage extraction, and answer extraction.

The index files for the text collection[9] occupy 1.15 GB of disk space, and took about 4 hours

[9]The text collection occupies around 9 GB of disk space, in over 600,000 files.

to build. The load time is around 1 minute, and the time to process 200 questions is less than 1 minute. These values correspond to tests using a machine with an AMD Athlon 64 processor (2.21 GHz), with 4GB of RAM, running Windows XP. These results support the achievement of one of the goals of the system, efficiency.

III

IdSay

Introduction to Part III

In Chapter 6 the algorithms of all components used in the search for an answer are proposed. IdSay uses a classical architecture of Document Retrieval, Passage Retrieval, Answer Extraction and Answer Validation, with all components being original and fully described in his chapter. It includes a new strategy for removing the most frequent word in the question and making another pass through the components, in the case no answers could be produced for the current question.

In Chapter 7 the results of IdSay at QA@CLEF 2008 are first analysed globally, followed by a question based analysis. The developed web application is presented, that was a valuable tool in the understanding of the origins of failures.

In Chapter 8 several improvements are added to IdSay. A new method for the construction and usage of word/entity synonyms based on Wikipedia redirecting pages, the WES base, Wikipedia Entity Synonyms, is proposed. The usage of the WES base, together with the TeP base, a manually built Thesaurus for the Portuguese Language, allows the matching to be done using synonyms, not only for words but also at entity level.

A method for the incorporation in Question Answering of ontological knowledge using the Wikipedia category structure is proposed.

Improvements were made to the numerical values normalization, avoiding the decimal separator to be mistaken with a separator/terminator character. Abbreviations and acronyms normalization was introduced, eliminating the use of terminator character in these cases. Regarding date treatment, support of ancient dates including Roman notation was introduced. A scoring mechanism giving higher importance to passages with less equivalences was introduced, together with a better method for lemma selection in case of ambiguity.

6 IdSay Components

6.1 Introduction

In this chapter we will describe the IdSay components, from the reception of a question to the delivery of an answer, as illustrated in Figure 2.2, which we reproduce again here.

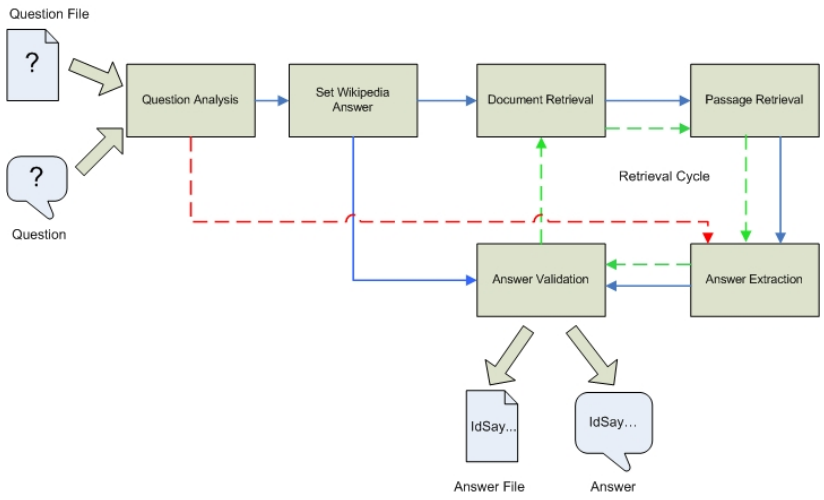


Figure 6.1: IdSay system architecture

Algorithm 6.1 is responsible to start IdSay for a given question, and it integrates the global components and deals with cluster questions. In the beginning, the Question Analysis component is called, to analyse the question and set some global variables relative to the question that are identified from the question text. These variables are described in Table 6.1. Level 1 settings are applied to the question text, and they transform all words to integers (returned as $questionInt$). This function will also find the numeric equivalent of qRE, question reference entity, returned as $qREInt$.

This algorithm defines the procedure followed for cluster questions: If we are working with cluster questions, we need to save information from the first cluster question to the other cluster

questions, or use information from the answer to the first cluster question. That is done by saving the question reference entity, and add it to the other questions in the same cluster, in $UpdateCluster$.

After the Question Analysis component, if the question is a definition question ($qTEC = D$), and a question reference entity is found, we use the Set Wikipedia Answer, SWAN, component. If that component does not return any answers, then the normal procedure is done, following the retrieval cycle.

Before entering the retrieval cycle, we transform the words to its roots. The retrieval cycle returns answers, the supporting passages and the corresponding document identification to those answers, which are IdSay final result.

Algorithm 6.1 Ask Question

Parameters:
Input: *question*
Output: *answers*
Using: $beginCluster, clusterQuestionInt, clusterqREInt$

1: $questionInt, qREInt, qTE, qTEC, qTET \leftarrow QuestionAnalysis(question)$
2: **if** $beginCluster$ **then**
3: $clusterQuestionInt \leftarrow questionInt$
4: $clusterqREInt \leftarrow qREInt$
5: **else**
6: $questionInt, qREInt \leftarrow UpdateCluster(questionInt, qREInt, clusterQuestionInt, clusterqREInt)$
7: **end if**
8: **if** $qTEC = D$ **and** $qREInt \neq ()$ **then**
9: $questionInt \leftarrow qREInt$
10: $answers \leftarrow WikipediaDefinition(questionInt)$
11: **if** $answers \neq \{\}$ **then**
12: **return**
13: **end if**
14: **end if**
15: $questionInt \leftarrow questionInt + qREInt$
16: $questionRoot \leftarrow ApplyL3Settings(questionInt)$
17: $answers \leftarrow RetrievalCycle(questionRoot, qREInt)$

Algorithm 6.2 updates the $questionInt$ and $qREInt$ of each of the subsequent questions in the cluster, with information of the topic of the cluster. Pronouns in the question are replaced by the question cluster reference entity ($clusterqREInt$). If in the first question of the cluster no reference entity was found, the full text of the first cluster question is joint to the question. Function $Pronoun$ will return true if $word$ is in the following list: ele, ela, eles, elas, dele, dela,

deles, delas, lhe, lhes, seu, sua, seus, suas, nessa, nessas, nesse, nesses.

For example, with pronoun replacement the question: "Quando é que ele nasceu?" [When was he born?], for a cluster that had as topic Afonso Henriques[1], the resulting question would be: "Quando é que Afonso Henriques nasceu?" [When was Afonso Henriques born?]. If there is no pronoun, the *clusterqREInt* is just added to the end of the question. It is important to ensure that *qREInt* exists in the supporting passages, in order to relate it to the topic of the cluster. Adding *clusterqREInt* to the end of the question will ensure that.

Algorithm 6.2 Update Cluster

Parameters:

Input: $questionInt, qREInt, clusterQuestionInt, clusterqREInt$

Output: $questionInt, qREInt$

Using:

$Pronoun(word)$ function, that return true if *word* is a pronoun

$Replace(sentence, find, replace)$ function, that return *sentence* after replacing $find$ by *replace*

```
1:  if clusterqREInt ≠ {} then
2:    for word ∈ questionInt do
3:      if Pronoun(word) then
4:        questionInt ← Replace(questionInt, word, clusterqREInt)
5:        return
6:      end if
7:    end for
8:    qREInt ← qREInt + clusterqREInt
9:  end if
10: questionInt ← questionInt + clusterQuestionInt
```

The SWAN component (Set Wikipedia Answer) determines if there is a Wikipedia entry for the reference entity of the question. In the positive case, and if we are looking for a definition, the beginning of the corresponding Wikipedia page is considered as an answer.

Algorithm 6.3 explores the several reference entity equivalences, and verifies the ones with a wikipedia page. An answer is returned for each wikipedia page, as a definition. Each answer is a pair with the answer and a set of supports. Each support is a pair (document, passage).

Algorithm 6.4 implements the main loop, the Retrieval Cycle, stopping only when enough answers are found, or *questionRoot* is empty. In each cycle the *DocumentRetrieval* is called for the words in *questionRoot*. The retrieved documents are forwarded to the *PassageRetrieval*

[1]The first king and founder of Portugal.

Algorithm 6.3 SWAN - Set Wikipedia Answer

Parameters:

Input: $questionInt$

Output: $answers$

Using:

$WikiDoc(entities)$ function returns the documents of wikipedia with title equal to one element of $entities$

$FindEntities(sentence)$ function returns the set of entities in the $sequence$

$ExtractWikipediaDefinition(doc)$ function extract the first paragraph of the wikipedia document given

1: $answers \leftarrow \{\}$
2: $resultDocs \leftarrow WikiDoc(FindEntities(ApplyL3Settings(questionInt)))$
3: **for** $doc \in resultDocs$ **do**
4: $\quad answers \leftarrow answers \cup \{(doc, ExtractWikipediaDefinition(doc))\}$
5: **end for**

that looks for passages with all words in *questionRoot* and reference entity. If there are no documents containing all the words and reference entity, no results are produced and we proceed to a new cycle, after removing the most frequent word in *questionRoot*. The passages produced by *PassageRetrieval* are treated by *AnswerExtraction* that extracts answers from them. Since the words could occur in the same document but disperse, and passages have a maximum size limit, it is possible that no passages are returned. In that case the system proceeds to the next cycle. For a new cycle the most frequent word is removed. The *resultPassages* and *answers* are updated and not replaced by *PassageRetrieval*, allowing the accumulation of answers from several cycles. When the cycle produces enough answers, or *questionRoot* is empty, we call *AnswerValidation* to select answers. In the case that we were not able to find any answers the systems returns NIL.

The proposed method for removing the most frequent word in each cycle, is a new strategy, that prevents words from being discarded unless we do not have enough answers. In that case, it could happen that an highly frequent word in the question is preventing a valid passage from being retrieved, and its removal allows those passages to be returned. We believe this be a better use of the concept behind stop lists, since we remove the words only if he cannot find passages, and use them if they can help to locate more refine passages.

Algorithm 6.4 Retrieval Cycle

Parameters:
Input: $questionRoot, qREInt$
Output: $answers$
Using: $numberAnswers$
$RemoveHighestFreqWord(sentence)$ function return $sentence$ without the word that has the highest document frequency

```
1: answers, resultPassages ← {}
2: while #answers < numberAnswers and questionRoot ≠ {} do
3:   resultDocs ← DocumentRetrieval(questionRoot, qREInt)
4:   if resultDocs ≠ {} then
5:     resultPassages ← PassageRetrieval(resultPassages, resultDocs, questionRoot, qREInt)
6:     if resultPassages ≠ {} then
7:       answers ← AnswerExtraction(answers, questionRoot, resultPassages)
8:     end if
9:   end if
10:  questionRoot ← RemoveHighestFreqWord(questionRoot)
11: end while
12: if answers ≠ {} then
13:   answers ← AnswerValidation(answers, resultPassages)
14: end if
```

6.2 Question Analysis

In this component the question is analysed, and a set of global variables is initialized. Table 6.1 lists the names of these variables and their description.

Table 6.1: Global Variables set in Question Analysis

Variable	Name	Description
qRE	Reference Entity	The reference entity in the question, that can be a single one or more than one.
qTE	Target Entity	Information on the entity we are searching for as an answer.
qTET	Target Entity Type	The type of target entity we are looking for.
qTEC	Question Category	The question category.

This component was built based on heuristic rules for questions in Portuguese that included a close analysis of the questions for Portuguese at QA@CLEF from years from 2004 to 2007.

The variables to be assigned have the following values

- First we expect that if the question was written in Natural Languages, it probably follows the usual conventions in this case, which is to for instance to capitalize the names of people the question is about. Based on that assumption we inspect the question string and if there are capitalized words, or words between quotation marks or guillemots, those words are considered as qRE. The words in the qRE will be treated as entities, or as units that cannot be taken out if there are more than one retrieval cycle.
- the question category, $qTEC$, that can have one of the following values: F - Factoid, D - Definition, L - List according to the classification of QA@CLEF.
- the Target Entity Type $qTET$. For this variable the following possibilities are considered: 0 - Person, 1 - Date/Time, 2 - Location, 3 - Organization, 4 - *not used*, 5 - Count/Measure, 6 - Object and 7 - Other.
- qTE, that may contain additional information on the entity we are targeting for as an answer. This variable is used when the target entity type is measure. This variable stores the names of the units acceptable for the measure we determine we are looking for, from the analysis of the question text. We store comprehensive information on the several systems of measures and the corresponding units in authority lists. The concept is described in (Prager 2006) but we do not make extensive use of it as a gazetteer, only in this case that relates to information that is limited and unlikely to suffer from frequent changes. The authority list we use are described in Table 8.3, where we make a description of the measures considered, in the form they occur in the question and the form used to classify the answers. If the question contains the former, qTE will be set to the latter. The variable qTE can be used to store slightly different kinds of information, according to the question content, that we will detail in the curse of this section.

In Algorithm 6.5 we describe the component Question Analysis. This algorithm uses other algorithms also described in this section, that will have access to the variables in Table 6.1, for reading and update, but the variables are not present in the parameters of all algorithms, for simplicity sake.

In the first 3 lines we initialize the variables to be returned. If no information can be drawn,

the target entity and target entity type will be empty, the reference entity will be a null sequence, and question category will be a factoid by default.

In line 4 we call *SetDelimiters* function, that will insert delimiters in capitalized words, if they exist, to allow the recognition of the reference entity. An example is question Quem é Nelson Mandela? [Who is Nelson Mandela?] that becomes Quem é "Nelson Mandela"?.

In lines 5 to 7 we call *ExtractqRE* function, that will remove the text between delimiters from the *question*, attributing it to *qRE*. For instance, in the above example, Quem é "Nelson Mandela"? becomes becomes Quem é " "?, with *qRE* =Nelson Mandela.

In lines 8 and 9 to *question* and *qRE* level 1 settings are applied, resulting in *questionInt* and *qREInt* respectively. In line 10 we check if any of the words in the question does not belong to the data collection. If that is the case, we return immediately without further processing.

The main propose of function *Find* is to verify if an expression is in the question, and it can accept simple regular expressions, and can remove the expression or not depending on the argument *remove*. This function will be used by all algorithms in this section. The three functions used here, *SetDelimiters*, *ExtractqRE*, and *Find* are not defined since its implementation is straightforward.

Line 13 uses function *Find* to remove the expression é que, that is quite common in questions in Portuguese, but which adds no information that is worth being processed. For example, in question Com que idade é que Michael Jackson começou a cantar no grupo "Jackson Five"? [At what age did Michael Jackson start to sing in the "Jackson Five" band?], the expression "é que" can be removed without any loss.

Line 14 determines if the question is in the imperative mode, ending with a full stop as for instance in the case Nomeie uma das sete maravilhas do mundo. [Name one of the seven wonders of the world.]. In that case, the procedure *ProcessFullStopQ* will be called.

If the question ends with a punctuation mark, a terminator, the terminator is removed (lines 17-19). *ProcessInterrogationQ* is called in line 20, and does the main processing of a question. In the end of the *QuestionAnalysis*, in lines 21 to 23, we check if the target entity type expects a Date/Time or a Count/Measure answer, and in that case we remove the pronoun "um" from the question, since it can be both an article or the number 1, and if is kept in the question, the answers returned cannot contain the number 1. For example, in question Quando é que houve

um golpe de estado em Chipre? [When did a coup d'etat happen in Cyprus?] the article "um" would remain without this rule, preventing dates with 1 to be considered, which would mean that the correct answer, 1974, would not be extracted.

Algorithm 6.5 Question Analysis

Parameters:

Input: $question$

Output: $questionInt, qREInt, qTE, qTEC, qTET$

Using:

$SetDelimiters(question)$ insert delimiters in capitalized words.

$ExtractqRE(question, delimiterLeft, delimiterRight)$ If $question$ as the delimiter, the qRE is assigned and the content cleaned

$Find(sentence, expression, remove)$ returns true if $expression$ is find in $sentence$, that can include regular expressions. If $remove$ is true, the expression is also removed from the sentence, if it exists.

1: $qTE, qTET \leftarrow \{\}$
2: $qREInt \leftarrow ()$
3: $qTEC \leftarrow F$
4: $question \leftarrow SetDelimiters(question)$
5: $question, qRE \leftarrow ExtractqRE(question,\text{“,”})$
6: $question, qRE \leftarrow ExtractqRE(question,\text{’,’})$
7: $question, qRE \leftarrow ExtractqRE(question,\text{«,»})$
8: $qREInt \leftarrow ApplyL1Settings(qRE)$
9: $questionInt \leftarrow ApplyL1Settings(question)$
10: **if** $Find(questionInt + qREInt, \{\}, false)$ **then**
11: **return**
12: **end if**
13: $Find(questionInt,$ “é que”$, true)$
14: **if** $questionInt(\#questionInt) =$ “.” **then**
15: $ProcessFullStopQ()$
16: **end if**
17: **if** $questionInt(\#questionInt) = terminator$ **then**
18: $questionInt \leftarrow (questionInt(1), ..., questionInt(\#questionInt - 1))$
19: **end if**
20: $ProcessInterrogationQ()$
21: **if** $qTET \cap \{1, 5\} \neq \{\}$ **then**
22: $Find(questionInt,$ “um”$, true)$
23: **end if**

Algorithm 6.6 processes the questions in the imperative mode, and the result will also undergo the process of questions in the interrogative mode. The first word is removed, since it is a verb without useful information. In line 3 we remove the words that ask a name in singular, for example Diga o nome de um jornal francês. [Indicate the name of a french newspaper.] becomes um jornal francês. [a french newspaper.]

The question can also be in the plural, and in that case, the test in line 4 will be successful and the question category variable is assigned to a list type in line 5. For example, the question Diga os nomes dos três Beatles que ainda estão vivos. [Indicate the names of the three Beatles that are still alive.] is identified as a list question and becomes três Beatles que ainda estão vivos. [three Beatles that are still alive.].

Algorithm 6.6 Process Full Stop Q

Using: $questionInt$

$Find(sentence, expression, remove)$ same as Question Analysis.

```
1: if #questionInt > 0 then
2:   questionInt ← (questionInt(2), ..., questionInt(#questionInt))
3:   Find(questionInt, "[o|um] nome [de|da|do]", true)
4:   if Find(questionInt, "[o|os] [nome|nomes] [de|da|do|das|dos]", true) then
5:     qTEC ← L
6:   end if
7: end if
```

In Algorithm 6.7 we process the questions ended with a question mark, and also the imperative questions after they were processed by $ProcessFullStopQ$. In the first line we set a local variable with the forms of verb to be in the third person, plural and singular, and in line 2 we process the expression "que" followed by one form of the verb to be, by calling $ProcessMatchExpression$. For example, the question O que era o Granma? [What was Granma?] would have a positive match. If the question has a positive match with the expression, the expression is removed from the question, and question category is set as a definition question if the rest of the question have not more than 2 words, otherwise the question category remains as a factoid. Using the above example again the question would became o Granma?, and the question category is set to 'D' definition type.

In line 3 we call $ProcessTimeQ$ to check for date/time questions, and in line 4 we check if a "que" remains in the question. In that case, we check in line 5 if a measure is in the question, we insert the units allowed for that measure, using the corresponding authority list, and set the question category for Count/Measure type. For example, the question A que altitude está a camada de ozono? [At what altitude is the ozone layer?] contains the name of a measure "altitude" [altitude], so the units related to this measure will be inserted in qTE. The variables with the authority lists *measures* and *units* are described in Table 8.3. In line 6 we process the authority list for locals, but this authority list contains only the generic names used for locations, not a complete list of locations. If qTE is not set to any of the previous authority

lists, the word next to the word "que" is used as qTE (lines 7-9).

In line 11 we process the expression "quem" [who] followed by one of the stored forms of the verb to be, by calling $ProcessMatchExpression$, and in that case the question type is set to person or organization. For example, Quem é Andy Warhol? [Who is Andy Warhol?] becomes Andy Warhol? with the question type a person or organization.

The word "qual" [what] is tested in line 12, and in the positive case the authority lists *measures* and *units* are used, as well the *locals* and *categoryLocals*. For example, que question Qual a área da Selva Lacandona? [What is the area of the Lacandon Jungle?] match with the measure área [area], setting qTE to the appropriate units.

If the question contains the word "quais" [which] the question category is set as a list type, like the question Quais são as três repúblicas bálticas? [Which are the three Baltic States?]. The words "[quanta|quanto]" [how much] is tested in line 19, and the authority lists *measures* and *units* are processed, like in question Quanto custou o Túnel da Mancha? [How much cost the Channel Tunnel?] that have a measure match custou [cost].

Next are the words "[quantas|quantos]" [how much] but in this case the authority list *units* is used also in the question and in the answer. For example, the question Quantas toneladas pesava o telhado de cimento de um supermercado em Nice que se partiu? [How many tons did the concrete roof of a supermarket in Nice that was broken weight?] have the units toneladas [tons] of a measure peso [weight]. In this case, if no unit is found, the target entity is the first word (after the words "[quantas|quantos]", since they are removed from the question), and the question category type of Count/Measure type. This case is more common, for instance, Quantos budistas há em Espanha? [How many Buddhists are there in Spain?] the target entity is set to budistas [Buddhists].

Finally, in lines 29 to 34, if the question contains the words "quando" [when] or "onde" [where], the target entity type will be set to Date/Tame and Location respectively. For instance, the questions Quando foi a independência de Cabo Verde? [When was the Cape Verde independency?] and Onde se vende haxixe nas «coffee shops»? [Where is hashish sold in «coffee shops»?] will set have the target entity type set to Date/Tame and Location respectively.

Algorithm 6.7 Process Interrogation Q

Using: $questionInt, qREInt, qTE, qTEC, qTET$
$Find(sentence, expression, remove)$ same as Question Analysis.
$NextQue()$ return the word next to "que" word.

1: $formsToBe \leftarrow$"[é|são|foram|era|eram|será|serão]"
2: $ProcessMatchExpression(\text{"}que\text{"} + formsToBe, \{\})$
3: $ProcessTimeQ()$
4: **if** $Find(questionInt, \text{"}que\text{"}, true)$ **then**
5: $ProcessAuthorityList(measures, units, 5)$
6: $ProcessAuthorityList(locals, categoryLocals, 2)$
7: **if** $qTE = \{\}$ **then**
8: $qTE \leftarrow qTE \cup \{NextQue()\}$
9: **end if**
10: **end if**
11: $ProcessMatchExpression(\text{"}quem\text{"} + formsToBe, \{0, 3\})$
12: **if** $Find(questionInt, \text{"}qual\text{"}, true)$ **then**
13: $ProcessAuthorityList(measures, units, 5)$
14: $ProcessAuthorityList(locals, categoryLocals, 2)$
15: **end if**
16: **if** $Find(questionInt, \text{"}quais\text{"}, true)$ **then**
17: $qTEC \leftarrow L$
18: **end if**
19: **if** $Find(questionInt, \text{"}[quanto|quanta]\text{"}, true)$ **then**
20: $ProcessAuthorityList(measures, units, 5)$
21: **end if**
22: **if** $Find(questionInt, \text{"}[quantos|quantas]\text{"}, true)$ **then**
23: $ProcessAuthorityList(units, units, 5)$
24: **if** $qTE = \{\}$ **then**
25: $qTET \leftarrow qTET \cup \{5\}$
26: $qTE \leftarrow qTE \cup \{questionInt(1)\}$
27: **end if**
28: **end if**
29: **if** $Find(questionInt, \text{"}quando\text{"}, true)$ **then**
30: $qTET \leftarrow qTET \cup \{1\}$
31: **end if**
32: **if** $Find(questionInt, \text{"}onde\text{"}, true)$ **then**
33: $qTET \leftarrow qTET \cup \{2\}$
34: **end if**

Algorithm 6.8 try to match an expression in the question, and in that case it change the question category to definition if the remaining question have only two words, or left the question category in factoid type. If a positive match is found, the $qTETvalues$ are also added to the target entity types, and if the $qREInt$ is empty, is set to the $questionInt$ without articles.

Algorithm 6.8 Process Match Expression

Parameters:
Input: $expression, qTETvalues$
Using: $questionInt, qREInt, qTEC, qTET$
$Find(sentence, expression, remove)$ same as Question Analysis.

1: **if** $Find(questionInt, expression, true)$ **then**
2: **if** $\#questionInt \leq 2$ **then**
3: $qTEC \leftarrow D$
4: **end if**
5: $qTET \leftarrow qTET \cup qTETvalues$
6: **if** $qREInt = ()$ **then**
7: $qREInt \leftarrow questionInt$
8: $questionInt \leftarrow ()$
9: $Find(qREInt$,"[o|a|os|as|um|uma|uns|umas]", $true)$
10: **end if**
11: **end if**

In Algorithm 6.9 the time questions are verified. If the question contains "em que" [in which] the several units of time measure are verified, and if one of them belongs to the question, Date/Time is added to $qTET$. For example, the question Em que século foi fundada a Skoda? [In which century was Skoda founded?] has the time unit século [century] giving a positive match.

Algorithm 6.9 Process Time Q

Using: $questionInt, qTET$
$Find(sentence, expression, remove)$ same as Question Analysis.

1: **if** $Find(questionInt, "em\ que", false)$ **then**
2: **for** $i = 1 \rightarrow \#units(time)$ **do**
3: **if** $Find(questionInt, units(time)(i), true)$ **then**
4: $Find(questionInt, "em\ que", true)$
5: $qTET \leftarrow qTET \cup \{1\}$
6: **end if**
7: **end for**
8: **end if**

Algorithm 6.10 processes authority lists, and it has two arguments, the $authorityListQuestion$, and the $authorityListAnswer$. If the target entity type is empty, the first authority list is checked to verify if any of its values exist in the question. If so, the second

authority list is joint to the target entities, and also the target entity types are added the given $qTETvalue$. This function will mainly be called with authority list *measures* and *units*, resulting in checking if any measure is in the question, and if one is found, the corresponding units of that measure are added to the target entity variable.

Algorithm 6.10 Process Authority List

Parameters:
Input: $authorityListQuestion, authorityListAnswer, qTETvalue$
Using: $questionInt, qTE, qTET$
$Find(sentence, expression, remove)$ same as Question Analysis.

```
1:  if qTET = {} then
2:    for i = 1 → #authorityListQuestion do
3:      for j = 1 → #authorityListQuestion(i) do
4:        if Find(questionInt, authorityListQuestion(i)(j), false) then
5:          qTET ← qTET ∪ {qTETvalue}
6:          qTE ← qTE ∪ authorityListAnswer(i)
7:          return
8:        end if
9:      end for
10:   end for
11: end if
```

The Question Analysis component presented is specific to the Portuguese language, and its rules are based in the questions of QA@CLEF. This is the component to change for new questions types, or for implementing a new language. The framework proposed can be easily improved by adding new rules or authority lists. From the results of Chapter 7, we can state that it archived a very high accuracy.

6.3 Document Retrieval

The Document Retrieval component is described in Algorithm 6.11. We intersect the documents index of each word, and also the document index for each entity present in the question. Since the document indexes are sorted by document number, the intersection is done in linear time with the number of documents, contrary to the quadratic time it would take if the lists were not ordered by document number but ordered by frequency of the term within the document.

This procedure have a complexity time of $O(Z \times N)$[2].

[2] Z: question size, N: number of documents.

Algorithm 6.11 Document Retrieval

Parameters:
Input: *questionRoot*
Output: *resultDocs*
Using: *docs*(*word*) returns the documents that contains *word*, structure updated in level 2.
EntityDocs(*entity*) returns the documents that contains *entity*, structure updated in level 3.
FindEntities(*sentence*) function that returns the set of entities in the *sequence*

1: $resultDocs \leftarrow docs(questionRoot(1))$
2: **for** $i = 2 \rightarrow \#questionRoot$ **do**
3: $\quad resultDocs \leftarrow resultDocs \cap docs(questionRoot(i))$
4: **end for**
5: $entities \leftarrow FindEntities(questionRoot)$
6: **for** $i = 1 \rightarrow \#entities$ **do**
7: $\quad resultDocs \leftarrow resultDocs \cap EntityDocs(entities(i))$
8: **end for**

The Document Retrieval module uses a boolean search, that returns documents without any scoring. Contrary to what is common in retrieval there are words in the query that are not searched using the "bag of words approach", but a specific index is considered for groups of words that **must occur together and preserving the specified order**, the entities. Ensuring that the entities appear in the documents, and afterwards passages, has a favourable impact in the number of documents retrieved. Also no passage will be returned that does not contain all entities. This can be done since the function *FindEntities*, that is described in Algorithm 6.12, can extract entities in a sentence efficiently.

This algorithm has two modes of extracting entities, the first one is used for finding entities of small size, that are the most frequent. It makes use of function *Entity* to check if sequence of words is an entity, that has constant time complexity due to the use of hash tables. The verification is done for all sequence of words in *sentence* with sizes between 1 and 5, which results in a time complexity of $O(\#sentence)$ for this phase . The second phase is for least frequent entities, for it would be a wast of time to verify all sequences of 6 or more words, whether they are an entity or not, when the number of entities of this size is relatively small. It is faster in this case to test if the entities exist in the sentence, and for that an auxiliary structure was created with all entities with size greater than 5. We cycle through that list and verify if the entity is in the sentence. The time complexity of this phase is $O(\#sentence \times \#highSizeEntity)$, but the constant number 5 can be selected in order to keep the size of the *highSizeEntity* low.

Algorithm 6.12 Find Entities

Parameters:
Input: *sentence*
Output: *entities*
Using:
$Entity(doc, start, end)$ return the number of entity if exists an entity $(doc(start), ..., doc(end))$.
$highSizeEntity$ a set with the entities with size greater than 5
$Find(sentence, what)$ return the position of *what* in *sentence*, or a negative number.

```
1:  entities ← {}
2:  for i = 1 → 5 do
3:    for j = 1 → #sentence − i do
4:      entity ← Entity(sentence, j, j + i)
5:      if entity ≥ 0 then
6:        entities ← entities ∪ {entity}
7:      end if
8:    end for
9:  end for
10: for i = 1 → #highSizeEntity do
11:   if Find(sentence, highSizeEntity(i)) ≥ 0 then
12:     entities ← entities ∪ {highSizeEntity(i)}
13:   end if
14: end for
```

6.4 Passage Retrieval

In Algorithm 6.13 the *PassageExtraction* is called to all documents received. The *resultPassages* and *docPassages* are updated (and not replaced) by the *PassageExtraction*. If the number of passages exceeds 10000, the algorithm does not process any further documents, returning the passages already extracted.

This stop criteria is necessary since *RetrievalCycle* can remove many words, and if few words remain that return a lot of documents, it is most likely that will be no answers. Searching with just a few words can be successful in some cases, and it allow us to not use an hard limit on the number of words in *RetrievalCycle*.

Algorithm 6.13 Passage Retrieval

Parameters:
Input: $resultPassages, resultDocs, questionRoot$
Output: $resultPassages$
Using:

1: $resultPassages \leftarrow \{\}$
2: **for** $i = 1 \rightarrow \#resultDocs$ **do**
3: $resultPassages \leftarrow PassageExtraction(resultPassages, questionRoot, resultDocs(i))$
4: **if** $\#resultPassages \geq 10000$ **then**
5: **return**
6: **end if**
7: **end for**

Algorithm 6.14 receives a document and extract passages. If a reference entity exists, The first step is to look for it and store in *occurrences* the positions in the document where the reference entity occurs. This will limit the passages extracted, since all of them must contain the reference entity.

The aim of this function is to produce all the passages the documents where the words we are looking for occur. The passage length should not exceed a given limit (currently, 60 words).

This is done in the following way: each document is searched once for the words of the question. Each time a word is found, its position in the document is registered. After storing this information for a newly found word, we check if all the words already have a position registered. If that is the case we check the total length of the passage by subtracting to the current position the minimum of the set of positions of all words. If the length does not exceed the limit we add this passage to the passage list.

In *questionPosition* vector, a number for each word in *questionRoot* is stored. It is the last position of that word. With this data, with a one time pass of the document, is possible to extract all passages with all words, with reference entity, and with smaller than a specific size. This procedure is later updated to accommodate the use of synonyms, requiring just an update in the test that compares a word in the question with a word in the document (Chapter8).

Algorithm 6.14 Passage Extraction

Parameters:
Input: $resultPassages, questionRoot, doc, resultPassages$
Output: $resultPassages$
Using: $maximalWordsInAnswer$

```
1:  occurrences ← {}
2:  if #qREInt>0 then
3:    for i = 1 → #doc do
4:      i ← Find(qREInt, doc, i)
5:      if i ≥ 0 then
6:        occurrences ← occurrences ∪ {i}
7:      end if
8:    end for
9:  end if
10: for i = 1 → #questionRoot do
11:   questionPosition[i] ← {}
12: end for
13: for i = 1 → #doc do
14:   for j = 1 → #questionRoot do
15:     if doc(i) = questionRoot(j) then
16:       questionPosition(j) ← i
17:     end if
18:   end for
19:   if i ∈ questionPosition and {} ∉ questionPosition then
20:     if questionPosition ∩ occurrences ≠ {} then
21:       size ← max(questionPosition) − min(questionPosition) + 1
22:       if size < maximalWordsInAnswer then
23:         min, max ← PassageAdjust(doc, min(questionPosition), max(questionPosition))
24:         resultPassages ← resultPassages ∪ {(doc, (doc(min), ..., doc(max)))}
25:       end if
26:     end if
27:   end if
28: end for
```

A little more detail is important here, since this procedure will be called for all documents returned. In the first part we locate occurrences of $qREInt$ in the document. This data allow to filter out the passages without $qREInt$ taking a complexity time of ($O(Ki \times \#qREInt)$ with Ki the size of a document i. This part have a lower complexity time then the rest of the algorithm, in exchanging with an increase of space complexity of $O(Ki)$ in worst case for allocating the local data structure *occurrences*. A vector *questionPosition* will then be initialized, and it contains the last position of each word in document. In line 19 we check if current position i was updated, meaning that a word in *questionRoot* is in position i, and all words were already

found in document, and in that case we will check if the passage located contains the entity (line 20), and if the size of passages is smaller than the allowed passage size (lines 21, 22). If this is the case, a valid passage is found, and we adjust the start and end of the passage (line 23) and store it in the $resultsPassages$ (line 24), storing a pair $(document, passage)$.

The worst case complexity time is $O(Ki \times Z \times logKi)$, with Ki the document size, and Z the question size. This complexity is due line 13 ($O(Ki)$), line 20 ($O(Z \times logKi)$), since *occurrences* vector is sorted allowing binary search and $questionPosition$ have size Z, each position must check if is in *occurrences*. Note that $PassageAdjust$ have a constant time complexity, and $resultPassages$ update is done in constant time also, since results are not sorted.

This algorithm has a low complexity time and can be used in large documents. It is an efficient way to extract passages, comparable in efficiency to other approaches, but original, to the best of our knowledge.

Algorithm 6.15 implements a simple method to avoid a passage from starting in the middle of a sentence.

First it starts by calculating the words available, and split in two (variable delta). Only delta words will be used to adjust the passage to the left, and to the right. Then it moves the start of the passage (variable from) until a terminator is found, or delta words are reached. In that case the process revert until a separator is found. A similar procedure is done to the right, leading the passage to start/end in a terminator, if possible, otherwise it could start/end in a separator. If there are no terminators or separators, the passage is returned without changes.

The time complexity of this algorithm is $O(maximalWordsInAnswer)$, and since it is a constant, it drops to $O(1)$. The existence of separators / terminators words is crucial in this procedure, otherwise the start and end of a sentence could even change the meaning of the sentence, or kept out some answers.

Algorithm 6.15 Passage Adjust

Parameters:
Input: $doc, from, to$
Output: $from, to$
Using: $maximalWordsInAnswer, nent, terminator, separator$

```
1: delta ← (maximalWordsInAnswer − to + from − 1)/2
2: count ← 0 {adjust left}
3: while from > 0 and count < delta and
   doc(from) ≠ nent and doc(from) ≠ terminator do
4:    from ← from − 1
5:    count ← count + 1
6: end while
7: if count = delta then
8:    while count > 0 and doc(from) ≠ separator do
9:       from ← from + 1
10:      count ← count − 1
11:   end while
12: end if
13: count ← 0 {adjust right}
14: while to < #doc and count < delta and
   doc(to) ≠ nent and doc(to) ≠ terminator do
15:   to ← to + 1
16:   count ← count + 1
17: end while
18: if count = delta then
19:   while count > 0 and doc(to) ≠ separator do
20:      to ← to − 1
21:      count ← count − 1
22:   end while
23: end if
```

6.5 Answer Extraction

Algorithm 6.16 extracts answers from the passages. First task is to have an index of the passages, by increasing order of size. In improved version of IdSay of Chapter 8, the index will use the passage score, instead of passage size. We then process each passage by increasing order of size, and call *ExtractAnswersFromPassage* method. The resulting answers will be added to the existent ones, and if an answer already exists, the support is added to the answer. Of course, the answer with most supports will later on in Answer Validation selected as the best answer, so repeated answers cannot be discarded. Verifying if an answer already exists (line 7) is done in constant time using hash tables.

The structure of variable *answers* is a set of answers, but each answer is a pair, with the answer and the set of its supports. Each support is a pair (document, passage).

Algorithm 6.16 Answer Extraction

Parameters:
Input: $answers, questionRoot, resultPassages$
Output: $answers$
Using:
$SortByPassageSize(resultPassages)$ returns an index of passages sorted by increasing passage size

```
1:  idPassage ← SortByPassageSize(resultPassages)
2:  for i = 1 → #idPassage do
3:    support ← resultPassages(idPassage(i))
4:    answersFromPassage ← ExtractAnswersFromPassage(answers, support, questionRoot)
5:    for j = 1 → #answersFromPassage do
6:      answer ← answersFromPassage(j)
7:      if (answer, passages) ∈ answers then
8:        passages ← passages ∪ support
9:      else
10:       answers ← answers ∪ {(answer, {support})}
11:     end if
12:   end for
13: end for
```

In Algorithm 6.17, depending of the $qTET$ and $qTEC$, determined by the Question Analysis component, one of the extract procedures is called. If we were not able to determine the type of answer, a generic answer procedure is called. An exception for definition questions, $qTEC$='D', where the generic answer procedure is always called.

Algorithm 6.17 Extract Answers From Passage

Parameters:
Input: $answers, support, questionRoot$
Output: $answers$
Using: $qTET, qTEC$

1: **if** $\#qTET = 1$ **and** $qTET(0) = 5$ **then**
2: $answers \leftarrow ExtractNumericAnswersFromPassage(answers, support)$
3: **else if** $\#qTET = 1$ **and** $qTET(0) = 5$ **then**
4: $answers \leftarrow ExtractDateTimeAnswersFromPassage(answers, support)$
5: **else**
6: **if** $qTEC = D$ **then**
7: $answers \leftarrow ExtractDefinitionAnswersFromPassage(answers, support, questionRoot)$
8: **end if**
9: $answers \leftarrow ExtractGenericAnswersFromPassage(answers, support, questionRoot)$
10: **end if**

Algorithm 8.4 extracts numeric answers from a passage and will be described in Chapter 8. It starts by overwriting all words neither numeric nor in the target entity (units expected). After that it just return all sets of not overwritten words in variable $answers$. The complexity time is $O(\#support)$.

Algorithm 6.18 extracts possible dates from a passage, and join them to $answers$ variable. It is based on procedure $CheckDate$ to verify if a date can start at position i. Its complexity time is $O(\#support)$ since $CheckDate$ has constant time complexity.

Algorithm 6.18 Extract Date Time Answers From Passage

Parameters:
Input: $answers, support$
Output: $answers$
Using:

1: **for** $i = 0 \rightarrow \#support$ **do**
2: **if** $IsNumber(support[i])$ **then**
3: $words, answer \leftarrow CheckDate(support, i)$
4: **if** $words > 0$ **then**
5: $answers \leftarrow answers \cup \{answer\}$
6: **end if**
7: $i \leftarrow i + words$
8: **end if**
9: **end for**

Algorithm 6.19 verifies if a date can start in *position* from passage *support*. A date is a sequence of numbers, separated by a separator. The numbers can be hours, minutes, seconds,

and days, months and years. The limits in the list *limits* must be satisfied in order for the sequence to be considered a date. The procedure processes the numbers in positions 0, 2, 4 and so on, and it removes the lowest limit. Since the order of the numbers is not normalised, in this way we can exclude sequence of numbers that cannot be a date since they do not respect the limits, but accept dates with any order of numbers. The procedure has a constant time complexity $O(1)$, since number of words in a date is limited to 6.

Algorithm 6.19 Check Date

Parameters:
Input: $support, position$
Output: $words, answer$
Using: *separator*

1: $words \leftarrow 0$
2: $answer \leftarrow ()$
3: $limits \leftarrow (12, 24, 31, 59, 59, 10000)$
4: **while** $IsNumber(support[position])$ **do**
5: **if** $min_i support(position) \leq limits(i)$ **then**
6: $limits \leftarrow (limits(0), ..., limits(i-1), limits(i+1), ...)$
7: **if** $answer \neq ()$ **then**
8: $words \leftarrow words + 1$
9: $answer \leftarrow answer + support(position - 1)$
10: **end if**
11: $words \leftarrow words + 1$
12: $answer \leftarrow answer + support(position)$
13: $position \leftarrow position + 2$
14: **else**
15: **return**
16: **end if**
17: **end while**

Algorithm 6.20 extracts definitions from passages. The procedure locates the reference entity, and then extracts the sentence after/before the reference entity. Since the least common entities are always explained when referenced, this procedure will extract those short explanations from the text.

Algorithm 6.21 returns a generic answer from passage. It starts by calling $ExtractEntityAnswersFromPassage$, and then, if target entity (qTE) is empty, it uses a generic method for extracting the less frequent words in the support sentence. It starts by clearing the words in the support passage that are in the question, or are terminators, separators or delimiters. Then it calculates the minimal document frequency word, and set the maximal frequency ($maxFreq$), as a ratio of the minimal frequency, specified in global configuration

Algorithm 6.20 Extract Definition Answers From Passage

Parameters:
Input: $answers, support, questionRoot$
Output: $answers$
Using: $separator, delimiter, terminator$

1: $sep \leftarrow \{separator, delimiter, terminator\}$
2: $position \leftarrow Find(support, questionRoot)$
3: **if** $position \geq 0$ **then**
4: $answer \leftarrow support(position - 1) + questionRoot$
5: $position \leftarrow position - 2$
6: **while** $position \geq 0$ **and** $support(position) \notin sep$ **and not** $IsNumber(support[position])$ **do**
7: $answer \leftarrow support(position) + answer$
8: $position \leftarrow position - 1$
9: **end while**
10: $answers \leftarrow answers \cup \{answer\}$
11: **end if**
12: $position \leftarrow Find(support, questionRoot)$
13: **if** $position \geq 0$ **then**
14: $position \leftarrow position + \#questionRoot$
15: $answer \leftarrow questionRoot + support(position)$
16: $position \leftarrow position + 1$
17: **while** $position \geq 0$ **and** $support(position) \notin sep$ **and not** $IsNumber(support(position))$ **do**
18: $answer \leftarrow answer + support(position)$
19: $position \leftarrow position + 1$
20: **end while**
21: $answers \leftarrow answers \cup \{answer\}$
22: **end if**

variable $supportToAnswer$. The words with document frequency greater than $maxFreq$ are cleaned also, and then the remaining sequences of not cleaned words are returned as answers. This method could ensure answers even for not recognized entities.

Algorithm 6.22 extract entity answers from passage, by starting to extracting entities from passage, and then verify if the entities are of type qTE and $qTET$. Functions IsA, $IsAPerson$, $IsAOrganization$, $IsALocation$ are described in Chapter 8, and use the Wikipedia to support the decision if the entity is or not of a specific category.

Algorithm 6.21 Extract Generic Answers From Passage

Parameters:
Input: $answers, support, questionRoot$
Output: $answers$
Using: $supportToAnswer$
$WordFreq(word)$ returns the document frequency of $word$

1: $answers \leftarrow ExtractEntityAnswersFromPassage(answers, support, questionRoot)$
2: **if** $qTE = \{\}$ **then**
3: **for** $i = 1 \rightarrow \#support$ **do**
4: **if** $support(i) \in questionRoot \cup \{terminator, separator, delimiter\}$ **then**
5: $support(i) \leftarrow \{\}$
6: **end if**
7: **end for**
8: $minFreq \leftarrow min\{WordFreq(support(i)) : support(i) \neq \{\}\}$
9: $maxFreq \leftarrow minFreq \times (1 + supportToAnswer)$
10: **for** $i = 1 \rightarrow \#support$ **do**
11: **if** $WordFreq(support(i)) > maxFreq$ **then**
12: $support(i) \leftarrow \{\}$
13: **end if**
14: **end for**
15: $answer \leftarrow ()$
16: **for** $i = 1 \rightarrow \#support$ **do**
17: **if** $support(i) \neq \{\}$ **then**
18: $answer \leftarrow answer + support(i)$
19: **else if** $answer \neq ()$ **then**
20: $answers \leftarrow answers \cup \{answer\}$
21: $answer \leftarrow ()$
22: **end if**
23: **end for**
24: **end if**

6.6 Answer Validation

Algorithm 6.23 will join all answers by order of scoring. A first step of joining the support of identical answers is done in function $JointAnswers$ and then the answers are sorted by number of answers, the most frequent first. The $resultAnswers$ is then built following this order, adding answers that does not have words in common with the previous answers, and return the first $numberAnswers$. This method allow us to return only answers with no words in common.

Algorithm 6.24 joins answers i with j if the best support of answer i contains the answer of j close to answer i. The answers will be joint including the missing words in the support, as well the support sets, improving the score of the answer in the $AnswerValidation$ function. This

Algorithm 6.22 Extract Entity Answers From Passage

Parameters:
Input: $answers, support, questionRoot$
Output: $answers$
Using:

1: $entities \leftarrow FindEntities(support)$
2: **for** $i = 1 \rightarrow \#entities$ **do**
3: $validate \leftarrow false$
4: **if** $\#qTE > 0$ **then**
5: **for** $j = 1 \rightarrow \#qTE$ **do**
6: **if** $IsA(entities(i), qTE(j))$ **then**
7: $validate \leftarrow true$
8: **end if**
9: **end for**
10: **else if** $\#qTET > 0$ **then**
11: **if** $Find(qTET, person) \geq 0$ **and** $IsAPerson(entities(i))$ **then**
12: $validate \leftarrow true$
13: **end if**
14: **if** $Find(qTET, organization) \geq 0$ **and** $IsAOrganization(entities(i))$ **then**
15: $validate \leftarrow true$
16: **end if**
17: **if** $Find(qTET, location) \geq 0$ **and** $IsALocation(entities(i))$ **then**
18: $validate \leftarrow true$
19: **end if**
20: **end if**
21: **if** $validate$ **then**
22: $answers \leftarrow answers \cup \{(entities(i), support)\}$
23: **end if**
24: **end for**

method allows two related answers to be joint at this point, like “Primeiro Ministro”, if answers “Primeiro” and “Ministro” are collected separated, both will share most of the support sentences, and appear with identical score, and has no meaning isolated. In this function the two answers can joint together.

The two proposed methods join answers with the same support and close to each other, and return only answers with all words distinct, and are fully described, which is not usual in the QA literature.

Algorithm 6.23 Answer Validation

Parameters:
Input: $answers, resultPassages$
Output: $resultAnswers$
Using: $numberAnswers$

1: $resultAnswers \leftarrow \{\}$
2: $answers \leftarrow JointAnswers(answers)$
3: $index \leftarrow SortByNumberAnswers(answers)$
4: $i \leftarrow 0$
5: **while** $i < \#index$ **and** $\#resultAnswers < numberAnswers$ **do**
6: $different \leftarrow true$
7: $(answer, supportSet) \leftarrow answers(index(i))$
8: **for** $j = 0 \rightarrow i - 1$ **do**
9: $(answerj, supportSetj) \leftarrow answers(index(j))$
10: **if** $answer \cap answerj \neq \{\}$ **then**
11: $different \leftarrow false$
12: **end if**
13: **end for**
14: **if** $different$ **then**
15: $resultAnswers \leftarrow resultAnswers + (answer, supportSet)$
16: **end if**
17: **end while**

Algorithm 6.24 Joint Answers

Parameters:
Input: $answers$
Output: $answers$
Using: $numberAnswers$

1: $index \leftarrow SortByNumberAnswers(answers)$
2: **for** $i = 1 \rightarrow \#index$ **do**
3: $(answer, supportSet) \leftarrow answers(index(i))$
4: $support \leftarrow AnswerBestSupport(supportSet)$
5: $position \leftarrow Find(answer, support)$
6: **for** $j = i + 1 \rightarrow numberAnswers \times 10$ **do**
7: $(answerj, supportSetj) \leftarrow answers(index(j))$
8: **if** $|Find(answerj, support) - position| < 2$ **then**
9: $answer \leftarrow (answer, ..., answerj)$
10: $supportSet \leftarrow supportSet \cup supportSetj$
11: $answers(index(i)) \leftarrow (answer, supportSet)$
12: **end if**
13: **end for**
14: **end for**

6.7 Conclusions

This chapter described the algorithms for all components used in the search for an answer by IdSay. This QA system uses a classical structure, but with a new strategy of removing the most frequent words only if there are no answers, allowing several retrieval cycles. In question analysis, a framework is used, allowing the implementation of the most common rules related to questions in Portuguese, using authority lists and match of regular expressions, allowing us to achieve a high level of accuracy in question classification and extraction of search keywords for document retrieval.

Document retrieval returns documents with all words and also entities in the question, using a new approach. The proposed passage extraction method has a low time complexity that allows a question based passage identification to be done dynamically, returning all passages containing both words and entities. It includes passage adjustment that prevents sentences to be cut in the middle, based on existing punctuation marks. In answer extraction and answer validation, several procedures are proposed, and as far as we know none was published before in the literature.

7 Evaluation of IdSay at QA@CLEF2008

7.1 Introduction

IdSay was submitted to the monolingual Portuguese task of the Question Answering track of the Cross-Language Evaluation Forum 2008 (QA@CLEF) for the first time.

In the present chapter we analyse the results obtained by IdSay system, according to the evaluation metrics that describe the overall performance of the system, and proceed with a more detailed question based analysis.

7.2 IdSay Results

Table 7.1 summarizes the results of IdSay system at QA@CLEF 2008, that are considered to be very good since it was a new system and was classified better than some veteran systems.

Table 7.1: IdSay results overview

Accuracy over the first answer	Accuracy over all answers	MRR
32.500%	42.500%	37.083%

IdSay has different approaches according to different criteria, for instance, specific procedures regarding question category and type. In the present section we analyse our results, covering different characteristics of the questions.

7.2.1 Question Category

Three question categories are considered in QA@CLEF, namely *F* (factoids), *D* (definitions) and *L* (closed list questions).

These values are provided by the organizers, according to its classification in the three categories above[1] . The results obtained by IdSay are summarized in Table 7.2.

Table 7.2: Results by category

Question Category	Total Questions	Right	Wrong	ineXact	Unsupported	Accuracy
F	162	47	100	7	8	29.012%
D	28	18	10	0	0	64.286%
L	10	0	9	1	0	0%

These results show a stronger ability for the system to answer definition questions than factoids, which was expected due to the valuable aid of the having an encyclopaedic data collection, since its aim is precisely to present definitions. The value obtained for list questions is not a surprise, because we did not have the time to treat this category of questions, so these are treated as factoids. In the case of the answer assessed as inexact, the list of three possibilities was given as the three answers to the question (Question#154 Por que estados corre o Havel?) [For which states does the Havel run?].

We start by making a more detailed analysis of definition questions, for which the type of question is not used and proceed with an analysis by question type for factoids.

7.2.2 Definition Questions

This type of question generally occurs in the form: "O que é X?" [What is X?] or "Quem é X?" [Who is X?], in which we consider X the reference entity. IdSay starts by searching for the reference entity in Wikipedia, looking for a page for this concept. If such a page is found, the beginning of the page is returned as the answer.

There were 22 definitions of the type "O que ser X?" [What to be X?][2], of which IdSay answered 11 correctly based in Wikipedia pages. There are a few cases in which the reference entity has an entry in Wikipedia, but there are several ways of referencing the same entity and therefore we do not get to the right page straight away. That is, for instance, the case of (Question#35 O que era o A6M Zero?) [What was the A6M Zero?]. However, we got to

[1] We do not know the classification of the questions by category of the organization, so in the rest of the text we use our classification, therefore the values can be slightly different.

[2] We use the lemmatized form of the verb to cover the several tenses occurring in the questions.

Wikipedia web page via retrieval of content, and manage to produce the correct answer, in this case but not in others. There is an advantage in dealing with the different ways of naming the same entity in Wikipedia, and we did not do that for the current version of the system only because of lack of time. A similar problem happens in (Question#127 O que são os forcados?) [What are forcados?] which in Wikipedia is identified by the singular form and not the plural of the concept. We tried to get the definition by lemmatization, however the word retrieved is the lemma of the verb, and unluckily there is a town in Spain that corresponds to a form of that verb, and that is the definition returned. In this case, we would have to consider besides of the alternative ways of naming the same entity in Wikipedia, the plural/singular variation. If Wikipedia does not provide a definition, we follow the default procedure of searching the data collection in search for occurrences of the reference entity. An example of a correct definition found via the default procedure is (Question#66 O que é o jagertee?) [What is jagertee?]. There is only one occurrence of this entity in the data collection, in a sentence "o jaggertee é chá com adição de rum" [jaggertee is tee with addition of rum], which allows the system to retrieve the correct answer "chá com adição de rum" [tee with addition of rum]. However many definitions are not found in these circumstances: that is the case of (Question#80 O que é o IPM em Portugal?) [What is the IPM in Portugal?] for which the acronym of this institution is found many times together with other institution names, which does not allow for the correct answer to be retrieved.

There were 7 definition questions of the type "Quem é X?" [Who is X?], of which IdSay answered 5 correctly based on Wikipedia pages. The two questions that were wrong (Question#23 Quem é FHC?) [Who is FHC?] and (Question#41 Quem é Narcís Serra?) [Who is Narcis Serra?] the first corresponding to a Wikipedia page that is not found (again) because the keyword FHC is not the name of the page for former Brazilian President Fernando Henrique Cardoso (but rather a redirect), and the second case because there is no Wikipedia page and although in this case two news articles are found with the information on Narcis Serra, the answers are wrong due to extraction problems.

7.2.3 Factoids Questions

We consider the following types of questions: P - person/organization, D - date/time, L - location, C - count, M - measure, O - Other. We will start by analysing the results for the types

for which we developed special procedures because they involved numeric values: *C*, *M* and *D*. We consider the assessment of the question to be the best answer, using the following priority: 'R', 'U', 'X' and 'W'[3].

Table 7.3 presents the results of IdSay for the type of factoids count, measure and date. We proceed with an analysis of these results.

Table 7.3: Results by question type

Question Type	# Questions	Right	Wrong	Unsupported	ineXact
Count	19	13 (68.4%)	5 (26.3%)	1 (5.3%)	0 (0%)
Measure	12	9 (75.0%)	2 (16.7%)	1 (8.3%)	0 (0%)
Date	24	11 (45.8%)	12 (50.0%)	0 (0%)	1 (4.2%)

For each type, we try to analyse an example to draw conclusions about the strength and weakness of the system in that type, except for type other, because the questions can be so different from one another that an almost per question analysis would be required.

7.2.3.1 Factoids - Count

These questions usually start by "Quantos/as X ..." [How many X ...]. X usually represents what we are trying to count, followed by the rest of the phrase. The general form of the question is usually a number followed by X and this is achieved in question analysis by setting $qTE = X$. There were 20 count questions, with very diverse instances of X, namely esposas, faixas, províncias, repúblicas, actos, atletas, estados, filhos, filmes, gêneros, habitants, jogadores, ossos, refugiados, votos [wives, stripes, provinces, republics, acts, athletes, states, sons, movies, gender, inhabitants, players, bones, refugees, votes].

An example of a correct answer is (Question#70 Quantas províncias tem a Ucrânia?) [How many provinces does Ukraine have?]. In the question, the reference entity Ukraine was identified and the identification of the unit to look for was provinces. The search string was in this case "província ter a ucrânia" [provinces to have ukraine] for which 71 documents were retrieved, where the correct answer was found: 24 provinces. The case of (Question#10 Quantas províncias tem a Catalunha?) [How many provinces does Catalonia have?] is similar, with 51 documents

[3]For example, if a question has three answers judged 'W', 'X' and 'U' we consider the 'U' answer.

retrieved that produced the answer "4 provinces" supported by more than one passage. However the answer was considered unsupported, due to the choice of the shortest passage. As example of a question that produced wrong answers, we can look at (Question#18 Quantos ossos têm a face?[sic]) [How many bones do the face have?]. Although the question is incorrectly formulated in terms of concordance in number of the verb, which is in the plural instead of singular as it should, the Lemmatization took care of that and produced the search string "bone to have face". However, the answers produced were incorrect (number of bones of parts of the face, as the nose, returned) because the correct answer occurred in a phrase using the construction "é constituída por" [consists of] instead of the verb "ter" [to have].

7.2.3.2 Factoids - Measure

This type of question is similar to the previous one, and generally occurs if the form of "Qual/ais ...o/a X de ..." [What...the X of ...] in which X is a measure, which can have several units. The answer is generally a numerical value in the correct units for the measure. There were several cases of measures in the question set: altura, área, dotação, envergadura, largura, temperatura, comprimento [height, area, money value, bulkiness, width, temperature, length]. IdSay supports several systems of measures, and the corresponding units implemented in the manner of the authority lists already mentioned in 6, that will be added to qTE. It allows the search of the answers of the correct type. The addition of new measures is an easy process; however, the implemented lists should be stable.

An example of a correct answer is (Question#142 Qual é a área da Groenlândia?)[What is the area of Greenland?]. For area measure we have cm 2, m 2, km 2, hectar, hectares, ha that is the content of qTE after question analysis. The occurrences of "km2", "km2" or other forms commonly used are normalized to "km 2" in the pre-processing of the data collection occurring at level1. For this question, only the value of the area "2 170 600 km 2" is returned and in the same passage there are other numbers, that would also be returned if we did not check the area units.

The normalization of the unit measures and numerical values is part of the pre-processing of the data collection occurring at level 1. The incorrect answers were given for questions that supposedly should produce NIL answers. The unsupported answer is (Question#163 Qual o comprimento da Ponte do Øresund?)[What is the length of the bridge of Øresund?] for which the

support that was taken from a news article was considered insufficient.

7.2.3.3 Factoids - Date

The most common form of occurrence for this type of question is in questions starting by "Quando ..." [When ...], tough there are also 4 question staring by "Em que ano ..." [In which year ...]. IdSay has a specific treatment of dates, starting with the pre-processing of the texts, and also in the extraction of the answer. However this treatment is not fully developed, for instance the temporal restrictions are not taken into account. Therefore, the results achieved for this type are worse that for the preceding two types. The low percentage of accuracy in temporally restricted questions, 18.750%, can also be interpreted in light of this limitation.

As an example of a correct answer is (Question#86 Quando é que ele tomou posse?)[When was he empowered?], and is an example of a question that belongs to a cluster with first question (Question# 85 Quantos votos teve o Lula nas eleições presidenciais de 2002?)[How many votes had Lula in the presidential election of 2002?]. Although Question#85 was not successfully answered, the reference to Lula (Brazilian President Luiz Inácio Lula da Silva) is correctly resolved in Question#86 (reference resolution based on the question, not the answer). The system produces the correct answer twice, the first answer is just the year, 2003, which is more frequent, but the support was not accepted because it said something like "President from 2003 to actuality", however the second answer, that indicated January of 2003, was considered right. As for the 12 wrong answers there are several aspects that contribute to that, there are questions about periods that were not treated by the system, and the need to treat date information from Wikipedia in a more practical way, for instance, the listed items in such pages are not terminated, so events tend to be mixed up in the resulting text. Since we had no notion of the dates, each numeric value can be a potential date, also with qualifiers like a.c., which increases the difficulty to answer this kind of question, namely the decision that the answer should be NIL. An example of a wrong answer is (Question#17 Em que ano é que Ernie Els venceu o Dubai Open?)[In which year did Ernie Els win the Dubai Open?] for which the system produced several incorrect numeric values, tentatively, from the only news article found with information on the subject, but in which the year of 1994 could only be produced if one took into account the date of the news article itself and not its text.

7.2.3.4 Factoids - Person

This type of question generally appears in a form starting by "Quem ..." [Who ...], but that is not always the case. There are 37 questions in total starting by "Quem ..." [Who ...], of which 7 are definitions and were already covered in the corresponding section, and 5 occurring with different forms. The results for this type had an overall accuracy of 34%, which is according to the general performance of the system. Examples of correct answers were (Question#92 Quem fundou a escola estóica?) [Who founded the stoic school?] (Question#143 Quem foi a primeira mulher no espaço?) [Who was the first woman in the space?] for which the system gives the correct answers, respectively Zenão de Cítio and Valentina Tershkova, but they are accompanied by wrong second and third answers, that have different information related to the subject. We must therefore find a way to filter entities of type person. The search strings were respectively "fundar escola estóico" [to found school stoic] and "primeiro marido no espaço" [first husband in space] and the number of documents retrieved were respectively 11/2 and 1991/75 before entities/after entities. The case of Question#143 shows an example of the utility in combining the search by single word with the search for entities,were the number of documents considered drops almost two orders of magnitude from 1991 to 75.

An example of a question that produced only wrong answers was (Question#160 Quem realizou "Os Pássaros"?) [Who directed "The Birds"?]. Since the title of the film is composed of words that occur very often in other contexts, other documents were retrieved that had no relation to the subject of the question. The search string was in this case "realizar o pássaro" [to direct the bird] and the number of documents retrieved was 317 (no entities were found in the question text).

A last example of questions of this type is (Question#15 Quem escreveu Fernão Capelo Gaivota?) [Who wrote Jonathan Livingston Seagull?] in which there were 6 documents each with several passages supporting the correct answer of "Richard Bach", but since the shortest passage was chosen as support, it was assessed unsupported.

7.2.3.5 Factoids - Location

These questions generally begin by "Onde ..." [Where ...]. IdSay does not make a special treatment for location questions. The results for this kind of question are close to the overall

performance of the system. The strong incidence of this type of question in clusters also contribute to raise its difficulty. We believe that better results can be achieved through a better treatment of this kind of information from Wikipedia.

7.2.4 NIL Accuracy

About the NIL accuracy, the reported value of 16.667% (2 right answers out of 12) for IdSay indicates the need of improvement in our mechanism to determine how well a passage supports the answers, to minimize the effect of the retrieval cycle in relaxing constraints. We cannot however make a very accurate analysis because the 12 questions to which the NIL answer should be returned have not been divulged, and also there are some questions whose formulation makes it unclear if they should be "corrected" or a NIL answer is expected. An example of such questions is (Question#152 Qual é a capital de Dublin?) [What is the capital of Dublin?]. We considered that these questions should not be intentional, because we are in the context of an evaluation campaign, and the aim is to compare results from several systems. For this goal to be achieved there should be as little ambiguity as possible. If there is the interest in finding out how systems react to strange situations it could be done at the CLEF Workshop interactively, using frozen versions of the system.

7.3 IdSay Web Application

We built a web application[4] based on the version of IdSay whose results were submitted to QA@CLEF 2008.

The purpose of this web application is manifold: this way we are able to reproduce online the results obtained at the CLEF campaign, but we can also use the system for any other question. The interface is very easy to use and the entry page of the application is reproduced in Figure 7.1.

For the sake of new comers, we give some information on the project (Figure 7.2), and we have links for the question set of the Portuguese monolingual task of QA@CLEF 2008, as well as for the answers submitted by IdSay. The questions correspond to Tables C.1 to C.5 and IdSay results to Tables C.6 and C.6 in Appendix C.

[4] Available at http://www.idsay.net.

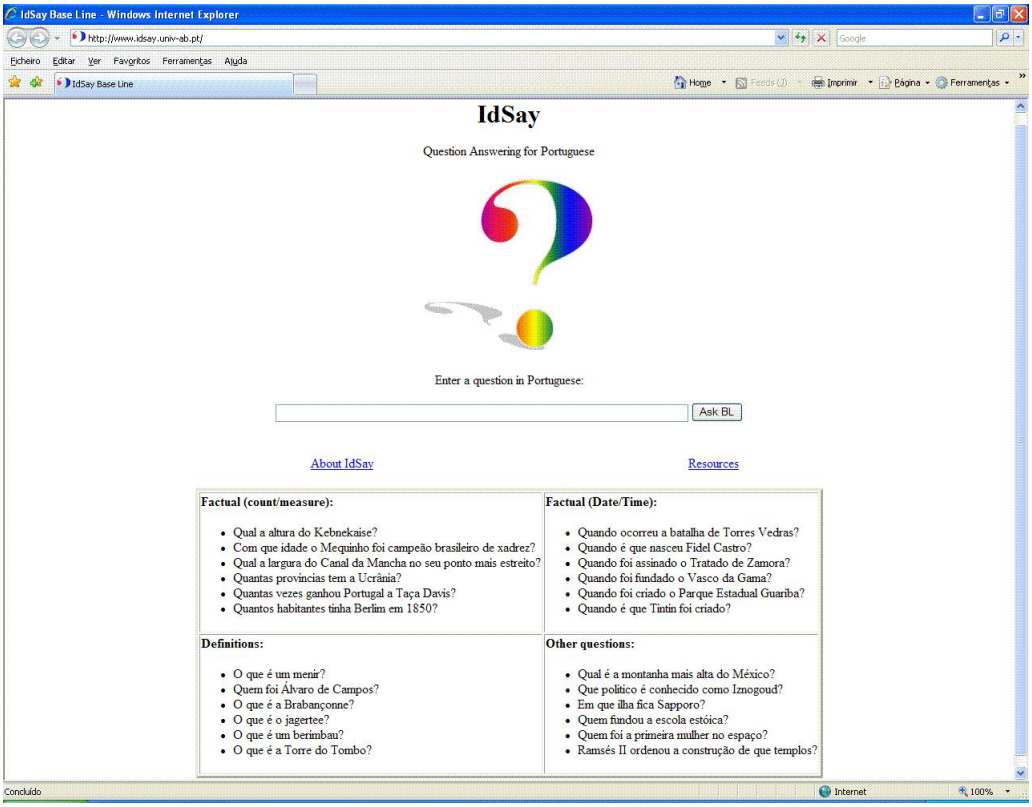


Figure 7.1: IdSay Web Application

We also use this web site to make available a resource that we compiled during the development of this project, the WES base, which we will describe in Section 8.2.

The user can type a question in Portuguese, or chose one of the questions given as example. These sample questions were successfully answered in the QA@CLEF 2008 campaign. IdSay works in the mode of file processing or single question manual input. In this later case we provide no functionality for cluster question treatment. The Web Application is based on the manual question interface, therefore some of the questions given in the example table of the main page, because they belonged to a cluster, were subject to small adaptations to make them usable in standalone mode.

We also added new debugging functionalities that makes it easier to understand the inner workings of the system and the answers that it produces, that we will exemplify later in this section.

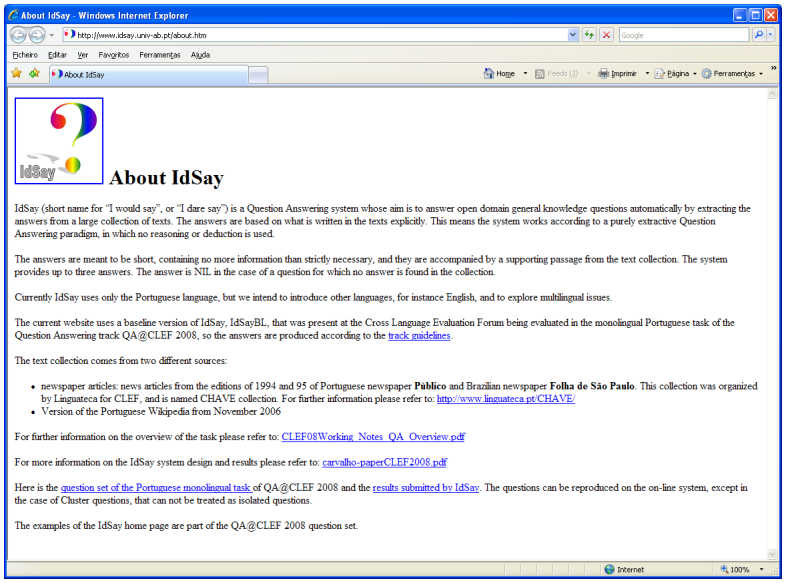


Figure 7.2: About IdSay Web Application

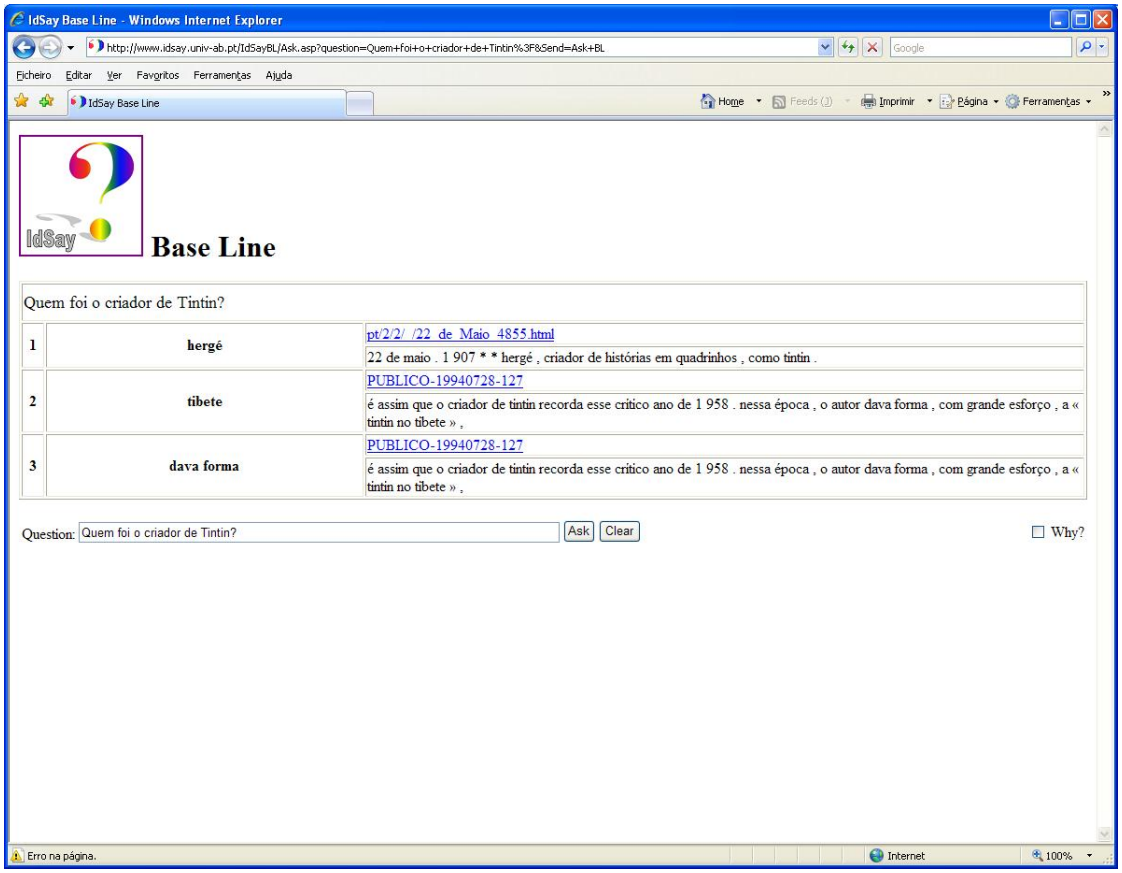


Figure 7.3: IdSay Web Application Results for Question#2

For example the results for Question#2 Quem foi o criador de Tintin? [Who was the creator of Tintin?] are presented in Figure 7.3

If we click over the sixth question in the table of examples in the main page, category **Factual(Date/Time)** we have the standalone equivalent of Question#3 Quando é que Tintin foi criado? [When was Tintin created?]. The results are presented in Figure 7.4. They exactly match does presented in Table C.6, emphasising the good co-reference resolution for this cluster question.

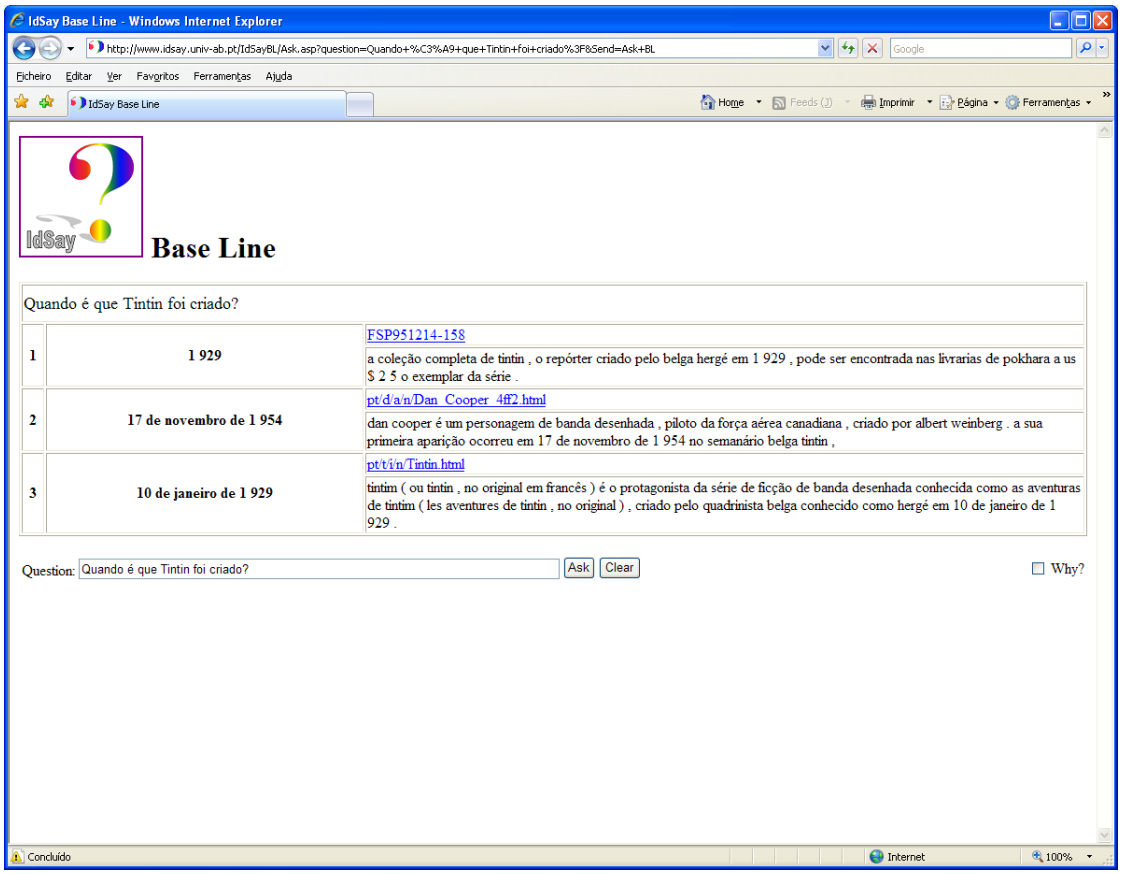


Figure 7.4: IdSay Web Application Results for the equivalent of Question#3

Each answer is accompanied by the supporting passage it was extracted from, as well as the reference to the document the passage belongs to. It is possible to see the whole document, with the supporting passage highlighted, clicking in the document reference. Figure 7.5 shows this functionality for the first answer to Question#3. The document reference is FSP951214-158, which means it is a news article coming from the Brazilian newspaper Folha de São Paulo, from

the 14th of December 1995, and it is the 158 news article of that newspaper edition.

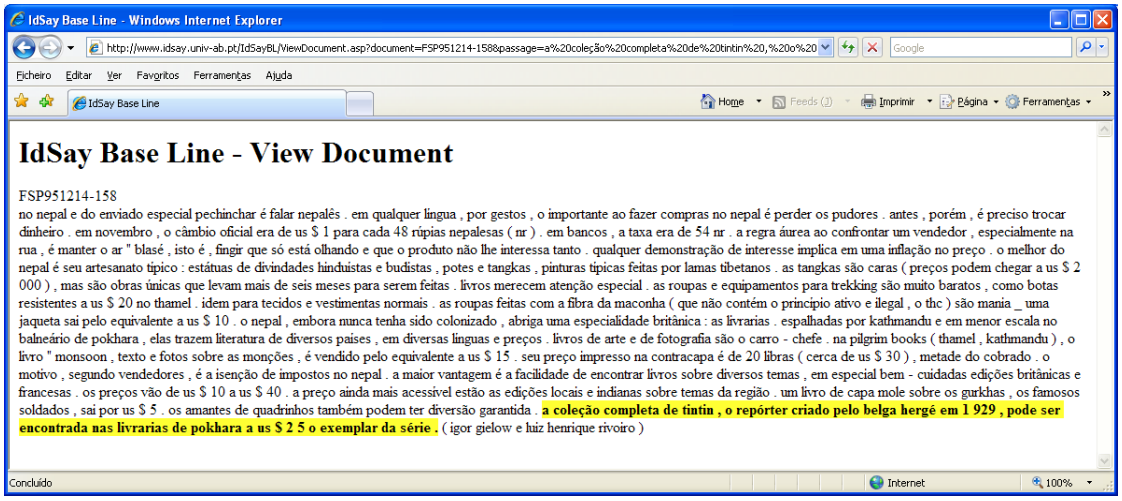


Figure 7.5: Full Document view for the correct answer to Question#3, with the supporting passage highlighted.

In the bottom right hand side of the screen presented in Figure 7.4, there is a "Why?" check box, that allows access to more information on how the answers were obtained for that question. When checked the system presents the information depicted in Figure 7.6.

It presents the architecture of IdSay system (a slightly different figure from that presented currently, for instance in Chapter 2, Figure 2.2, but the basic modules are the same). The information resulting from the Question Analysis module is presented on top. The information of each four modules in the retrieval cycle, is presented in next in collapsed form, but it can be detailed per module, with a double click in the corresponding region. The information is preceded by two words indicating the initials of the module it belongs to. We present the information for the four modules in the retrieval cycle, for Question#3, in Figure 7.7 to Figure 7.10.

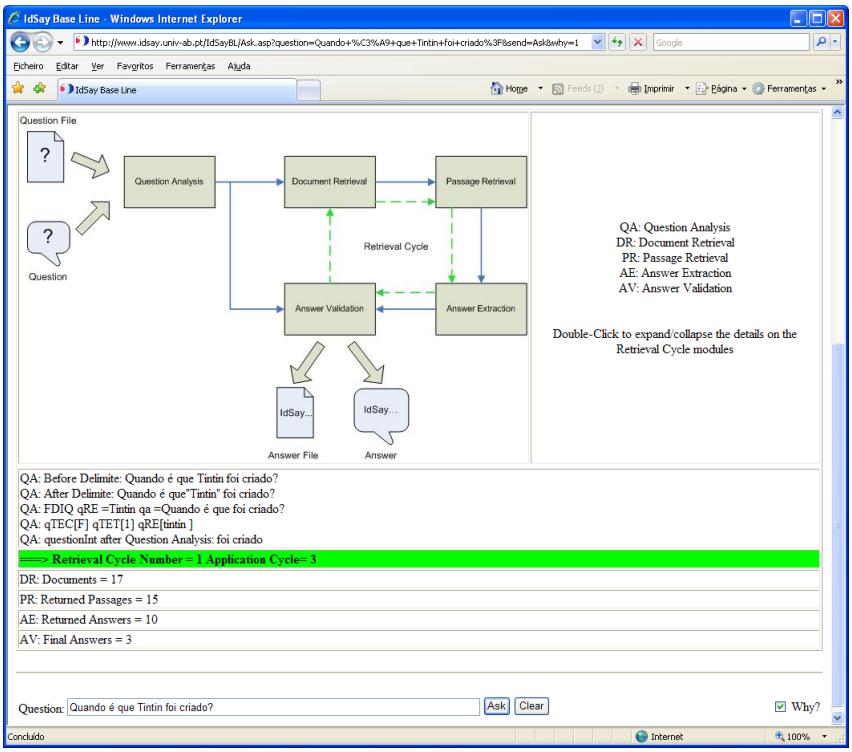


Figure 7.6: "Why?" check box explanatory screen for Question#3.

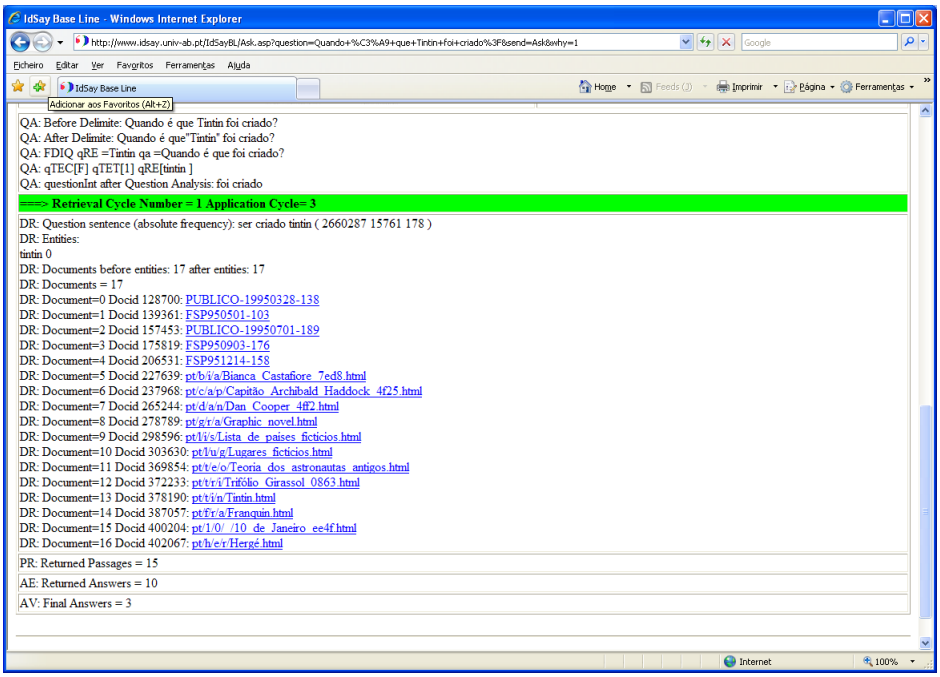


Figure 7.7: Document Retrieval (DR) module information for Question#3.

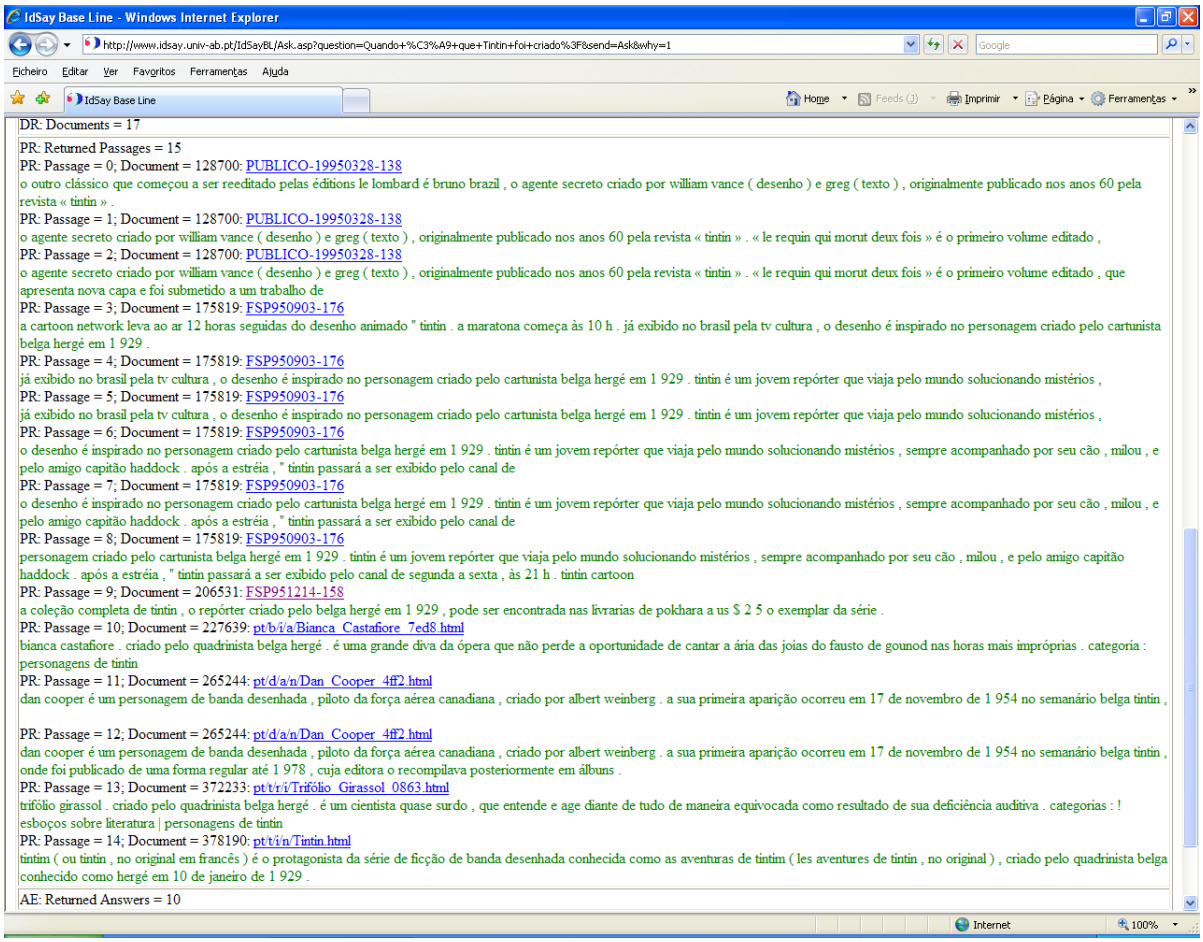


Figure 7.8: Passage Retrieval (PR) module information for Question#3.

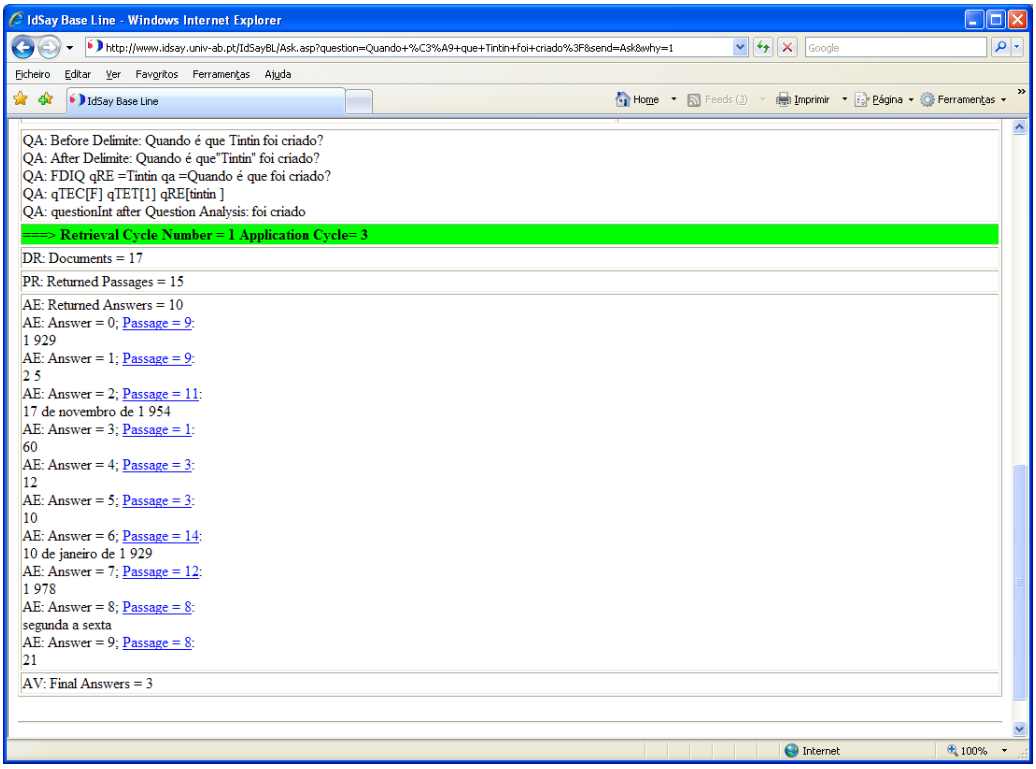


Figure 7.9: Answer Extraction (AE) module information for Question#3.

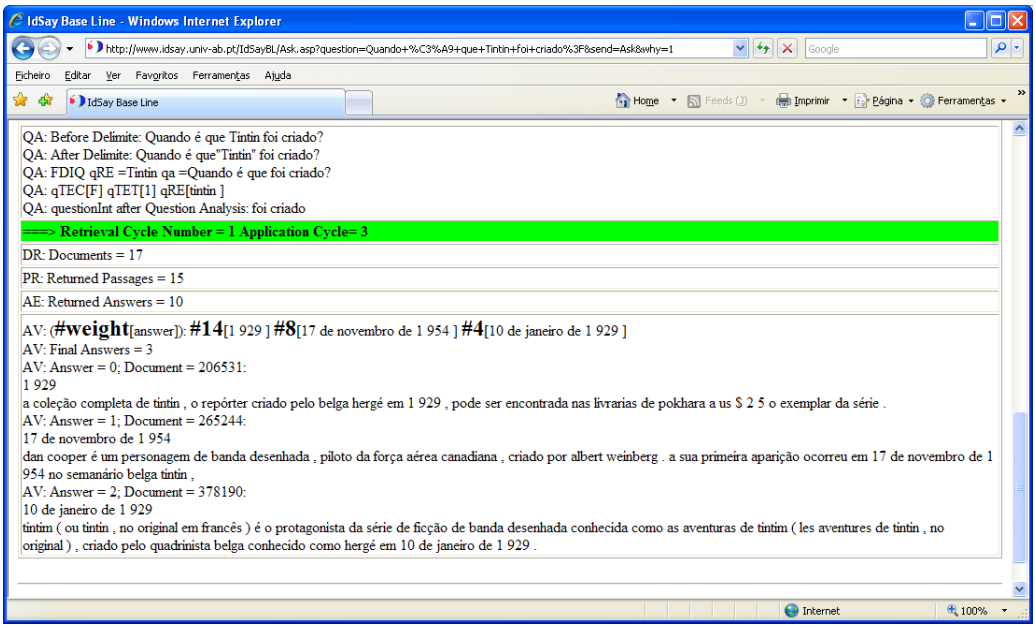


Figure 7.10: Answer Validation (AV) module information for Question#3.

The information in Figure 7.8 shows that there were more passages supporting the answer 1929, than the more detailed, also correct answer 10 de Janeiro de 1929 [10th of January of 1929]. It can be seen that the system chose as supporting passage the shortest one. We can see a passage in the context of the document it comes from. Figure 7.11 presents this information for passage 14, the one containing the more detailed correct answer. In this case the document is a Wikipedia document corresponding to the entry for Tintin The same information could be obtained from Figure 7.4, since it corresponds to the third answer produced by the system.

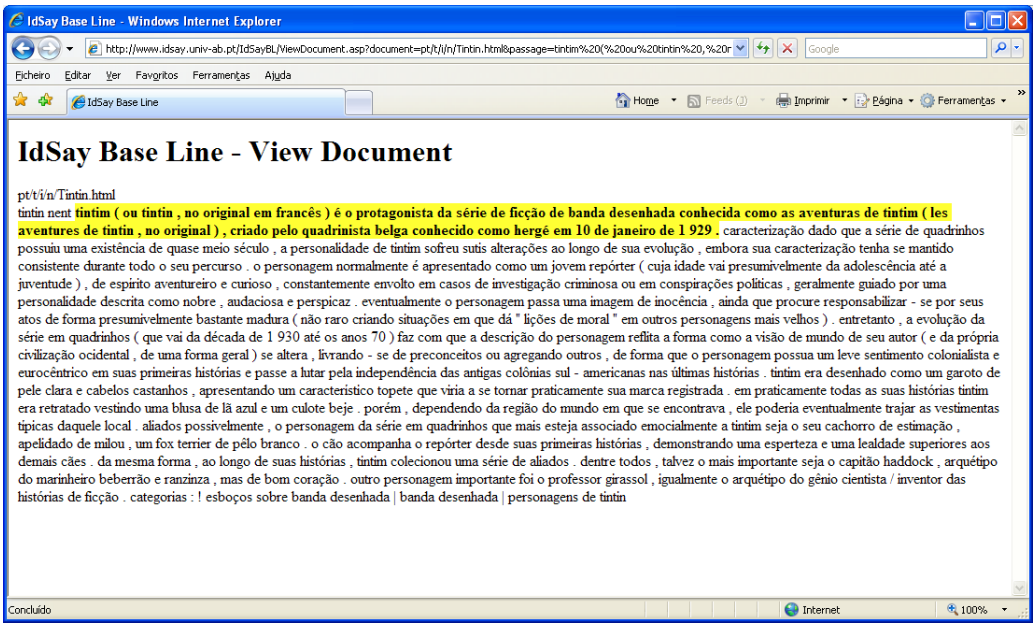


Figure 7.11: Full Document view for passage 14 (or 3rd answer) of Question#3, with the supporting passage highlighted.

For Question#3 the correct answer is found with the first query string tried, which corresponds to a single retrieval cycle, but that is not always the case. For instance for Question#33 Que político é conhecido como Iznogoud? [What politician is known as Iznogoud?] (Figure 7.12) the correct answer is found in the third retrieval cycle.

More information can be obtained using the "Why?" check box, with each cycle highlighted in green. The additional information for Question#33 in presented form Figure 7.13 to Figure 7.16.

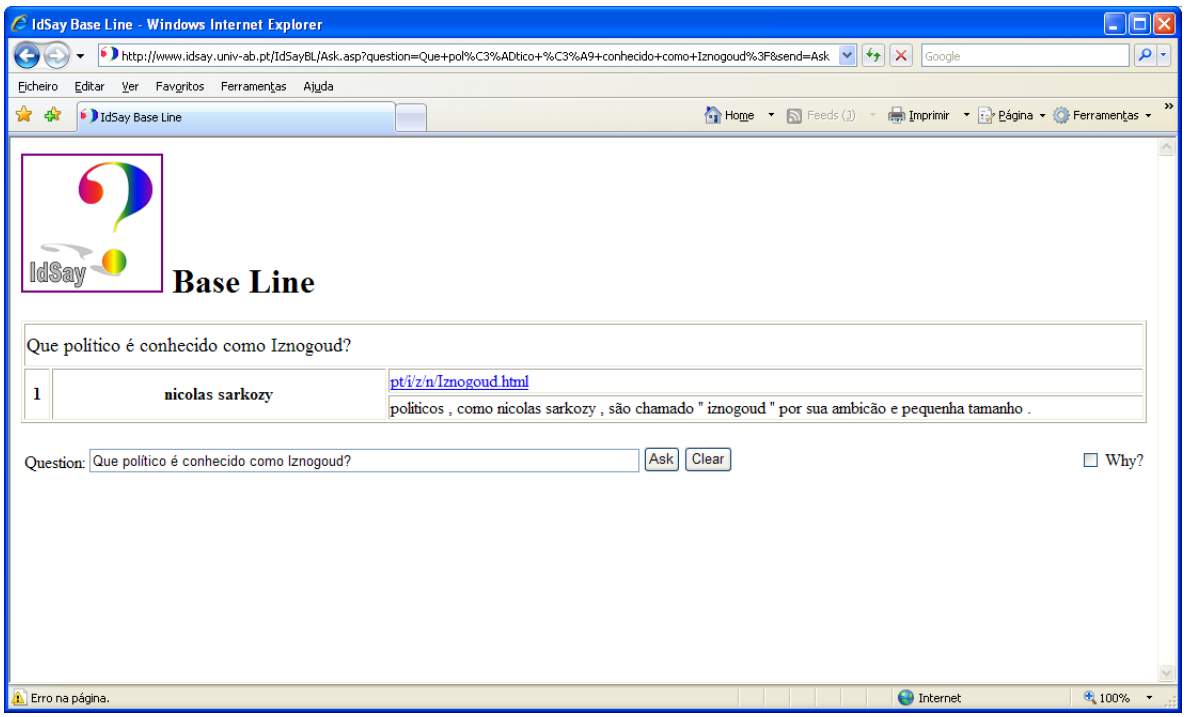


Figure 7.12: IdSay Web Application Results for Question#33

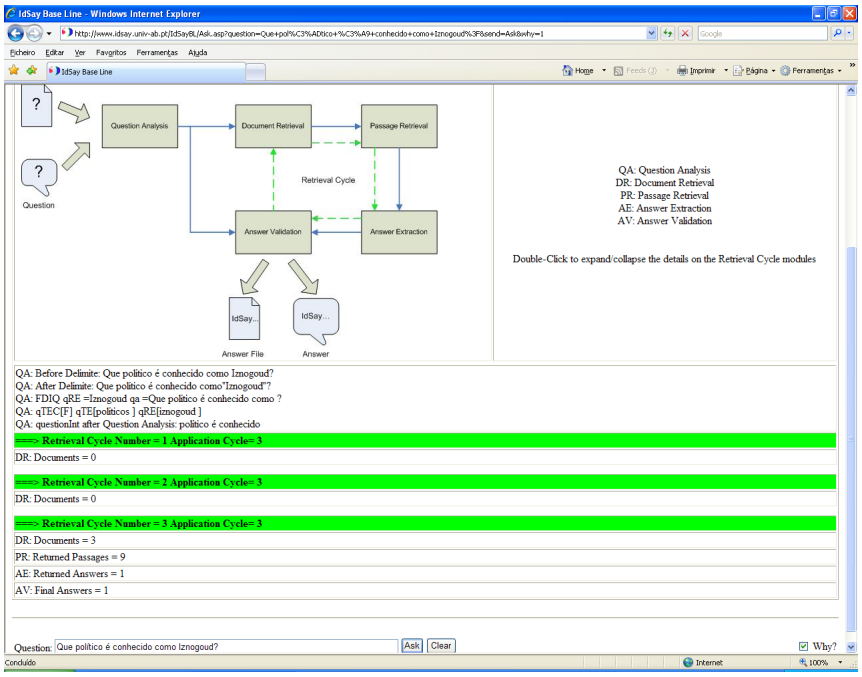


Figure 7.13: "Why?" check box explanatory screen for Question#33.

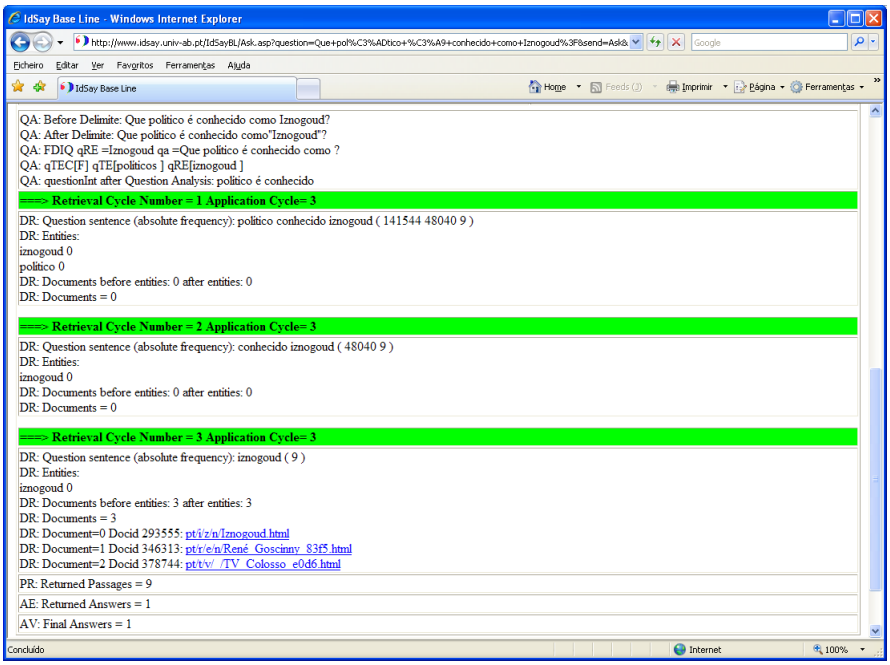


Figure 7.14: Document Retrieval (DR) module information indicating that three cycles were done for Question#33.

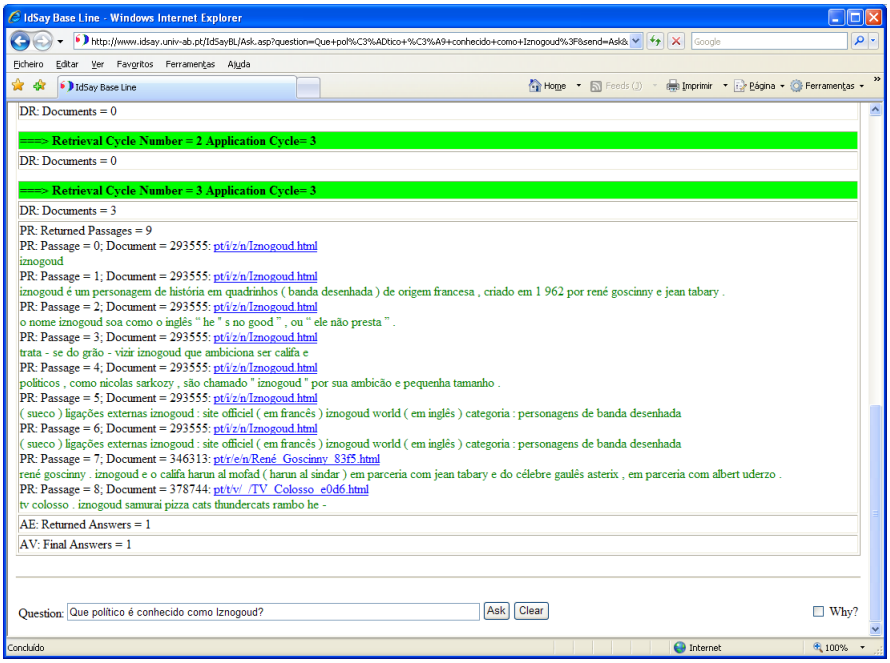


Figure 7.15: Passage Retrieval (PR) module information for Question#33 (3rd cycle).

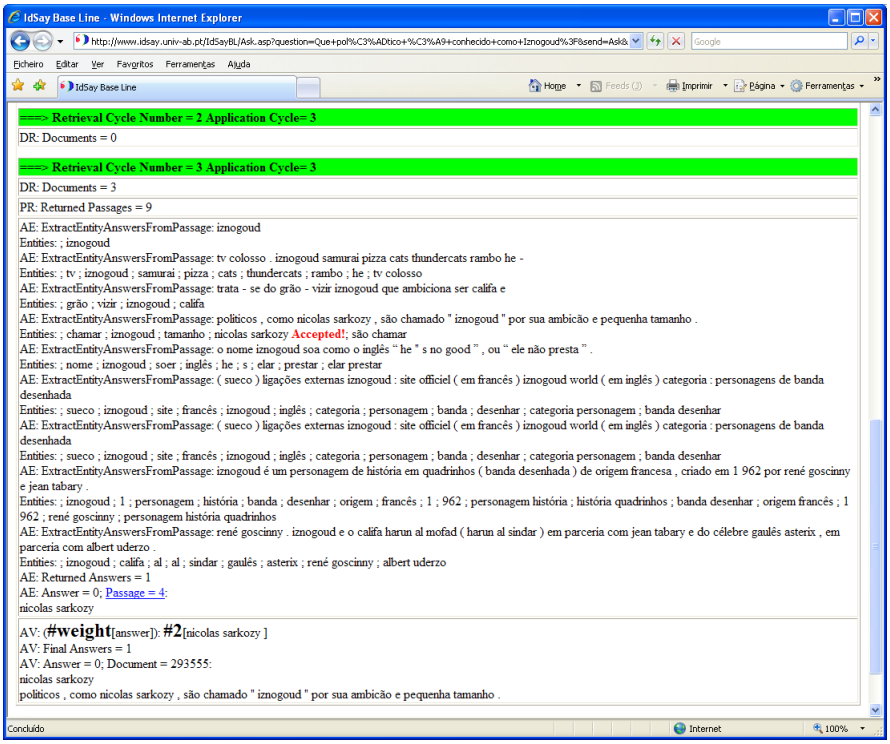


Figure 7.16: Answer Extraction (AE) and Answer Validation (AV) modules information for Question#33 (3^{rd} cycle).

This interface was very useful in question based analysis, since question analysis, answer extraction and answer validation components are difficult to analyse separately from other IdSay components, and this way we are able to know the reason of each failure and adjust better rules in question analysis or collect examples of passages were answers are not extracted correctly, providing valuable information for support future improvements.

7.4 Comparative Analysis

In the present section we make a comparative analysis of IdSay results at QA@CLEF with other systems, with two kinds of goals:

- to determine for the wrong answers the cause of failure, specifically which module in the global architecture can be deemed responsible for the lack of success;
- to make an analysis of the answers submitted by the systems participating in the same

task, so that we are able to compare our results with those of the other systems, and try to learn from them;

This analysis will indicate the orientation for making improvements to IdSay, that will be described in the next chapter, followed by the results of the improvements.

In addition to IdSay, the remaining participants in the Portuguese monolingual task in 2008 were Priberam(Amaral et al. 2009), Senso (Saias & Quaresma 2008a), Esfinge (Costa 2009), QA@L2F (Coheur et al. 2009), and Raposa (Sarmento et al. 2009). In 2008, the score of a hypothetical combination run that chose the best answers among the six participating systems for Portuguese would have a score of 168 right answers for all 200 questions, giving an accuracy over the first answer of 84%.

Table 7.4 summarizes the results of the systems participating in the Portuguese Monolingual task at QA@CLEF 2008.

Table 7.4: Results Overview of Portuguese Monolingual Task at QA@CLEF 2008

System	First Answer		All Answers		MRR
	R	accuracy	R	accuracy	
Priberam	127	63.5%	139	69.5%	66.3%
Senso	93	46.5%	93	46.5%	46.5%
IdSay	65	32.5%	85	42.5%	37.1%
Esfinge	47	23.5%	61	30.5%	26.6%
QA@L2F	40	20.0%	41	20.5%	20.3%
Raposa	29	14.5%	29	14.5%	14.5%
Combination	168	84.0%	174	87.0%	85.4%

In this section we try to take benefit from the valuable information that are the answers of the systems to the individual questions. This information is detailed in Appendix C.

The good results for the Portuguese language are especially highlighted by the score of the hypothetical combination run that chose the best answers among the six participating systems for Portuguese, and that would lead to an accuracy over the first answer of 84%. Considering the success of this combination, we made a detailed analysis of the results of the other five participants in the Portuguese monolingual task of QA@CLEF 2008, to find the strengths and weaknesses of each system and make a comparison with IdSay. Since it was a baseline, we expected this analysis to help us decide which improvements and new functionality were more

liable to enhance our results, keeping in mind the two key characteristics we want to keep in the system: efficiency in terms of response time (the 200 questions are answered in less than 1 minute) and robustness because we plan to be able to treat different types of data.

We analysed all answers submitted at QA@CLEF 2008 in the Portuguese exercise, which represented around 3600 answers. The analysis of IdSay results was done in more detail since we had information that allowed us to determine the exact causes of failure. However, the analysis of other systems' answers also allowed us to have a better insight into their behaviour.

To reflect the performance of IdSay relatively to other systems, we present a table that highlights the global results of all systems for Portuguese (Table 7.5).

Table 7.5: Comparison of IdSay results and those of other systems

		IdSay	0 sys	1 sys	2 sys	3 sys	4 sys	5 sys
First Answer	R	65	3	17	24	12	7	2
	Not R	135	32	40	38	19	6	0
All (3) answers	R	85	3	20	35	12	11	4
	Not R	115	26	32	36	15	6	0

We make a comparison using the results considering only the first answer (the 2 top rows), and considering all the three answers that could be returned per question (the 2 bottom rows). The first column shows the number of questions that IdSay got right (R) or otherwise (not R, which includes Unsupported, ineXact and Wrong answers), given the condition of the corresponding row (1st answer or all three answers). In the rest of the row, each column indicates the number of systems that got right answers.

For instance, taking the first row of the table, we are considering only the first answers, so we start with the number of questions IdSay got right in those circumstances which is 65, and for these 65 we check the number of systems that also got first right answers. So the value 3, next to 65, means that from the 65 questions, 3 were only answered correctly by IdSay (or 0 other systems got them right). On the other end of that row, the value 2 indicates that of the 65 questions, 2 have been answered correctly by all systems.

In the next row we consider the questions that were not answered correctly by IdSay (135), and we make a similar analysis of the number of systems that got them right. Of these questions, 32 were not answered correctly by any of the systems, while on the other end of the row we can

see that the number of questions that IdSay was not able to answer correctly but that all the other systems were able to give right answers is 0.

In general, a higher number of questions on the left hand side of the table represents a better performance of IdSay relative to the other systems.

In a row where we are considering right answers from IdSay, the questions that are included on the right of this row can be considered easy questions, and those included on the left side are questions in which IdSay proved to make a positive contribution when compared to other systems.

In a row where we consider the questions that IdSay was not able to answer correctly, the questions on the left are those generally not well covered by the systems, and can be considered difficult for the current state of the art. On the right hand side of this row are the questions which we take into special consideration, because they reveal weaknesses of our system, and that all or almost all the other systems were able to address successfully.

For a characterization of the relative performance of a QA system, we created a condensed form of Table 7.5, first answers only. We call it a results quadrant and its meaning is explained in Table 7.6. In each quadrant is the percentage of questions that meet the criteria in the corresponding upper and left border.

Table 7.6: Interpretation of a Results Quadrant

	0 or 1 systems	2, 3, 4 or 5 systems
R	innovative	good coverage of easier questions
Not R	needs to invest in new techniques to cover more difficult questions	liable to improve with the experience of others

We built a results quadrant for each of the six systems, which we will present in the next tables, with a summary of the most significant conclusions.

Priberam (Table 7.7) is the only system classified as "innovative" (33.5%), also having the highest value in the "good coverage of easier questions" (30.0%). The value of 12.5% in the fourth quadrant indicates that the system is already a very optimised one.

Table 7.7: Results Quadrant for Priberam

Priberam	0 or 1 systems	2, 3, 4 or 5 systems
R	33.5%	30.0%
Not R	24.0%	12.5%

Table 7.8: Results Quadrant for Senso

Senso	0 or 1 systems	2, 3, 4 or 5 systems
R	19.0%	27.5%
Not R	32.0%	21.5%

The results quadrants for Senso and IdSay are presented in Table 7.8 and in Table 7.9, respectively.

Senso (32.0%) and IdSay (36.0%) are in the quadrant "needs to make an investment to cover more difficult questions" with the first being more innovative (19.0% versus 10.0%) and having a better coverage for easier questions than IdSay (27.5% versus 22.5%), while IdSay has more room to learn from the others (31.5% against 21.5%).

Esfinge (Table 7.10), QA@L2F (Table 7.11), and Raposa (Table 7.12) all have higher percentages in the "liable to learn with the experience of others" quadrant, respectively 40.0%, 42.5%, and 48.0%.

Taking into account these results, we made a detailed question-based analysis, focusing on the questions in the quadrant "liable to improve with the experience of others", in order to identify the enhancements that could lead to better results. These improvements are described

Table 7.9: Results Quadrant for IdSay

IdSay	0 or 1 systems	2, 3, 4 or 5 systems
R	10.0%	22.5%
Not R	36.0%	31.5%

Table 7.10: Results Quadrant for Esfinge

Esfinge	0 or 1 systems	2, 3, 4 or 5 systems
R	4.5%	19.0%
Not R	36.5%	40.0%

Table 7.11: Results Quadrant for QA@L2F

QA@L2F	0 or 1 systems	2, 3, 4 or 5 systems
R	4.5%	15.5%
Not R	37.5%	42.5%

Table 7.12: Results Quadrant for Raposa

Raposa	0 or 1 systems	2, 3, 4 or 5 systems
R	5.0%	9.5%
Not R	37.5%	48.0%

in the next section.

7.5 Conclusions

Generally, the types of questions that are better answered by IdSay system are measure factoids, count factoids and definitions, but there is still work to be done in these areas, as well as in the treatment of time. List questions, location and people/organization factoids are the types of question with more room for evolution.

The ordering (ranking) of answers by frequency means that we generate as first answer, the answer that appears most frequently in the context of the data collection. However, this procedure may lead to difficulties in the support of the answer, because the same answer can occur several times in the collections, with different supports. Since in the CLEF evaluation campaign only one support is allowed, we chose the shortest one found by the system. This may lead, and it did in many cases, to the support not being the best possible (and in some cases answers were considered unsupported by the assessor). This is the case of the answers mentioned in the factoid-count analysis (Question#10), in the factoid-measure analysis (Question#163), and in the factoid-person analysis (Question#15).

Although the lemmatization process produces some "strange" results in the search strings, it provides an efficient search, with this phase of the process generally finding the related documents. We consider lemmatization a good choice as a whole, with just one case of a definition being wrong on its account (Question#127). The extraction part is less efficient and is generally responsible for the wrong answers produced by the system.

IdSay web application makes it easy to identify the reasons of failure, and proved to be very useful in the deep question based analysis, providing insight for future improvements.

The comparative analysis with other systems in the same track is another source that provides orientation for improvements. A method that summarizes the information of the analysis based on results quadrants is proposed, that allows a practical identification for each system of its strengths and weaknesses when compared to the other systems. The Results quadrant for a system highlights information such as its degree of innovation, if it has a good coverage of easy questions or even to what extent is it liable to improve with other QA systems.

8 Improving IdSay

8.1 Introduction

In the last chapter we described our participation at QA@CLEF Portuguese monolingual task. The participation had the aim to validate our baseline version (henceforth, in the present chapter, identified as IdSayBL) or in other words the basic stones of the system: its architecture and the approach we used to IR. As mentioned we considered the results encouraging, meaning that our approach was validated, and we identified the areas (types of questions) in which we were more successful.

In this section we use are guided by the previous analysis and implemented several improvements in IdSay, but always keeping in mind to respect the key factors we have decided for our system, namely:

- efficiency in terms of time and space;
- robustness to treat other types of data.

The last fact was important in our option to introduce semantic information instead of opting for the use of linguistic tools.

The results of the analysis validate the architecture of IdSay, which was presented in Figure 2.2, and that was essentially left unaltered.

The results of the analysis seem to validate our option to use an IR module as the base component for search. This option is common among QA Systems, that rely on the ranking of documents both to treat a small number of documents and to score answers. Our approach in IdSay is different: we aim at processing all documents that have occurrences of the words and entities (two or more words co-occurring together in the same order) in the question. For that we developed a retrieval module that has two indexes: one for words and the other for entities.

We tried to keep the entire process as simple and fast as possible, especially the modules in the retrieval cycle, so that it would be possible to process a large amount of information (many documents) a large number of times (many cycles). We focus our ranking efforts on the answers, taking redundancy as a key factor.

In our analysis we identified, as main sources of other systems' success, the use of semantic information resources that would influence the retrieval and validation of answers, and the use of linguistic analysis tools that accounted for better question treatment and answer extraction. Since one of the goals of IdSay is robustness, in the sense that it may be used with noisy data, as that produced automatically by speech transcription or machine translation, we opted for the introduction of semantic information instead of deep linguistic processing.

The current version of IdSay also benefited from changes and enhancements in other parts of the system, namely: the scoring mechanism of the answers, that in IdSayBL relied almost exclusively in the number of occurrences of the answer in the collection, was substantially changed. It now takes into account information that starts at passage level, reflecting the percentage of words that were presented in the question versus the ones that were obtained through equivalences, and reflects the method that was used to obtain the answer, for instance valuing information that was obtained directly from the Wikipedia web page, or the number of retrieval cycle in which the answer was obtained, among other factors. We improved the treatment of acronyms and abbreviations, normalizing them, and this change also had benefits in terms of dates with centuries with AC[BC] qualification. The treatment of numeric values was also taken into account with normalization taking place in terms of milliard separator and decimal separator. The passage retrieval module was improved in a way that they become closer to meaningful sentences, and list questions were considered for the first time. Some minor adjustments were also made in the Question Analysis module, especially regarding co-reference resolution in cluster questions.

8.2 Semantic Relations - Equivalences

Given the design options described for IdSay, it is natural to give priority to improving the recall of our baseline by searching for words and entities that are semantically equivalent to those present in the question.

The introduction of semantic equivalences between words and entities used two different

sources. Since we privilege the use of public free resources, for equivalences at a word level for Portuguese we used the TeP base (Dias-da-Silva et al. 2000; Maziero et al. 2008), and for equivalences at an entity level, we built a resource based on Wikipedia, which we called Wikipedia Entity Synonyms (WES base[1]) that we compiled using the Portuguese Wikipedia Version that is also part of the text collection of QA@CLEF. The WES base is obtained from the Wikipedia using context pages names and redirect files. The format is compatible with TeP base: the synset[2] starts by the canonical name of the entity in the Wikipedia (the name of the file that has content) followed by the existing alternative names of the entity (names of redirecting files to the content file). The alternative names are separated by a comma surrounded by spaces, since the names may themselves contain commas, but never surrounded by spaces on both sides. The file has 46 586 synsets, ordered alphabetically by canonical names, and the names are kept exactly as they appear in Wikipedia, without any processing.

An Example is:

15713. [Wikipedia] {*Fernando Henrique Cardoso , Presidente Fernando Henrique Cardoso , Fernando Henrique , FHC*}

The WES base proved to be a valuable resource to the system and helped answering several questions. Using the example given, for (Question#23 "Quem é FHC"?) [Who is FHC?], the equivalence between FHC and "Fernando Henrique Cardoso" allowed the retrieval of the correct answer with information on the former Brazilian President.

To store this information we added another level to our indexes file, Level 4 - synonyms. The indexes structure represented in Figure 2.3 now becomes that of Figure 8.1.

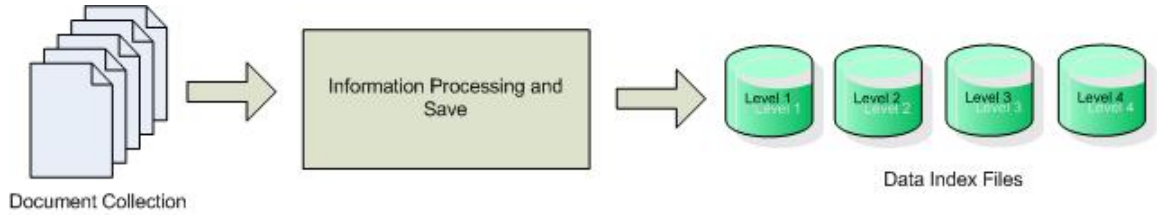


Figure 8.1: Information Indexing in IdSay - Synonyms

This new data structure allow changes in several components of the IdSay. It can be accessed

[1] Available at the resources section of IdSay web page.
[2] Synset is an entry representing a semantic equivalence between entities, in this case.

by the functions $WordSyn$ and $EntitySyn$, that receive an word or an entity, and return the set of words/entities that are synonyms of the word/entity.

In the document retrieval component the Algorithm 6.11 is updated to the Algorithm 8.1. The difference is the use of $SynDocsUnion$ that returns the union of the inverted index of the synonyms of the word considered, and not only the documents where the word belongs, and all other words with the same root. With synonyms we consider sentences where each word can be replaced by any synonym, word or entity, that could even represent distinct word meanings. In the example above, it allows to return the wikipedia page "Fernando Henrique Cardoso" since it is a synonym of "FHC".

Algorithm 8.1 Document Retrieval

Parameters:
Input: $questionRoot$
Output: $resultDocs$
Using:
$FindEntities(sentence)$ function returns the set of entities in the $sequence$

1: $resultDocs \leftarrow SynDocsUnion(questionRoot(1), \{\})$
2: **for** $i = 2 \rightarrow \#questionRoot$ **do**
3: $\quad resultDocs \leftarrow resultDocs \cap SynDocsUnion(questionRoot(i), \{\})$
4: **end for**
5: $entities \leftarrow FindEntities(questionRoot)$
6: **for** $i = 1 \rightarrow \#entities$ **do**
7: $\quad resultDocs \leftarrow resultDocs \cap SynDocsUnion(\{\}, entities(i))$
8: **end for**

The Algorithm 8.2 returns the documents that contains a word or entity, or any other synonym.

In passage extraction, the Algorithm 6.14 have two small changes, that can be explained without reproducing the algorithm here. Its main change is in the word match, $doc[i] = questionRoot(j)$ is extended to the use of word synonyms: $questionRoot(j) \in WordSyn(doc(i))$. Now, for a passage match to the question, is required that all words in the questions to have in the passage at least one synonym. This is done making use of the change in Document Retrieval, that now returns all documents with the words or synonyms in the question.

Another change is in the $resultPassage$ structure, where each passage is a pair with the document and the passage $(doc, (doc(min), ..., doc(max)))$, and now the information of the number of synonyms used in the retrieval of the passage is also returned

Algorithm 8.2 Syn Docs Union

Parameters:
Input: *word*, *entity*
Output: *synDocs*
Using:
docs(*word*) return the documents that contains *word*, structure updated in level 2.
EntityDocs(*entity*) return the documents that contains *entity*, structure updated in level 3.

```
1: if word ≠ {} then
2:    synDocs ← docs(word)
3:    for synWord ∈ WordSyn(word) do
4:       synDocs ← synDocs ∪ docs(synWord)
5:    end for
6:    for synEntity ∈ EntitySyn(word) do
7:       synDocs ← synDocs ∪ EntityDocs(synEntity)
8:    end for
9: else if entity ≠ {} then
10:   synDocs ← EntityDocs(entity)
11:   for synEntity ∈ EntitySyn(entity) do
12:      synDocs ← synDocs ∪ EntityDocs(synEntity)
13:   end for
14:   for synWord ∈ WordSyn(entity) do
15:      synDocs ← synDocs ∪ docs(synWord)
16:   end for
17: end if
```

$(doc, (doc(min), ..., doc(max)), synonyms)$. This allow us to use a scoring mechanism sensible to the number of synonyms used, since the more synonyms used less chances exist to get a passage related to the question, since each synonym can represent a different meaning of the word.

In answer extraction, the Algorithm 6.20 for extract definition answers from passage, instead of processing only the entity, it now process all synonym entities, detailed Algorithm 8.3 with the function $ExtractDefinitionAnswer$ as the Algorithm 6.20.

In Algorithm 6.21 for extracting a generic answer from passage, instead of cleaning the words that are in the *questionRoot*, a set with the synonyms of all words in *questionRoot* is used. This way we prevent from a synonym in the question appear as an answer, expanding the previous option of not answer with the question.

Algorithm 8.3 Extract Definition Answers From Passage

Parameters:
Input: $answers, support, questionRoot$
Output: $answers$
Using: $separator, delimiter, terminator$
$entity(i)$ return the entity of index i

$sep \leftarrow \{separator, delimiter, terminator\}$
$synQuestion \leftarrow \{questionRoot\}$
if $\#questionRoot = 1$ **then**
  $synQuestion \leftarrow synQuestion \cup \{(WordSyn(questionRoot(1)))\}$
  for $entID \in EntitySyn(questionRoot(1))$ **do**
    $synQuestion \leftarrow synQuestion \cup \{entity(entID)\}$
  end for
else if $Entity(questionRoot) \neq \{\}$ **then**
  for $entID \in EntitySyn(Entity(questionRoot))$ **do**
    $synQuestion \leftarrow synQuestion \cup \{entity(entID)\}$
  end for
  $synQuestion \leftarrow synQuestion \cup \{(WordSyn(Entity(questionRoot)))\}$
end if
for $i = 1 \rightarrow \#synQuestion$ **do**
  $answers \leftarrow ExtractDefinitionAnswer(answers, support, synQuestion(i))$
end for

8.3 Ontological Knowledge

Both Priberam and Senso use a proprietary ontology. Priberam's ontology is described in (Amaral et al. 2004), and mentions that it was obtained by translating 195,000 English entries. These entries are organised by senses, in an hierarchical four level structure that lead from 28 top nodes to 3387 end nodes. Senso's ontology (Saias & Quaresma 2008b) has 3500 concepts connected by relations such as *isA*, *usedFor*, *locatedAt*, *capableOf* and *madeOf*. Although the two systems make different uses of this information, our analysis suggested that it is an important component for the overall results, so we opted for incorporating ontological knowledge into IdSay, to be used during answer extraction, and to validate if a candidate answer has the type determined by the Question Analysis module. The most natural way to do that was once more to use Wikipedia[3], which can be viewed as a simple ontology if one considers its category structure (Yu et al. 2007).

We implemented a generic IsA filter, for which the category (qTE) is extracted from the question text directly. If the category or a synonym is in the category string of the

[3]In accordance with our preference for using Community Knowledge Bases that require minimum manual maintenance from single researchers/groups.

Wikipedia page, then return true. We build also filters for $IsAPerson$, $IsAOrganization$, and $IsALocation$. Since these filters do not correspond to exact categories of Wikipedia, we manually built authority lists for words that can be considered as part of that category, and we search for them in the first 50 words of the Wikipedia pages. If one of the words in the authority lists are there, the result will be true, otherwise will be false.

For the $IsAPerson$ filter we use not only the authority lists, but also the following heuristic: if in the first 50 words of the wikipedia page, exists one or two dates, accept the entity as a person.

Although we were limited to the fact that only entities in the Wikipedia could be checked, and had to rely on Wikipedia's sometimes incomplete or tangled category structure, we consider the filters implemented very helpful. The filters that yielded better results were generic IsA and $IsAPerson$. $IsAOrganization$ and $IsALocation$ proved more difficult to filter, but nevertheless there was an improvement in answering this type of questions.

8.4 Numeric Values

8.4.1 Pre-processing

If we did not treat the numeric values, the numbers would be indexed exactly as they appear in the original text. This could present several problems:

- Since we index punctuation marks and they are used to separate sentences and phrases in the text, as well as to separate milliards and decimal parts of numeric values, our Passage Extraction module could produce incoherent passages staring or ending in the middle of a numeric value;
- The same numeric value can be written in several formats: with or without milliard separators; grouping or not digits; or using European number format [4] or Anglo-saxonic number format [5]. This could prevent the match of the same numeric value existing in different formats;

[4] Point as milliard separator and comma as decimal separator.
[5] Comma as milliard separator and point as decimal separator, as used in this Thesis.

- Sometimes a list of distinct numbers are separated by commas, with the risk of being misunderstood for a milliard/decimal separator;
- Finally, since we use a dictionary of words, for each number appearing in the text an entry in the dictionary would be created, leading to a possible explosion in the number of words in the dictionary.

In order to solve these problems, we normalize the numeric values in the following way: the numbers are organized into three digit groups, with the milliard separator normalized to a single white space[6]. We make this grouping after checking if a numeric value is written using the European number format or the Anglo-saxonic number format. The decimal part separator is also checked, according to both formats, and a special separator, which we named "numeric comma", is used for this purpose. The numbers are organized into three digit groups both to the right and the left of the "numeric comma", if it exists. In this way we have in our word dictionary a maximum (1000+100+10) values that are numeric words (corresponding respectively to the number of groups of 3, 2 and 1 digit). If we add the "numeric comma" we have at most 1111 numeric words in our word dictionary. Finally, to separate lists of numeric values we use the semi-colon.

Next, we present formally the normalization algorithm for numeric values using regular expression following the notation used in (Sudkamp 2006). The procedure consists of the following rules, applied sequentially by rule number order:

1. Identification of a number in a text

   $Number = [0-9][0-9,.\ ]^*$

   This expression allows the detection of the biggest sequence of digits commas, points and spaces.

2. Removal of the ending non digit characters

   $Number = [0-9][0-9,.\ ]^*[,.\ ] \rightarrow [0-9][0-9,.\ ]^*$

[6] As explained in Chapter 5 the space is not a word in the dictionary, we assume a space exists between two words when printing a text. Therefore this milliard separator appears in the output, but is not stored in the index.

If the number ends with spaces, commas or points, they are removed. The most common case we expect to handle with this rule is a single space after the numeric value. After this step, we guarantee that the number starts and ends with a digit.

3. Removal of spaces from the middle of the number

   $Number = [0-9][0-9,.\ ]^* \rightarrow [0-9][0-9,.]^*$

   All spaces are removed to facilitate the latter grouping of digits (groups of three digits), independently of the original form it was written. Example: a telephone number, written as 21 345 11 22, would become 213451122.

4. Single milliard separator

   $Number = [0-9]\{1,3\}[,.][0-9]\{3\} \rightarrow [0-9]\{1,3\}[0-9]\{3\}$

   If we have a number with a single point or comma followed on the right hand side for a group of three digits, we consider it a milliard separator.

   Examples: $12,456 \rightarrow 12456$ (meaning 12 456); $123.456 \rightarrow 123456$ (meaning 123 456)

   In these cases the separator can be a decimal separator and not a milliard separator, but we have no way to know what is intended, so we opted for considering it a milliard separator. Note that the cases with number of the right hand side different from three, we consider the single separator as a decimal separator.

5. European number format with decimal part

   $Number = [0-9]^+(.[0-9]^+)^*, [0-9]^+ \rightarrow [0-9]^+([0-9]^+)^+0, [0-9]^+$

   If the number has zero or more points in the left hand side, and a comma in the right hand side, the number is in the European number format with decimal part. We remove the milliard separators and convert the decimal separator to the "numeric comma", henceforth represented as "**0,**".

   Example: $10.008.555,98 \rightarrow$ 10008555 **0,** 98

6. Anglo-saxonic number format with decimal part

   $Number = [0-9]^+(,[0-9]^+)^*.[0-9]^+ \rightarrow [0-9]^+([0-9]^+)^+0, [0-9]^+$

   If the number has zero or more commas in the left hand side, and a point in the right hand side, the number is in the Anglo-saxonic number format with decimal part. We remove the milliard separators and convert the decimal separator to the "numeric comma" "**0,**".

Example: $10,008,555.98 \rightarrow 10008555$ **0,** 98

7. Numbers without decimal part

   $Number = [0-9][0-9.]^+ \mid [0-9][0-9,]^+ \rightarrow [0-9][0-9]^+$

   If the number has only points or commas, then they must be removed since they are milliard separators. If the number has only one comma or point it would have been treated in one of the previous two rules.

8. List of numeric values

   $Number = [0-9][0-9,.]^* \rightarrow [0-9][0-9;0,]^*$

   If there are several commas and points then we consider that we are in the presence of a list of numbers, and that points are decimal separators and commas are number separators. Points are replaced by "numeric comma" "**0,**" and commas are replaced by semi colon.

   Example: $1.4, 3.6, 77.3 \rightarrow 1$ **0,** $4;3$ **0,** $6;77$ **0,** $3;\ 1, 4.3, 6.77, 3 \rightarrow 1;4$ **0,** $3;6$ **0,** $77;3$

9. Grouping digits three by three

   $Number = [0-9]\{1,3\}([0-9]\{3\})^* \rightarrow [0-9]\{1,3\}\ \ ([0-9]\{3\}\ \ )^*$

   $Number = [0-9]\{1,3\}([0-9]\{3\})^*0, ([0-9]\{3\})^*[0-9]\{1,3\} \rightarrow$
   $[0-9]\{1,3\}\ \ ([0-9]\{3\}\ \ )^*0, (\ \ [0-9]\{3\})^*[0-9]\{1,3\}$

   We arrange the digits in groups of three keeping the number of numeric words in the dictionary low, and at the same time keeping (or increasing) the readability of the number. The first rule is for numbers without decimal part, and the second one for numbers with a decimal part.

   Example: a telephone number 213451122 would become 213 451 122;
   The number 10008555 **0,** 98 → 10 008 555 **0,** 98

8.4.2 Intervals and uncertainty

The systems now treats intervals and uncertainty for instance (cerca de[around]).

For dealing with intervals and uncertainty in numeric answers we use the three following rules:

1. Uncertainty

   $Number$ = (“cerca de” | “perto de” | “menos de” | “mais de” | “aproximadamente” | “em média” | “quanto muito” | “pelo menos”) $Number$

   This rule adds words specifying uncertainty, before the number itself like “around 25”.

2. Interval Range

   $Number$ = $Number$ (“a” | “-”) $Number$

   This rule joins two numbers separated by a linking word like “to” in “1 to 5”.

3. Interval Between

   $Number$ = “entre” $Number$ “e” $Number$

   This rule has a similar function to the previous rule, but with a prefix to the interval, for example “between” in “between 10 and 15”

If these expressions are present in the text, it would not be correct to return only the numeric value without them, for instance returning the answer 15 when the text specifies “around 15”.

These rules allowed us to answer correctly to the following question:

Q#8: Quanto pesa um beija-flor? [What is the weight of a hummingbird?] Answer 19 a 21 gramas[19 to 21 grams] was obtained because of the enhancements in measure quantities that allow uncertainty and intervals.

8.4.3 Numbers written out as phrases

Numeric values written as phrases, for instance, oito milhões trezentos e cinquenta e oito mil duzentos e dezassete[eight million three hundred and fifty-eight thousand and two hundred and seventeen] will be transformed to numeric values because of the root of the words that have a numeric meaning, that are presented in Table 8.1 and for large numbers in Table 8.2.

The number above becomes 8 1000000 300 e 50 e 8 1000 200 e 17[8 1000000 300 and 50 and 8 1000 200 and 17][7]. The transformation of these value to the correct value 8 358 217 is left for

[7]Between [] is the literal translation of the Portuguese text before it. Applying the method to the English number “eight million three hundred and fifty eight thousand and two hundred and seventeen” would yield “8 1000000 3 100 and 50 8 1000 and 2 100 and 17”.

Table 8.1: Numeric Words' Roots

Word	Root	Word	Root	Word	Root	Word	Root
zero	0						
um	1	onze	11			cem | cento | centena | centenas	100
dois	2	doze	12	vinte	20	duzentos	200
três	3	treze	13	trinta	30	trezentos	300
quatro	4	catorze	14	quarenta	40	quatrocentos	400
cinco	5	quinze	15	cinquenta	50	quinhentos	500
seis	6	dezasseis	16	sessenta	60	seiscentos	600
sete	7	dezassete	17	setenta	70	setecentos	700
oito	8	dezoito	18	oitenta	80	oitocentos	800
nove	9	dezanove	19	noventa	90	novecentos	900
dez	10						

Table 8.2: Numeric Words' Roots - Large Numbers

Word	Root
mil | milhar | milhares	1000
milhão | milhões	1000000
bilião | biliões | bilhão | bilhões	1000000000
trilião | triliões | trilhão | trilhões	1000000000000

future work. The advantage of this method is that we are able to match round numbers, and also match two numbers, both written in words. Besides, more important, this way of treating the numbers written as words allows them to be identified as numeric values and be returned as answers to questions requiring numeric answers. Still we cannot match a generic number written in words to the corresponding numeric value, unless they have only one word in Table 8.1.

8.4.4 Extraction of Numeric Answers

The extraction of numeric answers occurs in two different situations depending on the question. This is detected in the Question Analysis module, and a variable qTE (**q**uestion**T**arget**E**ntity) is used to identify words that qualify the numeric answer, i.e. the expected answer must be a number together with the words in qTE. The two situations are:

- Count Questions

These questions require a number of occurrences of the "item" specified in the question. In this case qTE contains the item mentioned, or is empty if we are unable to identify the item.

An example of this type of question is Quantos habitantes tem Lisboa? which results in a qTE=habitantes.

- Measure Questions

  In the case of this questions, once the type of measure is identified, the corresponding authority list is copied to qTE. Besides of being a list, for instance the several units allowed to express a length, an element of a list can be composed of more than one word, as the case of square kilometres that are identified as "km 2" (two separate words), since this is the way we use to normalize this unit of area in the pre-processing stage.

  An example of this type of question is Qual a área de Lisboa? which results in a qTE=cm 2 | km 2 | m 2 | hectares | hectar | ha.

  We present the authority lists for the measures used in Table 8.3.

The Algorithm 8.4 to extract numeric answers from a passage is done in two steps:

1. **Go through the passage words and mark as {} all words that are not numbers, and that do not belong to qTE.** Lines 1-6.

   This step removes words that are not eligible as answers, and delimits sequences of words that can be answers. The "numeric comma" "**0,**" is not deleted because it is considered a number, keeping a number with decimal part as a whole.

2. **Go through the passage a second time, and for all non {} word sequences, if qTE is empty accept the sequence as a possible answer, otherwise accept the sequence only if it ends with a possible value of qTE.** Lines 7-17.

   This guarantees that all longest numeric answers that respect the qTE restrictions are extracted as required.

Table 8.3: Authority Lists

Measure	Forms in the Question	Units in the Answer
linear	comprimento | largura | altura | altitude | profundidade | distância | diâmetro | raio | mede | medida | envergadura	metros | centrímetros | quilómetros | metro | centrímetro | quilómetro | km | m | cm | mm | dm | pés | jardas | milhas | polegadas
area	área | superfície	cm2 | km2 | m2 | hectares | hectar | ha
volume	volume	litros | litro | l
weight	pesa | peso	quilos | gramas | toneladas | quilo | grama | tonelada | kg | g | ton
money value	custa | custo | dotação | preço | montante | salário | orçamento	euros | escudos | contos | pesetas | francos | liras | dollars
temperature	temperatura	graus negativos | graus centígrados negativos | graus centígrados | graus celsius negativos | ^{o}c | graus celsius | graus delisle negativos | ^{o}de | graus delisle | graus fahrenheit negativos | ^{o}f | graus fahrenheit | graus newton negativos | ^{o}n | graus newton | graus rankine negativos | ^{o}r | ^{o}ra | graus rankine | graus réaumur negativos | graus réaumur | graus kelvin | ^{o}k | k | graus
time	tempo | dura | demora | leva	séculos | anos | meses | dias | semanas | horas | minutos | segundos | século | ano | mês | dia | semana | hora | minuto | segundo
age	idade	anos | ano
speed	velocidade	km/h
percentage	percentagem | probabilidade	% | por cento | pontos percentuais

8.5 Abbreviations and Acronyms

Acronyms were not normalized in the baseline version and were kept as they appeared in the plain text. This led to two kinds of problems:

- an acronym could be missed if it was identified differently in the collection and in the question; the same could happen with two different occurrences of the same acronym in the collection,
- an acronym could introduce false punctuation marks that affected passage segmentation.

Algorithm 8.4 Extract Numeric Answers From Passage

Parameters:
Input: $answers, support$
Output: $answers$
Using: qTE

```
1: numbers ← support
2: for i = 1 → #numbers do
3:    if not IsNumber(numbers(i)) and numbers(i) ∉ qTE then
4:       support(i) ← {}
5:    end if
6: end for
7: answer ← {}
8: for i = 1 → #numbers do
9:    if numbers(i) ≠ {} then
10:      answer ← answer + numbers(i)
11:   else if answer ≠ {} then
12:      if qTE = {} or qTE ∩ answer ≠ {} then
13:         answers ← answers ∪ {answer}
14:      end if
15:      answer ← {}
16:   end if
17: end for
```

The normalization is done after the pre-processing steps of lowercase conversion and splitting the words by punctuation marks, with the punctuation marks becoming words themselves. For example an occurrence of EUA - Estados Unidos da América [USA - United States of America] in the form of "E.U.A." would become "e . u . a .", after the pre-processing steps referred above.

We then do the normalization, that consists of the two following rules:

1. Identification of a letter in an acronym

   $Acronym = [a-z]\ . \rightarrow [a-z].$

   With this rule the letters followed by a point (two words) will be replaced by a single word composed of the letter and the point together, eliminating the possible confusion of the point as the end of a sentence.

   For example "e . u . a ." (six words) will become "e. u. a." (three words).

2. Fusion of letters of acronyms

   $Acronym = ([a-z].)^{+} \rightarrow ([a-z])^{+}$

   This rule joins all the letters in the acronym, removing the points.

The example “e. u. a.” (three words) finally becomes “eua” (one word).

3. Fusion of full stops

   $.(.)^{+} \rightarrow \ldots$

   If more than one full stop occur together, we replace them by an ellipsis, independently of the number of full stops being 2, 3 or 20. The several words become one, with the full stop as a root, since the ellipsis has the function of terminating a sentence. This rule is intended to prevent the extraction of incorrect passages.

8.6 Dates

The only pre-processing done, as far as dates are concerned, is attributing the roots 1 to 12 to two different forms of identifying the month names. These are presented in Table 8.4.

Table 8.4: Month Names' Roots

Month Name	Short Form	Root
janeiro	jan	1
fevereiro	fev	2
março	mar	3
abril	abr	4
maio	mai	5
junho	jun	6
julho	jul	7
agosto	ago	8
setembro	set	9
outubro	out	10
novembro	nov	11
dezembro		12

There is no short form in the table for the month of Dezembro[December] because its short form dez coincides with the number ten, which already has 10 as root. This pre-processing is part of level 2, and the rest of the treatment of dates is done at extraction time.

The identification of dates is done through the following steps:

1. Identification of a possible date

   $Date = Number(Separator\ Number)\{0,5\}$

A *Separator* is any element (word) from the following list: , ; – : ( ) { } [ ] | • / *de e ou*

This step identifies 1 to 6 numbers, corresponding to year, month, day, hour, minute and second, separated by an element of the above list.

The numbers can also be words, as long as its root is a number, to cover the names of a month. The separator must be the same used several times, except the last one, supposedly separating the year

An example is the date *8 de setembro , 1 970*. This date has three numbers, 8, 10 and 1970 that are separated by the word "de" and a comma. The word September has the root 10, allowing it to be matched to the above expression.

2. Reject big numbers

   $Number = [0-9]\{1,3\}\ [0-9]\{3\}\ ([0-9]\{3\})^{+}$

   If any of the numbers between the separators has more than two words (set of 3 digits), then it is rejected as a date, since there is no meaning in such big numbers is a date, for instance 1 000 000.

3. Check if the numbers can be a date

   After having identified the 1 to 6 numbers that can potentially be a date we check if they can really be a date by going through the list of upper bounds on the six numbers. The list is presented in Table 8.5, ordered by increasing value of upper bound.

   If there is a one to one correspondence between the numbers in a potential date and a number meaning, satisfying its upper bound, then the date is accepted, otherwise it is not.

Table 8.5: Upper Bounds on Date Numbers

Number meaning	Upper Bound
month	12
hour	24
day	31
minute	59
second	59
year	10 000

Having the table ordered as indicated allows the checking of the numbers to be done through the following algorithm:

For each number in the potential date we go through the list and select the smallest upper bound that is bigger or equal than the number being checked, and remove that line and number meaning from the list. If, for a number, we are not able to remove a line with the above criterium we reject the date as a whole.

For example the date *8 de setembro , 1 970*, that has three numbers, 8, 10 and 1970, the 8 will remove the month line, the 10 will remove the hour line and 1970 will remove the year line. All number were successfully assigned to an upper bound, therefore the date is accepted. The correspondence of the number meaning is not the correct one, since 10 is the month and it removes the hour line, however that is not the aim of the procedure, we want only to check a possible assignment to a date.

4. Add ancient dates prefix/suffix

   $Date = [século \mid séc \mid séc. \mid seculo \mid sec \mid sec.]\ Date$

   $Date = Date\ [ac \mid bc \mid dc]$

   There are words that may occur together with a date, either as prefixes or suffixes, in the case of ancient dates. Since they belong to the date, they are added to it.

   An example is 3 000 that is identified as a date, but if it is followed by AC it is concatenated to *3 000 ac*.

   AC and *DC* are the normalized forms of for instance *A.D.* and *D.C.*, that have been treated by the rules of acronyms described in Section 8.5.

5. Add Roman ancient dates

   $Date = [século \mid séc \mid séc. \mid seculo \mid sec \mid sec.]\ [ivxlcdm]^+$

   $Date = [ivxlcdm]^+\ [ac \mid bc \mid dc]$

   Sometimes ancient dates use the Roman numeration system, as common with centuries. We can only identify these dates if they are accompanied by a prefix/suffix. That is usually the case when using Roman numeration for dates.

   An example is *sec. xii* that will be identified as a date by the first rule.

8.7 Scoring mechanism

The main aim of our system is to return a list of ranked answers, preferably with the correct answer (or the correct answers) at the top of the list. Our baseline scoring mechanism was based on the source of the answer being the wikipedia for some definition answers and on redundancy in all other cases. The number of occurrences of an answer in the collection, improved with a merging step in the AV module, was the basis on which the list of answers were ordered. A score value between 0 and 1 is useful for instance for the calculation of such metrics for the system as Confidence weighted score, CWS. Therefore one of the improvements to the baseline was the creation of such a score. It is more a formal improvement than a direct improvement in answering new questions (verify with examples if the ordering of answers were changed for the best in the baseline and current version)

The scoring is based in the following:

- Redundancy continues to be the main source of scoring;
- A Passage score was introduced that is propagated from PR → AE → AV based on the number of words/entities of the question/query that it contains (that are highly valued) and the number of words/entities in the passage that were obtained through equivalence relations (that are less valued, namely for our lack of Word Sense Disambiguation mechanism, which make them possibly unreliable);
- The validation of the filters for the type of the answer instead of working on a binary accepting/excluding philosophy, introduces a validation scale that is reflected in the score of the answer;
- A boost factor is attributed to answers coming from directly from the wikipedia page of the entity that was identified as reference entity of the question;
- The number of the retrieval cycle in which the answer was found is considered in the score, valuing answers found earlier cycles

From the Answer Extraction component, each answer have a score, that is greater if the extraction was done with less uncertainty, like in the wikipedia page of the reference entity. The score of an answer is then split in half if *RetrievalCycle* is not in the first cycle, to give more

importance to the answers in the first cycle. After that the score is divided by the number of synonyms used plus one, to allow reducing in a low score the passages where many synonyms are used. Also, if there are more than one terminator in the passage, the score is reduced since the passage have at least two sentences, and the probability of having something useful is lower.

The answers are processed in Answer Validation by the sum of the score of the supporting passages, but the score of the answer shown is the score of the best supporting passage.

8.8 Roots for the Portuguese Language

The use of lemmatization in level 2 was also changed. For all terms we check if the string version of the term is in the lexicon, using binary search. If it is we return its lemma. Since no Morpho-Syntactical disambiguation is considered by IdSay, if a term has more than one lemma listed in the lexicon the lemma to be used is chosen in the following order:

1. If only one equal word has the cat="n" (name), use that lemma
   Example: forcados : cat="n" forcado*; cat="v" forcar The word forcados appears twice in the lexicon: once with cat="n" (name) the lemmas is forcado* and another with cat="v" (verb) with lemma forcar. The chosen form (marked with *) is first one corresponding to a name
   English: forcados are a group of men who participate in a bullfight on foot, organised in a straight line, with the objective of grabbing the bull by the horns and neutralize it. Forcado is the singular of the name and forcar is the verb related to that activity, of which forcados is the past participle.

2. If two or more words has the cat="n" (name), use the lemma with the smallest editing distance from the original word
   Example: mulheres : cat="n" mulher* ; cat="n" marido ; cat="n" homem
   English: mulheres[women], mulher[woman] marido[husband] (women in the sense of wives) and homem[man]

3. If none of the words has the cat="n", use the the lemma with the smallest editing distance from the original word
   Example: varia : cat="v" variar*; cat="v" varear
   English: varia[varies] with two alternative written forms of the verb "to vary"

8.9 Results of Improved Version

Table 8.6 summarizes the results of IdSay system after the improvements that were described.

Table 8.6: IdSay results overview after improvements

Accuracy over the first answer	Accuracy over all answers	MRR
50.500%	62.500%	55.500%

The results of IdSay after the improvements described in the last chapter, in comparison with the other systems results, are presented in Table 8.7.

Table 8.7: Comparison of IdSay results and those of other systems

		IdSay	0 systems	1 system	2 systems	3 systems	4 systems	5 systems
First Answer	R	101	5	23	37	26	8	2
	Not R	99	30	34	25	5	5	0
All (3) answers	R	124	9	27	48	23	13	4
	Not R	76	20	25	23	4	4	0

Although IdSay still has room for improvement, our approach seems to be validated, keeping efficiency (the 200 questions are answered in less than 2 minutes). IdSay produced original results in some cases, as denoted by the 3 questions that it was the only system that was able to answer in the BL version, that increased to 5 and 9 in the current version, considering the first answer only or all the three possible answers, respectively.

Since the changes are made only to IdSay, an improvement means that a question moves from the row of "not R" to the above "R" row.

In Table 8.8, we present the changes according to the quadrant of Table 7.6.

Table 8.8: Final Results Quadrant for IdSay

Results Quadrants IdSay (IdSayBL)	0 or 1 systems	2, 3, 4 or 5 systems
R	14% (10%)	37% (23%)
Not R	31% (35%)	18% (32%)

8.10 Conclusions

An entity synonym base was built based on Wikipedia redirection pages (WES - Wikipedia Entity Synonyms). This information was joined with the information of TeP base for the introduction of a level 4 in IdSay, with equivalences between words and entities. Document Retrieval, Passage Extraction, and Answer Extraction were adapted to accommodate semantic information, keeping efficiency and improving the performance of already good novel procedures.

Ontological knowledge is taken into account using the Wikipedia category structure. The information is dynamically verified depending on the question, using the category string of Wikipedia web pages. Authority lists are used to filter for person/organization/location question/answer types. As far as we are aware, this is a new procedure in QA systems, and it proves to work in several questions.

Numerical values are now normalized, with the introduction of a special word for decimal point, avoiding it to be misinterpreted as a sentence separator/terminator character. Numbers written as full words have numeric roots allowing them to be returned when numeric answers are sought. When looking for numeric values for a specific measure, the corresponding units are considered, filtering out numbers with the wrong units. Intervals and uncertainty are also considered by the system.

Abbreviations and acronyms are now normalised and do not use terminator characters, improving passage segmentation and answer extraction. Dates now support ancient forms, including Roman numbers. The scoring mechanism, which was initially based exclusively on the number of occurrences of the answer in the collection, was significantly enhanced. It now values higher answers originated in passages with less equivalences (more original terms from the question), answers found in the first cycle, or in Wikipedia Pages of the reference entity page and coming from passages with only one or zero terminator characters. A method for root selection when several lemma options are possible was introduced. It values higher the “name category”, then the “verb category”, and in case of ties the editing distance is used.

The improvements made resulted in an increase in IdSay performance, allowing it to achieve an accuracy over the first answer of 50%.

IV Case Study and Conclusions

Introduction to Part IV

Chapter 9 starts with a survey on QA over speech transcripts, describing the Question Answering over Speech Transcripts, QAST task of QA@CLEF. Since no Data Collection Speech Transcripts and Question Set was available for Portuguese, both were developed. The system is tested considering two version of the ASR system, the use of Wikipedia and two punctuation marks options.

Chapter 10 confirms the achievement of the main objective of the thesis: to study and develop new components of IR and QA, to build a complete QA system efficient and robust, that can compete with the current state-of-art QA systems. A list of 21 distinguishing features of the system is presented that we consider are the more relevant ones for the success of our approach. The chapter ends discussing direction for future improvements and new areas of research.

9 Case Study: Question Answering over speech transcripts

9.1 Introduction

This chapter is motivated from our concern, from the beginning, to build a robust system that could perform with different types of data. Here we put that concern to practice and have IdSay system actually working with data that has different characteristics than that it was designed and tested with, i.e. the CLEF data collection.

We had several options to expand our system, ranging from its application to a specific domain collection, as happened with the CLEF campaigns ResPubliQA of 2009 and 2010 (currently in course) that uses the JRC Acquis parallel corpus of European legislation, but the characteristics of the task were less challenging than the previous campaigns, in the sense that the exact answer was not required, only the already segmented paragraph that contained the answer, there was no co-reference between questions (i.e. no cluster question) just to name the most important changes.

Our decision was to try our system with data obtained through automatic speech recognition (ASR). These kind of task had been tried in IR environments both at CLEF and TREC and also in QA at CLEF, QAST (Question Answering in Speech Transcripts) track that run in 2007 (Turmo et al. 2007), 2008 (Turmo et al. 2009) and 2009 (Turmo et al. 2009), but never with the Portuguese language as an option.

9.2 Related Work

The QAST track had its first edition in 2007, and it took place until 2009 (Moreau et al. 2010) (Lamel et al. 2008).

The QAST task in QA@CLEF track had several aspects similar to the main task, especially as far as question types were concerned, assessment and evaluation measures. The main

difference was the corpus being made of speech transcripts instead of written text.

We start by a brief summary of the rules that were kept through all the three editions, following by a description of the corpus and other specificities of each edition.

For each task, for the same speech collection there was always a collection with manual transcripts and a collection with automatic transcripts, so that comparisons could be made.

A development set was made available for participants. Each system could submit up to five answers per question and up to two runs per task. The answers were assessed manually, being attributed one of the four possible judgements:

- 0 - Correct
- 1- Incorrect
- 2 - Non-exact
- 3 - Unsupported

The measures used to evaluate runs were the same for all editions, and they were

- Mean Reciprocal Rank (MRR), and
- Accuracy (over the first answer)

The 2007 QAST edition started with the English language, and along the three pilot editions three different languages were covered, European English in the first edition, with French and Spanish tasks being also offered in the following two editions, besides of English.

In the 2007 QAST edition there were two corpus both for European English:

- one consisted of lectures on *speech and language processing* (the CHIL corpus). Each lecture lasted for about one hour, and the corpus had around 25 hours of duration. The manual transcripts were done by ELDA, while the automatic transcripts were produced by LIMSI and had an Word Error Rate (WER) of about 20%.

- the other one consisted of meetings on the subject of *design of television remote controls* (the AMI corpus). It totalled around 100 hours of conversation corresponding to 168 meetings. The automatic transcripts were produced by the University of Edinburgh with around 38% WER.

In 2007 edition the questions were all factoids belonging to 11 question types (*Person, Location, Organization, Language, System, Method, Measure, Time, Colour, Shape* and *Material*) and they were 100 for each corpus.

In the 2008 edition two other languages were introduced besides of English: French and Spanish. This was done by adding to other corpus in addition of the ones used in the previous edition. Additionally for the new corpus, three different ASR outputs were considered, each with a different WER.

The new corpus introduced were:

- A corpus from broadcast news in French (the ESTER corpus). The news came from different sources (France Inter, Radio France International, Radio Classique, France Culture, Radio Television du Maroc) and the corpus had around 10 hours of duration. The manual transcripts were done by ELDA, while the automatic transcripts were produced with WER of 11.0%, 23.9% and 35.4%.
- A corpus of the European Parliament sessions (the TC-STAR05 EPPS corpus) in English and in Spanish. It totalled around 3 hours of recordings for each languages. The manual transcripts were produced by ELDA, with automatic transcripts with 10.6%, 14% and 24.1% WER for English and 11.5%, 12.7% and 13.7% WER for Spanish.

The types of questions were no longer factoids alone, but there were also definition questions (around 20%) and NIL answers (around 10%) bringing QAST closer to the main task in this regard. The definition questions could be of type *Person, Organization, Object* or *Other*, while for factoids the 11 types of the previous year were kept.

The new addition to 2009 edition was the introduction of oral questions in addition to the written ones that had been in use the previous editions. The objective was to get spontaneous questions. The questions were obtained from people that were given 2 to 4 passages randomly selected from the collection and that had to produce questions that were recorded. These

questions were filtered by ELDA to eliminate questions not within the permitted types. The resulting questions produced the test and the evaluation set of questions. The written questions were obtained by manually transcribing the oral questions, stripping them of disfluencies.

The types of questions were reduced from 11 to 6 for factual questions (Person, Organization, Location, Time, Measure and Language), while the definitions types were unchanged[1].

In 2009 the speech collections used in the first edition were no longer used, and the corpus used were the corpus introduced in the previous edition, with the focus definitively moving from semi-structured speech to prepared speech.

Overall QAST had eight participant groups:

- The UPC (Universitat Politècnica de Catalunya) from Spain co-organised the track and participated in the three editions
- LIMSI (Laboratoire d'Informatique et de Mécanique des Sciences de L'Ingénieur) from France co-organised the track and participated in the three editions with the Ritel System
- INAOE (Instituto Nacional de Astrofísica, Óptica y Electrónica) from México participated in the 2008 and 2009 editions
- Tokyo Institute of Technology from Japan participated in the 2007 and 2009 editions
- Universidad de Alicante from Spain participated in the 2008 edition
- CUT (Chemnitz University of Technology) from Germany participated in the 2008 edition
- The Centre for Language Technology, Macquarie University's Sydney, Australia participated in the 2007 edition with its AnswerFinder QA system, which had also participated in the TREC-QA editions from 2003 to 2007 and a Cross-lingual attempt from Dutch as source language and English as target language, at QA@CLEF 2007
- DFKI from Germany participated in the 2007 edition with its cross-lingual open-domain system that has been used in the QA@CLF main tasks from 2003 to 2008

[1] However, the types *Language* for factual questions and *Object* for definition questions did not occur among the generated questions.

For Portuguese, Priberam also has an on line version of their system working over script transcripts of Portuguese Broadcast news, that is basically the same system that participated at QA@CLEF, but with an additional effort to correct spelling mistakes that come out of the ASR system. The efficacy of the system suffers a drop of performance because of WER, but no further details are known.

As far as we know there are no other systems working in this area for Portuguese.

9.3 Data Collection and Question Set for Portuguese

The data we use are transcripts of Broadcast News from the Portuguese national TV station RTP, Radio Televisão Portuguesa, [Portuguese Radio and Television]. Appendix E have detailed data of the case study.

To build our data collection we considered the editions of the two daily evening Broadcast News from the two channels of RTP, Telejornal from RTP 1, and Jornal 2 from RTP 2. The period used was from the 1st of June to the 11th September of 2008, a period that included the Olympic Games of Beijing 2008. It gave a total of 103 days and 206 editions of Broadcast News. We manually transcribed audio excerpts and built a collection of 100 questions that were answered in these excerpts. We used a variety of themes from 17 editions of Broadcast News from the data collection.

We tried to build a collection that represented the different possible situation of audio that occur in Broadcast News, with a percentage close to the frequency of occurrence of each situation. We did not make an empirical measure of each situation but our experience as listeners, as well as cases picked at random helped in determining the constitution of our data collection. In terms of gender, our data collection is predominantly spoken by female voices: 61 of the question were based upon excerpts spoken by a female voice, while the other 39 excerpts present a male voice. There are 29 different speakers, of whom the vast majority are professionals: television journalists. There are a few cases of well known personalities and also some experts being interviewed sporadically. The corpus is made of 99 European Portuguese Speakers and 1 Brazilian Portuguese Speaker (Lula da Silva, The Brazilian President as of 2008). Table 9.1 contains the information about the speakers by gender.

The conditions of the environment around the speaker were also taken into account, and we

Table 9.1: Data Collection: Speakers Information

	Female Speakers	**Male Speakers**
Total Number	61	39
Different	16	13
Journalists	• ALR (Ana Luísa Rodrigues) • AMF (Alberta Matos Fernandes) • AR (Ana Ribeiro) • AS (Ana Santos) • CC (Cecília Carmo) • CE (Cristina Esteves) • DS (Daniela Santiago) • FF (Fernanda Fernandes) • ILS (Isabel Loução Santos) • JS (Judite de Sousa) • LB (Luísa Bastos) • LS (Laura Santos) • PG (Patricia Gallo) • SF (Sandra Felgueiras) • SMS (Sandra Machado Soares) • TN (Teresa Nicolau)	• AC (António Carneiro) • ASF (Armando Seixas Ferreira) • EP (Eduardo Pestana) • JAC (José Alberto Carvalho) • JB (João Botas) • JRS (José Rodrigues dos Santos) • PS (Paulo Solipa) • VG (Vitor Gonçalves)
Other		• GS (Comissário Gonçalo Simões) Unidade Especial de Polícia • JS (José Sócrates) Primeiro Ministro de Portugal • LS (Lula da Silva) Presidente do Brasil • MS (Manuel Salgado) Vereador da CML • VF (Comandante Vitor Felisberto) Grupo de inactivação de explosivos do Exército

tried to make the corpus balance in that regard. We also aimed at presenting some challenging situations, such as a spontaneous interview conducted over the telephone. The characteristics of the environment for the audio excerpts used are described in Table 9.2.

This data has a number of specificities that makes it very different from the written data we had used before. Its scope are news which could be considered similar to the news paper articles, but they have a lot of differences: one is that they are much shorter that news paper articles, and they are intended for a shorter attention span from the listener, so there are a lot of a anaphoric references that make the news shorter and quicker, without the risk of the listener

Table 9.2: Data Collection: Audio Excerpt Environment

Environment type	Description	Number
A	Anchor	44
A + J	Anchor over a jingle background	7
A + N	Anchor voice in a noisy environment	5
A + M	Anchor voice over music	4
P	Prepared speech by a journalist	2
P + N	Prepared speech in a noisy environment	27
P + M	Prepared speech over music	4
S	Spontaneous speech or interview	4
S + T	Spontaneous speech or interview over the telephone	1
O	Outside piece by journalist	2

getting confused, but which makes task of automatic processing much more difficult. Also the style of language is quite different from the written news style: it tries to be more appealing to keep the listener interested (and it tries to prevent them from zapping to another station) so the language is not very straightforward, and uses many images and other stylistic features, which again makes it more difficult for an automatic system to understand its meaning. Another characteristic of the style is that we have a stronger presence of direct speech in broadcast news as compared to newspaper articles that have a few, and encyclopaedic resources that have even less, if any. The lack of confidence in punctuation marks makes the task more difficult.

Since the transcripts come from video, we lack the aid the image sometimes provides. That limitation is especially important in the case of a foreign language speaker that is subtitled in the video, information that we do not have available in our data transcripts, which makes these situations unusable, similar to a white noise area. This kind of situation occurs frequently, around 3 times in average for a newsreel. Portuguese news emissions tend to be quite long, with an average duration of about one hour, twice daily, and it resources to a lot of outside interviews, on spot coverage, and images from worldwide or local (foreign) companies. Another example of this kind is when you have someone being interviewed and you do not have a clue who that person is because that information is given in written form. For instance in the context of an accident we hear somebody talk and we do not have the information if it is somebody who was involved in the accident, an witness a technician talking about the security conditions, or a minister responsible for the area.

Besides of the language specificities of the data (broadcast news), one as to consider the

inherent difficulties of automatic data transcription, that are more present in this case, for instance: - many different speakers, with different accents and specific vocal characteristics - noisy outdoor environments - man in the street interviews: speakers with different talking aptitudes - interviewing people in the sequence of events where people tend to be in an altered emotional state, either good (e.g. after winning an Olympic gold medal) , or bad (e.g. after being rescued from a train derailment), which has implications in their discourse

The information has a lot of anaphoric temporal references related to the current date and time, with a very high dependence on the meta data, for instance references to today, where the day can only be determined a posteriori using the meta data.

The news sometimes have a very short validity span, for instance what is assumed at noon, can no longer hold in the same evening, so it becomes more difficult to rely in many occurrences of the same piece of information to indicate its accuracy or veracity: either because if a fact is true it is not going to be repeated many times because it would be boring (even for positive events), or because it can no longer be true, in which case the information changes.

Multi-language would be a very useful feature in the transcription of Broadcast news, since there are many local interviews conducted in local languages. For instance, in the broadcast news form 2008_07_26-21_59_02-Jornal2-2.avi, the piece of news about the then candidate to the Presidency of the United States, Barack Obama, and his seven days visit to the middle east and Europe, the most interesting part of the piece, that are the declarations of Barack Obama, since they are in English, are not in the transcripts. The transcripts, being thought to aid people with hearing impairments, are fully responsive to their aim since they allow the automatic creation of sub-tiles for the Portuguese spoken news. But the transcript as a piece of news in itself, with the aim of for instance being used for searching in video streams, should contain these important pieces that come from other languages. The best option would be to have an ASR system that recognized multiple languages and transcribed them, connected to an automatic translation component at some part of the processing chain. However since these pieces are being sub-titled manually, an option could be to add the subtitles to the ASR result, so that the information of the Broadcast News would be complete.

Priberam in its version of the system that works over script transcripts has the only difference the fact that it tries to do spelling check and correction to names in the corpus, while in the version that works for non noisy data it tries to do that only for the question text. The biggest

challenge we have for IdSay is to combine the knowledge sources that are not part of the corpus with the information in the corpus, and to find a support in the corpus for some information we may have found in the complementary knowledge sources, like wikipedia (it can be thought of as having the web as an aid).

For the use of the data collection in IdSay we have a couple of options that allow us to make some tests. First we use two versions of the recognizer, one in 2008 (Meinedo et al. 2008) (Batista et al. 2008) with an word error rate (WER) of 21,5%, and other in 2010 (Meinedo et al. 2010) (Batista et al. 2010) with an error rate of 18,4%, that allow us to study the relative importance of the improvement to the QA system performance. We have from the recognizer punctuation marks but with an high error rate, so we study two options for the punctuations marks in IdSay: set all punctuation marks to commas; use full stops and commans. Since the full stops can have an higher impact in passage retrieval in IdSay, these two options will allow us to know if for IdSay is better or not to treat full stops as commas. Finally, since IdSay have several characteristics dependent on Wikipedia, we test the use of Wikipedia with the data collection, and this way we can verify if the use of Wikipedia is good or not for IdSay performance, even considering the questions does not have the answer in the Wikipedia.

9.4 Results of IdSay

In these section we present the results obtained by the Case Study tests. We will make an analysis starting with the global results, and moving to a more detailed analysis, focusing the usage of punctuation marks, the inclusion of Wikipedia, and the improvements between the 2008 and 2010 versions of the Broadcast news speech recognition system.

Table 9.3 presents a summary of results of the Case Study tests. From the results shown, we can see that the usage of Wikipedia consistently increased the accuracy through all test configurations. The increase is around the double relatively to the corresponding version of without Wikipedia.

Regarding the punctuation marks we have two configurations, using full stops and commas as provided by the recognition system, and transforming all punctuation marks to commas. We can see from the table that using full stops and commas is better than using just commas, except for the 2010 version of the system, using Wikipedia.

Table 9.3: Summary of Results
A - Transcriptspv (full stops and commas)
B - Transandwikipv (full stops and commas, Wikipedia)
C - Transcripts (only commas)
D - Transandwiki (only commas, Wikipedia)

	2008				2010			
	A	B	C	D	A	B	C	D
Accuracy at First	15%	30%	12%	27%	10%	23%	10%	26%
Accuracy All (3 answers)	20%	42%	15%	36%	16%	29%	15%	31%

Comparing the results of 2008 and 2010 we can see a decrease of the results in 2010. While in the results just with commas are more or less the same, the decrease in the situation of using fullstops and commas, suffers a considerable decrease in around 30%. These results are surprising since the 2008 results are always better than those of the 2010 version, except for case C, in the accuracy over all indicator, in which the results are equal.

The best results were obtained for the 2008 version, using full stops and commas and Wikipedia, in which the accuracy over the first answer is 30% and the accuracy over the first three answers is 42%, that we consider a good result taking into account that it is in line with the results obtained at QAST, in which the best performance reported is 41% (but accuracy over the first five answers is considered) and our system performance droped only 20% when compared to the tests using formal written text.

Table 9.4 presents the results in a more detailed way, in terms of number of questions per each possible assessment value.

The first line of this Table is equal to the first line of Table 9.3, while the second line of the latter Table presents the sum of the first 3 lines of Table 9.4 (total number of right questions in the three answers). Since the number of questions is 100, the number of questions or percentages have the same value.

A question is assessed as -2 when the right answer was not obtained due to inexact extraction, but for which the right passage was found. There are many questions with this assessment, more

Table 9.4: Summary of Results per Assessment Value
A - Transcriptspv (full stops and commas)
B - Transandwikipv (full stops and commas, Wikipedia)
C - Transcripts (only commas)
D - Transandwiki (only commas, Wikipedia)

		2008				2010			
		A	B	C	D	A	B	C	D
1st Answer Right	1	15	30	12	27	10	23	10	26
2nd Answer Right	2	3	4	3	6	2	2	2	2
3rd Answer Right	3	2	8	0	3	4	4	3	3
Passage Right, Inexact Extraction	-2	13	26	13	26	19	33	20	32
Unsupported	-1	0	0	0	0	0	2	0	1
NIL Answer	-3	40	6	38	5	43	10	45	7
Wrong	0	27	25	34	33	22	26	20	29

than in the case with formal texts. Adding these questions with the correct ones, we will have the set of questions with correct passage extracted. These results are presented in Table 9.5.

Table 9.5: Right Passages
A - Transcriptspv (full stops and commas)
B - Transandwikipv (full stops and commas, Wikipedia)
C - Transcripts (only commas)
D - Transandwiki (only commas, Wikipedia)

	2008				2010			
	A	B	C	D	A	B	C	D
Right Passages	33	68	28	62	35	62	35	63

These values emphasize the usefulness of the system composed of ASR and Question Answering Systems, as a search mechanism for video streams, where the search can be posed as a question and the relevant video stream is returned. Under these circumstances the performance achieved is more than 60%. To have the correct answer automatically under these circunstances would require the refining of the answer extraction module, right now the correct answer is given in about 30% of the questions.

When no answer is returned by the system, the question is assessed as NIL (-3). We can see

in Table 9.4 that there is a large number of questions with NIL answers in test configurations A and C, that are dramatically reduced through the use of Wikipedia (columns B and D). We once more validate the usage of Wikipedia. From the global values we suspect that these questions moved to correctly answered questions or questions where the right answer was not obtained due to inexact extraction, but for which the right passage was found, since in most of the cases the wrong answers are sensibly the same.

As far as punctuation marks are concerned, we can conclude that there are no significant changes in the distribution of questions that were not answered correctly.

Regarding the versions of 2008 and 2010, in 2010 there was an increase in for the number of questions assessed as -2 and -3, consistently across all configurations, and for wrong answers a decrease in all configurations except B, that increases by one question.

9.5 Statistical Validation of Conclusions

To validate the conclusion on a statistical basis we considered each of the three research questions separately and performed a paired test between all the matched questions for each of the two possible values of the research question being studied. Specifically we performed the three following tests:

- Test T1, in which the type of punctuation marks is analysed. In this test we considered all the situations of the ASR text with both commas and full stops, against the corresponding situation but with the ASR text with all punctuation marks considered as commas.
- Test T2, in which the usage of Wikipedia is evaluated. In this test considered all the situations in which the answers were obtained using the Wikipedia, versus the corresponding situation without the Wikipedia.
- Test T3, that determined if the results obtained with the ASR version of 2008 were different than those obtained using the 2010 ASR version. All situations using the 2008 ASR version were tested against the corresponding situation using the 2010 version.

We performed the non parametric tests indicated for these situations, so that no assumption on the nature of the distribution function of our samples is assumed. We began by calculating the

Sign Test, that only takes into account which of the situations is better, and then we calculated the Wicoxon Signed Rank test, which takes into account the magnitude of the difference between the two situations, whether it corresponds to an improvement or deterioration. The latter test is more powerful than the former, since it takes more information into account, however it makes the further assumption of a symmetric probability distribution (Wilcoxon 1948). For the sign rank test to be meaningful, we performed a conversion from the values that we have been using to the corresponding ordinal scale, with an interval of 1 between values, keeping the same order we have been using in the present chapter [2]. The conversion is presented in Table 9.6.

Table 9.6: Conversion Table for Ranked Scale

Description	Value	Ranked Value
1st Answer Right	1	7
2nd Answer Right	2	6
3rd Answer Right	3	5
Passage Right, Inexact Extraction	-2	4
Unsupported	-1	3
NIL Answer	-3	2
Wrong	0	1

We also performed the parametric counter part of these test, a matched Student's t-Test, for this test is considered to conduct to meaningful results according some authors, despite its assumptions in terms of the normal distribution of the population. For all the tests we assume the effect of the measurement scale originating from a nominal scale to be negligible and still produce stable results. We also assume the randomness and independence of the sample.

For all tests a 95% confidence leval was considered, using a bi-caudal distribution. The tests used the R implementation of the methods. The values obtained to the tests are presented in Table 9.7.

[2]This conversion also reflects the fact that in this chapter we give a slightly different importance for the cases of an inexact answer and an unsupported answer than that used in previous chapters; since the baseline version of IdSay that participated at QA@CLEF 2008 campaign presented a problem in identifying the best passage for supporting a (right) answer, we considered this unsupported score to be the best for the "not right scores", in the present chapter we value the situation in which the right passage is identified, emphasizing in this way the search in video streams of the QA system used with ASR text, and also because the described problem with passage selection has been improved via a better score mechanism for passages.

Table 9.7: Statistical Tests: Results

Description	Sign Test	Wicoxon Signed Rank Test	Student's t-Test
T1 - Punctuation	number of successes = 39, number of trials = 57, p-value = 0.007508 alternative hypothesis: true probability of success is not equal to 0.5 95 percent confidence interval: 0.5475702 0.8009499 sample estimates: probability of success 0.6842105	V = 1103, p-value = 0.02622	t = 2.0995, df = 399, p-value = 0.0364 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: 0.00795219 0.24204781 sample estimates: mean of the differences 0.125
T2 - Wikipedia	number of successes = 174, number of trials = 252, p-value = 1.386e-09 alternative hypothesis: true probability of success is not equal to 0.5 95 percent confidence interval: 0.6294119 0.7469938 sample estimates: probability of success 0.6904762	V = 25935.5, p-value <2.2e-16	t = 9.0518, df = 399, p-value <2.2e-16 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: 0.8493522 1.3206478 sample estimates: mean of the differences 1.085
T3 - ASR version	number of successes = 57, number of trials = 116, p-value = 0.926 alternative hypothesis: true probability of success is not equal to 0.5 95 percent confidence interval: 0.3973676 0.5858417 sample estimates: probability of success 0.4913793	V = 3585, p-value = 0.591	t = 0.7571, df = 399, p-value = 0.4494 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.1037765 0.2337765 sample estimates: mean of the differences 0.065

The values for the three tests are compatible and present statistical evidence of the same conclusions that we had already devised in the previous section, namely:

- T1 - The p-value is below 0.05 in all three testes, therefore the null hypothesis is rejected for a 95% confidence level, meaning that the use of punctuation marks with full stops is relevant because the differences point that way in the tests.
- T2 - The p-value for this case is extremely low in all three tests, therefore the null hypothesis is rejected, and we conclude that the use of Wikipedia is relevant with a high degree of certainty.
- T3 - The p-values are very high so we do not have statistical evidence for the rejection of the null hypothesis in all this cases, meaning that we cannot find significant differences between the two ASR versions in these tests.

9.6 Question Based Analysis

The results of our tests concerning the research questions being investigated were drawn in the last sections, analysing the results globally. In the present section we will further investigate the causes of this behaviour of the system, on a question based analysis, to help shed some light on the results obtained.

To help in this task we have built the following tables, one per each topic being investigated, in which we identify the change in behaviour of the answers to the questions in the corresponding tests.

Table 9.8 reflects the behaviour of the system regarding punctuation marks.

In the columns of the matrix we consider results for the situation where both the full stops and commas added to the transcripts are used (test cases A and B, for 2008 and 2010), while in the rows we have the results for the situation where all punctuation marks are considered to be commas (tests C and D, 2008 and 2010). We consider all the possible outcomes for the question assessment, and make a comparison of the outcomes of the paired test, for instance considering the first column and line in the matrix, we count the number of questions that were assessed right in the first answer in both situations (66 questions). In the second line of the first column, appears the number of questions that produce a first answer right using full stops

Table 9.8: Answer Assessment Changes: Punctuation Marks in Transcripts - T1

		AB								
		1	2	3	-2	-1	-3	0		
CD	1	**66**	3	0	3	0	1	2	9	-9
	2	5	**6**	0	0	1	0	1	2	3
	3	0	0	**9**	0	0	0	0	0	0
	-2	3	0	5	**79**	0	0	4	4	4
	-1	0	0	0	0	**1**	0	0	0	0
	-3	1	0	0	0	0	**91**	3	3	-2
	0	3	2	4	9	0	7	**91**	**18**	25
		12	2	9	9	0	7	**39**		
		12	-1	9	6	-1	6	-10		**21**

and commas, but for which the answers using only commas is only given in the second answer (5 questions). If we follow the first row we have the 66 questions that were answered right at first answers in both situations, and in the second column we can see that there are 3 questions that were answered correctly in the first answer in the commas only test, that were only able to be answered correctly in the second answer of the corresponding test using both full stops and commas. We can see that in the inferior half of the matrix we have the questions which benefited from the use of full stops and commas, while on the superior half of the matrix we have the questions that produced better results using just commas.

The last two rows in the table are summary rows, and make an overall account of all questions that benefited from using both full stops and commas: the first summary row has the number of questions that improved, and the last row discounts the number of questions that had worse results (subtracting to the above row the sum of the corresponding column in the superior part of the matrix). The last position of the first of these summary rows presents the total number of questions that improved using full stops and commas, compared to the only commas situation, which were 39. The same procedure was applied to obtain the values of the two summary columns at the end of the table. We can see that there were 18 questions that produced better results using only commas than using full stops and commas. These values are the values used for the sign test in test T1,in which we compare the number of better results in the full stops and commas (39 questions) situations, in 57 cases where the tests produced different results (39 + 18). Globally we can see in the last cell of the table that the usage of full stops and commas produced better results than using just commas in 21 cases, obtained either

by subtracting the number of worse cases to the number of better cases (39 - 18), or adding the values of the last row or last column of the table.

Table 9.9 presents the same informations for test T2, in which the usage of Wikipedia together with the ASR data is compared with the situation in which only ASR data is used. In this table the cases to be compared are B and D 2008 and 2010 in columns, and cases A and C, 2008 and 2010 in rows. We can see in this case that the number of questions that benefited from the addition of the Wikipedia, is much larger than in the previous case of test T1. It should be notice that this difference could be even higher (148 instead of 96) if we considered answers assessment -3 (NIL answer) equal to 0 (Wrong answer), since 52 questions that were assessed as -3 without the use of Wikipedia drop to 0 when using Wikipedia, are counted as improvements.

Table 9.9: Answer Assessment Changes: Wikipedia - T2

		BD								
		1	2	3	-2	-1	-3	0		
AC	1	**32**	2	0	6	0	3	4	15	-15
	2	2	**6**	0	1	0	0	1	2	0
	3	3	0	**1**	1	0	0	4	5	-2
	-2	21	0	**5**	35	0	0	4	4	22
	-1	0	0	0	0	**0**	0	0	0	0
	-3	31	5	2	48	3	**25**	52	52	37
	0	17	1	10	26	0	0	**49**	**78**	54
		74	6	17	74	3	0	**174**		
		74	4	17	66	3	-3	-65		**96**

The results of the T3 test, in which the versions of the 2008 and 2010 systems were compared, are presented in the same way in Table 9.10. In this table the cases to be compared are A to D 2008 in columns, and A to D 2010 in rows. The numbers are identical as expected, but about half of the questions are better in one system than the other. This could indicate the importance of the recognition system, but in this case the 2010 version does not result only in advantages, so a carefully analysis of the each change isolated could indicate what is best/worst for integration with IdSay.

We will base the analysis on the details of the assessment per question that is presented in Appendix E, divided in three tables, to allow a better legibility, from Table E.18 (Assessment Value per Question: Part 1 of 3 (Questions 1-35)) to Table E.20 (Assessment Value per Question: Part 3 of 3 (Questions 71-100)). We also consider in our analysis the full detail of answers per

Table 9.10: Answer Assessment Changes: 2008 vs. 2010 ASR versions - T3

		2008								
		1	2	3	-2	-1	-3	0		
2010	1	**64**	1	1	1	0	1	1	5	-5
	2	1	**5**	0	2	0	0	0	2	-1
	3	3	2	**6**	3	0	0	0	3	2
	-2	4	3	5	**52**	0	0	40	40	-28
	-1	1	0	0	0	**0**	0	2	2	-1
	-3	3	3	0	5	0	**87**	7	7	4
	0	8	2	1	15	0	1	**70**	**59**	27
		20	10	6	20	0	1	**57**		
		20	9	5	14	0	0	-50		**-2**

question, presented in the Appendix, in Table E.21.

9.6.1 Punctuation

We start by giving an example that shows the influence of the punctuation in the QA system performance. For that we will look at what happens with Question #81. The question, together with the transcripts is presented in Table 9.11.

Table 9.11: Transcripts for Question #81

Question		
0081	2664	Qual a última paragem da visita de Barack Obama à Europa?

Manual Transcript
Barack Obama terminou a visita ao Médio Oriente e à Europa. A última paragem foi em Londres.

2008 Transcript	**2010 Transcript**
Barack Obama terminou visita ao Médio Oriente. E a Europa à última paragem foi em Londres.	Para que Obama terminou visita ao Médio Oriente e a Europa à última paragem foi Londres (. . .)

The automatic transcriptions were obtained from the XML output of the SSNT system. An example of this information is presented in Figure 9.1 for the 2008 version of the system and the excerpt corresponding to Question #81.

```
<TranscriptSegment>
  <TranscriptGUID>230</TranscriptGUID>
  <AudioType start='71096' end='71375' conf='0.703200'>Clean</AudioType>
  <Speaker name='Mulher' gender='F' id_conf='0.056500' gender_conf='0.668100' known='F' id='2014'/>
  <SpeakerLanguage native='T'>PT</SpeakerLanguage>
  <Time reasons='' start='71096' sns_conf='0.984900' end='71375'/>
  <TranscriptWordList>
    <Word start='71114' cap='Barack' end='71143' conf='0.655'>barack</Word>
    <Word start='71144' cap='Obama' end='71175' conf='0.678'>obama</Word>
    <Word start='71176' end='71219' conf='0.949'>terminou</Word>
    <Word start='71220' end='71249' conf='0.960'>visita</Word>
    <Word start='71250' end='71258' conf='0.912'>ao</Word>
    <Word start='71259' cap='Médio' end='71303' conf='0.934'>médio</Word>
    <Word start='71304' cap='Oriente' end='71359' conf='0.989' punct='.'>oriente</Word>
  </TranscriptWordList>
</TranscriptSegment>
<TranscriptSegment>
  <TranscriptGUID>231</TranscriptGUID>
  <AudioType start='71376' end='71642' conf='0.620000'>Clean</AudioType>
  <Speaker name='Mulher' gender='F' id_conf='0.304500' gender_conf='0.834600' known='F' id='2014'/>
  <SpeakerLanguage native='T'>PT</SpeakerLanguage>
  <Time reasons='' start='71376' sns_conf='0.995900' end='71642'/>
  <TranscriptWordList>
    <Word start='71387' cap='E' end='71398' conf='0.611'>e</Word>
    <Word start='71399' end='71409' conf='0.388'>a</Word>
    <Word start='71410' cap='Europa' end='71441' conf='0.775'>europa</Word>
    <Word start='71442' end='71452' conf='0.823'>à</Word>
    <Word start='71453' end='71486' conf='0.855'>última</Word>
    <Word start='71487' end='71539' conf='0.979'>paragem</Word>
    <Word start='71540' end='71558' conf='0.979'>foi</Word>
    <Word start='71559' end='71566' conf='0.900'>em</Word>
    <Word start='71567' cap='Londres' end='71629' conf='0.946' punct='.'>londres</Word>
  </TranscriptWordList>
</TranscriptSegment>
```

Figure 9.1: XML Information for Question 81 - 2008 Version

The correct answer to this question, Londres(London), was given in the second place for the tests in which full-stops and commas were considered (A and B), and the tests that used all punctuation as commas (C and D) produced wrong results, considering the 2008 version of the system. This can be explained because a better punctuation in the cases A and B allows the selection of the passage: “barack obama terminou visita ao médio oriente . e a europa à última paragem foi em londres .” that is almost correct in terms of punctuation (the full stop in the middle of the sentence is misplaced). In this situation the correct answer is extracted. In the case of considering all punctuation marks as commas (C and D tests) the passage extracted is of much worse quality, in terms of sentence boundaries. The passage extracted in this cases is “fica cumprido , um último adeus , barack obama terminou visita ao médio oriente , e a europa à última paragem foi em londres , o se essa viagem foi visível em tróia mas para obama isso até poderá levar ,”.

Table 9.12: Extended Transcripts for Question #81

Question		
0081	2664	Qual a última paragem da visita de Barack Obama à Europa?

Manual Transcript
Barack Obama terminou a visita ao Médio Oriente e à Europa. A última paragem foi em Londres. **O sucesso da viagem foi visível e notório, mas para Obama isso até poderá levar a uma queda nas sondagens**.

2008 Transcript	**2010 Transcript**
Barack Obama terminou visita ao Médio Oriente. E a Europa à última paragem foi em Londres. **O se essa viagem foi visível em Tróia mas para Obama isso até poderá levar. A uma queda nas sondagens.**	Para que Obama terminou visita ao Médio Oriente e a Europa à última paragem foi Londres **o SEF a viagem foi visível em tory, mas para Obama isso até poderá levar uma queda nas sondagens.**

The first part of the passage "fica cumprido , um último adeus ," is still part of the previous news, about the funeral of dead troopers, that was separated by a full stop in cases A and B, which was converted to a comma in cases C and D, causing the system to be totally "mislead". At the end of the passage there is also text that was added, this case from the same piece, but that made it more difficult for the system to extract the correct answer. The 2010 tests produced all NIL results because there was a problem in the extraction of the entity in the question "Barack Obama" by the 2010 ASR version that did not allow the system to retrieve the correct passage. Also there is no correspondence in the WES base between "Obama" and "Barack Obama" since the information was derived from a 2006 version the Wikipedia, and Barack Obama was then a relatively unknown Senator for the State of Illinois.

To better understand the passages we present in Table 9.12 an extended transcript for the question, with the extra text (that was not strictly necessary to answer the question) in bold.

The XML information corresponding to Table 9.12, for the 2010 SSNT version this time, is depicted in Figure 9.2.

```
<TranscriptSegment>
  <TranscriptGUID>230</TranscriptGUID>
  <AudioType start="71097" end="71376" conf="0.703200">Clean</AudioType>
  <Time reasons="" start="71097" sns_conf="0.985200" end="71376" />
  <Speaker name="Mulher" gender="F" id_conf="0.056700" gender_conf="0.668100" known="F" id="2014" />
  <SpeakerLanguage native="T">PT</SpeakerLanguage>
  <TranscriptWordList>
    <Word start="71115" cap="Para" end="71136" conf="0.963243">para</Word>
    <Word start="71137" end="71144" conf="0.930305">que</Word>
    <Word start="71145" cap="Obama" end="71176" relev="true" conf="0.852515">obama</Word>
    <Word start="71177" end="71221" conf="0.995615">terminou</Word>
    <Word start="71222" end="71250" conf="0.984467">visita</Word>
    <Word start="71251" end="71260" conf="0.949324">ao</Word>
    <Word start="71261" cap="Médio" end="71304" relev="true" conf="0.991805">médio</Word>
    <Word start="71305" cap="Oriente" end="71366" relev="true" conf="0.998032">oriente</Word>
  </TranscriptWordList>
</TranscriptSegment>
<TranscriptSegment>
  <TranscriptGUID>231</TranscriptGUID>
  <AudioType start="71377" end="71643" conf="0.620000">Clean</AudioType>
  <Time reasons="" start="71377" sns_conf="0.995800" end="71643" />
  <Speaker name="Mulher" gender="F" id_conf="0.304500" gender_conf="0.834600" known="F" id="2014" />
  <SpeakerLanguage native="T">PT</SpeakerLanguage>
  <TranscriptWordList>
    <Word start="71386" end="71399" conf="0.896664">e</Word>
    <Word start="71400" end="71408" conf="0.808699">a</Word>
    <Word start="71409" cap="Europa" end="71443" relev="true" conf="0.975354">europa</Word>
    <Word start="71444" end="71452" conf="0.910206">à</Word>
    <Word start="71453" end="71486" conf="0.979905">última</Word>
    <Word start="71487" end="71540" conf="0.998398">paragem</Word>
    <Word start="71541" end="71565" conf="0.991537">foi</Word>
    <Word start="71566" cap="Londres" end="71630" relev="true" conf="0.991105">londres</Word>
  </TranscriptWordList>
</TranscriptSegment>
 <TranscriptSegment>
  <TranscriptGUID>232</TranscriptGUID>
  <AudioType start="71644" end="72104" conf="0.608700">Clean</AudioType>
  <Time reasons="" start="71644" sns_conf="0.997000" end="72104" />
  <Speaker name="Mulher" gender="F" id_conf="0.743800" gender_conf="0.776300" known="F" id="2014" />
  <SpeakerLanguage native="T">PT</SpeakerLanguage>
  <TranscriptWordList>
    <Word start="71657" end="71666" conf="0.981054">o</Word>
    <Word start="71667" cap="SEF" end="71699" relev="true" conf="0.909442">sef</Word>
    <Word start="71700" end="71704" conf="0.952012">a</Word>
    <Word start="71705" end="71759" conf="0.992466">viagem</Word>
    <Word start="71760" end="71817" conf="0.986935">foi</Word>
    <Word start="71821" end="71884" conf="0.982057">visível</Word>
    <Word start="71885" end="71899" conf="0.920987">em</Word>
    <Word start="71900" end="71933" conf="0.872886" punct=",">tory</Word>
    <Word start="71934" end="71950" conf="0.973757">mas</Word>
    <Word start="71951" end="71963" conf="0.984552">para</Word>
    <Word start="71964" cap="Obama" end="71993" relev="true" conf="0.896598">obama</Word>
    <Word start="71994" end="72009" conf="0.913060">isso</Word>
    <Word start="72010" end="72027" conf="0.931396">até</Word>
    <Word start="72028" end="72053" conf="0.991233">poderá</Word>
    <Word start="72054" end="72088" conf="0.985122">levar</Word>
  </TranscriptWordList>
</TranscriptSegment>
 <TranscriptSegment>
  <TranscriptGUID>233</TranscriptGUID>
  <AudioType start="72105" end="72301" conf="0.354400">Clean</AudioType>
  <Time reasons="" start="72105" sns_conf="0.983500" end="72301" />
  <Speaker name="Mulher" gender="F" id_conf="0.787700" gender_conf="0.896500" known="F" id="2014" />
  <SpeakerLanguage native="T">PT</SpeakerLanguage>
  <TranscriptWordList>
    <Word start="72122" end="72143" conf="0.962417">uma</Word>
    <Word start="72144" end="72190" conf="0.978576">queda</Word>
    <Word start="72196" end="72215" conf="0.956369">nas</Word>
    <Word start="72216" end="72282" conf="0.997975" punct=".">sondagens</Word>
  </TranscriptWordList>
</TranscriptSegment>
```

Figure 9.2: XML Information for Question 81 - 2010 Version

As far as the use of punctuation is concerned we can present a case in which our experiments in replacing all full stops for commas (derived by a concern that a misplaced full stop would produce worse results than a misplaced comma, since the full stop is a terminator and the comma is a separator) produced better results than keeping the full stop. That is the case of Question #24, that we present in Table 9.13.

Table 9.13: Transcripts for Question #24

Question		
0024	2617	O que é a Liga Sagres?

Manual Transcript
(...) É um caso que vamos conhecer na segunda parte do Telejornal, onde vamos olhar também para a estreia da Liga Sagres, o campeonato nacional de futebol, que está de regresso à RTP. Até já.

2008 Transcript	**2010 Transcript**
(...) Foi o que fez na segunda parte do Telejornal onde vamos olhar também para à estreia da Liga sáez. O campeonato nacional de futebol que está de regresso à RTP.	(...) Fomos bater na segunda parte do Telejornal onde vamos olhar também para à estreia da Liga Sagres. O campeonato nacional de futebol que está de regresso à RTP.

In this case the right passage could not be identified in the 2008 version because the name of the entity we are looking for is incorrectly recognized in the passage that contained the definition (the answer to the question). However that was not the case for the 2010 version in which "Liga Sagres" is correctly identified in the passage that contains the sought after answer. Since the answer is a definition that appears enclosed by commas, and in the transcript the punctuation is incorrect, with the full stop separating the name from the definition, we are only able to select the correct passage in the test that replaces full stops by commas. The correct answer was extracted with the aid of Wikipedia because it has the required answer identified as an entity, "Campeonato Nacional de Futebol" the Portuguese Soccer Premiere League. We were therefore able to produce a valid answer for test D 2010.

9.6.2 Wikipedia

The advantage of using the Wikipedia can be seen in Questions #19 and #20 that are cluster questions, whose answers are in the same transcript. The questions and corresponding transcripts are presented in Table 9.14.

Table 9.14: Transcripts for Questions #19 and #20

Questions		
0019	2613	De quem é o projecto do museu Iberê Camargo?
0020	2613	Em que cidade fica?

Manual Transcript
Foi ontem inaugurado em Porto Alegre o museu de Iberê Camargo. O projecto do arquitecto Siza Vieira que já ganhou uma medalha de ouro na Bienal de Veneza, e que no Brasil já é considerado um dos edifícios contemporâneos mais bonitos do País.

2008 Transcript	2010 Transcript
Frente inaugurado em Portalegre o Museu de iberê Carmago um projecto do arquitecto Siza Vieira, já ganho uma medalha de ouro na Bienal de Veneza em que no Brasil, já é considerado um dos edifícios contemporâneos mais bonitos do país.	Foi ante inaugurado em Portalegre o Museu de ereira é Carmago projecto do arquitecto Siza Vieira já ganho uma medalha de ouro na Bienal de Veneza em que no Brasil já é considerado um dos edifícios contemporâneos mais bonitos do país.

None of the tests without the Wikipedia (A and C, 2008 and 2010) managed to produce any answers (NIL answers in all cases) for both questions. However, all situations with Wikipedia managed to give the correct answer to the first question, which is the Architect Siza Vieira, because it is recognized as an entity with the help of Wikipedia.

The transcript form 2010 does not recognize correctly the entity Iberê Camargo, as in the 2008 transcript, therefore the correct passage is only identified for the 2008 tests. In the case of the first question the answer for 2010 comes directly from the Wikipedia.

In the case of the second question, Question #20, the anaphoric reference is correctly solved but the answer given for the 2008 case is Portalegre (in Portugal) instead of Porto Alegre (in Brazil) but since it was what the recognizer produced it is considered correct, as in the QAST evaluation. The correct name of the city Porto Alegre also occurs in a different passage, but in this case it is incorrectly extracted (confusion with Portuguese city Porto). However without the Wikipedia no answers are produced. Because of the above mentioned recognition error for Iberê Camargo, no right answers are produced for the 2010 case.

There are numerous cases of questions that were answered only in the case of using the additional knowledge of Wikipedia, for instance (Question #9 Que prémio ganhou João Ubaldo Ribeiro em 2008?) [Which prize did João Ubaldo Ribeiro win in 2008?] in which Wikipedia help in finding the right answer Prémio Camões [Camões Prize, the higher prize for writers in the

Portuguese language]. Other examples are (Question #62 Em que país fica o Rio Bengo?) [In which country is the Bengo river located?] with the correct answer, Angola, (Question #76 Que movimento iniciou João Gilberto?) [Which movement did João Gilberto start?] with the correct answer, Bossa-Nova, and (Question #82 Que festa se realiza na Quinta da Atalaia?) [Which party takes place in Quinta da Atalaia?] with the correct answer, Festa do Avante.

Besides the examples given so far, the Wikipedia is especially useful for definitional questions, as in (Question #45 Quem é António Guterres?) [Who is António Guterres?], (Question #57 Quem é Carlos do Carmo?) [Who is Carlos do Carmo?] or (Question #71 Quem é Jorge Nuno Pinto da Costa?) [Who is Jorge Nuno Pinto da Costa?] or information closely related as the case of (Question #60 Quando nasceu João Ubaldo Ribeiro?) [When was João Ubaldo Ribeiro born?].

Examples of questions were the inclusion of Wikipedia had a negative effect are (Question #13 Quem é Paulo Rangel?) [Who is Paulo Rangel?], who is a Portuguese Politician whose political career started at about the time of our data, 2008. In this case only the situations in which the transcripts were considered alone, manage to produce correct answers, be it o rosto que vai liderar os deputados do psd [the face that is going to lead the psd parliamentarians] or o novo líder da bancada parlamentar do psd [the new lider of psd in parliament] or even deputado pela primeira vez nesta legislatura [deputy for the first time in this legislative session]. However when Wikipedia is taken into account, no valid answer is obtained because there is no entry for Paulo Rangel. There are several incorrect answers that correspond to names of better known politicians (that can be checked via Wikipedia) that appear in the context, near the name of Paulo Rangel, and also a Brazilian Actor by the name of Pedro Paulo Rangel, who has a Wikipedia entry.

Another example is (Question #15 Onde estiveram reunidos os ministros das finanças do G-8?) [Where did the G-8 Finance Ministers meet?] relating to the 2008 G-8 Finance Ministers Meeting in Osaka, Japan. In this case the Wikipedia did not help, because there were so many occurrences of G-8, that they led the system astray.

9.6.3 Numeric Values

Our QA system achieved very good results with questions of categories count and measure, or more generically with questions whose answer was a numeric value, since we had developed an extensive number of rules to treat this case, that were described in 8.4. However, as we said in the

paragraph "Numbers written out as phrases" of that section, we did not develop the algorithm to transform a number that was written out as a phrase into the corresponding number. The lack of such algorithm had a very big impact in these sort of questions with the speech transcripts since all numbers obtained that way are written out (except for one digit number or round numbers as twenty). Some examples of questions that were not answered correctly for this reason in the extraction are:

- Question #3 Quantos feridos provocou o descarrilamento? [How many hounded were there in the derailment?] (of the Tua train line) in which one of the valid answers, 37, trinta e sete [thirty seven] was given partially as 7, sete [seven];
- Question #6 Quantos anos tem a Linha do Tua? [How many years has the Tua line?] in which the answer, 120, cento e vinte [one hundred and twenty] was partially given as 20, vinte [twenty];
- Question #29 Quantos dólares custa o barril de petróleo em Nova Iorque? [How much, in dollars, is the cost of the oil in New York?] in which possible answers were 123.5, cento e vinte e três dólares e meio[one hundred and twenty three dollars and a half(fifty cents)] was given as 3, três dólares [three dollars] (24th July 2008 data), or (slightly above) 107, ligeiramente acima dos cento e sete dólares [slightly above one hundred and seven dollars], given as 7, sete dólares [seven dollars] (4th September 2008 data);
- Question #30 Quantos dólares custa em Londres? [How much, in dollars, does it cost in London?], in a cluster with the previous one, was able to solve the reference but the answers suffered from the same problem of the previous question. In this case answers as 144, a rondar os cento e quarenta e quatro dólares[around one hundred and forty four dollars] were given as 4, quatro dólares[four dollars] (30th June 2008 data) and 97, abaixo dos noventa e sete dólares [under ninety seven dollars] was given as 7, sete dólares[seven dollars](11th September 2008 data);
- Question #53 Quantos milhões de crianças podem ficar órfãs em África, segundo as Nações Unidas? [How many millions of children in Africa can become orphans, according to the United Nations?] the answer mais de cinquenta e três milhões [over fifty three millions]in this case is extracted as três milhões [three million]

9.6.4 Answers made valid by the ASR system

There are several questions that were considered right following the QAST rules, that were in fact "free interpretations of the ASR system". One of such cases was already identified in the context of Questions #19 and #20 presented in Table 9.14, where the correct answer, the city of "Porto Alegre" in Brazil, was converted into the city of "Portalegre" in Portugal. Another case, this one corresponding to a somewhat more inspired performance of the ASR, can be observed in Question #21, Table 9.15

Table 9.15: Transcripts for Question #21

Question		
0021	2614	Onde fica a fundação José Saramago?
Manual Transcript		
A sede da fundação José Saramago vai ser na Casa dos Bicos, em Lisboa.		
2008 Transcript		**2010 Transcript**
Na sede da Fundação José Saramago vai ser na casa dos bicos em Lisboa.		A sede da Fundação José Saramago vai ser na casa dos bicos em Lisboa.

According to the table the answer to the question is Lisboa, or "Casa dos Bicos, em Lisboa", with the first one returned by the system in several occasions. However the system found an alternative passage from a previous broadcasting news show that had an alternative valid answer: "Azinhaga" where a local pole of Fundação José Saramago also exists, since "Azinhaga" is the village where José Saramago was born. The system also returned this answer once. But it also returns the answer in a slightly different version "Zequinha Agha" which is phonetically similar and is the output of the recognizer in a different audio environment. The transcripts of the alternative passages for Questions #21 are presented in Table 9.16. We have so far treated cases of Portuguese names, but we come across the same kind of phenomenon, with more frequency, when dealing with foreign language names, as will be shortly discussed.

Table 9.16: Alternative Passages for Question #21

Question		
0021	2614	Onde fica a fundação José Saramago?
Alternative Answer		
Azinhaga		
Transcript Document: 2008_06_01-21_59_02-Jornal2-2.avi		
Time	**Environment**	**Speaker**
24:06 to 24:16	A	Cecília Carmo
Manual Transcript		
José Saramago e a mulher Pilar del Rio estiveram ontem na aldeia natal do escritor, a **Azinhaga** no Alentejo. O prémio Nobel da Literatura inaugurou a sede local da fundação que tem o seu nome.		
2008 Transcript		**2010 Transcript**
José Saramago e a mulher Pilar del Rio estiveram ontem na aldeia natal do escritor **Azinhaga** no Alentejo os prémio Nobel da Literatura inaugurou a sede local da fundação que tem o seu nome.		José Saramago e a mulher Pilar del Rio estiveram ontem na aldeia natal do escritor **Azinhaga** no Alentejo. Prémio Nobel da Literatura inaugurou a sede local da fundação que tem o seu nove.
Time	**Environment**	**Speaker**
24:28 to 24:38	P+M	Alberto Serra
Manual Transcript		
Há festa na aldeia. O filho ilustre já tinha há muito uma rua. Agora que é a mulher. Pilar del Rio, que dá nome à antiga rua da estação de **Azinhaga**.		
2008 Transcript		**2010 Transcript**
Há festa na Aldeia. Filho ilustre já tinha há muito uma rua. Agora que é mulher. Pilar del Rio o que dá nome. A antiga rua da estação de **azinho água**.		Há festa na aldeia. O filho ilustre já tinha há muito uma rua. Agora porque é mulher Pilar del Rio que dá o nome. À antiga rua da estação de **ao senhor guerra**.
Time	**Environment**	**Speaker**
24:57 to 25:05	P+N	Alberto Serra
Manual Transcript		
O prémio Nobel veio também a **Azinhaga** para inaugurar a sede local da fundação José Saramago, que tem a biblioteca, computadores (. . .).		
2008 Transcript		**2010 Transcript**
O prémio Nobel veio também **à vizinha água** para inaugurar a sede local da fundação José Saramago, que tem Muteka, condutores (. . .)		O prémio Nobel veio também **Zequinha Agha** para inaugurar a sede local da fundação José Saramago que tem a bedeteca computadores.

9.6.5 Transcription of Foreign Language Names

Since the ASR system is prepared to recognize the Portuguese language, we have a problem when other languages come into play. We are not considering the case when a piece of news have reporters or interviews speaking a different language, but we come across the same kind of problem when we have entities with foreign names. When that occurs the ASR have problems, and so does the QA system. We will give two examples with entities that have foreign names (in French). In question #23 we want some information concerning the "Cirque du Soleil", but the system was unable to find the information because the transcribed name had quite a different spelling from the original French. The results were similar for 2008 and 2010 ASR versions, namely NIL answers for the tests without Wikipedia, because the entity did not have any matches, and for the tests with Wikipedia the entity was found but the system could not find the expected answer. The transcripts for this question are shown in Table 9.17.

Table 9.17: Transcripts for Question #23

Question		
0023	2616	Onde nasceu o Cirque du Soleil?

Manual Transcript
Ao quartel-general em Montreal, no Canadá, lugar onde nasceu o **Cirque du Soleil** há quase 30 anos, chegam pessoas muito diferentes, de diferentes lugares, para serem desafiadas.

2008 Transcript	**2010 Transcript**
Ao quartel-general em Montreal no Canadá do Carmo nasceu **si roçou lei** é quase trinta anos. Segundo pessoas muito diferentes e diferentes lugares, para serem sorteadas.	Ao quartel-general em Montreal no Canadá o caro, mas eu **sirvo pessoa** é quase trinta anos. Sigam pessoas muito diferentes diferentes lugares para serem sorteadas.

In question #11, whose transcripts are in Table 9.18, the answer we are looking for is the city of Neuchâtel in Switzerland. In this case the foreign named entity appears in the answer, not in the question as in the previous example, and although the name is not correctly transcribed, at least the system is able to identify correctly the passage where the answer occurs. This happens only in the tests where Wikipedia is used, because it helps identifying other possible locations, although not the name of the city. The tests that do not use the Wikipedia return NIL answers.

The previous examples were given for the French language, but we have an extensive number of English names occurring in the data collection, as well as names in other languages. We chose

Table 9.18: Transcripts for Question #11

Question		
0011	2606	Qual a primeira cidade onde vai permanecer a selecção nacional durante o europeu de futebol?

Manual Transcript
A selecção nacional de futebol já está na Suíça. A equipa partiu a meio da tarde do aeroporto de Lisboa e aterrou há cerca de duas horas. Depois seguiu viagem até **Neuchâtel**, a primeira cidade onde vai permanecer durante o Europeu de Futebol.

2008 Transcript	2010 Transcript
A Selecção Nacional de futebol já está na Suíça. A equipa partiu a meio da tarde do aeroporto de Lisboa e aterrou há cerca de duas horas depois e a viagem até **nos à Tel**. A primeira cidade onde vai permanecer durante o Europeu de futebol.	A selecção nacional de futebol já está na Suíça a equipa partiu a meio da tarde do aeroporto de Lisboa e aterrou há cerca de duas horas depois seguiu viagem até **nos à Tel** a primeira cidade onde vai permanecer durante o Europeu de futebol.

these examples because they were rare occurrences in the collection. With names that occurs frequently, the system is able to identify them correctly, as the case of Barack Obama in Question #81 (Table 9.11). However even names that were recurrent in our data collection, as Michael Phelps (the data collection included the 2008 Beijing Olympic Games) presented some challenges for the ASR system. There were 6 questions in the test collection related to Michael Phelps, based on information that came from 4 transcripts.

Table 9.19 summarizes this information, with the results obtained.

There are three questions in which the name of Michael Phelps appears in the question: the full name in Question #7, as the reference ele [he] in the cluster connected Question #8, and as Phelps in question #38. The other three questions (#18, #37, and #59) require Michael Phelps as the answer.

The manual transcript that lead to the first three questions, T#6, was read by the anchor, but the sound was over a jingle, to announce the piece, that was going to be further developed at a later point in the broadcast. Unfortunately, as shown in Table 9.20, both in the 2008 and 2010 transcripts the audio was cleaned, as it was considered a filler. However due to redundancy of information in the corpus, the system managed to produce some correct answers, based on different passages.

In the case of Question #7, Quantas medalhas de ouro ganhou Michael Phelps em Pequim?

Table 9.19: Information on Transcripts and Questions related to Michael Phelps

T.#	**Time Interval**	**Document**
6	28:23 to 28:31	2008_08_22-19_59_02-Telejornal-1.avi
Q.#	**C.#**	**Question**
7	2603	Quantas medalhas de ouro ganhou Michael Phelps em Pequim?
8	2603	Para que país foi ele passar férias?
18	2612	Quem é a maior estrela dos Jogos Olímpicos de Pequim?
T.#	**Time Interval**	**Document**
28	33:44 to 33:59	2008_08_16-19_59_02-Telejornal-1.avi
Q.#	**C.#**	**Question**
37	2627	Quem é o nadador mais bem pago da história?
T.#	**Time Interval**	**Document**
29	30:21 to 30:29	2008_08_16-21_59_02-Jornal2-2.avi
Q.#	**C.#**	**Question**
38	2628	Com que idade se tornou Phelps milionário?
T.#	**Time Interval**	**Document**
44	30:41 to 30:51	2008_08_16-19_59_02-Telejornal-1.avi
Q.#	**C.#**	**Question**
59	2645	Quem igualou o recorde de medalhas de Mark Spitz?

[How many gold medals did Michael Phelps win in Beijing?] the system always answers correctly. This is due to the fact that, since Michael Phelps won the large quantity of eight gold medals, there are several correct answers in the corpus, for he was consistently wining gold medals across several days. However amongst the correct occurrences of Michael Phelps, we can find other transcripts for his name as "Michael of Health" in the supporting passage for the 2nd and 3rd answer in test A 2010, and "Fel" in the supporting passage for the 1st answer of test B 2010. For the related cluster question, Question #8, Para que país foi ele passar férias? [For which country did he go on holidays?], the system was not able to give the answer in any of the tests, not for problems of anaphoric resolution nature (since Michael Phelps appears in the supporting passages) but because the answer was a difficult one to extract, since it required a country "Portugal" that appeared in the T#6 (that was not transcribed at all) and in the corpus there were several occurrences of "Algarve" but few of Portugal, connected to Michael Phelps holidays.

Question #18, Quem é a maior estrela dos Jogos Olímpicos de Pequim? [Who is the greatest star of the Beijing Olympic Games?], seems to be more affected by the missing transcript of the supporting passage that originated the question, as no test was able to produce a valid result. However tests D of 2008 and 2010 and test B2010 provide a tentative "Field" as answer (from a

Table 9.20: Transcript #6 (Questions #7, #8 and #18)

Questions		
0007	2603	Quantas medalhas de ouro ganhou Michael Phelps em Pequim?
0008	2603	Para que país foi ele passar férias?
0018	2612	Quem é a maior estrela dos Jogos Olímpicos de Pequim?

Manual Transcript
A maior estrela dos jogos olimpicos de Pequim está de férias no Algarve. Michael Phelps depois das 8 medalhas o descanso em Portugal.

2008 Transcript	**2010 Transcript**
[clean] (audio cleaned because it is understood as a filler.)	[clean] (audio cleaned because it is understood as a filler.)

transcript of "Michael Field" instead of "Michael Phelps") but the supporting passage classifies him a estrela do desporto mundial [international sports star] and mentions the Beijing Olympic Games, but does not explicitly says he is the "greatest star of the Beijing Olympic Games". It is quite a close attempt, that comes from a piece of news from the 16th of August 2008, that we will look at in detail.

The other three transcripts that originated questions about Michael Phelps (T#28, T#29, and T#44) come from the above mentioned piece from the 16th of August 2008 that was integrally broadcast in both Telejornal 1 and Jornal 2, with the only difference being the anchor. The piece has around 6 minutes of duration and is composed of 6 segments, with the characteristics presented in Table 9.21.

Table 9.21: Piece of August 16th 2008 on Michael Phelps

Telejornal 1			**Jornal 2**		
Duration	**Time**	**Environment**	**Duration**	**Time**	**Environment**
Speaker			**Speaker**		
S1 - Segment 1 Introductory piece on the victory of Michael Phelps in the 100 meters butterfly in the Beijing 2008 Olympic Games, which provided him with his seventh gold medal.					
10 sec	30:41 to 30:51	A	11 sec	25:41 to 25:52	A
Judite de Sousa			Alberta Matos Fernandes		
S2 - Segment 2 Report of the 100 meters butterfly swimming event in the Beijing 2008 Olympic Games emphasising the victory of Michael Phelps in the last stroke to the Serbian swimmer Milorad Čavić.					
58 sec	30:52 to 31:50	P+N	58 sec	25:53 to 26:51	P+N
Rui Martins					
S3 - Segment 3 News piece centred on the number of gold medals that Michael Phelps won (seven, so far) in the Beijing Olympic Games, and introducing a report on Mark Spitz who won the same number of gold medals medals in the 1972 Munich Olympic Games.					
19 sec	31:51 to 32:10	A	20 sec	26:52 to 27:12	A
Judite de Sousa			Alberta Matos Fernandes		
S4 - Segment 4 Report on the Mark Spitz performance in the 1972 Munich Olympic Games, with his seven gold medals, with a comparison to Michael Phelps, who is in the position of breaking the record of seven gold medals won in the swimming competition of a single edition of the Olympic Games.					
1 min 31 sec	32:11 to 33:42	P+N	1 min 31 sec	27:13 to 28:44	P+N
António Nabo					
S5 - Segment 5 News introduction to a report on the financial aspects to the victories and popularity of Michael Phelps.					
15 sec	33:44 to 33:59	A	14 sec	28:45 to 28:59	A
Judite de Sousa			Alberta Matos Fernandes		
S6 - Segment 6 Report on the financial aspects to the victories and popularity of Michael Phelps, focusing his earnings in advertisements, his high revenue since an early age that, enables him to be part of the Forbes list of richest men, and the comparison with other sports stars.					
2 min 40 sec	34:00 to 36:40	P+M+N	2 min 40 sec	29:00 to 31:40	P+M+N
Laura Santos					

Table 9.22, lists all the occurrences of Michael Phelps in the piece and the corresponding transcripts, emphasising in green the transcripts that led to the remaining three questions. There was a total of 20 occurrences in the piece, both in the form of "Michael Phelps" and "Phelps", in the two news programs. Of those 80 instances, 6 were correctly transcribed, and there was a multitude of different variations, that when returned as answers were considered correct.

Table 9.22: Occurrences of Michael Phelps in the Piece of August 16th 2008

		Telejornal 1		**Jornal 2**	
#	**Occurrence**	**T2008**	**T2010**	**T2008**	**T2010**
S1 - Segment 1					
1	Michael Phelps	T#44 Michael Fel	Michael Fel	✓	✓
S2 - Segment 2					
5	Phelps	fel	fel	fel	fel
	Phelps	Céu Lopes	FEUP	Céu Lopes	FEUP
	Phelps	✓	✓	✓	help
	Phelps	Cassell	Celso	Cassell	Celso
	Phelps	Fell	rap se	Fell	rap se
S3 - Segment 3					
1	Michael Phelps	Michael Field	Michael Field	Mike Oldfield	Michael Fel
S4 - Segment 4					
4	Michael Phelps	muito forte	Michael Field	muito forte	Michael Field
	Phelps	[noise]	[noise]	[noise]	[noise]
	Michael Phelps	marcam few que	Michael felt	marcam few que	Michael felt
	Michael Phelps	[noise]	[noise]	[noise]	[noise]
S5 - Segment 5					
1	Michael Phelps	T#28 [noise]	[noise]	Michaelsen	Michael Sal
S6 - Segment 6					
8	Michael Phelps	Michael Field	Michael Field	Michael Field	Michael Field
	Michael Phelps	English mais o véu	mais o véu	Maio o	Mais . . .
	Phelps	fuel	Sel	fuel	Fel
	Phelps	Celso (Java)	Fiel	Telles	Fiel
	Phelps	Céu	Help	T#29 ✓	Help
	Phelps	L	pelo	pelo	pelo
	Phelps	el	vel	Céu	vel
	Phelps	fiel	pelo	Eiffel	apelo

The blue shaded area of Table 9.22, corresponds to an advertisement featuring Michael Phelps (Figure 9.3). The advertisement is in English, and it includes a music with lyrics also in English, therefore the results of the transcript of that part have little meaning, as shown in

Table 9.23, but they are part of the text corpus anyway.

Figure 9.3: Michael Phelps Advertisement

The same piece of news also provide other foreign names that are interesting to look at. For instance the name of Mark Spitz occurs 12 times, as presented in Table 9.24. It has a higher rate of correct transcriptions, 19 in 48, possibly because of it distinctive phonetic characteristics. The correct transcriptions are predominantly given by the 2010 version of the recognizer (13 out of 19).

Table 9.23: Transcripts for Michael Phelps English Advertisement

Manual Transcript
Michael Phelps isn't part dolphin, or fish or amphibian. He doesn't have gills or flippers ...

Translation to Portuguese
Michael Phelps não é meio-golfinho, nem peixe nem anfíbio. Não tem guelras ou barbatanas ...

2008_08_16-19_59_02-Telejornal-1.avi	
2008 Transcript	**2010 Transcript**
Mais o véu de dizendo oito del fandango Pablo fez também. Não ter a minha. Vida não me deu muitos confraternizam também por cento.	Um mais o véu de dizendo oito tal Fandango tão ou fez bem ou antimíssil. O líder do ramo deu muitos. Confraternização também doce cento.

2008_08_16-21_59_02-Jornal2-2.avi	
2008 Transcript	**2010 Transcript**
Em Maio o nível de pobreza entre bóias del fandango Pable fez também não abona muito a minha. Vida não me deu muitos confraternizam também torcem.	Tanto mais louvável de dizendo oito tão houver um cunho tão ou fez também não é muito tímida. O líder do ramo deu muitos. Confraternização também Ocidente.

Table 9.24: Occurrences of Mark Spitz in the Piece of August 16th 2008

		Telejornal 1		**Jornal 2**	
#	**Occurrence**	**T2008**	**T2010**	**T2008**	**T2010**
S1 - Segment 1					
1	Mark Spitz	✓	✓	✓	Mac Spitz
S2 - Segment 2					
1	Mark Spitz	Make it	✓	Make it	✓
S3 - Segment 3					
1	Mark Spitz	✓	✓	✓	✓
S4 - Segment 4					
8	Mark Spitz	English [noise]	[noise]	[noise]	[noise]
	Mark Spitz	Mar, e	Mark e	marque e se	✓
	Mark Spitz	Mare e que	✓	Marte e que	✓
	Mark Spitz	Marte e	✓	mar de Pete	✓
	Mark Spitz	Marco	Marte	marketing	Marte e
	Spitz	-	e	e	e
	Mark Spitz	marketing	Marques PIDE	marketing	✓
	Mark Spitz	[noise]	[noise]	[noise]	[noise]
S6 - Segment 6					
1	Mark Spitz	✓	✓	✓	✓

The Serbian swimmer Milorad Čavić is also present 5 times in the piece, in segment S4, and as shown in Table 9.25 it has a lower rate of correct transcriptions, 2 in 20, than Mark Spitz, but a higher rate than Michael Phelps. In this case the correct transcriptions of his name occur in the 2008 version of the recognizer.

Table 9.25: Occurrences of Milorad Čavić in the Piece of August 16th 2008

		Telejornal 1		**Jornal 2**	
#	**Occurrence**	**T2008**	**T2010**	**T2008**	**T2010**
S2 - Segment 2					
5	Milorad Čavić	✓	Milan dá cá vistas	✓	Milorad cá vistos
	Čavić	que ávidos	cá vidros	que ávidos	ca havidos
	Čavić	que ávido	ca Vito	que ávido	ca Vito
	Čavić	que há vidros	cave e das	que há vidros	cave e dos
	Čavić	que ávidos	cá Vito	que ávidos	cá Vito

Table 9.26 presents occurrences of other well known sports stars, all in single occurrences in the piece, and probably because they are so well known they have a high degree of correct transcriptions by the system, except for the case of Tiger Woods. A small notice to the name of Fernando Alonso, who is transcribed as Alonzo in the 2010 version of the recognizer, also a fairly common name, especially in English texts.

Table 9.26: Occurrences of other names in the Piece of August 16th 2008

		Telejornal 1		**Jornal 2**	
#	**Occurrence**	**T2008**	**T2010**	**T2008**	**T2010**
S6 - Segment 6					
1	Fernando Alonso	✓	Fernando Alonzo	✓	Fernando Alonzo
1	Ronaldinho Gaúcho	✓	✓	✓	✓
1	David Beckham	✓	✓	✓	✓
1	Tiger Woods	Aveiro	Tiger Bhutto	Aveiro	Tiger Bhutto

Coming back to the question analysis, we present in Table 9.27 the transcripts for Question#37 Quem é o nadador mais bem pago da história? [Who is the best paid swimmer of history?]. It corresponds to the manual transcript T#28, that coincides with the segment S5, Telejornal 1 part, of the the 16th of August 2008 piece. This is another unfortunate example as far as the automatic transcripts are concerned, because since the name of Michael Phelps occurs in the beginning of the sentence, after another news report, it is missed by the recognizer. There is however in segment S6 a phrase that closely matches the one in the transcript T#28. Michael Phelps is identified in this phrase as "Phelps" and it corresponds to the 4th occurrence in segment S6. It is natural then that the correct answers of tests B2008 and D2008 is "Celso" or "Celso Java". "Celso" is the transcription for "Phelps" and the full name "Celso Java" comes from a connection with the next syllables (já ba) of the phrase that states: "... mesmo que não consiga chegar às oito medalhas de ouro, Phelps já bateu um outro recorde: já é o nadador mais bem pago da história." ["... even if he does not manage to reach the eight gold medals, Phelps has already beaten another record: he is already the best paid swimmer of history."]. The 2010 tests do not produce meaningful answers since the transcripts for Phelps in that region are: "Fiel" [faithful] and "Help", words less liable to be considered person names by the system.

Table 9.27: Transcripts for Question #37

Question		
0037	2627	Quem é o nadador mais bem pago da história?

Manual Transcript
Michael Phelps soma já 13 medalhas de ouro conquistadas em Jogos Olímpicos, é também o nadador mais bem pago da história, mas afinal quanto vale em euros a estrela dos Jogos Olímpicos de Pequim?

2008 Transcript	**2010 Transcript**
[noise] Soma já três medalhas de ouro conquistadas em Jogos Olímpicos, é também o nadador mais bem pago da história. Mas afinal quanto vale em euros. A estrela dos Jogos Olímpicos de Pequim.	[noise] Só mas já três medalhas de ouro conquistadas em Jogos Olímpicos é também o nadador mais bem pago da história, mas afinal quanto vale em euros a estrela dos Jogos Olímpicos de Pequim

As far as Question#38 is concerned, Com que idade se tornou Phelps milionário? [At what age did Phelps became a millionaire?], it is based on transcript T#29, which is part of segment S6, from the Jornal 2 broadcast, as appears highlighted in green in Table 9.22. It can be seen that in the 2008 version of the recognizer, Phelps is correctly transcribed, which makes the system

produce the correct answer to this question in all the four tests. The full texts for T#29 are given in Table 9.28. In the 2010 tests that is no longer the case, with Phelps being transcribed as "Help". The correct passage is never found, and a meaningless answer, "90 anos" [ninety years], is given based on a Wikipedia article that has the name Phelps, but regarding the inventor of the bra (Mary Phelps Jacob), not the swimmer[3].

Table 9.28: Transcripts for Question #38

Question		
0038	2628	Com que idade se tornou Phelps milionário?

Manual Transcript
Phelps tornou-se milionário e atleta profissional da natação poucos meses antes de fazer 16 anos.

2008 Transcript	**2010 Transcript**
Phelps tornou-se milionário de atleta profissional de natação, poucos meses antes de fazer dezasseis anos.	Help tornou-se milionário de atleta profissional de natação poucos meses antes de fazer dezasseis anos.

As for the last question regarding Michael Phelps, Question#59 Quem igualou o recorde de medalhas de Mark Spitz? [Who has equalled the medal record of Mark Spitz?], it is based on transcript T#44, corresponding to segment S1 of Telejornal 1. The full transcripts for this question are displayed in Table 9.29. The system is able to provide a valid answer, or a passage containing an answer in all cases, however the answer is never "Michael Phelps" but some recognized variation of his name. In this case it is more important for the system to be successful that the name of Mark Spitz is correctly identified, as it is the lead entity in the question. The tests without the Wikipedia all produce the same answer, "fel", based on the same passage. The passage is the ending part of the transcript T#44, segment S1 of Telejornal 1 (in which the name of Mark Spitz is correctly recognized), and the beginning of segment S2, in which the first occurrence of Phelps is recognized as "fel", hence the answer.

In the case of the tests with the Wikipedia, the answers are based on a different passage, this one belonging to segment S6, in which the name of Mark Spitz is also correctly recognized in the four versions (Telejornal 1 and Jornal 2 of 2008 and 2010), with all of them appearing as supporting passage in answers of these tests. The same passage of the tests A and D is

[3]Our version of the Wikipedia is from 2006, and consulting the page for Michael Phelps in the Portuguese version of the Wikipedia, one can see that the page was created in March 2007.

Table 9.29: Transcripts for Question #59

Question		
0059	2645	Quem igualou o recorde de medalhas de Mark Spitz?

Manual Transcript
Michael Phelps voltou a deslumbrar. O norte-americano venceu os 100 metros mariposa e igualou Mark Spitz ao conquistar a sétima medalha de ouro nos Jogos Olímpicos.

2008 Transcript	2010 Transcript
Michael fel que voltou a deslumbrar o norte-americano venceu os cem metros mariposa, igualou Mark spitz, ao conquistar a sétima medalha de ouro, nos Jogos Olímpicos.	Michael fel voltou a deslumbrar o norte-americano venceu os cem metros mariposa. Igualou Mark Spitz ao conquistar a sétima medalha de ouro nos Jogos Olímpicos.

also returned as supporting passage. However in tests B and D the system presents extraction problems. The most common answer in these tests is extracted from the excerpt of segment S6 presented in Table 9.30. It accounts for answers as "norte - americano jay" and "james alô".

Table 9.30: Transcripts for Excerpt of Segment S6 for Question#59, Tests B and D

Manual Transcript
Em Pequim o norte-americano já igualou o feito do lendário Mark Spitz (. . .)

2008_08_16-19_59_02-Telejornal-1.avi	
2008 Transcript	**2010 Transcript**
Em Pequim o norte-americano Jay golo feito do lendário Mark spitz (. . .)	Em Pequim o norte - americano James alô feito do lendário Mark Spitz (. . .)

2008_08_16-21_59_02-Jornal2-2.avi	
2008 Transcript	**2010 Transcript**
Em Pequim o norte-americano Jay golo feito do lendário Mark spitz (. . .)	Em Pequim o norte - americano Jay o aluno feito do lendário Mark Spitz (. . .)

Just to finish we can give the example of a very different language, Russian, where the name of the city Tiblissi became, in an transcription for Question #45 (Table 9.31) the Portuguese word utilíssimo [extremely useful] that is phonetically close, who was selected as answer by the system, for the 2010 tests without the Wikipedia.

9.6.6 Transcription of Brazilian Portuguese

We included in the question set a question from a transcript, T#52, from a Brazilian native speaker, namely the former President of Brazil, Lula da Silva. The question was quite a tough one

Table 9.31: Alternative Transcripts for Question #45

Question		
0045	2634	Quem é António Guterres?
Transcript Document: 2008_08_19-19_59_02-Telejornal-1.avi		
Time	**Environment**	**Speaker**
18:03 to 18:14	P+N	José Rodrigues dos Santos
Manual Transcript		
Se a retirada se confirmar fica mais fácil a tarefa do alto comissário da ONU para os refugiados. António Guterres veio a **Tiblissi**, e segue agora para Moscovo para negociar os corredores humanitários.		
2008 Transcript		**2010 Transcript**
Se a retirada se confirmar, fica mais fácil a tarefa do alto comissário da ONU para os Refugiados. António Guterres veio **Tiblissi**, e segue agora para Moscovo para negociar. Os corredores humanitários.		Se a retirada se confirmar fica mais fácil a tarefa do alto comissário da ONU para os Refugiados. António Guterres foi **utilíssimo** e segue agora para Moscovo para negociar os corredores humanitários.

to answer, since the country name Brazil appeared before, and separated from the central core of the transcript. That fact, allied with the poor quality of the transcripts, that are presented in Table 9.32, produced NIL answers in all test cases.

Table 9.32: Transcripts for Question #69

Question		
0069	2654	Quantos aviões da EMBRAER está comprando o governo do Brasil?
Manual Transcript		
Os aviões da EMBRAER são de tamanha qualidade que até o governo está comprando dois aviões.		
2008 Transcript		**2010 Transcript**
O de Agosto em breve são de extrema qualidade e, que até, o governo para compra de dois aviões.		O deixou de em breve se onde tem maior qualidade. Quetta. O Governo para compra de dois aviões.

9.6.7 When a wrong transcript can help the QA system

We present in Table 9.33 the transcripts for Questions 43 and 44, a cluster of questions related to the transcript T#32. The first question of the cluster is additionally a list question, and quite e difficult one because the form of the question does not necessarily mean that the answer is

a list of two people. The system is not able to find the correct answer, but in the tests with the Wikipedia, a partially correct answer is provided (just the first person) with the passage containing the answer being correctly identified. The Wikipedia helps in identifying "Amália", the famous fado singer, as an entity, so it is not removed form the keywords to search as a common word.

As far as the second question in the cluster is concerned, the 2008 version of the system is able to produce the correct answer. It is aided by the wrong translation of "Alain Oulman" in the 2008 transcript, transcribed as "Alan um ano". This word "ano" is accidentally part of the question, which makes the passage to provide a better match for the question.

Table 9.33: Transcripts for Questions #43 and #44

Questions		
0043	2633	Quem deu vida à "Formiga" cantada por Amália?
0044	2633	Em que ano?

Manual Transcript
Em 1969 Fontes Rocha e Fernando Alvim deram vida à formiga cantada por Amália, uma adaptação de **Alain Oulman** do poema de O'Neil velha fábula em Bossa Nova.

2008 Transcript	**2010 Transcript**
(...), e sessenta e nove fontes Rocha e Fernando Alvim, deram vida à Formiga cantada por Amália. Uma adaptação de **Alan um ano** o problema do o Neill, velha fábula em bolsa nove um nove um.	(...) e sessenta e nove fontes Rocha e Fernando Alvim deram vida à Formiga cantada por Amália. Uma adaptação de **Alain num humano** do poema de o Neill, velha fábula bossa-nova num nem um.

9.7 Conclusions

The robustness of IdSay was validated since the results obtained for text originated from ASR were at the level of the best systems in the literature.

To accomplish this goal a resource was created that consists of about three months of automatically transcribed broadcast news from the Portuguese public network, RTP, along with a question set of 100 questions. The questions are based upon manual transcripts of the pieces of the broadcast news, and follow the guidelines defined for QA@CLEF, including the distribution of categories and types.

The results confirm the adequacy of the use of Wikipedia to increase the performance of Question Answering systems over speech transcripts, since a significant improvement in results is achieved in this situation.

Using both full stops and commas leads to better results for the QA application than considering all punctuation marks as commas. The validation of this conclusion, unlike the case of the Wikipedia usage, is obtained with a tight margin, and would not stand should we require a higher degree of certainty.

There is no statistical evidence on improvements through the use of the 2010 ASR version instead of the 2008 ASR version, since both produce similar results. This fact is surprising, since the WER for the Recognizer and SER for the identification of both full stops and commas are all smaller for the 2010 version. To explain these results, we must bear in mind that the data of our experiments is data from 2008, and therefore the ASR 2008 models are better trained for our data. This effect may function in the opposite direction, and cancel the improvements achieved for the 2010 version described in the literature.

Even tough the Wikipedia is used as part of the data collection to be searched in the tests conducted in this chapter, a different situation is possible, where the ontological information derived from the Wikipedia could be used alone, instead of the Wikipedia's full text. It could make sense if the main goal was to search within the video stream. However using the Wikipedia as part of the searchable test did not introduce much noise as a whole, as indicated by the high number of questions with correct passage extracted (Table 9.5). This situation presents, however obvious advantages as far as definitional questions are concerned. This is a question to be addressed in the future, depending on the specific usage.

10 Conclusions and Future Work

10.1 Conclusions

We presented a complete Question Answering (QA) system for Portuguese, developed from scratch, that proved to be competitive with the best QA systems, being classified in third place among the six systems participating at the monolingual Portuguese task of QA@CLEF 2008 with 32.5% of accuracy over the first answer, ahead of several veteran systems. The results obtained place IdSay in 5^{th} place among the total of 21 participating systems for the 11 languages offered at the evaluation.

The improvements made after the evaluation resulted in a considerable improvement of accuracy over the first answer to 50.5% corresponding to second place in terms of ranking. The system is efficient in indexing and answering, taking only 4 hours to index the whole text collection of the QA@CLEF 2008, and about 2 minutes to answer to the corresponding 200 question set.

It was tested in a case study with QA over speech transcripts data. The test data had to be fully organised itself, as the Portuguese language was not among the options of the QAST task of QA@CLEF, resulting in a resource that can be used to test QA systems for Portuguese over speech transcripts.

IdSay proved to be robust and even in the presence of several mistranscribed words and misplaced punctuation marks, the system obtains 30% accuracy over the first answer, 42% accuracy over all three answers, and more than 60% accuracy if we are only interested in passage retrieval. These values are in line with the best reported results at QAST.

Since the main goal of this work is to study and develop new components for IR and QA, to build a QA system complete, efficient and robust, that can compete with the current state-of-the-art QA systems, we consider it accomplished.

Regarding the use of Portuguese as the working language, we confirm that the results ob-

tained in the QA field for this language are at the top level when compared with those obtained for other languages. The results validate the concept of corpus of written text in which both the European and Brazilian variants are present, without the need to discriminate between them. As far as spoken language is concerned, however, the results indicate that special attention should be given to each of these two different variants of Portuguese.

We present the list of the distinguishing features, accomplishments and conclusions that contribute to the success of our approach. The list is organised in order of appearance in the thesis, referencing the corresponding chapter.

1. Method to **create stop lists** (SL3).

   The method is proposed in Chapter 3 and it considers words that have document frequency above a certain value for the stop list. It is dependent on the collection and simple to obtain. It proves to be better than other two stop lists considered, for a cut-off value of 10, with statistical significance.

2. In pre-processing, best results were obtained by **converting letters to lowercase** and by **removing punctuation marks**, leading to an increase in document retrieval efficiency, but there is no statistical evidence that using **stop lists**, **lemmatization** or **stemming** improves document retrieval for QA.

   These results, obtained in Chapter 3, are somewhat surprising, since most of the techniques are assumed to be relevant. This study shows that stop lists, lemmatization and stemming do not always produce better results, and their usefulness for a specific application of a given IR system needs to be validated. The results were obtained using a metric especially thought up for an IR component working as the core piece of a QA system.

3. Data structure for **storing document data**, using one number per word, instead of strings.

   This proposal was one of the most important ones of Chapter 5, and responsible for the overall efficiency of the system, leading to advantages in both space and time consumption, when compared to a string data structure for documents. The use of hash tables for words allows the achievement of the number of a word in constant time, essential to the usability of this data structure. With words mapped to numbers, Passage Extraction, Answer

Extraction and Answer Validation deal with just one number per word, instead of string manipulation.

4. **Calculation of entities of any word size, by frequency** , using a low space complexity.

   This method proposed in Chapter 5 has a low space and time complexity, allowing the calculation of entities of any word size, that would be impossible with counters in a n-dimensional array. It makes use of hash tables on entities, like the ones for words, that allows the entity number to be determined in constant time, essential to keep time complexity at a low level.

5. **Punctuation marks are kept, but separated from words**, improving document retrieval efficacy.

   This technique from Chapter 5 separates punctuation marks from letters and digits, thus allowing a better matching for words occurring close to punctuation marks. In the tests of Chapter 3 leaving punctuation marks as they appear in the text often created words with a single occurrence, which are not detected during retrieval and have a reduced use. It has the additional advantage of keeping the punctuation marks of the original text, that are useful for passage segmentation, preventing sentences to be cut in the middle.

6. Number normalization **grouping digits in words of 3 digits**.

   This form of number normalization described in Chapter 5 makes it possible to keep the number of distinct words at a low value, improving the usability of the word hash table. This method does not imply any loss in generality in the representation of numeric values, but prevents the number of distinct words in the collection to increase or even escalate to unmanageable values due to the occurrence of many different numeric values in a text collection.

7. Document Retrieval that returns documents with all **words in the query**, and also with all **entities in the query**.

   This is a core proposal of Chapter 6, since it improves the efficacy of the IR system. All entities in the question are extracted, and with an inverted index for entities and another for words, the resulting intersection will produce documents more relevant to the question, and not only documents with the query words in the question scattered around.

8. **Retrieval cycle strategy for selectively removing words from the question**, if no satisfactory results have been produced so far.

   This is part of the architecture of IdSay system, and is explained in detail in Chapter 6. This can be seen as an alternative for stop word removal since the most frequent word is chosen to be removed in the next cycle, if results were not satisfactory so far. It has the advantage of being selective in the sense that the words to be removed are not chosen blindly and beforehand at indexing time, but instead they are removed at run time depending on the collection and question being processed. The fact that we keep entities apart from other words, with the latter being preferential candidates for removal, adds another criterium for question based selection of which words to be removed or kept.

9. Fast algorithm for Passage Extraction, that extracts all passages containing all the **words**, and also the **reference entity** of the question.

   This procedure of Chapter 6 has a low time complexity since it starts by locating the positions of the reference entity in the document, and then it makes a single pass through the document storing the last position of each word in the question, selecting passages at the same time. This way passages are defined in run time instead as in pre-processing time.

10. Method to **join answers** with the same best support, that have at most one word in the middle.

    This proposal of Chapter 6 prevents problems in Answer Extraction, that could lead to answers with the parts of an entity separated.

11. Strategy for preventing **answers that have words in common** to be returned.

    This strategy is proposed in Chapter 6 and allows an increase in the variety of the answers provided. Otherwise answers that were very close to each other could be generated, leaving out important different answers.

12. **Result Quadrant** format for analysing systems results in perspective with other systems' results.

    This format for representing a system's results taking into account the results of other systems is described Chapter 7 and allows the identification, for each system, of the degree

of innovation, coverage of easy questions, and to what extent it is liable to learn from the other systems. This information helps to guide efforts when improving a system.

13. Entity synonyms base built from the **Wikipedia redirection pages** (WES base).

    This proposal is from Chapter 8 and makes use of Wikipedia redirection pages to extract equivalences between entities. This is an automatically built resource based upon a source that is constantly being updated and edited, Wikipedia. This ensures that the information is extended, enhancing the quality through time. This synonym base is useful for the Document Retrieval, Passage Retrieval and Answer extraction modules.

14. Document Retrieval that returns documents with all **words or corresponding synonyms in the query**, and also with all **entities or corresponding synonyms in the query**.

    The use of synonyms was introduced to the Document Retrieval module in Chapter 8, allowing the documents retrieved to have at least one word or a synonym for any of the words in the question, and one entity or a synonym for the entities in the query. This increased the number of documents returned, keeping its high relevance to the question.

15. Passage Extraction, with one passage through the document, extract all passages containing all the **words or corresponding synonyms in the query**, and also the **reference entity or one of its synonyms**.

    This component was also adapted to the use of synonyms in Chapter 8, allowing this way the extraction of a greater variety of passages, but all related with all words and reference entities, by the word/reference entities itself or by a synonym.

16. Ontological knowledge based on the **category string of Wikipedia pages**.

    This is another use of Wikipedia pages proposed in Chapter 8, that allows us to check if an entity belongs to a category using Wikipedia category structure, via the category string. This is a way of having ontological knowledge, that is dependent on the categories in Wikipedia. This information still presents some shortcomings, but it is nonetheless helpful, with its quality tending to increase over time.

17. Using a **special word for numerical separator** when performing number normalization.

This is an important improvement done in number normalization, fully described in Chapter 8. It allows fractional numbers to be correctly stored and represented, with the further advantage of avoiding interference with passage segmentation due the use of common punctuation marks for both distinct tasks. We use punctuation marks with their usual functions in text, and a special numerical separator to replace them if they are used within numeric values.

This, together with the other improvements in number normalization described in Chapter 8, allowed the correct extraction of most numbers occurring in written text.

18. **Abbreviations and acronyms normalization**, occupying only one word and not using full stops.

    This normalization procedure is described in Chapter 8, allowing different forms of an abbreviation to match, and removing full stops that, if used for this purpose, could lead to a passage being segmented in the middle of an abbreviation.

19. Method for lemma selection giving more importance to the **name category**, followed by the **verb category**, with ties solved by means of the **edit distance**.

    This proposal of Chapter 8 allows the lemmatization to be more effective than in the baseline version of the system, with a rule-based determination of the root form selection in case of ambiguity, reducing the problem of attributing words to a different normalization class than the one they belong to.

20. Validation of the **robustness** of the system in a corpus of **speech transcripts**

    A case study was conducted to validate the robustness of the system that required the construction of a corpus of automatically transcribed text in Portuguese, with the corresponding questions obtained from manual transcripts. The case study is described in Chapter 9, with results of accuracy over the first answer similar to the best systems of QAST, thus validating our goal.

21. **Wikipedia** proved to be relevant for answer accuracy in **QA over speech transcripts**

    Even when looking for answers that occur within a speech transcripts corpus, the importance of using the Wikipedia is shown in Chapter 9, with results doubling in accuracy over the first answer in most tests.

10.2 Future Work

This section concludes the thesis by discussing two directions of future of work: firstly, we outline additional possibilities to enhance our current system and secondly we identify further areas of application of the developed work.

10.2.1 Directions for Improvement of IdSay

We would like to enhance our current system along the following guidelines:

- Improve **Answer Extraction and Answer Validation** components

  Extracting the success and failures in these components, can help to develop new strategy to improve its performance.

- Improve **Question Analysis** component

  Analysing the results of the end of this module, could be a relative easy task, and since we have a framework that is easy to change and build new rules, containing authority lists and match with regular expressions, this module can be improved with some effort.

- Usage of **meta-data and context information**

  The system do not use any meta-data or context information, but it could be studied the best way to make the use of that information, when it is available.

- Treatment of **time**

  We need to implement a more thorough treatment of time, with the normalization of time values and the resolution of relative time values.

- Improve the **web application** developed with the **most updated version of IdSay and Wikipedia** pages

  This would not only make easy to look deep in the results of IdSay, and this way better identify opportunities for improvement, but also, since is open to questions from the users, to collect pairs of questions/answers and study new types of questions.

10.2.2 Further Areas of Application

We believe in the utility of both the aim of QA systems and our approaches in the context of growing volume of unstructured text information namely in the web. This text information comes in different languages, so to address the problem of aiding the users in accessing that information it makes sense to develop multi-language systems.

A lot of attention is given to this kind of system, both from industry and the academic community, but methodologies are at a considerable early stage of development. Although the CLEF initiative centres its attention in cross language and multi-language aspects, it also provides monolingual task which are far more participated and with better results than the cross language ones. Hence our option to first develop a monolingual system, for the language we understand better, our native language Portuguese, with sound principles and techniques that are developed in a way that facilitate the introduction of another language in the future.

Now that we have achieved that goal, to extend our system for other languages is the next step in our research horizon, and the next language would be English. Since we have developed components with multi language in mind, we foresee this process to be a smooth one. However, modifications are needed to extend the system to English, and we identify the main ones as the adaptation of the question analysis module and answer extraction patterns, as well as the authority lists that would have to be checked. Also resources would have to be different, for instance a lexical resource for lemmatization and a thesaurus for English would be required. As far as entities are concerned, our frequency based method for extraction of entities from the corpus is language independent, and our methods for the identification of entities and their synonyms from Wikipedia would just require the use of the English Wikipedia, which makes them straightforward steps. The process of adapting the system to English is also adequate to introduce machine learning methodologies, that can be reused in the introduction of further languages.

As for cross language operation, our intention is to explore the developed mechanism of equivalences, applied to cross-language resources. In other words, we intend to start by investigating how far can we go working on the resources side, keeping unchanged the processes we currently have.

Our system was designed to work with questions that constitute the current state of research

in the area. These questions are relatively short in length and usually require short answers. Most questions in the questions sets presented at the appendixes deal with just one reference entity. Other types of more complex questions have also been subject of attention by the research community. Such is the case of the complex interactive Question Answer track, ciQA (Kelly & Lin 2007), offered in TREC in 2006 and 2007. In this track more complex questions are considered, but the entities involved are identified, and a narrative accompanies the topic/question to help guide the type of answers the user/analyst is interested in. An example taken from the 2006 evaluation is:

topic *What [financial relationships] exist between [Greece] and [Cyprus]?*

narrative *The analyst would like to know what financial relationships exist between Greece and Cyprus. This is intended to include trade between the two countries as well as direct financial grants.*

The interactive side to the task is described by the organizers [1] in the following way:

"On "interactive": Participants will have the opportunity to deploy a fully-functional Web-based QA system for evaluation. For each topic, a human assessor will spend five minutes interacting with each system.".

The objective in this task is to gather a collection of relevant facts that contribute to built a nugget of information that responds to the question. The process includes the help of the user feedback on what information to select to integrate the final nugget, in an interactive and iterative fashion. Even tough the track in no longer offered at TREC it raises interesting research directions that we could integrate in our work. The main challenge would be to make the transition from a philosophy of selecting facts based on several occurrences to selecting facts that are novel.

Another future research direction to follow emerges from the case study and has to do with search in audio/video streams using textual information obtained automatically.

Finally, and given the author's professional functions and interest in teaching and on-line learning, to explore the educational potential of QA constitutes another area of application to be explored, through the creation of automated, innovative methodologies and tools designed to promote the students' learning process.

[1] http://www.umiacs.umd.edu/~jimmylin/ciqa/

Bibliography

Abney, S., M. Collins, & A. Singhal (2000, April). Answer Extraction. In *Proceedings of the 6th Applied Natural Language Processing Conference, Seattle, Washington, USA, April 29-May 4, 2000*, pp. 296–301. ACL - Association for Computational Linguistics. DOI: http://dx.doi.org/10.3115/974147.974188. 113

Aït-Mokhtar, S., J.-P. Chanod, & C. Roux (2001). A multi-input dependency parser. In *Proceedings of the Seventh IWPT (International Workshop on Parsing Technologies), Beijing, China, 17-19 October, 2001*. http://www.xrce.xerox.com/Research-Development/Publications/2001-0153/(language)/eng-GB. 48

Alves, M. A. (2002). Engenharia do Léxico Computacional: princípios, tecnologia e o caso das palavras compostas. Master's thesis, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Lisboa, Portugal. 54, 76, 135

Amaral, C., A. Cassan, H. Figueira, A. Martins, A. Mendes, P. Mendes, J. Pina, & C. Pinto (2008). Priberam's question answering system in QA@CLEF 2008. In *Working Notes of the 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008*. http://www.clef-campaign.org/2008/working_notes/amaral-paperCLEF2008.pdf. 46

Amaral, C., A. Cassan, H. Figueira, A. Martins, A. Mendes, P. Mendes, J. Pina, & C. Pinto (2009). Priberam's Question Answering System in QA@CLEF 2008. In *Evaluating Systems for Multilingual and Multimodal Information Access: 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers, LNCS Series Volume 5706*, pp. 337–344. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-642-04447-2_39. 22, 46, 200

Amaral, C., A. Cassan, H. Figueira, A. Martins, A. Mendes, P. Mendes, C. Pinto, &

D. Vidal (2007). Priberam's question answering system in QA@CLEF 2007. In *Working Notes of the 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007*. http://clef-campaign.org/2007/working_notes/AmaralCLEF2007.pdf. 45

Amaral, C., H. Figueira, A. Martins, A. Mendes, P. Mendes, & C. Pinto (2005). Priberam's question answering system for Portuguese. In *Working Notes of the 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria, September 21-23, 2005*. http://clef-campaign.org/2005/working_notes/workingnotes2005/amaral05.pdf. 38

Amaral, C., H. Figueira, A. Martins, A. Mendes, P. Mendes, & C. Pinto (2006). Priberam's Question Answering System for Portuguese. In *Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria, September 21-23, 2005, Revised Selected Papers, LNCS Series Volume 4022*, pp. 410–419. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/11878773_46. 38, 137

Amaral, C., H. Figueira, A. Mendes, P. Mendes, & C. Pinto (2004). A Workbench for Developing Natural Language Processing Tools. In *Pre-proceedings of the 1st Workshop on International Proofing Tools and Language Technologies, Patras, Greece, July 1-2, 2004*. http://www.priberam.pt/docs/WorkbenchNLP.pdf. 42

Amaral, C., D. Laurent, A. Martins, A. Mendes, & C. Pinto (2004). Design and Implementation of a Semantic Search Engine for Portuguese. In ELRA (Ed.), *Proc. of the 4th Language Resources and Evaluation Conference - LREC 2004, Lisboa, Portugal, May 24-30, 2004*, pp. 247–250. http://www.lrec-conf.org/proceedings/lrec2004/pdf/378.pdf. 38, 42, 212

Ayache, C., B. Grau, & A. Vilnat (2005). Campagne d'évaluation EQueR-EVALDA : Evaluation en Question-Réponse. In *Actes de l'Atelier EQueR-EASY de TALN 2005, Dourdan, France*. 19

Ayache, C., B. Grau, & A. Vilnat (2006). EQueR: the French Evaluation campaign of Question-Answering Systems. In ELRA (Ed.), *Proceedings of the 5th Language Resources*

and Evaluation Conference - LREC 2006, Genoa, Italy, May 22-28, 2006. http://www.lrec-conf.org/proceedings/lrec2006/pdf/653_pdf.pdf. 19

Banerjee, S. & T. Pedersen (2003). The Design, Implementation, and Use of the Ngram Statistics Package. In *"Computational Linguistics and Intelligent Text Processing" Proceedings of the 4th Conference on Intelligent Text Processing and Computational Linguistics - CICLing, Mexico City, Mexico, February 16-22, 2003, LNCS Series Volume 2588*, pp. 370–381. Springer-Verlag Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/3-540-36456-0_38. 33

Batista, F., D. Caseiro, N. Mamede, & I. Trancoso (2008, October). Recovering capitalization and punctuation marks for automatic speech recognition: Case study for Portuguese broadcast news. *Speech Communication 50*(10), 847–862. DOI: http://dx.doi.org/10.1016/j.specom.2008.05.008. 241

Batista, F., H. Moniz, I. Trancoso, H. Meinedo, A. I. Mata, & N. Mamede (2010). Extending the punctuation module for European Portuguese. In *Interspeech 2010, Makuhari, Chiba, Japan, September 26-30, 2010*. ISCA. http://www.inesc-id.pt/pt/indicadores/Ficheiros/6467.pdf. 241

Bick, E. (2000). *The Parsing System "PALAVRAS": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Ph. D. thesis, University of Aarhus, Denmark. 29, 36, 52

Bick, E. (2003a). A Constraint Grammar Based Question Answering System for Portuguese. In *Progress in Artificial Intelligence EPIA 2003, 2003, LNCS Series Volume 2902*, pp. 414–418. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-540-24580-3_47. 52

Bick, E. (2003b). Multi-level NER for Portuguese in a CG Framework. In *Proceedings of the 6th International Conference on Computational Processing of the Portuguese Language, PROPOR 2003, Faro, Portugal, June 26-28, 2003, LNCS Series Volume 2721*, pp. 118–125. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/3-540-45011-4_18. 52

Bilotti, M. W., B. Katz, & J. Lin (2004). What Works Better for Question Answering: Stemming or Morphological Query Expansion? In *Proceedings of the Work-*

shop IR4QA in the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, UK, July 25-29, 2004. http://nlp.shef.ac.uk/ir4qa04/bilotti-katz-lin-sigir04-ir4qa.pdf. 70

Bilotti, M. W. & E. Nyberg (2008). Improving text retrieval precision and answer accuracy in question answering systems. In *IRQA '08 Coling 2008: Proceedings of the 2nd workshop on Information Retrieval for Question Answering*, pp. 1–8. ACL - Association for Computational Linguistics. http://www.aclweb.org/anthology-new/W/W08/W08-1801.pdf. 115

Bobrow, D. G. (1964). A question-answering system for high school algebra word problems. In *Proceedings of the October 27-29, 1964, AFIPS American Federation of Information Processing Societies fall joint computer conference, part I*, pp. 591–614. ACM. DOI: http://dx.doi.org/10.1145/1464052.1464108. 14

Bowden, M., M. Olteanu, P. Suriyentrakorn, J. Clark, & D. Moldovan (2006). LCC's PowerAnswer at QA@CLEF 2006. In *Working Notes of the 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006, Alicante, Spain, September 20-22, 2006.* http://clef-campaign.org/2006/working_notes/workingnotes2006/bowdenCLEF2006.pdf. 49

Bowden, M., M. Olteanu, P. Suriyentrakorn, T. d'Silva, & D. Moldovan (2007). Multilingual Question Answering through Intermediate Translation: LCC's PowerAnswer at QA@CLEF 2007. In *Working Notes of the 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007.* http://www.clef-campaign.org/2007/working_notes/bowdenCLEF2007.pdf. 49

Bowden, M., M. Olteanu, P. Suriyentrakorn, T. d'Silva, & D. Moldovan (2008). Multilingual Question Answering through Intermediate Translation: LCC's PowerAnswer at QA@CLEF 2007. In *Advances in Multilingual and Multimodal Information Retrieval: 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007, Revised Selected Papers, LNCS Series Volume 5152*, pp. 273 – 283. DOI: http://dx.doi.org/10.1007/978-3-540-85760-0_33. 49

Box, G. E., J. S. Hunter, & W. G. Hunter (2005). *Statistics for Experimenters: Design, Innovation and Discovery - 2nd Edition.* John Wiley & Sons. ISBN 978-0-471-71813-0. 81

Branco, A., L. Rodrigues, J. Silva, & S. Silveira (2008a). Real-Time Open-Domain QA on the Portuguese Web. In *Proceedings of the 11th Ibero-American Conference on Artificial Intelligence, IBERAMIA 2008, Lisbon, Portugal, October 14-17, 2008, LNCS Series Volume 5290*, pp. 322–331. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-540-88309-8_33. 51

Branco, A., L. Rodrigues, J. Silva, & S. Silveira (2008b). XisQuê: An Online QA Service for Portuguese. In *Proceedings of the 8th International Conference on Computational Processing of the Portuguese Language, PROPOR 2008, Curia, Portugal, September 8-10, 2008, LNCS Series Volume 5190*, pp. 232–235. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-540-85980-2_27. 51

Brill, E. (2003). Processing Natural Language without Natural Language Processing. In *In "Computational Linguistics and Intelligent Text Processing" Proceedings of the 4th Conference on Intelligent Text Processing and Computational Linguistics - CICLing, Mexico City, Mexico, February 16-22, 2003, LNCS Series Volume 2588*, pp. 179–185. Springer-Verlag Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/3-540-36456-0_37. 33

Buckley, C. & E. M. Voorhees (2004). Retrieval Evaluation with Incomplete Information. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, UK, July 25-29, 2004*, pp. 25–32. ACM. ISBN: 1-58113-881-4 DOI: http://dx.doi.org/10.1145/1008992.1009000. 72

Büttcher, S., C. L. A. Clarke, & G. V. Cormack (2010). *Information Retrieval: Implementing and Evaluating Search Engines*. MIT Press. ISBN: 978-0-262-02651-2. 81

Cabral, L. M., L. F. Costa, & D. Santos (2007). Esfinge at CLEF 2007: First steps in a multiple question and multiple answer approach. In *Working Notes of the 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007*. http://clef-campaign.org/2007/working_notes/CabralCLEF2007.pdf. 37

Canty, A. J. (2002). Resampling Methods in R: The boot Package. In *R News, Vol. 2/3, December 2002*, pp. 2–7. ISSN: 1609-3631. 82

Carvalho, G., D. M. de Matos, & V. Rocio (2007). Document Retrieval for Question Answering: A Quantitative Evaluation of Text Preprocessing. In *Proceedings of ACM first*

Ph.D. Workshop, PIKM 2007, in the 16th ACM Conference on Information and Knowledge Management, CIKM 2007, Lisboa, Portugal, November 5-10, 2007, pp. 125–130. ACM. ISBN: 978-1-59593-832-9 DOI: http://dx.doi.org/10.1145/1316874.1316894. 57

Cassan, A., H. Figueira, A. Martins, A. Mendes, P. Mendes, C. Pinto, & D. Vidal (2006). Priberam's question answering system in a cross-language environment. In *Working Notes of the 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006, Alicante, Spain, September 20-22, 2006*. http://clef-campaign.org/2006/working_notes/workingnotes2006/cassanCLEF2006.pdf. 42

Christ, O. (1994). A modular and flexible architecture for an integrated corpus query system. In *Proceedings of the 3rd Conference on Computational Lexicography and Text Research, COMPLEX'94, July 7-10, 1994, Budapest, Hungary*, pp. 23–32. 34

Cleverdon, C. W. (1991). The significance of the Cranfield tests on index languages. In *Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Chicago, Illinois, USA, October 13-16, 1991*, pp. 3–12. ACM. DOI: http://dx.doi.org/10.1145/122860.122861. 71

Coheur, L., A. Mendes, J. ao Guimarães, N. J. Mamede, & R. Ribeiro (2008). QA@L^{2}F, second steps at QA@CLEF. In *Working Notes of the 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008*. http://www.clef-campaign.org/2008/working_notes/coheur-paperCLEF2008.pdf. 48

Coheur, L., A. Mendes, J. Guimarães, N. J. Mamede, & R. Ribeiro (2009). Question Interpretation in QA@L2F. In *Evaluating Systems for Multilingual and Multimodal Information Access: 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers, LNCS Series Volume 5706*, pp. 377–384. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-642-04447-2_44. 200

Costa, L. (2004). First evaluation of Esfinge - a question answering system for Portuguese. In *Working Notes of the 5th Workshop of the Cross-Language Evaluation Forum, CLEF 2004, Bath, UK, September 15-17, 2004*. http://clef-campaign.org/2004/working_notes/WorkingNotes2004/48a.pdf. 33

Costa, L. (2005). 20th Century Esfinge (Sphinx) solving the riddles at CLEF 2005. In *Working Notes of the 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria, September 21-23, 2005*. http://clef-campaign.org/2005/working_notes/workingnotes2005/costa05.pdf. 35

Costa, L. (2006a). 20th Century Esfinge (Sphinx) Solving the Riddles at CLEF 2005. In *Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria, September 21-23, 2005, Revised Selected Papers, LNCS Series Volume 4022*, pp. 467–476. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/11878773_52. 137

Costa, L. (2006b). Esfinge - a modular question answering system for Portuguese. In *Working Notes of the 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006, Alicante, Spain, September 20-22, 2006*. http://clef-campaign.org/2006/working_notes/workingnotes2006/CostaCLEF2006.pdf. 36

Costa, L. (2006c). Esfinge - a Question Answering System in the Web using the Web. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2006, Trento, Italy, April 3-7, 2006*. http://aclweb.org/anthology/E/E06/E06-2011.pdf. 36

Costa, L. (2008). Esfinge at CLEF 2008: Experimenting with answer retrieval patterns. Can they help? In *Working Notes of the 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008*. http://www.clef-campaign.org/2008/working_notes/costa-paperCLEF2008.pdf. 37

Costa, L. F. (2009). Using Answer Retrieval Patterns to Answer Portuguese Questions. In *Evaluating Systems for Multilingual and Multimodal Information Access: 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers, LNCS Series Volume 5706*, pp. 361–368. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-642-04447-2_42. 200

Costa, L. F. & L. Sarmento (2006). Component Evaluation in a Question Answering System. In ELRA (Ed.), *Proceedings of the 5th Language Resources and Evaluation*

Conference - LREC 2006, Genoa, Italy, May 22-28, 2006. http://www.lrec-conf.org/proceedings/lrec2006/pdf/306_pdf.pdf. 36, 46

Croft, W. B., D. Metzler, & T. Strohman (2010). *Search Engines: Information Retrieval in Practice*. Pearson Education. ISBN: 978-0-13-6072224-9. 81

Davison, A. & D. Hinkley (1997). *Bootstrap Methods and their Application*. Cambridge University Press. ISBN: 978-0-521-57471-6. 81, 82

Dias-da-Silva, B. C., M. F. Oliveira, H. R. Moraes, R. Hasegawa, D. Amorim, C. Paschoalino, & A. C. A. Nascimento (2000). Construção de um Thesaurus Eletrônico para o Português do Brasil. In *V Encontro para o Processamento computacional da Língua Portuguesa Escrita e Falada, Atibaia, Brasil*, Volume 4, pp. 1–10. 209

Earley, J. (1970). An efficient context-free parsing algorithm. *Communications of the ACM 13*(2), 94–102. DOI: http://dx.doi.org/10.1145/362007.362035. 45

Efron, B. & R. J. Tibshirani (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC. ISBN: 0-412-04231-2. 81

Ferrés, D., S. Kanaan, E. González, A. Ageno, H. Rodríguez, M. Surdeanu, & J. Turmo (2004). TALP-QA System at TREC 2004: Structural and Hierarchical Relaxation Over Semantic Constraints. In *Proceedings of the Thirteenth Text REtrieval Conference (TREC 2004), Question Answering Track, held in Gaithersburg, Maryland, November 16-19, 2004*. NIST - National Institute of Standards and Technology. http://trec.nist.gov/pubs/trec13/papers/upc.qa.pdf. 39

Ferrucci, D., E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, N. Schlaefer, & C. Welty (2010). Building Watson: An Overview of the DeepQA Project. *AI Magazine 31*(3), 59–79. 15

Filho, P. P. B., V. R. de Uzêda, T. A. S. Pardo, & M. das Graças Volpe Nunes (2006). Using a Text Summarization System for Monolingual Question Answering. In *Proceedings of the 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006, Alicante, Spain, September 20-22, 2006*. http://www.clef-campaign.org/2006/working_notes/workingnotes2006/BalageCLEF2006.pdf. 49

Flores, F., V. Moreira, & C. Heuser (2010). Assessing the Impact of Stemming Accuracy on Information Retrieval. In *Proceedings of the 9th International Conference on Computational Processing of the Portuguese Language, PROPOR 2010, Porto Alegre, RS, Brazil, April 27-30, 2010, LNCS Series Volume 6001*, pp. 11–20. Springer-Verlag Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-642-12320-7_2. 70

Forner, P., A. Peñas, E. Agirre, I. Alegria, C. Forăscu, N. Moreau, P. Osenova, P. Prokopidis, P. Rocha, B. Sacaleanu, R. Sutcliffe, & E. T. K. Sang (2009). Overview of the CLEF 2008 Multilingual Question Answering Track. In *Evaluating Systems for Multilingual and Multimodal Information Access: 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers, LNCS Series Volume 5706*, pp. 262–295. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-642-04447-2_34. 22

Fox, C. (1989). A stop list for general text. *ACM SIGIR Forum 24* (1-2 Fall 89/Winter 90), 19–21. DOI: http://dx.doi.org/10.1145/378881.378888. 75

Giampiccolo, D., A. Peñas, C. Ayache, D. Cristea, P. Forner, V. Jijkoun, P. Osenova, P. Rocha, B. Sacaleanu, & R. Sutcliffe (2007). Overview of the CLEF 2007 Multilingual Question Answering Track. In *Working Notes of the 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007*. http://www.clef-campaign.org/2007/working_notes/giampiccoloCLEF2007_Overview.pdf. 28, 49

Giampiccolo, D., A. Peñas, C. Ayache, D. Cristea, P. Forner, V. Jijkoun, P. Osenova, P. Rocha, B. Sacaleanu, & R. Sutcliffe (2008). Overview of the CLEF 2008 Multilingual Question Answering Track. In *Working Notes of the 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008*. http://www.clef-campaign.org/2008/working_notes/CLEF08Working_Notes_QA_Overview.pdf. 28, 49

Giménez, J. & L. Màrquez (2004). SVMTool: A general POS tagger generator based on Support Vector Machines. In ELRA (Ed.), *Proc. of the 4th Language Resources and Evaluation Conference - LREC 2004, Lisboa, Portugal, May 24-30, 2004*, pp. 43–46. http://www.lrec-conf.org/proceedings/lrec2004/pdf/597.pdf. 44

Green, B. F., A. K. Wolf, C. Chomsky, & K. Laughery (1961). Baseball: an automatic question-answerer. In *In Papers presented at IRE-AIEE-ACM AIEE computer conference '61 (Western),May 9-11, 1961*, pp. 219–224. ACM. DOI: http://dx.doi.org/10.1145/1460690.1460714. 14

Harman, D. (2005). The History of IDF and Its Influences on IR and Other Fields . In *Charting a New Course: Natural Language Processing and Information Retrieval, The Kluwer International Series on Information Retrieval, 2005, Volume 16*, pp. 69–79. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/1-4020-3467-9_5. 110

Hull, D. (1993). Using statistical testing in the evaluation of retrieval experiments. In *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Pittsburgh, Pennsylvania, USA, June 27- July 01, 1993*, pp. 329–338. ACM. DOI: http://dx.doi.org/10.1145/160688.160758. 81

Kanis, J. & L. Skorkovská (2010). Comparison of different lemmatization approaches through the means of information retrieval performance. In *Proceedings of the 13th edition of the International Conference on Text, Speech and Dialogue, TSD 2010, Brno, Czech Republic, September 6-10, 2010, LNCS Series Volume 6231*, pp. 93–100. Springer-Verlag Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-642-15760-8_13. 70

Kelly, D. & J. Lin (2007). Overview of the TREC 2006 ciQA Task. *ACM SIGIR Forum 41*(June 2007), 107–116. DOI: http://dx.doi.org/10.1145/1273221.1273231. 283

Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of the Tenth Machine Translation Summit, Phuket, Thailand, September 12-16, 2005*, pp. 76–86. http://www.mt-archive.info/MTS-2005-Koehn.pdf. 44

Lamel, L., S. Rosset, C. Ayache, D. Mostefa, J. Turmo, & P. R. Comas (2008). Question Answering on Speech Transcriptions: the QAST evaluation in CLEF. In *Proceedings of the 6th Language Resources and Evaluation Conference - LREC 2008, Marrakech, Morocco, May 26 - June 1, 2008*, pp. 1995–1999. http://www.lrec-conf.org/proceedings/lrec2008/pdf/511_paper.pdf. 233

Lopes, J. G., N. M. C. Marques, & V. Rocio (1994). POLARIS: A POrtuguese Lexicon Acquisition and Retrieval Interactive System. In *Proceedings of the Second International Conference on the Practical Applications of Prolog, London, UK*. 30, 76

Magnini, B., D. Giampiccolo, L. Aunimo, C. Ayache, P. Osenova, A. Peñas, M. de Rijke, B. Sacaleanu, D. Santos, & R. Sutcliffe (2006). The Multilingual Question Answering Track at CLEF. In ELRA (Ed.), *Proceedings of the 5th Language Resources and Evaluation Conference - LREC 2006, Genoa, Italy, May 22-28, 2006*. http://www.lrec-conf.org/proceedings/lrec2006/pdf/816_pdf.pdf. 18

Magnini, B., D. Giampiccolo, P. Forner, C. Ayache, V. Jijkoun, P. Osenova, A. Peñas, P. Rocha, B. Sacaleanu, & R. Sutcliffe (2006). Overview of the CLEF 2006 Multilingual Question Answering Track. In *Working Notes of the 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006, Alicante, Spain, September 20-22, 2006*. http://www.clef-campaign.org/2006/working_notes/workingnotes2006/magniniOCLEF2006.pdf. 28, 49

Magnini, B., S. Romagnoli, A. Vallin, J. Herrera, A. Peñas, V. Peinado, F. Verdejo, & M. de Rijke (2003). The Multiple Language Question Answering Track at CLEF 2003. In *Working Notes of the 4th Workshop of the Cross-Language Evaluation Forum, CLEF 2003, Trondheim, Norway, August 21-22, 2003*. http:http://www.clef-campaign.org/2003/WN_web/36.pdf. 19

Magnini, B., S. Romagnoli, A. Vallin, J. Herrera, A. Peñas, V. Peinado, F. Verdejo, & M. de Rijke (2004). The Multiple Language Question Answering Track at CLEF 2003. In *Comparative Evaluation of Multilingual Information Access Systems: 4th Workshop of the Cross-Language Evaluation Forum, CLEF 2003, Trondheim, Norway, August 21-22, 2003, LNCS Series Volume 3237*, pp. 471–486. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-540-30222-3_46. 19

Magnini, B., A. Vallin, C. Ayache, G. Ebach, A. Peñas, M. de Rijke, P. Rocha, K. Simov, & R. Sutcliffe (2004). Overview of the CLEF 2004 Multilingual Question Answering Track. In *Working Notes of the 5th Workshop of the Cross-Language Evaluation Forum, CLEF 2004, Bath, UK, 15-17 September 2004*. http://www.clef-campaign.org/2004/working_notes/WorkingNotes2004/35.pdf. 19, 28, 49

Manning, C. D., P. Raghavan, & H. Schütze (2008). *Introduction to Information Retrieval*. Cambridge University Press. ISBN: 978-0-521-86571-5. 131

Maziero, E. G., T. A. Pardo, A. D. Felippo, & B. C. Dias-da-Silva (2008). A Base de Dados Lexical e a Interface Web do TeP 2.0 - Thesaurus Electrônico para o Português do Brasil. In *VI Workshop em Tecnologia da Informação e da Linguagem Humana (TIL)*, pp. 390–392. 209

Medeiros, J. C. D. (1995). Análise Morfológica e Correcção Ortográfica do Português. Master's thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, Portugal. 48

Meinedo, H., A. Abad, T. Pellegrini, J. Neto, & I. Trancoso (2010). The L2F Broadcast News Speech Recognition System. In *FALA 2010: VI Jornadas en Tecnología del Habla and II Iberian SLTech Workshop, Vigo, Spain, November 10-12, 2010*, pp. 93–96. http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/VI/pdfs/0018.pdf. 241

Meinedo, H., M. Viveiros, & J. Neto (2008). Evaluation of a Live Broadcast News Subtitling System for Portuguese. In *Interspeech 2008, Brisbane, Australia, September 22-26, 2008*. ISCA. http://www.inesc-id.pt/pt/indicadores/Ficheiros/5251.pdf. 241

Mendes, A., L. Coheur, N. J. Mamede, L. R. ao, J. ao Loureiro, R. Ribeiro, F. Batista, & D. M. de Matos (2007). QA@L^{2}F@QA@CLEF. In *Working Notes of the 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007*. http://clef-campaign.org/2007/working_notes/mendesCLEF2007.pdf. 47

Moldovan, D., S. Harabagiu, M. Paşca, R. Mihalcea, R. Goodrum, R. Gîrju, & V. Rus (1999). Lasso: A Tool for Surfing the Answer Net. In *Proceedings of the Eighth Text REtrieval Conference (TREC-8), Question Answering Track, held in Gaithersburg, Maryland, November 17-19, 1999*. NIST - National Institute of Standards and Technology. http://trec.nist.gov/pubs/trec8/papers/smu.pdf. 113

Moreau, N., O. Hamon, D. Mostefa, S. Rosset, O. Galibert, L. Lamel, J. Turmo, P. R. Comas, P. Rosso, D. Buscaldi, & K. Choukri (2010). Evaluation Protocol and Tools for Question-Answering on Speech Transcripts. In *Proceedings of the 7th Language Resources*

and Evaluation Conference - LREC 2010, La Valletta, Malta, May 19-21, 2010, pp. 2769–2773. http://www.lrec-conf.org/proceedings/lrec2010/pdf/372_Paper.pdf. 233

Orengo, V. M. & C. Huyck (2001). A Stemming Algorithmm for the Portuguese Language. In *Proceedings of the Eighth International Symposium on String Processing and Information Retrieval SPIRE'01, Laguna de San Rafael, Chile, November 13-15, 2001*, pp. 186–193. IEEE Computer Society. DOI: http://dx.doi.org/10.1109/SPIRE.2001.10024. 35

Paşca, M. A. & S. M. Harabagiu (2001). High performance question/answering. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, Louisiana, USA, September 9-13, 2001*, pp. 366–374. ACM. DOI: http://dx.doi.org/10.1145/383952.384025. 113

Pardo, T. A. S. (2002). GistSumm: Um Sumarizador Automático Baseado na Idéia Principal de Textos. Technical Report NILC-TR-02-13, Núcleo Interinstitucional de Lingüística Computacional - NILC, Instituto de Ciẽncias Matemáticas e de Computação (ICMC) of Universidade de São Paulo (USP), São Carlos, Brazil. http://www.icmc.usp.br/~taspardo/NILCTR0213-Pardo.pdf. 49

Pardo, T. A. S., L. H. M. Rino, & M. G. V. Nunes (2003). GistSumm: A Summarization Tool Based on a New Extractive Method. In *Proceedings of the 6th International Conference on Computational Processing of the Portuguese Language, PROPOR 2003, Faro, Portugal, June 26-28, 2003, LNCS Series Volume 2721*, pp. 210–218. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/3-540-45011-4_34. 49

Paulo, J. L., M. Correia, N. J. Mamede, & C. Hagège (2002). Using Morphological, Syntactical, and Statistical Information for Automatic Term Acquisition. In *Proceedings of the PorTAL - Portugal for Natural Language Processing Faro, Portugal, June 23-26, 2002, LNCS Series Volume 2389*, pp. 219–227. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/3-540-45433-0_31. 48

Ponte, J. M. & W. B. Croft (1998). A Language Modeling Approach to Information Retrieval. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Melbourne, Australia, August 24-28, 1998*, pp. 275–281. ACM. DOI: http://dx.doi.org/10.1145/290941.291008. 108, 110

Prager, J., E. Brown, A. Coden, & D. Radev (2000). Question-answering by predictive annotation. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Athens, Greece, July 24-28, 2000*, pp. 184–191. ACM. DOI: http://dx.doi.org/10.1145/345508.345574. 39

Prager, J. M. (2006). Open-Domain Question-Answering. *Foundations and Trends in Information Retrieval 1*(2), 91–231. DOI: http://dx.doi.org/10.1561/1500000001. 39, 54, 158

Quaresma, P., L. Quintano, I. Rodrigues, J. Saias, & P. Salgueiro (2004). The University of Évora approach to QA@CLEF-2004. In *Working Notes of the 5th Workshop of the Cross-Language Evaluation Forum, CLEF 2004, Bath, UK, 15-17 September 2004*. http://clef-campaign.org/2004/working_notes/WorkingNotes2004/48b.pdf. 28

Quaresma, P. & I. Rodrigues (2005). A logic programming based approach to the QA@CLEF05 track. In *Working Notes of the 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria, September 21-23, 2005*. http://clef-campaign.org/2005/working_notes/workingnotes2005/quaresma05.pdf. 31

Ribeiro, R., L. Oliveira, & I. Trancoso (2003). Using Morphossyntactic Information in TTS Systems: comparing strategies for European Portuguese. In *Proceedings of the 6th International Conference on Computational Processing of the Portuguese Language, PROPOR 2003, Faro, Portugal, June 26-28, 2003, LNCS Series Volume 2721*, pp. 143–150. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/3-540-45011-4_21. 48

Roberts, I. & R. Gaizauskas (2004). Evaluating Passage Retrieval Approaches for Question Answering. In *Advances in Information Retrieval: Proceedings of the 26th European Conference on IR Research, ECIR 2004, Sunderland, UK, April 5-7, 2004, LNCS Series Volume 2997*, pp. 72–84. Springer-Verlag, Berlin, Heidelberg. http://www.springerlink.com/content/33gwclw6tbgg65xl/. 71, 72

Robertson, S. & K. Spärck Jones (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science 27*(3), 127–146. 106, 110, 111

Robertson, S. & H. Zaragoza (2009). The Probabilistic Relevance Framework: BM25

and Beyond. *Foundations and Trends in Information Retrieval 3*(4), 333–389. DOI: http://dx.doi.org/10.1561/1500000019. 101, 107, 111

Robertson, S. E. & S. Walker (1994). Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland,July 3-6, 1994*, pp. 232–241. ACM. http://portal.acm.org/citation.cfm?id=188561. 107

Robertson, S. E., S. Walker, S. Jones, M. M. Hancock-Beaulieu, & M. Gatford (1994). Okapi at TREC-3. In *Proceedings of the Third Text REtrieval Conference (TREC-3) held in Gaithersburg, Maryland, November 2-4, 1994*. NIST - National Institute of Standards and Technology. trec.nist.gov/pubs/trec3/papers/city.ps.gz. 107

Roux, C. (1999). Phrase-Driven Parser. In *Proceedings of VExTAL - Venezia per il Trattamento Automatico delle Lingue, VEXTAL 1999, Venezia, Italy. November 22-24, 1999*. http://project.cgm.unive.it/events/papers/roux.pdf. 48

Saggion, H., R. Gaizauskas, M. Hepple, I. Roberts, & M. A. Greenwood (2004). Exploring the Performance of Boolean Retrieval Strategies For Open Domain Question Answering. In *Proceedings of the Workshop IR4QA in the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, UK, July 25-29, 2004*. http://nlp.shef.ac.uk/ir4qa04/ir4qa-saggion.pdf. 115

Saias, J. & P. Quaresma (2007). The Senso Question Answering approach to Portuguese QA@CLEF-2007. In *Working Notes of the 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007*. http://clef-campaign.org/2007/working_notes/saiasCLEF2007.pdf. 32

Saias, J. & P. Quaresma (2008a). The Senso Question Answering System at QA@CLEF 2008. In *Working Notes of the 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008*. http://www.clef-campaign.org/2008/working_notes/saias-paperCLEF2008.pdf. 22, 32, 200

Saias, J. & P. Quaresma (2008b). The University of Évora's Participation in QA@CLEF-2007. In *Advances in Multilingual and Multimodal Information Retrieval:*

8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007, Revised Selected Papers, LNCS Series Volume 5152, pp. 316–323. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-540-85760-0_38. 32, 212

Salton, G. & C. Buckley (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management 24*(5), 513–523. DOI: http://dx.doi.org/10.1016/0306-4573(88)90021-0. 97, 110

Salton, G. & Lesk (1968). Computer evaluation of indexing and text processing. *Journal of the ACM (JACM) 15*(1), 8–36. DOI: http://dx.doi.org/10.1145/321439.321441. 81

Salton, G., A. Wong, & C. S. Yang (1975). A vector space model for automatic indexing. *Communications of the ACM 18*(11), 613–620. DOI: http://dx.doi.org/10.1145/361219.361220. 96

Santos, D. & P. rocha (2001). Evaluating CETEMPúblico, a free resource for Portuguese. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, Toulouse, France, July 9-11, 2001*, pp. 442–449. ACL - Association for Computational Linguistics. DOI: http://dx.doi.org/10.3115/1073012.1073070. 34

Santos, D. & P. Rocha (2005). The Key to the First CLEF with Portuguese: Topics, Questions and Answers in CHAVE. In *Multilingual Information Access for Text, Speech and Images: 5th Workshop of the Cross-Language Evaluation Forum, CLEF 2004, Bath, UK, September 15-17, 2004, Revised Selected Papers, LNCS Series Volume 3491*, pp. 821–832. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/11519645_80. 18

Santos, D., N. Seco, N. Cardoso, & R. Vilela (2006). HAREM: An Advanced NER Evaluation Contest for Portuguese. In ELRA (Ed.), *Proceedings of the 5th Language Resources and Evaluation Conference - LREC 2006, Genoa, Italy, May 22-28, 2006*. http://www.lrec-conf.org/proceedings/lrec2006/pdf/59_pdf.pdf. 35

Sarmento, L. (2006a). BACO - A large database of text and co-occurrences. In ELRA (Ed.), *Proceedings of the 5th Language Resources and Evaluation Conference -*

LREC 2006, Genoa, Italy, May 22-28, 2006. http://www.lrec-conf.org/proceedings/lrec2006/pdf/195_pdf.pdf. 36

Sarmento, L. (2006b). Hunting answers with RAPOSA (FOX). In *Working Notes of the 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006, Alicante, Spain, September 20-22, 2006*. http://clef-campaign.org/2006/working_notes/workingnotes2006/sarmentoCLEF2006.pdf. 46

Sarmento, L. (2006c). REPENTINO - A Wide-Scope Gazetteer for Entity Recognition in Portuguese. In *Proceedings of the 7th International Conference on Computational Processing of the Portuguese Language, PROPOR 2006, Itatiaia, RJ, Brazil, May 13-17, 2006, LNCS Series Volume 3960*, pp. 31–40. Springer-Verlag Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/11751984_4. 35

Sarmento, L. (2006d). SIEMÊS - A Named-Entity Recognizer for Portuguese Relying on Similarity Rules. In *Proceedings of the 7th International Conference on Computational Processing of the Portuguese Language, PROPOR 2006, Itatiaia, RJ, Brazil, May 13-17, 2006, LNCS Series Volume 3960*, pp. 90–99. Springer-Verlag Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/11751984_10. 35

Sarmento, L. & E. Oliveira (2007). Making RAPOSA (FOX) smarter. In *Working Notes of the 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007*. http://clef-campaign.org/2007/working_notes/sarmentoCLEF2007.pdf. 47

Sarmento, L., J. Teixeira, & E. Oliveira (2008). Experiments with Query Expansion in the RAPOSA (FOX) Question Answering System. In *Working Notes of the 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008*. http://www.clef-campaign.org/2008/working_notes/sarmento-paperCLEF2008.pdf. 47

Sarmento, L., J. Teixeira, & E. Oliveira (2009). Assessing the Impact of Thesaurus-Based Expansion Techniques in QA-Centric IR. In *Evaluating Systems for Multilingual and Multimodal Information Access: 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers, LNCS*

Series Volume 5706, pp. 325–332. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-642-04447-2_37. 200

Simões, A. M. & J. J. Almeida (2001). jspell.pm : um módulo de análise morfológica para uso em processamento de linguagem natural. In *Actas da Associação Portuguesa de Linguística, 16, Coimbra, 2000*, pp. 485–495. Associação Portuguesa de Linguística. http://repositorium.sdum.uminho.pt/bitstream/1822/638/1/jspell.pm.pdf. 35

Simmons, R. F. (1965). Answering english questions by computer: a survey. *Communications of the ACM 8*(1), 53–70. DOI: http://dx.doi.org/10.1145/363707.363732. 13

Simmons, R. F. (1970). Natural language question-answering systems: 1969. *Communications of the ACM 13*(1), 15–30. DOI: http://dx.doi.org/10.1145/361953.361963. 15

Singhal, A. (2001). Modern Information Retrieval: A Brief Overview. *IEEE Data Engineering Bulletin 24*(4), 35–43. http://sites.computer.org/debull/A01DEC-CD.pdf. 100, 112

Singhal, A., C. Buckley, & M. Mitra (1996). Pivoted document length normalization. In *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Zurich, Switzerland, August 18-22, 1996*, pp. 21–29. ACM. DOI: http://dx.doi.org/10.1145/243199.243206. 98

Singhal, A., J. Choi, D. Hindle, D. D. Lewis, & F. Pereira (1998). AT&T at TREC-7. In *Proceedings of the Seventh Text REtrieval Conference (TREC-7) held in Gaithersburg, Maryland, November 09-11, 1998*, pp. 239. NIST - National Institute of Standards and Technology. http://trec.nist.gov/pubs/trec7/papers/att.pdf.gz. 100, 112

Smucker, M. D., J. Allan, & B. Carterette (2007). A Comparison of Statistical Significance Tests for Information Retrieval Evaluation. In *Proceedings of the 16th ACM Conference on Information and Knowledge Management, Lisboa, Portugal*, pp. 623–632. ACM. ISBN: 978-1-59593-803-9 DOI: http://dx.doi.org/10.1145/1321440.1321528. 81

Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation 28*(1), 11–21. http://www.soi.city.ac.uk/~ser/idfpapers/ksj_orig.pdf. 110

Sudkamp, T. A. (2006). *Languges and Machines: An Introduction to the Theory of Computer Science - 3rd Edition*. Addison-Wesley/Pearson. ISBN: 978-0-321-32221-0. 214

Tanev, H. (2006). Extraction of Definitions for Bulgarian. In *Working Notes of the 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006, Alicante, Spain, September 20-22, 2006*. http://clef-campaign.org/2006/working_notes/workingnotes2006/tanevCLEF2006.pdf. 33

Tellex, S., B. Katz, J. Lin, A. Fernandes, & G. Marton (2003). Quantitative Evaluation of Passage Retrieval Algorithms for Question Answering. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, Canada, July 28 - August 1, 2003*, pp. 41–47. ACM. DOI: http://dx.doi.org/10.1145/860435.860445. 77, 114

Thede, S. M. & M. P. Harper (1999). A second-order hidden Markov model for part-of-speech tagging. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, College Park, Maryland, USA, June 20-26, 1999*, pp. 175–182. ACL - Association for Computational Linguistics. DOI: http://dx.doi.org/10.3115/1034678.1034712. 44

Turmo, J., P. R. Comas, C. Ayache, D. Mostefa, S. Rosset, & L. Lamel (2007). Overview of QAST 2007. In *Working Notes of the 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007*. http://www.clef-campaign.org/2007/working_notes/turmoCLEF2007-QASToverview.pdf. 233

Turmo, J., P. R. Comas, S. Rosset, O. Galibert, N. Moreau, D. Mostefa, P. Rosso, & D. Buscaldi (2009). Overview of QAST 2009. In *Working Notes of the 10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009, Corfu, Greece, September 30 - October 2, 2009*. http://www.clef-campaign.org/2009/working_notes/QAST2009-overview.pdf. 233

Turmo, J., P. R. Comas, S. Rosset, L. Lamel, N. Moreau, & D. Mostefa (2009). Overview of QAST 2008. In *Evaluating Systems for Multilingual and Multimodal In-*

formation Access: 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers, LNCS Series Volume 5706, pp. 314–324. Springer-Verlag, Berlin, Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-642-04447-2_36. 233

Vallin, A., B. Magnini, D. Giampiccolo, L. Aunimo, C. Ayache, P. Osenova, A. Peñas, M. de Rijke, B. Sacaleanu, D. Santos, & R. Sutcliffe (2005). Overview of the CLEF 2005 Multilingual Question Answering Track. In *Working Notes of the 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria, September 21-23, 2005*. http://www.clef-campaign.org/2005/working_notes/workingnotes2005/vallin05.pdf. 28, 49

Voorhees, E. M. (2002). Overview of the TREC 2002 Question Answering Track. In *Proceedings of the Eleventh Text REtrieval Conference (TREC 2002) held in Gaithersburg, Maryland, November 19-22, 2002*, pp. 115–123. NIST - National Institute of Standards and Technology. http://trec.nist.gov/pubs/trec11/papers/QA11.pdf. 18

Voorhees, E. M. (2007). Overview of TREC 2007. In *Proceedings of the Sixteenth Text REtrieval Conference (TREC 2007) held in Gaithersburg, Maryland, November 5-9, 2007*, pp. 1–16. NIST - National Institute of Standards and Technology. http://trec.nist.gov/pubs/trec16/papers/OVERVIEW16.pdf. 16

Voorhees, E. M. & D. M. Tice (2000). Building a question answering test collection. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Athens, Greece, July 24-28, 2000*, pp. 200–207. ACM. DOI: http://dx.doi.org/10.1145/345508.345577. 16

Wilcoxon, F. (1948). Individual comparisons by ranking methods. *Biometrics Bulletin 1*(6), 80–83. http://links.jstor.org/sici?sici=0099-4987%28194512%291%3A6%3C80%3AICBRM%3E2.0.CO%3B2-P. 245

Witten, I. H., A. Moffat, & T. C. Bell (1999). *Managing Gigabytes - Compressing and Indexing Documents and Images - 2nd Edition*. Morgan Kaufmann Publishers, INC. ISBN: 978-1-55860-570-1. 131

Yu, J., J. A. Thom, & A. Tam (2007). Ontology evaluation using wikipedia categories for browsing. In *Proceedings of the 16th ACM Conference on Information and Knowledge*

Management, Lisboa, Portugal, pp. 223–232. ACM. DOI: http://dx.doi.org/10.1145/1321440.1321474. 212

Zobel, J., S. Heinz, & H. E. Williams (2001, December). In-memory Hash Tables for Accumulating Text Vocabularies. *Information Processing Letters 80*(6), 271–277. DOI: http://dx.doi.org/10.1016/S0020-0190(01)00239-3. 123

V Appendices

A Pre-Processing Experimental Results (Portuguese)

A.1 Question Set

Table A.1: Questions: Part 1 of 5 (Questions 1-36)

Q.#	**Question**
1	Em que cidade se encontra a prisão de San Vittore?
2	Onde era o campo de concentração de Auschwitz?
3	Quem foi o autor de "Mein Kampf"?
4	Qual é a capital da Rússia?
5	Quem foi o primeiro presidente dos Estados Unidos?
6	Como morreu Jimi Hendrix?
7	Com quem se casou Michael Jackson?
8	Em que género musical se distingue Michael Jackson?
9	O que é a Mossad?
10	Quantos crimes são atribuídos ao Monstro de Florença?
11	Quantos desempregados há na Europa?
12	Quantas religiões monoteístas há no mundo?
13	O que é a UNICEF?
14	Nomeie uma pessoa acusada de pedofilia.
15	Quem é Jean-Bertrand Aristide?
16	Quem escreveu "Ulisses"?
17	Onde se situa o CERN?
18	Quem é Yves Saint-Laurent?
19	Em que dia calha o solstício de verão?
20	Onde fica o Museu do Hermitage?
21	De que são feitos os cabos de fibra óptica?
22	Qual o nome da mulher de Kurt Cobain?
23	Quando foi lançada a sonda espacial Ulisses?
24	O que é a maçonaria?
25	Quem foi o último czar da Rússia?
26	Qual o acrónimo da Amnistia Internacional?
27	Onde se entregam os Óscares?
28	Onde fica o arquipélago de Svalbard?
29	Onde se realizou a Conferência Mundial da Mulher?
30	Indique uma companhia de fast-food.
31	Quem foi Rosa Chacel?
32	Quem é Christo?
33	O que são as FARC?
34	Qual a abreviatura do Exército Popular de Libertação do Sudão?
35	Esmirna fica em que país?
36	Qual a localização de Tipaza?

Table A.2: Questions: Part 2 of 5 (Questions 37-72)

Q.#	**Question**
37	Como se chama a filha do líder chinês Deng Xiaoping?
38	O que é o FSK?
39	Em que país fica Vukovar?
40	Em que cidade americana se encontra o Museu Warhol?
41	Quem é Andy Warhol?
42	Quem descobriu o vírus da sida?
43	Quem é a ministra do Ambiente alemã?
44	Mencione um bonecreiro.
45	Quem é o presidente da UEFA?
46	O que é a MTV?
47	O que é a NASA?
48	Em que equipa de básquete joga Shaquille O'Neill?
49	Quem é o presidente da Câmara dos Representantes americana?
50	Em que ano foi assassinado o presidente chileno Salvador Allende?
51	Quem é Marvin Minsky?
52	Como se intitula a autobiografia de Nelson Mandela?
53	Como se chama a viúva do falecido presidente de Moçambique, Samora Machel?
54	Quem é João Havelange?
55	O que significa a abreviatura OUA?
56	Onde fica Hyde Park?
57	Quantos assinantes tem a MSN?
58	O que é a UNICE?
59	Quem foi forçado a demitir-se de governador da Caríntia em 1991?
60	Qual o lucro do grupo electrónico e de telecomunicações finlandês Nokia em 1994?
61	O que é o CERN?
62	Quantos estados-membros tem o CERN?
63	Quem é Kevin Mitnick?
64	Quando foi criado o CERN?
65	Qual o monte mais alto do mundo?
66	Onde fica Halifax?
67	Quem é Umberto Bossi?
68	Onde fica o La Scala?
69	Onde fica a sede da UNESCO?
70	Quem é o realizador de "Nikita"?
71	O que é o GIA?
72	Qual o cargo de Redha Malek em 1994?

Table A.3: Questions: Part 3 of 5 (Questions 73-108)

Q.#	Question
73	Que quantia exige o FC Sevilha de Diego Maradona?
74	Como se chama o ministro das Finanças polaco?
75	Como morreu Juvénal Habyarimana?
76	Qual a nacionalidade do tenista Sergi Bruguera?
77	De que país é a escritora Taslima Nasreen?
78	Qual o cargo de Albert Reynolds na Irlanda?
79	Qual a taxa de desemprego nos Estados Unidos no final de 1994?
80	Quando tiveram lugar as eleições europeias de 1994?
81	Onde fica a Esfinge de Gizé?
82	Onde é Izhevsk?
83	Onde fica o Estádio José Alvalade?
84	Onde vive José Saramago?
85	Onde nasceu Nelson Mandela?
86	Qual o maior satélite de Júpiter?
87	Onde fica Turku?
88	Qual a antiga capital da Polónia?
89	Em que distrito fica Paredes de Coura?
90	Onde fica Sosnovy Bor?
91	Em que ilha fica Ponta Delgada?
92	Onde é que nasceu Álvaro Cunhal?
93	Onde é o hospital Júlio de Matos?
94	Qual é o país mais pequeno da União Europeia?
95	Onde fica Gabrovo?
96	Qual o estado mais setentrional dos EUA?
97	Qual é a capital da Bielorússia?
98	Onde desagua o rio Cubango?
99	Em que cidade o Mosela encontra o Reno?
100	Em que estado do Brasil fica Campo Grande?
101	Onde se situa Tianjin?
102	Onde é a Ilha do Diabo?
103	Quem inventou o saxofone?
104	Quem escreveu ”O Principezinho”?
105	Quem é o recordista mundial do salto à vara?
106	Quem é a ”diva dos pés descalços”?
107	Quem é o secretário-geral do PCP?
108	De quem é filha Martine Aubry?

Table A.4: Questions: Part 4 of 5 (Questions 109-144)

Q.#	**Question**
109	Quem é o Presidente da Câmara de Lisboa?
110	Quem é o Presidente da Câmara de Lamego?
111	Quem é o embaixador de Portugal em França?
112	Com quem casou Whoppi Goldberg?
113	Quem foi o primeiro presidente dos Estados Unidos?
114	Quem é o ministro-presidente da Renânia-Palatinado?
115	Quem foi o último governador de Timor Leste?
116	Quem era o marido de Vieira da Silva?
117	Quem é o capitão do FC Porto?
118	Quem é o imã da mesquita de Lisboa?
119	Quem realizou o filme "Lisbon Story"?
120	Quem é a ministra sueca do ambiente?
121	Como se chama a rainha da Dinamarca?
122	Quem é o padroeiro de Penafiel?
123	Que grupo matou Aldo Moro?
124	De que grupo é vocalista Teresa Salgueiro?
125	Que equipa venceu a Taça CERS em hóquei em patins?
126	De que clube é treinador Bobby Robson?
127	Que empresa tem uma refinaria em Leça da Palmeira?
128	A que partido pertence Duarte Lima?
129	Quem financia as IPSS?
130	Quantos submarinos tem a marinha portuguesa?
131	Quantos municípios há em Portugal?
132	Qual o comprimento da Ponte do Freixo?
133	Quantos anos tem Inês de Medeiros?
134	Qual a distância de Braga a Guimarães?
135	Qual a altura do K2?
136	Qual o valor da dívida da Eurotunnel?
137	Qual a área da Baixa-Saxónia?
138	Quantos habitantes tem a República Dominicana?
139	Quantos golos marcou Eusébio na sua carreira?
140	A que velocidade viaja a luz?
141	Quando foram criadas as FPLM (Forças Populares de Libertação de Moçambique)?
142	Quando foi a independência de Cabo Verde?
143	Quando estreia o filme "Lisbon Story"?
144	Quando foi aprovada a Declaração Universal dos Direitos do Homem?

Table A.5: Questions: Part 5 of 5 (Questions 145-180)

Q.#	Question
145	Quando morreu Salvador Allende?
146	Quando morreu Simão Bolívar?
147	Em que dia se comemora a independência do Brasil?
148	Quando se tornou "A Portuguesa" hino nacional?
149	Em que ano ocorreu o 25 de Abril?
150	Em que embateu o Titanic?
151	Qual o símbolo de liderança da Volta a Itália?
152	O que foi erguido em 13 de Agosto de 1961?
153	A que era alérgico Mel Blanc?
154	Que país é campeão do mundo de futebol?
155	Como morreu Pasolini?
156	Como se tornou o Brasil tetracampeão mundial de futebol?
157	Qual foi o primeiro filme sonoro português?
158	Que vende Fausto ao Diabo?
159	Que animal é o símbolo da Namíbia?
160	Qual o pseudónimo de Álvaro Cunhal?
161	Qual é a nacionalidade de Yordan Letchkov?
162	Qual a nacionalidade de Hercule Poirot?
163	O que era Napoleão III a Napoleão Bonaparte?
164	De que material são os frisos do Parténon?
165	Qual a patente de Alfred Dreyfus?
166	De que cor é a neve?
167	Qual é a moeda iraquiana?
168	Qual o endereço da Livraria Barata?
169	Quem é Leonor Beleza?
170	Quem é Arnold Ruutel?
171	Quem é Wim Duisenberg?
172	Quem é Rocha Vieira?
173	Quem é Guilherme da Fonseca?
174	Quem é Fernando Gomes?
175	Quem é Valentina Terechkova?
176	Quem é Jorge Amado?
177	O que é o PC do B?
178	O que é o PSN?
179	O que é o CSKA?
180	O que é a Vigor?

A.2 Test Results

Table A.6: Pre-Processing Test Results

Question #	Test0	Test1	Test2	Test3	Test4	Test5	Test6	Test7	Test8
1	-1	9	4	5	5	5	7	3	5
2	695	2	10	6	17	20	4	11	8
3	-1	-1	5	5	10	10	5	5	5
4	-1	-1	-1	-1	-1	-1	-1	-1	-1
5	-1	-1	-1	-1	-1	-1	-1	-1	-1
6	3	5	4	4	5	5	4	6	4
7	2	3	2	1	4	2	1	1	1
8	-1	135	109	115	93	100	115	121	119
9	-1	26	27	27	18	21	21	27	27
10	-1	522	11	12	16	11	12	30	32
11	69	19	6	6	3	4	6	23	11
12	-1	-1	-1	-1	-1	-1	-1	-1	-1
13	-1	50	17	16	14	15	21	16	16
14	27	38	39	39	44	52	32	6	7
15	-1	537	234	119	90	94	101	119	118
16	-1	-1	27	27	14	10	39	35	27
17	-1	796	102	93	131	136	86	23	448
18	137	31	31	1	2	2	1	1	1
19	-1	-1	-1	-1	-1	-1	-1	485	-1
20	-1	326	4	4	19	22	6	4	3
21	18	16	18	18	15	16	14	57	62
22	7	2	6	6	5	2	5	7	7
23	10	1	1	1	2	4	1	2	1
24	-1	-1	-1	-1	-1	-1	-1	-1	-1
25	7	11	12	11	4	4	14	10	10
26	251	217	197	197	246	100	211	235	225
27	-1	-1	-1	-1	-1	-1	-1	566	-1
28	-1	70	4	4	6	4	3	8	4
29	-1	-1	694	-1	111	129	632	566	449
30	-1	-1	-1	79	73	62	68	77	83

Question #	Test0	Test1	Test2	Test3	Test4	Test5	Test6	Test7	Test8
31	528	1	1	2	1	1	1	1	2
32	-1	6	5	5	2	2	5	5	5
33	-1	3	3	3	3	3	3	3	3
34	-1	-1	-1	-1	-1	-1	-1	-1	-1
35	1	1	3	3	4	4	4	4	3
36	-1	2	2	2	2	2	2	3	3
37	209	49	69	72	54	56	99	80	76
38	-1	3	2	2	3	3	3	2	2
39	-1	81	3	3	6	3	4	4	3
40	111	1	1	1	2	2	2	2	1
41	29	9	19	19	10	10	18	19	19
42	-1	-1	-1	-1	-1	-1	-1	-1	-1
43	171	273	452	460	273	290	484	471	472
44	-1	-1	1	1	1	1	1	1	-1
45	-1	146	111	113	467	162	63	115	113
46	-1	68	61	59	55	60	80	62	59
47	-1	116	259	259	279	253	269	259	259
48	279	168	182	176	215	134	182	201	196
49	-1	-1	-1	-1	-1	-1	-1	-1	-1
50	27	17	19	19	19	16	19	19	18
51	17	8	6	6	2	2	6	6	6
52	1	1	1	1	1	1	1	1	1
53	261	2	2	2	10	6	2	2	2
54	-1	8	13	13	2	2	12	13	13
55	-1	-1	59	59	75	76	66	67	63
56	17	9	10	10	7	7	9	3	10
57	77	10	4	5	4	4	3	5	4
58	-1	3	3	3	3	3	3	3	-1
59	-1	-1	-1	76	11	11	35	209	84
60	4	4	5	5	1	5	2	2	2
61	-1	77	34	35	34	34	34	34	148
62	-1	-1	-1	2	3	3	3	4	3
63	531	1	1	1	1	1	1	1	1
64	-1	3	2	2	3	3	2	2	1
65	153	301	307	314	175	81	298	694	684
66	-1	37	3	4	14	14	3	1	4

Question #	Test0	Test1	Test2	Test3	Test4	Test5	Test6	Test7	Test8
67	42	30	23	36	34	17	40	36	39
68	-1	-1	188	110	449	510	264	64	128
69	-1	-1	229	227	65	67	149	233	228
70	-1	-1	367	362	550	591	152	393	393
71	-1	73	49	50	61	68	53	71	50
72	11	9	5	5	6	3	5	5	6
73	2	1	2	2	1	1	1	2	1
74	-1	-1	227	120	119	127	128	52	40
75	-1	2	5	5	7	9	5	7	5
76	822	245	130	130	133	116	124	104	108
77	61	70	72	72	83	83	80	74	74
78	-1	-1	-1	-1	-1	-1	-1	-1	-1
79	-1	-1	-1	-1	-1	-1	-1	-1	-1
80	-1	-1	-1	-1	-1	-1	-1	-1	-1
81	3	4	4	4	4	5	6	4	4
82	-1	1	1	1	1	1	1	1	1
83	-1	-1	618	611	306	331	908	768	611
84	-1	197	37	38	49	53	41	47	38
85	6	1	1	1	3	3	5	1	1
86	-1	-1	-1	-1	-1	-1	-1	-1	-1
87	-1	106	6	6	14	14	7	8	6
88	275	321	468	462	365	131	456	-1	-1
89	89	72	33	33	29	29	23	25	55
90	14	1	1	1	1	1	1	1	1
91	44	24	34	32	38	16	27	39	21
92	-1	4	8	8	28	29	10	8	8
93	-1	97	5	5	3	4	5	31	9
94	388	171	457	464	256	701	210	943	962
95	-1	1	1	1	1	1	1	1	1
96	2	3	1	1	2	1	1	1	1
97	-1	-1	-1	-1	-1	-1	-1	-1	-1
98	50	1	4	4	8	9	4	3	4
99	-1	-1	-1	-1	-1	-1	-1	-1	-1
100	55	86	123	135	239	93	80	228	225
101	-1	309	4	4	5	6	2	6	4
102	334	299	87	85	18	21	90	103	111

Question #	Test0	Test1	Test2	Test3	Test4	Test5	Test6	Test7	Test8
103	-1	375	243	244	401	326	209	345	318
104	-1	-1	-1	-1	-1	-1	-1	-1	5
105	33	30	18	18	9	11	24	21	23
106	-1	-1	95	96	80	83	113	125	140
107	-1	-1	-1	935	-1	-1	748	942	-1
108	2	2	2	2	3	3	2	2	2
109	-1	-1	-1	-1	-1	378	-1	-1	-1
110	-1	5	14	13	28	7	21	13	13
111	201	74	117	114	494	364	112	107	127
112	-1	143	5	2	1	1	2	4	2
113	-1	-1	-1	-1	-1	-1	-1	-1	-1
114	-1	-1	-1	7	8	4	6	8	8
115	37	-1	37	14	14	16	10	23	14
116	-1	-1	63	65	20	20	109	140	69
117	947	988	-1	-1	982	-1	-1	-1	-1
118	52	269	286	281	619	213	149	299	288
119	143	117	11	8	10	11	10	6	10
120	-1	422	19	13	6	6	5	25	27
121	-1	507	87	103	75	91	70	136	106
122	11	12	4	3	2	2	2	4	4
123	250	20	9	9	9	8	9	4	9
124	28	12	11	11	9	5	7	12	12
125	25	27	59	-1	42	46	59	62	60
126	-1	-1	-1	-1	-1	-1	-1	-1	-1
127	26	5	4	4	4	4	5	3	4
128	-1	-1	-1	-1	-1	-1	-1	-1	-1
129	109	1	1	1	1	1	1	1	1
130	1	1	1	1	1	1	2	1	1
131	-1	-1	-1	-1	-1	-1	-1	-1	-1
132	12	1	2	2	2	2	2	2	2
133	-1	42	58	58	225	145	51	99	77
134	-1	-1	-1	512	549	177	740	577	577
135	-1	1	1	1	2	2	2	1	1
136	-1	7	11	11	5	17	4	11	11
137	-1	-1	-1	284	235	95	177	726	588
138	-1	-1	-1	-1	-1	-1	-1	-1	981

Question #	Test0	Test1	Test2	Test3	Test4	Test5	Test6	Test7	Test8
139	24	39	11	11	21	30	25	23	23
140	124	194	89	31	24	34	17	149	32
141	1	1	1	1	1	1	1	1	1
142	605	853	577	517	417	441	360	649	548
143	-1	-1	27	27	21	23	24	11	11
144	75	29	38	37	37	41	16	110	70
145	-1	39	12	11	19	19	10	13	30
146	-1	-1	-1	-1	-1	-1	-1	-1	-1
147	-1	454	184	222	382	426	249	272	271
148	629	230	157	156	300	223	153	62	190
149	-1	-1	-1	-1	-1	-1	-1	-1	-1
150	108	3	2	2	1	1	1	2	2
151	614	-1	-1	-1	-1	-1	-1	-1	-1
152	22	24	16	16	7	8	8	35	20
153	29	8	1	1	4	2	1	6	2
154	-1	-1	779	743	328	368	486	760	760
155	683	4	3	3	1	2	4	5	3
156	624	601	-1	-1	-1	-1	-1	588	-1
157	434	466	774	794	836	632	-1	-1	-1
158	3	5	3	3	4	4	4	4	3
159	-1	-1	-1	-1	-1	-1	-1	-1	-1
160	-1	-1	-1	-1	-1	-1	-1	-1	-1
161	115	6	5	4	3	2	3	4	4
162	51	3	3	3	3	3	3	3	3
163	53	7	6	8	3	3	4	16	8
164	-1	-1	-1	-1	-1	-1	-1	-1	-1
165	707	1	2	2	2	2	2	2	2
166	-1	-1	-1	-1	-1	-1	-1	835	793
167	149	190	5	5	4	3	5	29	29
168	-1	-1	-1	-1	-1	-1	-1	-1	-1
169	536	141	97	94	162	113	106	94	94
170	79	1	1	1	1	1	1	1	1
171	-1	5	6	6	1	1	5	6	6
172	-1	-1	-1	-1	-1	-1	-1	-1	-1
173	675	248	70	70	8	11	39	70	75
174	-1	-1	-1	-1	-1	-1	-1	-1	-1

Question #	**Test0**	**Test1**	**Test2**	**Test3**	**Test4**	**Test5**	**Test6**	**Test7**	**Test8**
175	40	1	3	3	2	2	3	3	3
176	-1	183	28	31	21	21	31	33	33
177	-1	-1	-1	-1	-1	-1	-1	-1	-1
178	-1	40	20	21	16	16	12	21	21
179	-1	30	50	52	38	42	45	48	52
180	-1	1	1	1	1	1	1	1	2

B Small Portuguese Corpus based on Famous Poems

The Poem Text Collection used as an example of a small Portuguese Corpus in Chapter 5, is composed of nine verses from nine poems from well known Portuguese and Brazilian poets. Since the corpus must be small enough to be fully treated manually we had to choose a small number of documents, each of them containing only a few words. Instead of creating the phrases in the example, we decided to "borrow" them from some well known Poems, of equally well known Poets from Portuguese Speaking Community. The corpus should have multiple occurrences of the same words, for the example to be clear about the points we were explaining, so the words chosen to get together the lines form the poems, were the small and frequently occurring words, eu [I] and não [no]. In this appendix we identify the Poets of each document, and present the complete text of the poems.

We reproduce here the original version of the texts, as shown an Table 5.2.

Table B.1: Poem text collection: Original Documents

Doc #	Document
1	Eles não sabem que o sonho
2	Assim devera eu ser se não fora não querer.
3	Eu te amo porque te amo,
4	Ai que prazer não cumprir um dever
5	Eu amo o longe e a miragem
6	É um não querer mais que bem querer;
7	Eu não sou eu nem sou o outro,
8	Se as coisas são inatingíveis... ora! Não é motivo para não querê-las...
9	Eu te peço perdão por te amar de repente

António Gedeão
alias **Rómulo Vasco da Gama de Carvalho**
Lisboa, 24 de Novembro de 1906 – Lisboa, 19 de Fevereiro de 1997

In "Movimento Perpétuo", 1954

Pedra Filosofal

Eles não sabem que o sonho
é uma constante da vida
tão concreta e definida
como outra coisa qualquer,
como esta pedra cinzenta
em que me sento e descanso,
como este ribeiro manso
em serenos sobressaltos,
como estes pinheiros altos
que em verde e oiro se agitam,
como estas aves que gritam
em bebedeiras de azul.

Eles não sabem que o sonho
é vinho, é espuma, é fermento,
bichinho álacre e sedento,
de focinho pontiagudo,
que fossa através de tudo
num perpétuo movimento.

Eles não sabem que o sonho
é tela, é cor, é pincel,
base, fuste, capitel,
arco em ogiva, vitral,
pináculo de catedral,
contraponto, sinfonia,
máscara grega, magia,
que é retorta de alquimista,
mapa do mundo distante,
rosa-dos-ventos, Infante,
caravela quinhentista,
que é Cabo da Boa Esperança,
ouro, canela, marfim,
florete de espadachim,
bastidor, passo de dança,
Colombina e Arlequim,
passarola voadora,
pára-raios, locomotiva,
barco de proa festiva,
alto-forno, geradora,
cisão do átomo, radar,
ultra-som, televisão,
desembarque em foguetão
na superfície lunar.

Eles não sabem, nem sonham,
que o sonho comanda a vida.
Que sempre que um homem sonha
o mundo pula e avança
como bola colorida
entre as mãos de uma criança.

DOC #1

Figure B.1: António Gedeão - Doc 1

Alexandre O'Neill
Lisboa, 19 Dezembro 1924 – Lisboa, 21 Agosto 1986

In "Feira Cabisbaixa", 1965

Velha Fábula em Bossa Nova

Minuciosa formiga
não tem que se lhe diga:
leva a sua palhinha
asinha, asinha.
Assim devera eu ser
e não esta cigarra
que se põe a cantar
e me deita a perder.
Assim devera eu ser:
de patinhas no chão,
formiguinha ao trabalho
e ao tostão.
Assim devera eu ser
se não fora
não querer.
(- Obrigado, formiga!
Mas a palha não cabe
onde você sabe...)

DOC #2

Figure B.2: Alexandre O'Neill - Doc 2

Carlos Drummond de Andrade
Itabira do Mato Dentro, Minas Gerais (MG), 31 de Outubro de 1902
— Rio de Janeiro (RJ), 17 de Agosto de 1987

In "Amar se Aprende Amando", 1985

As Sem-razões do Amor

Eu te amo porque te amo,
Não precisas ser amante,
e nem sempre sabes sê-lo.
Eu te amo porque te amo.
Amor é estado de graça
e com amor não se paga.

Amor é dado de graça,
é semeado no vento,
na cachoeira, no eclipse.
Amor foge a dicionários
e a regulamentos vários.

Eu te amo porque não amo
bastante ou demais a mim.
Porque amor não se troca,
não se conjuga nem se ama.
Porque amor é amor a nada,
feliz e forte em si mesmo.

Amor é primo da morte,
e da morte vencedor,
por mais que o matem (e matam)
a cada instante de amor.

DOC #3

Figure B.3: Carlos Drummond de Andrade - Doc 3

Fernando Pessoa
Lisboa, 13 de Junho de 1888 — Lisboa, 30 de Novembro de 1935

In "Liberdade", 1935

Ai que prazer
Não cumprir um dever,
Ter um livro para ler
E não o fazer!
Ler é maçada,
Estudar é nada.
O sol doira
Sem literatura.

O rio corre, bem ou mal,
Sem edição original.
E a brisa, essa,
De tão naturalmente matinal,
Como tem tempo não tem pressa...

Livros são papéis pintados com tinta.
Estudar é uma coisa em que está indistinta
A distinção entre nada e coisa nenhuma.

Quanto é melhor, quando há bruma,
Esperar por D. Sebastião,
Quer venha ou não!

Grande é a poesia, a bondade e as danças...
Mas o melhor do mundo são as crianças,
Flores, música, o luar, e o sol, que peca
Só quando, em vez de criar, seca.

O mais do que isto
É Jesus Cristo,
Que não sabia nada de finanças
Nem consta que tivesse biblioteca...

DOC #4

Figure B.4: Fernando Pessoa - Doc 4

José Régio
Vila do Conde, 17 de Setembro de 1901 — Vila do Conde,
22 de Dezembro de 1969

In "Poemas de Deus e do Diabo", 1925

Cântico Negro

"Vem por aqui" — dizem-me alguns com os olhos doces
Estendendo-me os braços, e seguros
De que seria bom que eu os ouvisse
Quando me dizem: "vem por aqui!"
Eu olho-os com olhos lassos,
(Há, nos olhos meus, ironias e cansaços)
E cruzo os braços,
E nunca vou por ali...
A minha glória é esta:
Criar desumanidades!
Não acompanhar ninguém.
— Que eu vivo com o mesmo sem-vontade
Com que rasguei o ventre à minha mãe
Não, não vou por aí! Só vou por onde
Me levam meus próprios passos...
Se ao que busco saber nenhum de vós responde
Por que me repetis: "vem por aqui!"?

Prefiro escorregar nos becos lamacentos,
Redemoinhar aos ventos,
Como farrapos, arrastar os pés sangrentos,
A ir por aí...
Se vim ao mundo, foi
Só para desflorar florestas virgens,
E desenhar meus próprios pés na areia inexplorada!
O mais que faço não vale nada.

Como, pois, sereis vós
Que me dareis impulsos, ferramentas e coragem
Para eu derrubar os meus obstáculos?...
Corre, nas vossas veias, sangue velho dos avós,
E vós amais o que é fácil!
Eu amo o Longe e a Miragem,
Amo os abismos, as torrentes, os desertos...

Ide! Tendes estradas,
Tendes jardins, tendes canteiros,
Tendes pátria, tendes tetos,
E tendes regras, e tratados, e filósofos, e sábios...
Eu tenho a minha Loucura !
Levanto-a, como um facho, a arder na noite escura,
E sinto espuma, e sangue, e cânticos nos lábios...
Deus e o Diabo é que guiam, mais ninguém!
Todos tiveram pai, todos tiveram mãe;
Mas eu, que nunca principio nem acabo,
Nasci do amor que há entre Deus e o Diabo.

Ah, que ninguém me dê piedosas intenções,
Ninguém me peça definições!
Ninguém me diga: "vem por aqui"!
A minha vida é um vendaval que se soltou,
É uma onda que se alevantou,
É um átomo a mais que se animou...
Não sei por onde vou,
Não sei para onde vou
Sei que não vou por aí!

DOC #5

Figure B.5: José Régio - Doc 5

Luís Vaz de Camões
Constância 1524? — 1580

Amor é fogo que arde sem se ver,
é ferida que dói, e não se sente;
é um contentamento descontente,
é dor que desatina sem doer.

É um não querer mais que bem querer;
é um andar solitário entre a gente;
é nunca contentar-se de contente;
é um cuidar que ganha em se perder.

É querer estar preso por vontade;
é servir a quem vence, o vencedor;
é ter com quem nos mata, lealdade.

Mas como causar pode seu favor
nos corações humanos amizade,
se tão contrário a si é o mesmo Amor?

DOC #6

Figure B.6: Luís Vaz de Camões - Doc 6

Mário de Sá Carneiro
Lisboa, 19 Maio 1890 — Paris, 26 Abril 1916

"Eu não sou eu nem sou o outro,
Sou qualquer coisa de intermédio:
Pilar da ponte do tédio
Que vai de mim para o outro."

DOC #7

Figure B.7: Mário de Sá Carneiro - Doc 7

Mário Quintana
Alegrete, Rio Grande do Sul (RS), 30 de Julho de 1906 — Porto Alegre, Rio Grande do Sul (RS), 5 de Maio de 1994

In "Espelho Mágico", 1951

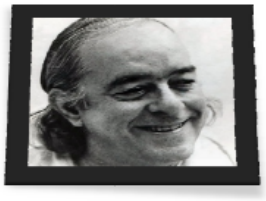
Das Utopias

Se as coisas são inatingíveis... ora!
Não é motivo para não querê-las...
Que tristes os caminhos se não fora
A mágica presença das estrelas!

DOC #8

Figure B.8: Mário Quintana - Doc 8

Vinicius de Moraes
Rio de Janeiro (RJ), 19 de Outubro de 1913 — Rio de Janeiro (RJ),
9 de Julho de 1980



In "Antologia Poética", 1954

TERNURA

Eu te peço perdão por te amar de repente
Embora o meu amor seja uma velha canção nos teus ouvidos
Das horas que passei à sombra dos teus gestos
Bebendo em tua boca o perfume dos sorrisos
Das noites que vivi acalentado
Pela graça indizível dos teus passos eternamente fugindo
Trago a doçura dos que aceitam melancolicamente.
E posso te dizer que o grande afeto que te deixo
Não traz o exaspero das lágrimas nem a fascinação das promessas
Nem as misteriosas palavras dos véus da alma…
É um sossego, uma unção, um transbordamento de carícias
E só te pede que repouses quieta, muito quieta
E deixes que as mãos cálidas da noite encontrarem sem fatalidade
[o olhar extático da aurora.

DOC #9

Figure B.9: Vinicius de Moraes - Doc 9

C IdSay Evaluation at QA@CLEF 2008 Evaluation Campaign (Portuguese)

C.1 Question Set

Table C.1: Questions: Part 1 of 5 (Questions 1-40)

Q.#	**C.#**	**Question**
1	2600	Que animal é o Cocas?
2	2601	Quem foi o criador de Tintin?
3	2601	Quando é que ele foi criado?
4	2601	Como se chama o cão dele?
5	2601	De que raça é o cão?
6	2602	Diga uma escola de samba fundada nos anos 40.
7	2603	Em que ano houve um terramoto no Irão?
8	2604	Quanto pesa um beija-flor?
9	2605	Onde ficava a Gália Cisalpina?
10	2606	Quantas províncias tem a Catalunha?
11	2607	Qual é a montanha mais alta do México?
12	2607	E do Japão?
13	2608	Onde fica Saint-Exupéry?
14	2609	Qual a altura do Kebnekaise?
15	2610	Quem escreveu Fernão Capelo Gaivota?
16	2611	O que é um menir?
17	2612	Em que ano é que Ernie Els venceu o Dubai Open?
18	2613	Quantos ossos têm a face?
19	2614	Quando começou o Neolítico?
20	2615	Quando nasceu Thomas Mann?
21	2615	E quando morreu?
22	2616	A que partido pertence Zapatero?
23	2617	Quem é FHC?
24	2618	Quem foi Álvaro de Campos?
25	2618	Diga uma das suas obras.
26	2619	Os tucanos são membros de que partido?
27	2620	Quem foi Pierre Larousse?
28	2621	O que é a Brabançonne?
29	2622	Onde vivem as aves tucanos?
30	2623	Em que estado brasileiro habitam os tucanos?
31	2624	Quem disse ”alea iacta est”?
32	2624	Ao atravessar que rio?
33	2625	Que político é conhecido como Iznogoud?
34	2626	O que é uma cítara?
35	2627	O que era o A6M Zero?
36	2628	Diga um gás nobre.
37	2628	E um não-metal
38	2629	Qual é o asteróide número 4?
39	2630	Qual a capital da Picardia?
40	2631	Quando reinou Isabel II de Castela?

Table C.2: Questions: Part 2 of 5 (Questions 41-80)

Q.#	C.#	Question
41	2632	Quem é Narcís Serra?
42	2633	Como se chamava o cavalo do Dom Quixote?
43	2634	Qual é a capital do estado de Nova York?
44	2635	Quais são as províncias da Irlanda?
45	2636	Que instrumento tocava Ringo Starr?
46	2637	Que papa sucedeu a Leão X?
47	2638	Quem é o pato mais rico do mundo?
48	2639	Quem são os sobrinhos do Pato Donald?
49	2639	E a namorada dele?
50	2639	Qual a profissão dele?
51	2640	O que é a paella?
52	2641	Que países abrange a Lapónia?
53	2642	O que é a açorda?
54	2643	O que é o feta?
55	2643	De que país é originário?
56	2644	Em que ano foi construída a sinagoga de Curação?
57	2645	Com que idade o Mequinho foi campeão brasileiro de xadrez?
58	2646	Quem dirigiu o Japão durante a Segunda Guerra Mundial?
59	2647	Quantas repúblicas formavam a URSS?
60	2648	Em que país fica a Ossétia do Norte?
61	2648	E a Ossétia do Sul?
62	2649	Qual a largura do Canal da Mancha no seu ponto mais estreito?
63	2650	Quem criou Descobridores de Catan?
64	2651	Quem é o santo patrono dos cervejeiros?
65	2651	E do pão?
66	2652	O que é o jagertee?
67	2653	Qual a envergadura de um milhafre-preto?
68	2653	Quanto é que ele pesa?
69	2653	Que tipo de ave é?
70	2654	Quantas províncias tem a Ucrânia?
71	2655	Que partido foi fundado por Amílcar Cabral?
72	2656	Quantos filhos teve a rainha Cristina da Suécia?
73	2657	Quem é o dono do Chelsea?
74	2658	Quantos habitantes tinha Berlim em 1850?
75	2658	Quantos tem hoje em dia?
76	2659	O que é o ICCROM?
77	2659	Quantos estados membros tinha em 1995?
78	2659	Onde tem a sua sede?
79	2660	Quantas vezes ganhou Portugal a Taça Davis?
80	2661	O que é o IPM em Portugal?

Table C.3: Questions: Part 3 of 5 (Questions 81-120)

Q.#	**C.#**	**Question**
81	2662	Quem foi o último rei de Portugal?
82	2662	Em que período foi ele rei?
83	2662	Em que barco ele embarcou para o exílio?
84	2663	Diga uma batalha ocorrida durante a Guerra dos Cem Anos
85	2664	Quantos votos teve o Lula nas eleições presidenciais de 2002?
86	2664	Quando é que ele tomou posse?
87	2665	Quem era o pai de Carlomano?
88	2666	Quem foi Baden Powell de Aquino?
89	2667	Quem escreveu o Livro da Selva?
90	2667	Quem é a personagem principal do livro?
91	2668	Em que ilha fica Sapporo?
92	2669	Quem fundou a escola estóica?
93	2670	Quais são as regiões da Bélgica?
94	2671	Qual é o 31º estado dos Estados Unidos?
95	2671	E o 37º?
96	2672	O que era a RSFSR?
97	2673	Quantos atletas participaram nos Jogos Olímpicos de 1976?
98	2673	Em que país se realizaram?
99	2673	E em que cidade?
100	2674	O que é um berimbau?
101	2675	Que países fazem fronteira com a Itália?
102	2676	Como se chama o xadrez japonês?
103	2677	Qual é a temperatura do zero absoluto?
104	2678	Quem era a deusa da sabedoria?
105	2679	Que rio banha Paris?
106	2680	Qual o comprimento do Spree?
107	2681	Qual é a capital do Cazaquistão?
108	2681	E a sua maior cidade?
109	2682	Quem é o actual presidente da Guatemala?
110	2682	Qual era o cargo dele em 1991?
111	2683	Quantas faixas tem a bandeira dos Estados Unidos?
112	2684	Quais as cores da bandeira da Hungria?
113	2685	Quando ocorreu a batalha de Torres Vedras?
114	2686	Quem é o papa dos Infiéis?
115	2687	O que é VRML?
116	2688	Onde está a Arca da Aliança?
117	2689	Como se chamava o Huambo durante a era colonial?
118	2690	Qual é a língua oficial do Egipto?
119	2691	Quais os submarinos da Marinha Brasileira?
120	2692	Em que guerra combateu Joana de Arc?

Table C.4: Questions: Part 4 of 5 (Questions 121-160)

Q.#	**C.#**	**Question**
121	2692	Onde é que ela foi queimada?
122	2692	Quando?
123	2692	Que idade tinha ela?
124	2693	Desde quando está Fidel Castro no poder?
125	2693	Quando é que ele nasceu?
126	2693	Quem é o irmão dele?
127	2694	O que são os forcados?
128	2695	Quando foi assinado o Tratado de Zamora?
129	2696	O que é o fogo de São Telmo?
130	2697	O que é que é um brigadeiro?
131	2698	Quem inventou o forno de microondas?
132	2699	Qual a nacionalidade de Nicole Kidman?
133	2700	Quem patenteou o primeiro telégrafo sem fios?
134	2701	Qual é a companhia francesa de caminhos-de-ferro ?
135	2702	O que é a Feplam?
136	2703	Qual a dotação do Prémio Cervantes?
137	2703	Quem é que ganhou o prémio em 1994?
138	2704	Quem são os co-príncipes de Andorra?
139	2705	Que tipo de tecido é o damasco?
140	2706	Quantos jogadores tem uma equipa de voleibol?
141	2707	Quando é que viveu Zenão de Eleia?
142	2708	Qual é a área da Groenlândia?
143	2709	Quem foi a primeira mulher no espaço?
144	2709	E a segunda?
145	2710	Diga um jornal libanês.
146	2711	Quantos refugiados haitianos estão na base de Guantanamo?
147	2712	Quando foi fundado o Vasco da Gama?
148	2712	Por quem foi fundado?
149	2713	Quando nasceu Vasco da Gama?
150	2713	Onde é que ele morreu?
151	2714	Em que distrito fica Sines?
152	2715	Qual é a capital de Dublin?
153	2716	Em que ano é que Halle Berry venceu o Óscar?
154	2717	Por que estados corre o Havel?
155	2718	Diga um escritor irlandês.
156	2719	Quem foi Carl Barks?
157	2719	Onde é que ele nasceu?
158	2719	Quem eram os pais dele?
159	2720	O que é um kilt?
160	2721	Quem realizou «Os Pássaros» ?

Table C.5: Questions: Part 5 of 5 (Questions 161-200)

Q.#	C.#	Question
161	2722	Quantos filmes realizou Jean Vigo?
162	2722	Diga um desses filmes.
163	2723	Qual o comprimento da Ponte do Øresund?
164	2724	Que companhia está baseada no Aeroporto Ben Gurion?
165	2725	Que navio americano foi afundado em Pearl Harbor in 1941?
166	2725	E que navio japonês?
167	2726	O que é o Crescente Fértil?
168	2727	Diga um clube de futebol de Campinas.
169	2727	E um de Belo Horizonte.
170	2728	Qual a capital do Mato Grosso?
171	2729	Quem foi o oitavo marido de Elizabeth Taylor?
172	2729	Quando é que eles se casaram?
173	2729	Qual é a nacionalidade dela?
174	2730	Quantos gêneros tem o alemão?
175	2730	E quantos tem o romanche?
176	2731	Quanto tempo reinou Ramsés II?
177	2731	Quando começou o seu reinado?
178	2731	Ele ordenou a construção de que templos?
179	2732	Que se passou a 9 de Novembro de 1991?
180	2733	Quantos actos tem a ópera Verdi da Aida?
181	2733	Quem escreveu o libretto dessa ópera?
182	2733	Quando é que estreou a ópera?
183	2734	Quem se tornou lider do Partido Quebequense em 2005?
184	2735	Qual é a maior cidade do Canadá?
185	2736	O que é o Gil Vicente FC?
186	2737	Quem foi Gil Vicente?
187	2738	Quem foi o ”pai do teatro português”?
188	2739	Qual a área do Parque Estadual Guariba?
189	2739	Quando foi criado o parque?
190	2740	O que é a Torre do Tombo?
191	2740	Onde fica?
192	2741	Que país faz fronteira com Cuba?
193	2742	Qual é o comprimento do metro de Coimbra?
194	2743	Quantas esposas tinha Ngungunhane?
195	2743	Como é que se chamava o filho dele?
196	2744	Qual é a capital de Cuba?
197	2745	Quem criou o primeiro alfabeto?
198	2746	Quando é que Porto Rico se tornou um estados dos EUA?
199	2747	Onde fica Livorno?
200	2748	O que são os iaques?

C.2 IdSay Answers

Table C.6: IdSay Answers and Support: Part 1 of 2 (Questions 1-100)

A#: Answer	Support
Question #1 - Que animal é o Cocas?	
língua	**pt/g/f/s/GFS_Marketplace_400_ca3b.html**: língua quíchua . vários vocábulos originais da língua entraram nas línguas modernas através do espanhol , tais como coca ,
mão	**FSP950706-070**: coca - cola na mão .
Question #2 - Quem foi o criador de Tintin?	
hergé	**pt/2/2/_/22_de_Maio_4855.html**: 22 de maio . 1 907 * * hergé , criador de histórias em quadrinhos , como tintin .
tibete	**PUBLICO-19940728-127**: é assim que o criador de tintin recorda esse crítico ano de 1 958 . nessa época , o autor dava forma , com grande esforço , a « tintin no tibete » ,
dava forma	**PUBLICO-19940728-127**: é assim que o criador de tintin recorda esse crítico ano de 1 958 . nessa época , o autor dava forma , com grande esforço , a « tintin no tibete » ,
Question #3 - Quando é que ele foi criado?	
1 929	**FSP951214-158**: a coleção completa de tintin , o repórter criado pelo belga hergé em 1 929 , pode ser encontrada nas livrarias de pokhara a us $ 2 5 o exemplar da série .
17 de novembro de 1 954	**pt/d/a/n/Dan_Cooper_4ff2.html**: dan cooper é um personagem de banda desenhada , piloto da força aérea canadiana , criado por albert weinberg . a sua primeira aparição ocorreu em 17 de novembro de 1 954 no semanário belga tintin ,
10 de janeiro de 1 929	**pt/t/i/n/Tintin.html**: tintim (ou tintin , no original em francês) é o protagonista da série de ficção de banda desenhada conhecida como as aventuras de tintim (les aventures de tintin , no original) , criado pelo quadrinista belga conhecido como hergé em 10 de janeiro de 1 929 .

Question #4 - Como se chama o cão dele?	
sempre acompanhado	**FSP950903-176**: tintin é um jovem repórter que viaja pelo mundo solucionando mistérios , sempre acompanhado por seu cão , milou , e pelo amigo capitão haddock .
jovem repórter que viaja pelo mundo solucionando mistérios	**FSP950903-176**: tintin é um jovem repórter que viaja pelo mundo solucionando mistérios , sempre acompanhado por seu cão , milou , e pelo amigo capitão haddock .
milou	**FSP950903-176**: tintin é um jovem repórter que viaja pelo mundo solucionando mistérios , sempre acompanhado por seu cão , milou , e pelo amigo capitão haddock .
Question #5 - De que raça é o cão?	
NIL	
Question #6 - Diga uma escola de samba fundada nos anos 40.	
gres estação primeira de mangueira	**pt/g/r/e/GRES_Estação_Primeira_de_Mangueira_5779.html**: gres estação primeira de mangueira . . aos poucos todos os outros blocos do morro foram se agregando e nos anos 30 e 40 , a mangueira já figurava no rol das " grandes " escolas de samba da cidade .
poucos todos os outros	**pt/g/r/e/GRES_Estação_Primeira_de_Mangueira_5779.html**: gres estação primeira de mangueira . . aos poucos todos os outros blocos do morro foram se agregando e nos anos 30 e 40 , a mangueira já figurava no rol das " grandes " escolas de samba da cidade .
rol	**pt/g/r/e/GRES_Estação_Primeira_de_Mangueira_5779.html**: gres estação primeira de mangueira . . aos poucos todos os outros blocos do morro foram se agregando e nos anos 30 e 40 , a mangueira já figurava no rol das " grandes " escolas de samba da cidade .
Question #7 - Em que ano houve um terramoto no Irão?	
16 de julho de 1 990	**PUBLICO-19950118-153**: richter no irão : 35 a 36 mil mortos . 16 de julho de 1 990 - - 1 641 pessoas morrem , 969 desaparecem e 3 441 ficam feridas em luçon , principal ilha das filipinas , após um terramoto de 7 7 graus .
1 991	**pt/a/b/b/Abbas_Kiarostami_1c17.html**: abbas kiarostami . foi um brilhante retrato do trágico terramoto que assolou o irão em 1 991 .
20 por cento	**PUBLICO-19940413-098**: algumas companhias poderão cair na insolvência devido ao terramoto e outras enfrentam prejuízos de vários milhões de dólares , de acordo com a a . m . best co . cerca de 20 por cento das reclamações de danos irão ser cobertos pela companhia « state farm » que anunciou prejuízos superiores a 1

Question #8 - Quanto pesa um beija-flor?	
1 g	**FSP951006-102**: entre os destaques da exposição , ovos de beija - flor (com apenas 11 milímetros de diâmetro e pesando menos de 1 g) e uma réplica perfeita de um ovo da extinta ave - elefante , com 35 cm de comprimento .
Question #9 - Onde ficava a Gália Cisalpina?	
itália	**pt/r/e/n/Renascença_italiana.html**: gália cisalpina renascimento dominação napoleónica risorgimento itália fascista itália republicana categoria : história da itália renascença italiana é como ficou conhecida a fase de
rio rubicão	**pt/r/i/o/Rio_Rubicão_fc5d.html**: rio rubicão . o rio ficou conhecido pelo fato de que o direito romano da época da república proibia qualquer general romano de atravessá - lo com suas tropas . o curso d ' água marcava então a divisa entre a província da gália cisalpina e
direito romano da época	**pt/r/i/o/Rio_Rubicão_fc5d.html**: rio rubicão . o rio ficou conhecido pelo fato de que o direito romano da época da república proibia qualquer general romano de atravessá - lo com suas tropas . o curso d ' água marcava então a divisa entre a província da gália cisalpina e
Question #10 - Quantas províncias tem a Catalunha?	
quatro províncias	**PUBLICO-19951121-045**: numa das quatro províncias , a que tem girona por capital , os socialistas catalães foram os mais votados . também neste caso parecem claras as transferências : os seis deputados que os socialistas perderam favoreceram a iniciativa pela catalunha ,
Question #11 - Qual é a montanha mais alta do México?	
pico de orizaba	**pt/p/i/c/Pico_de_Orizaba_8a80.html**: pico de orizaba pico de orizaba , a montanha mais alta do méxico elevação :
fiorde	**FSP940620-108**: no méxico . o fiorde mais longo - o braço do fiorde nordvest , no estreito de scoresby , na parte oriental da groenlândia , é considerado o maior do mundo com 313 km de extensão , do mar em direção à terra . um fiorde é um golfo estreito e profundo , entre montanhas altas .
aimé bonpland	**pt/a/i/m/Aimé_Bonpland_2f08.html**: aimé bonpland . a montanha mais alta da macaronésia . a estadia , embora curta , impressionou - os de tal maneira que dedicaram cerca de 60 páginas da descrição da sua viagem às canárias . poucos dias depois retomam a sua viagem com destino a havana e méxico ,
Question #12 - E do Japão?	
monte fuji	**pt/j/a/p/Japão.html**: a montanha mais alta do japão é o famoso monte fuji , com 3 776 m de altitude .

alpes japoneses	**pt/a/l/p/Alpes_japoneses.html**: alpes japoneses . as montanhas mais altas do japão ,
sobre as ilhas	**pt/j/a/p/Japão.html**: a montanha mais alta do japão é o famoso monte fuji , com 3 776 m de altitude . disputas territoriais . o japão reivindica a soberania sobre as ilhas etorofu ,
Question #13 - Onde fica Saint-Exupéry?	
puer aeternus	**pt/p/u/e/Puer_Aeternus_f588.html**: puer aeternus . saint - exupéry . mozart , como mostrado no filme amadeus , demonstra o aspecto padrão do arquétipo pueril . psicólogos analistas alegam que o arquétipo pueril pode levar a problemas psicológicos que ficam patentes pela manifesta imaturidade ,
guerra civil espanhola	**pt/g/u/e/Guerra_Civil_Espanhola_e3d2.html**: guerra civil espanhola . auden , os escritores franceses andré malraux e saint - exupéry e a matemática , católica de esquerda e ativista política , também francesa , simone weil . os governos da inglaterra e da França , anti - fascistas , optaram por ficar de
padrão do arquétipo	**pt/p/u/e/Puer_Aeternus_f588.html**: puer aeternus . saint - exupéry . mozart , como mostrado no filme amadeus , demonstra o aspecto padrão do arquétipo pueril . psicólogos analistas alegam que o arquétipo pueril pode levar a problemas psicológicos que ficam patentes pela manifesta imaturidade ,
Question #14 - Qual a altura do Kebnekaise?	
2 117 m	**pt/k/i/r/Kiruna.html**: o monte kebnekaise , no município de kiruna , é a montanha mais alta da suécia e tem 2 117 m de altitude .
2 103 metros	**pt/k/e/b/Kebnekaise.html**: o maciço do kebnekaise , que faz parte das montanhas escandinavas , tem dois picos , dos quais o mais a sul atinge 2 103 metros (ca .
150 quilómetros	**pt/k/e/b/Kebnekaise.html**: o kebnekaise situa - se na lapónia , a cerca de 150 quilómetros (ca .
Question #15 - Quem escreveu Fernão Capelo Gaivota?	
richard bach	**pt/r/i/c/Richard_Bach_daa5.html**: richard bach . . fernão capelo gaivota)
romance	**pt/g/a/i/Gaivota.html**: anton tchekhov fernão capelo gaivota , romance de richard bach categoria : desambiguação
admirável mundo novo	**PUBLICO-19950301-090**: para os outros , há agatha christie , aldous huxley (« o admirável mundo novo ») , richard bach (« fernão capelo gaivota ») , passada que está a fase dos chamados « livros para jovens » .

Question #16 - O que é um menir?	
menir - monumentos pré - históricos em pedras , cravadas verticalmente no solo (ortóstatos) , às vezes de tamanho bem elevado (megalito denominado menir) . a palavra menir foi adotada , através do francês , pelos arqueólogos do século xix com base nas palavras do bretão significando men = pedra e hir = longa (comparar com o gaélico : maen hir = pedra longa)	**pt/m/e/n/Menir.html**: menir - monumentos pré - históricos em pedras , cravadas verticalmente no solo (ortóstatos) , às vezes de tamanho bem elevado (megalito denominado menir) . a palavra menir foi adotada , através do francês , pelos arqueólogos do século xix com base nas palavras do bretão significando men = pedra e hir = longa (comparar com o gaélico : maen hir = pedra longa) . no bretão moderno usa - se a palavra peulvan .
Question #17 - Em que ano é que Ernie Els venceu o Dubai Open?	
20	**PUBLICO-19940131-013**: africano ernie els , de 24 anos de idade , uma das maiores esperanças do golfe mundial , venceu ontem o dubai open , terceira prova da temporada do circuito europeu de golfistas profissionais , dotada com cerca de 120 mil contos de prémios . els somou 20 pancadas abaixo do par (
1 108	**PUBLICO-19950104-013**: africano ernie els , que da 20 ª posição escalou até ao sexto lugar , depois de ter acumulado 1 108 pontos , mais quatro do que price . um ano inesquecível para o golfe africano , portanto . para além do seu triunfo no open dos eua , em junho , els venceu o dubai desert classic
24	**PUBLICO-19940131-013**: golfe els venceu no dubai o sul - africano ernie els , de 24 anos de idade , uma das maiores esperanças do golfe mundial , venceu ontem o dubai open ,
Question #18 - Quantos ossos têm a face?	
três ossos	**pt/o/u/v/Ouvido_médio.html**: ouvido médio . danos ao ramo horizontal durante uma cirurgia podem levar a uma paralisia parcial da face da pessoa . anatomia comparativa os mamíferos são os únicos que têm três ossos no ouvido .
um osso	**pt/o/u/v/Ouvido_médio.html**: ouvido médio . . . . . paralisia parcial da face da pessoa . anatomia comparativa os mamíferos são os únicos que têm três ossos no ouvido . a bigorna e o estribo desenvolvem de ossos da mandíbula , e permitem a melhor detecção do som . alguns mamíferos - como os gatos - tem um ouvido médio maior localizado em um osso bulboso e

Question #19 - Quando começou o Neolítico?	
3 500	**pt/d/u/n/Dundalk.html**: dundalk . . geografia história começou a ser habitada no período neolítico (3 500 a . c .) e mais tarde pelos os celtas (500 d .
500	**pt/d/u/n/Dundalk.html**: dundalk . . geografia história começou a ser habitada no período neolítico (3 500 a . c .) e mais tarde pelos os celtas (500 d .
1 941	**FSP940711-137**: comunista começou a desmoronar e , pouco depois , leningrado foi desrebatizada , voltando a se chamar são petersburgo . nos bálcãs , o futuro voltou ao passado , primeiro a 1 941 e , agora , a 1 912 ou mesmo antes . no cáucaso e na áfrica central o que está ocorrendo é um renascimento do neolítico ou
Question #20 - Quando nasceu Thomas Mann?	
5 de fevereiro de 1 943	**pt/m/i/c/Michael_Mann_53ac.html**: michael mann . michael thomas mann , cineasta estadunidense , nasceu na cidade de chicago , illinois , no dia 5 de fevereiro de 1 943 .
1 955	**PUBLICO-19940216-067**: thomas mann manteve intacta a glória do espírito alemão » , disse um dia françois mauriac . no limite , terá sido esse o drama do escritor , o drama que o manteria no exílio , para sempre - - da califórnia à cidade suíça de zurique , onde viria a morrer em 1 955 . nasceu a
90	**pt/t/h/o/Thomas_Mann_7f52.html**: notas biográficas filho do comerciante johann heinrich mann e da brasileira júlia da silva bruhns , nasceu em lübeck , uma cidade do norte da alemanha , onde mais de 90 % da população é protestante . a família de thomas mann detinha ali um negócio há várias gerações .
Question #21 - E quando morreu?	
1 955	**PUBLICO-19940216-067**: thomas mann manteve intacta a glória do espírito alemão » , disse um dia françois mauriac . no limite , terá sido esse o drama do escritor , o drama que o manteria no exílio , para sempre - - da califórnia à cidade suíça de zurique , onde viria a morrer em 1 955 . nasceu a
Question #22 - A que partido pertence Zapatero?	
NIL	
Question #23 - Quem é FHC?	
fhc	**FSP940714-030**: fhc .
plano	**FSP940731-039**: plano fhc (

governo	**FSP940713-020**: um governo fhc .
Question #24 - Quem foi Álvaro de Campos?	
álvaro de campos (1 890 - 1 935) é um dos heterónimos mais conhecidos de fernando pessoa . nascido em tavira , teve a educação de liceu comum de sua época , posteriormente foi para a escócia estudar engenharia mecânica , e depois engenharia naval	**pt/á/l/v/Álvaro_de_Campos_9b10.html**: álvaro de campos (1 890 - 1 935) é um dos heterónimos mais conhecidos de fernando pessoa . nascido em tavira , teve a educação de liceu comum de sua época , posteriormente foi para a escócia estudar engenharia mecânica , e depois engenharia naval . em férias fez uma viagem ao oriente onde escreveu o opiário . entre todos os heterónimos , campos foi o único a manifestar fases poéticas diferentes ao longo de sua obra . era um engenheiro de educação inglesa e origem portuguesa , mas sempre com a sensação de ser um estrangeiro em qualquer parte do mundo .
Question #25 - Diga uma das suas obras.	
fernando pessoa	**PUBLICO-19940126-129**: « saudação a walt whitman » , em « poesias de álvaro de campos » , nas obras completas de fernando pessoa (ática) .
alberto caeiro	**pt/f/e/r/Fernando_Pessoa_e0aa.html**: fernando pessoa . os três heterônimos mais conhecidos (e também aqueles com maior obra poética) foram álvaro de campos , ricardo reis e alberto caeiro .
ricardo reis	**pt/f/e/r/Fernando_Pessoa_e0aa.html**: fernando pessoa . . . . álvaro de campos e ricardo reis , sem contarmos ainda com o semi - heterónimo bernardo soares . a principal obra de
Question #26 - Os tucanos são membros de que partido?	
noite	**pt/p/a/r/Partido_da_Social_Democracia_Brasileira_4c41.html**: da sucursal de brasília membros do comando de campanha de fhc discutiam até a noite de ontem com o ministro - chefe da casa civil , henrique hargreaves , se o candidato tucano deveria ou
Question #27 - Quem foi Pierre Larousse?	
pierre athanase larousse (toucy , 23 de outubro de 1 817 - paris , 3 de janeiro de 1 875) foi um pedagogo , editor e enciclopedista francês	**pt/p/i/e/Pierre_Larousse_94be.html**: pierre athanase larousse (toucy , 23 de outubro de 1 817 - paris , 3 de janeiro de 1 875) foi um pedagogo , editor e enciclopedista francês . sua sepultura se encontra no cemitério de montparnasse . ligações externas business week online , november 11 , 2 002 , commentary : ' french publisher for sale : no foreigners , please ' (em inglês) categorias : !
Question #28 - O que é a Brabançonne?	
la brabançonne é o hino nacional da bélgica	**pt/h/i/n/Hino_nacional_da_Bélgica_4a7b.html**: la brabançonne é o hino nacional da bélgica .
land of the free	**pt/l/i/s/Lista_de_hinos_nacionais_e_regionais.html**: lista de hinos nacionais e regionais . la balanguera bangladesh amar shonar bangla barbados in plenty and in time of need país basco eusko abendaren ereserikia baviera gott mit dir du land der bayern bélgica la brabançonne belize land of the free benim l ' aube nouvelle bermudas god save the queen biafra land of the rising sun bielorrússia hino nacional da bielorrússia bolívia bolivianos ,

belga capital bruxelas	**pt/b/é/l/Bélgica.html**: bélgica . ” la brabançonne ” gentílico : belga capital bruxelas 50 ° 54 ’ 00 ” n 4 ° 32 ’ 00 ” e maior cidade bruxelas .
Question #29 - Onde vivem as aves tucanos?	
família	**pt/t/u/c/Tucano.html**: são designadas por tucano as aves da família ramphastidae que vivem nas florestas da américa central e américa do sul .
ramphastidae que vivem nas florestas da américa central	**pt/t/u/c/Tucano.html**: são designadas por tucano as aves da família ramphastidae que vivem nas florestas da américa central e américa do sul .
outras presas	**pt/r/a/m/Ramphastos.html**: ramphastos é o gênero de aves piciformes , família ramphastidae , que possui os tucanos sul - americanos , cuja principal caracterítica são seus enormes e coloridos bicos . todos eles se alimentam basicamente de frutas , mas podem comer insetos e outras presas . vivem normalmente em árvores ,
Question #30 - Em que estado brasileiro habitam os tucanos?	
NIL	
Question #31 - Quem disse ”alea iacta est”?	
rio rubicão	**pt/r/i/o/Rio_Rubicão_fc5d.html**: rio rubicão . segundo suetônio , césar teria então proferido a famosa frase alea iacta est (” a sorte está lançada ”) .
famosa frase	**pt/r/i/o/Rio_Rubicão_fc5d.html**: rio rubicão . segundo suetônio , césar teria então proferido a famosa frase alea iacta est (” a sorte está lançada ”) .
césar	**pt/r/i/o/Rio_Rubicão_fc5d.html**: rio rubicão . segundo suetônio , césar teria então proferido a famosa frase alea iacta est (” a sorte está lançada ”) .
Question #32 - Ao atravessar que rio?	
rio rubicão	**pt/r/i/o/Rio_Rubicão_fc5d.html**: rio rubicão . . segundo suetônio , césar teria então proferido a famosa frase alea iacta est (” a sorte está lançada ”) . o mesmo autor também descreve como césar parecia indeciso ao se aproximar do rio e atribui a decisão de atravessar a uma aparição sobrenatural .
Question #33 - Que político é conhecido como Iznogoud?	
nicolas sarkozy	**pt/i/z/n/Iznogoud.html**: politicos , como nicolas sarkozy , são chamado ” iznogoud ” por sua ambicão e pequenha tamanho .
Question #34 - O que é uma cítara?	
a cítara é um instrumento musical de várias cordas presas sobre um arco de madeira , com ou sem caixa de ressonância , que se tocavam com ambas as mãos	**pt/c/í/t/Cítara.html**: a cítara é um instrumento musical de várias cordas presas sobre um arco de madeira , com ou sem caixa de ressonância , que se tocavam com ambas as mãos . a lenda diz que o imperador nero queimou roma tocando uma cítara . composta por onze cordas de ressonância e sete que são tocadas , é muito leve , feita geralmente com duas cabaças , uma para o corpo e uma acoplada no braço do instrumento para servir apenas como ressonância . as cordas são feitas de cobre ou bronze . é afinada em quintas , entre os tons dó , dó # e ré .

Question #35 - O que era o A6M Zero?	
mitsubishi	**pt/m/i/t/Mitsubishi_A6M_Zero_8a24.html**: mitsubishi a 6 m zero
principal caça da marinha japonesa durante toda a segunda guerra mundial	**pt/m/i/t/Mitsubishi_A6M_Zero_8a24.html**: mitsubishi a 6 m zero o a 6 m zero foi o principal caça da marinha japonesa durante toda a segunda guerra mundial ,
corsair	**pt/f/6/f/F6F_Hellcat_6b43.html**: f 6 f hellcat . . ver também f 4 f wildcat f 4 u corsair a 6 m zero ki - 43 oscar categoria : aviões militares
Question #36 - Diga um gás nobre.	
hélio	**pt/h/é/l/Hélio.html**: hélio . . compostos dado que o hélio é um gás nobre ,
superfluidez	**pt/s/u/p/Superfluidez.html**: superfluidez . este último um gás nobre .
série química	**pt/u/n/u/Ununóctio.html**: ununóctio . 118 série química presumivelmente um gás nobre grupo ,
Question #37 - E um não-metal	
série química	**pt/f/ó/s/Fósforo.html**: fósforo . é um não - metal multi valente pertencente à série química do nitrogênio (
grupo	**pt/c/o/m/Composto_inorgânico.html**: composto inorgânico . compostos inorgânicos contêm metais ou hidrogênio combinado com um não - metal ou um grupo de não metais .
enxofre	**pt/e/n/x/Enxofre.html**: o enxofre , um não - metal insípido e
Question #38 - Qual é o asteróide número 4?	
apenas quando a sua órbita	**pt/l/i/s/Lista_de_asteróides.html**: lista de asteróides . . . como 4 179) . podem , opcionalmente , receber um nome também (como " toutatis ") . actualmente os asteróides recebem números sequencias apenas quando a sua órbita está documentada com precisão .
planeta anão	**pt/c/i/n/Cintura_de_asteróides.html**: cintura de asteróides . asteróides , e estima - se que o número alcance os milhões . cerca de 220 deles são maiores que 100 km . a massa total da cintura , contando com o planeta anão ceres é estimada em 3 0 3 6 predefinição : e , o que é cerca de 4 % da massa da
nome também	**pt/l/i/s/Lista_de_asteróides.html**: lista de asteróides . . . como 4 179) . podem , opcionalmente , receber um nome também (como " toutatis ") . actualmente os asteróides recebem números sequencias apenas quando a sua órbita está documentada com precisão .
Question #39 - Qual a capital da Picardia?	
amiens	**PUBLICO-19940506-109**: um pouco mais para sul , amiens capital da picardia foi a grande derrotada .

Question #40 - Quando reinou Isabel II de Castela?	
1 833	**pt/r/e/i/Reino_do_Algarve_afab.html**: reino do algarve . . . dado ter adquirido em 1 262 os restos do reino de niebla / algarve , situados já além do odiana - os demais reis de castela , e depois da espanha , até à subida ao trono da rainha isabel ii (1 833)
1 262	**pt/r/e/i/Reino_do_Algarve_afab.html**: reino do algarve . dado ter adquirido em 1 262 os restos do reino de niebla / algarve , situados já além do odiana - os demais reis de castela , e depois da espanha , até à subida ao trono da rainha isabel ii (
mar	**pt/r/e/i/Reino_do_Algarve_afab.html**: reino do algarve . . . . . castela , e depois da espanha , até à subida ao trono da rainha isabel ii (1 833) , continuaram a usá - lo entre os seus plúrices títulos . além - mar em portugal , o nome do reino algarvio (
Question #41 - Quem é Narcís Serra?	
ministro da defesa	**FSP950629-051**: narcís serra , e do ministro da defesa , julián garcía vargas .
deputado josé maria benegas	**PUBLICO-19950818-053**: narcís serra , e o deputado josé maria benegas .
julián garcía vargas	**FSP950629-051**: narcís serra , e do ministro da defesa , julián garcía vargas .
Question #42 - Como se chamava o cavalo do Dom Quixote?	
rocinante	**pt/r/o/c/Rocinante.html**: rocinante era o famoso cavalo de dom quixote de la mancha , personagem do romance de miguel de cervantes .
sancho pança	**pt/d/o/m/Dom_Quixote_9ddd.html**: as figuras de dom quixote , de sancho pança e do cavalo de dom quixote , rocinante , depressa conquistaram a imaginação popular .
diabo	**PUBLICO-19950102-077**: título : a cavalo no diabo autor : josé cardoso pires editor : dom quixote 206 pgs .
Question #43 - Qual é a capital do estado de Nova York?	
londres	**FSP941227-029**: para 31 dias (capital de giro) : entre 67 % e 78 % ao ano . no exterior feriados em nova york e em londres .
tóquio	**FSP950131-046**: em outras 11 capitais , de tóquio a nova york .
los angeles	**FSP940601-096**: em junho do ano que vem , esta exposição itinerante deverá ir a tóquio , e daí seguir para los angeles , chicago , nova york e washington . algumas capitais européias também estarão no circuito .
Question #44 - Quais são as províncias da Irlanda?	
condado condado de galway	**pt/b/a/l/Ballinasloe.html**: igreja de são joão , ballinasloe veja também lista de cidades na irlanda ballinasloe béal átha na slua província connacht condado condado de galway população (2 006) - pop .

norte	**pt/i/r/l/Irlanda_do_Norte_8715.html**: irlanda do norte . o ulster formou uma das províncias históricas da ilha da irlanda e consiste de 9 condados . três desses agora são parte da república da irlanda .
república	**pt/i/r/l/Irlanda_do_Norte_8715.html**: irlanda do norte . o ulster formou uma das províncias históricas da ilha da irlanda e consiste de 9 condados . três desses agora são parte da república da irlanda .
Question #45 - Que instrumento tocava Ringo Starr?	
sintetizador	**pt/g/e/o/George_Harrison_e25f.html**: george harrison . assim como george , ringo starr e eric clapton particparam do álbum usando pseudônimos . o álbum trazia somente composições próprias e as músicas eram instrumentais . george também produziu o álbum . eletronic sound é considerado um álbum experimental , várias músicas foram tocadas em sintetizador moog .
Question #46 - Que papa sucedeu a Leão X?	
concordata	**pt/j/o/ã/João_Calvino_0665.html**: joão calvino . sucedeu a luís xii . inicialmente moderado em matéria de religião , a postura deste rei foi endurecendo ao longo do seu reinado , terminando na perseguição declarada dos protestantes . pela concordata de bolonha , assinada no início do seu reinado , o papa leão x concedia ao rei da França o
Question #47 - Quem é o pato mais rico do mundo?	
a saga do tio patinhas	**pt/a/_/s/A_saga_do_Tio_Patinhas_b89c.html**: a saga do tio patinhas . . . ele se tornara o pato mais rico do mundo .
carl barks	**pt/t/i/o/Tio_Patinhas_0ad9.html**: tio patinhas . . . . . carl barks para construir uma extensa biografia do pato mais rico do mundo .
don rosa	**pt/p/a/t/Pato_Donald_cb07.html**: pato donald . . o tio patinhas , o pato mais rico do mundo ; professor pardal , o cientista maluco e gastão , o primo sortudo . outra figura que ajudou os quadrinhos do pato foi don rosa ,
Question #48 - Quem são os sobrinhos do Pato Donald?	
tio patinhas	**pt/t/i/o/Tio_Patinhas_0ad9.html**: tio patinhas . na história , patinhas convida seu sobrinho pato donald e
ducktales	**pt/m/a/g/Maga_Patalógika_d03d.html**: maga patalógika . . maga patalójika também era personagem semi - regular no seriado de animação ducktales , opondo - se a pato donald e seus sobrinhos quando não estão com patinhas .

maga patalógika	**pt/m/a/g/Maga_Patalógika_d03d.html**: maga patalógika . . maga patalójika também era personagem semi - regular no seriado de animação duck-tales , opondo - se a pato donald e seus sobrinhos quando não estão com patinhas .
Question #49 - E a namorada dele?	
margarida (banda desenhada)	**pt/m/a/r/Margarida_(banda_desenhada).html**: margarida (banda desenhada) . ela é a namorada do pato donald , embora o gastão também seja um pretendente .
superpato	**pt/s/u/p/Superpato.html**: superpato . ego é o conhecido pato donald . foi criado pela filial disney italiana . é uma paródia de batman , com todos os seus equipamentos inventados pelo professor pardal . poucos sabem sua identidade verdadeira . a namorada de
poucos sabem sua identidade	**pt/s/u/p/Superpato.html**: superpato . ego é o conhecido pato donald . foi criado pela filial disney italiana . é uma paródia de batman , com todos os seus equipamentos inventados pelo professor pardal . poucos sabem sua identidade verdadeira . a namorada de
Question #50 - Qual a profissão dele?	
chefe da segurança da loja	**pt/o/s/_/Os_Simpsons_5f6f.html**: os simpsons . nick riviera - o péssimo médico , que não entende nada da profissão e só tem clientela porque cobra mais barato . não suporta o dr . julius hibbert . don brodka - donald wilson brodka é o carrancudo chefe da segurança da loja trap -
brodka	**pt/o/s/_/Os_Simpsons_5f6f.html**: os simpsons . nick riviera - o péssimo médico , que não entende nada da profissão e só tem clientela porque cobra mais barato . não suporta o dr . julius hibbert . don brodka - donald wilson brodka é o carrancudo chefe da segurança da loja trap -
entende nada	**pt/o/s/_/Os_Simpsons_5f6f.html**: os simpsons . nick riviera - o péssimo médico , que não entende nada da profissão e só tem clientela porque cobra mais barato . não suporta o dr . julius hibbert . don brodka - donald wilson brodka é o carrancudo chefe da segurança da loja trap -
Question #51 - O que é a paella?	
llanito	**pt/l/l/a/Llanito.html**: llanito . . . paella tastes great llanito :
culinária de espanha	**pt/c/u/l/Culinária_de_Espanha_82e1.html**: culinária de espanha . internacionalmente , a paella ,
casa	**FSP951027-135**: a casa tem ainda a paella negra , com arroz , pimentão , tinta de lula e frutos do mar (r $ 50 00) .

Question #52 - Que países abrange a Lapónia?	
finlândia	**pt/m/u/r/Murmansk_(oblast).html**: murmansk (oblast) . . . geografia o oblast fica na península de kola e faz parte da lapónia , uma região que abrange quatro países . limita - se com a karelia , o condado de finnmark na noruega e a província da lapónia na finlândia .
noruega	**pt/m/u/r/Murmansk_(oblast).html**: murmansk (oblast) . . . geografia o oblast fica na península de kola e faz parte da lapónia , uma região que abrange quatro países . limita - se com a karelia , o condado de finnmark na noruega e a província da lapónia na finlândia .
Question #53 - O que é a açorda?	
açor	**PUBLICO-19950108-035**: raposo ; abreu ; acel ; acimol ; açor ; açorda ;
pap	**PUBLICO-19951226-103**: pap ' açorda .
arroz	**PUBLICO-19940423-074**: em açorda , carne de alguidar , borrego ou arroz de feijão .
Question #54 - O que é o feta?	
o feta é uma variedade de queijo típica da grécia , fabricado com leite de cabra ou de ovelha , exclusivamente . é um queijo branco , farelento e levemente salgado	**pt/f/e/t/Feta.html**: o feta é uma variedade de queijo típica da grécia , fabricado com leite de cabra ou de ovelha , exclusivamente . é um queijo branco , farelento e levemente salgado . ligações externas queijo feta (em inglês) categorias : !
Question #55 - De que país é originário?	
gouda	**pt/q/u/e/Queijo.html**: queijo . [1] tipos de queijos appenzeller brie camembert cheddar chèvre cottage edam feta gouda gorgonzola jarlsberg mussarela parmesão parmigiano -
grécia	**pt/f/e/t/Feta.html**: o feta é uma variedade de queijo típica da grécia , fabricado com leite de cabra ou de ovelha , exclusivamente .
Question #56 - Em que ano foi construída a sinagoga de Curação?	
NIL	
Question #57 - Com que idade o Mequinho foi campeão brasileiro de xadrez?	
13 anos	**FSP940205-144**: xadrez profissional) . mequinho foi campeão brasileiro aos 13 anos , em 1 965 , e campeão sul - americano dois anos depois .
dois anos	**FSP940205-144**: xadrez profissional) . mequinho foi campeão brasileiro aos 13 anos , em 1 965 , e campeão sul - americano dois anos depois .
14 anos	**pt/2/7/_/27_de_Dezembro_5328.html**: 27 de dezembro . o mequinho - torna - se aos 14 anos campeão brasileiro de xadrez 1 978 -

Question #58 - Quem dirigiu o Japão durante a Segunda Guerra Mundial?	
toshiro mifune	**pt/t/o/s/Toshiro_Mifune_6f4c.html**: toshiro mifune . que o dirigiu dezesseis vezes - e em produções americanas . mifune trabalhou como fotógrafo em xangai e serviu ao exército do japão durante a segunda guerra mundial .
serviu ao exército	**pt/t/o/s/Toshiro_Mifune_6f4c.html**: toshiro mifune . que o dirigiu dezesseis vezes - e em produções americanas . mifune trabalhou como fotógrafo em xangai e serviu ao exército do japão durante a segunda guerra mundial .
americanas	**pt/t/o/s/Toshiro_Mifune_6f4c.html**: toshiro mifune . que o dirigiu dezesseis vezes - e em produções americanas . mifune trabalhou como fotógrafo em xangai e serviu ao exército do japão durante a segunda guerra mundial .
Question #59 - Quantas repúblicas formavam a URSS?	
15 repúblicas	**FSP940911-062**: o balneário de stálin conflitos étnicos e caos econômico arrasam a geórgia do enviado especial à geórgia entre as 15 repúblicas que formavam a urss , a geórgia chamava a atenção por suas estações balneárias e por seus vinhos .
seis repúblicas	**pt/g/u/e/Guerra_da_Bósnia_96b6.html**: guerra da bósnia . nacionalismo com o fim dos regimes socialistas , a partir da desintegração da urss , emergem as diferenças étnicas , culturais e religiosas entre as seis repúblicas que formam a iugoslávia , impulsionando movimentos pela independência .
Question #60 - Em que país fica a Ossétia do Norte?	
geórgia	**pt/c/h/e/Chechênia.html**: geografia da geórgia . . . ossétia do norte - alania , inguchécia , chechénia , daguestão .
rússia	**pt/o/s/s/Ossétia_do_Norte-Alania_f13c.html**: repúblicas da rússia . ossétia do norte - alania 16 .
Question #61 - E a Ossétia do Sul?	
geórgia	**pt/g/e/ó/Geórgia.html**: geórgia . . ossétia do sul -
tskhinvali	**pt/t/s/k/Tskhinvali.html**: tskhinvali . cidade do sul da ossétia do sul , na geórgia .
república	**pt/o/s/s/Ossétia_do_Sul_0594.html**: ?????????? ?????? ??????? ?????????? ????? ?????? república da ossétia do sul (
Question #62 - Qual a largura do Canal da Mancha no seu ponto mais estreito?	
34 km	**pt/c/a/l/Calais.html**: calais está localizada no estreito de dover , no ponto mais estreito do canal da mancha com apenas 34 km de largura , sendo a cidade francesa mais próxima da inglaterra .

Question #63 - Quem criou Descobridores de Catan?	
jogo de tabuleiro	**pt/d/e/s/Descobridores_de_Catan_69e1.html**: descobridores de catan é um jogo de tabuleiro inventado por klaus teuber .
considerações	**pt/d/e/s/Descobridores_de_Catan_69e1.html**: considerações estratégicas tal como a secção anterior esta encontra - se em descobridores de catan , considerações estratégicas .
jogar a todos	**pt/d/e/s/Descobridores_de_Catan_69e1.html**: descobridores de catan . . produtos comerciais : descobridores de catan : o jogo standart , settlers of catan (1 995) , é requerido para jogar a todos os mapas .
Question #64 - Quem é o santo patrono dos cervejeiros?	
ele	**pt/a/r/n/Arnulfo_de_Metz_b694.html**: arnulfo de metz . . . . . ele é freqüentemente confundido nas lendas com arnoldo de soissons , que é outro santo patrono dos cervejeiros .
arnulfo de metz	**pt/a/r/n/Arnulfo_de_Metz_b694.html**: arnulfo de metz . arnulfo foi canonizado santo pela igreja católica romana e é conhecido como o santo patrono dos cervejeiros .
dia em uma dessa datas	**pt/a/r/n/Arnulfo_de_Metz_b694.html**: arnulfo de metz . . . . como o santo patrono dos cervejeiros . comemora - se seu dia em uma dessa datas : 18 de julho ou 16 de agosto . na iconografia , ele é retratado com um ancinho em sua mão . ele é freqüentemente confundido nas lendas com arnoldo de soissons , que é outro santo patrono dos cervejeiros .
Question #65 - E do pão?	
ele	**pt/a/r/n/Arnulfo_de_Metz_b694.html**: arnulfo de metz . . . . . ele é freqüentemente confundido nas lendas com arnoldo de soissons , que é outro santo patrono dos cervejeiros .
arnulfo de metz	**pt/a/r/n/Arnulfo_de_Metz_b694.html**: arnulfo de metz . arnulfo foi canonizado santo pela igreja católica romana e é conhecido como o santo patrono dos cervejeiros .
dia em uma dessa datas	**pt/a/r/n/Arnulfo_de_Metz_b694.html**: arnulfo de metz . . . . como o santo patrono dos cervejeiros . comemora - se seu dia em uma dessa datas : 18 de julho ou 16 de agosto . na iconografia , ele é retratado com um ancinho em sua mão . ele é freqüentemente confundido nas lendas com arnoldo de soissons , que é outro santo patrono dos cervejeiros .

Question #66 - O que é o jagertee?	
chá com adição de rum	**pt/c/h/á/Chá.html**: o jagertee é chá com adição de rum .
Question #67 - Qual a envergadura de um milhafre-preto?	
135 155 cm	**pt/m/i/l/Milhafre-preto.html**: o milhafre - preto mede cerca de 55 cm de comprimento e 135 155 cm de envergadura , para cerca de 1 kg de peso . a plumagem é de
55 cm	**pt/m/i/l/Milhafre-preto.html**: o milhafre - preto mede cerca de 55 cm de comprimento e 135 155 cm de envergadura , para cerca de 1 kg de peso . a plumagem é de
Question #68 - Quanto é que ele pesa?	
1 kg	**pt/m/i/l/Milhafre-preto.html**: o milhafre - preto mede cerca de 55 cm de comprimento e 135 155 cm de envergadura , para cerca de 1 kg de peso .
Question #69 - Que tipo de ave é?	
NIL	
Question #70 - Quantas províncias tem a Ucrânia?	
24 províncias	**pt/s/u/b/Subdivisões_da_Ucrânia_9457.html**: subdivisões da ucrânia . a ucrânia está subdividida em 24 províncias (óblasts) e em uma república autônoma (crimeia) , adicionalmente , duas cidades têm um status especial :
Question #71 - Que partido foi fundado por Amílcar Cabral?	
madrugada	**PUBLICO-19950223-068**: antonov continua no sal o avião de carga antonov que desde a madrugada de domingo se encontra no aeroporto internacional amílcar cabral ,
Question #72 - Quantos filhos teve a rainha Cristina da Suécia?	
NIL	
Question #73 - Quem é o dono do Chelsea?	
roman abramovich	**pt/r/o/m/Roman_Abramovich_efcb.html**: roman abramovich é um investidor russo , dono do clube inglês de futebol chelsea .
inglês de futebol	**pt/r/o/m/Roman_Abramovich_efcb.html**: roman abramovich é um investidor russo , dono do clube inglês de futebol chelsea .
britânico mais rico	**pt/r/o/m/Roman_Abramovich_efcb.html**: futebol chelsea . talvez por boris berezovski ser um asilado , e não um residente na grã - bretanha , seu antigo amigo e hoje rival roman abramovich é o residente britânico mais rico , dono de

Question #74 - Quantos habitantes tinha Berlim em 1850?	
300 000 habitantes	**pt/b/e/r/Berlim.html**: em 1 850 berlim já tinha 300 000 habitantes .
Question #75 - Quantos tem hoje em dia?	
dois anos	**pt/m/o/r/Moritz_Ludwig_Frankenheim_0e38.html**: richard dedekind . 1 850 com a idade de dezenove anos . seus principais orientadores foram moritz abraham stern (1 807 1 894) , gauss e wilhelm weber , o físico . deles recebeu uma completa base de cálculo , elementos de alta aritmética , alta geodésia , e física experimental . passou mais de dois anos em berlim ,
Question #76 - O que é o ICCROM?	
engenharia civil	**PUBLICO-19940601-014**: * doutor em engenheria civil , arc / iccrom * * geólogo * doutor em engenharia civil , arc / iccrom * * geólogo
doutor em engenheria	**PUBLICO-19940601-014**: * doutor em engenheria civil , arc / iccrom * * geólogo * doutor em engenharia civil , arc / iccrom * * geólogo
geólogo	**PUBLICO-19940601-014**: * doutor em engenheria civil , arc / iccrom * * geólogo * doutor em engenharia civil , arc / iccrom * * geólogo
Question #77 - Quantos estados membros tinha em 1995?	
91 estados	**PUBLICO-19951221-008**: financiado pelas contribuições anuais dos seus 91 estados membros , o iccrom é gerido por uma assembleia geral bienal cujos delegados examinam e aprovam o programa de funcionamento e respectivo orçamento .
Question #78 - Onde tem a sua sede?	
roma	**PUBLICO-19951221-008**: iccrom) , com sede em roma .
esta organização	**PUBLICO-19951221-008**: com sede em roma . esta organização , criada pela unesco em 1 956 , tem por missão reunir ou melhorar as condições que permitam a conservação dos bens culturais à escala mundial . financiado pelas contribuições anuais dos seus 91 estados membros , o iccrom é gerido por uma assembleia geral bienal
bienal	**PUBLICO-19951221-008**: com sede em roma . esta organização , criada pela unesco em 1 956 , tem por missão reunir ou melhorar as condições que permitam a conservação dos bens culturais à escala mundial . financiado pelas contribuições anuais dos seus 91 estados membros , o iccrom é gerido por uma assembleia geral bienal

Question #79 - Quantas vezes ganhou Portugal a Taça Davis?	
NIL	
Question #80 - O que é o IPM em Portugal?	
instituto português do património arquitectónico	**pt/i/n/s/Instituto_Português_do_Património_Arquitectónico_8a0f.html**: instituto português do património arquitectónico . . ipm) instituto português de conservação e restauro (ipcr) ligações externas site oficial categorias : instituições de portugal | arquitectura de portugal
miguel ângelo lupi	**pt/m/i/g/Miguel_Ângelo_Lupi_bf9d.html**: miguel ângelo lupi . ipm , 2 002 (isbn 972 776 124 0) . ligações externas miguel lupi no portugal - dicionário histórico gravura da obra d .
lista de museus	**pt/l/i/s/Lista_de_museus_de_Portugal_82f4.html**: lista de museus de portugal . ipm) tutela 29 museus (arte , arqueologia e etnologia) onde os visitantes podem encontrar muitas das peças incontornáveis do património de portugal .
Question #81 - Quem foi o último rei de Portugal?	
manuel de portugal	**pt/m/a/n/Manuel_de_Portugal_3bd4.html**: manuel de portugal . portugal , último rei de
ii	**pt/n/e/v/Nevada_Stoody_Hayes_fa10.html**: nevada stoody hayes . . manuel ii , último rei de portugal .
príncipe real	**pt/l/u/í/Luís_Filipe,_Príncipe_Real_de_Portugal_2c2f.html**: luís filipe , príncipe real de portugal . . manuel ii , e que viria a ser o último rei de portugal .
Question #82 - Em que período foi ele rei?	
religião	**pt/f/o/r/Forte_de_Âncora_3509.html**: henrique ii de frança . . seu fim trágico , inesperado , fará cair a frança no dramático período das guerras de religião . membro da dinastia de valois . rei de
moscóvia	**pt/g/u/e/Guerra_dos_Cem_Anos_973d.html**: moscóvia . o rei polonês wladyslaw iv , cujo pai e antecessor sigismundo iii foi eleito por boiardos russos como czar da rússia durante o período de
Question #83 - Em que barco ele embarcou para o exílio?	
baleeiro	**PUBLICO-19950710-102**: embarca num navio baleeiro e vai conhecer os mares do extremo sul .
Question #84 - Diga uma batalha ocorrida durante a Guerra dos Cem Anos	
iluminuras	**PUBLICO-19950104-092**: a recuperarem as iluminuras medievais das guerra dos cem anos de mistura com alguma epopeia à maneira de john ford .
ran	**PUBLICO-19950104-092**: fui poucas vezes ao cinema nos últimos dez anos . mas destacaria , dos filmes que vi , « ran » , de kurosawa com as suas golfadas de crueldade guerreira e colorido fortíssimo , a recuperarem as iluminuras medievais das guerra dos cem anos de

mundo moderno	**FSP950311-119**: mas no tempo da guerra dos cem anos a inglaterra e a frança eram dois países de civilização igual , pedras fundamentais sobre as quais se ergueria o que conhecemos como o mundo moderno .
Question #85 - Quantos votos teve o Lula nas eleições presidenciais de 2002?	
454 445 votos	**FSP940514-015**: brizola recebe 11 168 228 votos e fica em terceiro lugar na eleição presidencial . luiz inácio lula da silva (pt) tem 454 445 votos (0 5 %) a mais .
11 168 228 votos	**FSP940514-015**: brizola recebe 11 168 228 votos e fica em terceiro lugar na eleição presidencial . luiz inácio lula da silva (pt) tem 454 445 votos (0 5 %) a mais .
um voto	**FSP940622-014**: paulo - comparar as pesquisas do datafolha sobre a sucessão presidencial (publicada quinta - feira) e sobre a eleição em são paulo (divulgada sábado) equivale a enterrar a hipótese de um voto casado . voto casado seria o eleitor de , digamos , luiz inácio lula da silva (
Question #86 - Quando é que ele tomou posse?	
2 003	**pt/p/r/e/Presidente_do_Brasil_1a8f.html**: presidente do brasil . . 2 003 35 º luiz inácio lula da silva 2 003 - - atualidade * nota : os três presidentes eleitos mas que não tomaram posse não fazem parte da numeração , assim como as duas juntas militares .
janeiro de 2 003	**pt/g/o/v/Governo_do_Brasil_6c7b.html**: governo do brasil . luiz inácio lula da silva (partido dos trabalhadores) , como presidente e josé alencar (partido liberal) como vice - presidente da república . eles tomaram posse no dia 1 º de janeiro de 2 003 .
8 de junho de 2 006	**pt/m/a/r/Maria_Thereza_Rocha_de_Assis_Moura_b17f.html**: maria thereza rocha de assis moura . advogada há 26 anos , foi indicada pelo presidente da república , luiz inácio lula da silva , no dia 8 de junho de 2 006 , a partir de lista tríplice enviada pelo superior tribunal de justiça , stj . ela tomou posse dia 9 de
Question #87 - Quem era o pai de Carlomano?	
carlos martel	**pt/c/a/r/Carlomano,_filho_de_Carlos_Martel_e27c.html**: carlomano , filho de carlos martel . . com a morte de carlos em 741 , ele e seu meio irmão pepino o breve sucederam seu pai em seus cargos , pepino na nêustria e carlomano na austrásia .

luís iii de França	**pt/l/u/í/Luís_III_de_França_3a5f.html**: luís iii de França . quando o seu pai morreu , em 879 , luís tornou - se rei , juntamente com o seu irmão carlomano .
rei	**pt/l/u/í/Luís_III_de_França_3a5f.html**: luís iii de França . quando o seu pai morreu , em 879 , luís tornou - se rei , juntamente com o seu irmão carlomano .
Question #88 - Quem foi Baden Powell de Aquino?	
baden powell de aquino , violonista brasileiro , (varre - sai , rj , 6 de agosto de 1 937 - rio de janeiro , rj , 26 de setembro , de 2 000)	**pt/b/a/d/Baden_Powell_de_Aquino_e61d.html**: baden powell de aquino , violonista brasileiro , (varre - sai , rj , 6 de agosto de 1 937 - rio de janeiro , rj , 26 de setembro , de 2 000) . filho de dona adelina e do violinista lino de aquino , que deu - lhe esse nome por ser fã do criador do escotismo , general britânico robert stephenson smyth baden - powell . é pai do pianista e tecladista phillipe baden powell e do violonista louis marcel powell (ambos nascidos na França) e primo do violonista joão de aquino .
Question #89 - Quem escreveu o Livro da Selva?	
janete clair	**pt/j/a/n/Janete_Clair_effe.html**: janete clair . nos anos 70 escreve algumas de suas telenovelas de maior sucesso , como irmãos coragem (1 970 / 1 971) , selva de
rudyard kipling	**FSP950604-172**: rudyard kipling , autor dos ‘ ‘ livros da selva ” , que inspiraram o ‘ ‘ mogli ” de walt disney , não escrevia só literatura infanto - juvenil .
eugenia	**FSP950817-116**: a selva ” . maria eugenia é professora do departamento de teoria literária na universidade estadual de campinas (unicamp) . de fato , além de não escrever em linguagem acadêmica ,
Question #90 - Quem é a personagem principal do livro?	
nos estados unidos	**PUBLICO-19940215-099**: escreve henry roth no seu novo romance « mercy of a rude stream (vol . 1 : a star shines over mt . morris park) » , ao falar do adolescente ira stigman , personagem principal do livro publicado recentemente nos estados unidos .
novo romance	**PUBLICO-19940215-099**: escreve henry roth no seu novo romance « mercy of a rude stream (vol . 1 : a star shines over mt . morris park) » , ao falar do adolescente ira stigman , personagem principal do livro publicado recentemente nos estados unidos .
mercy	**PUBLICO-19940215-099**: escreve henry roth no seu novo romance « mercy of a rude stream (vol . 1 : a star shines over mt . morris park) » , ao falar do adolescente ira stigman , personagem principal do livro publicado recentemente nos estados unidos .

Question #91 - Em que ilha fica Sapporo?	
hokkaido	**FSP940214-034**: sapporo , capital da ilha de hokkaido . as mesmas cenas se repetiram no principal aeroporto internacional de tóquio , narita , onde ficaram presos 8 000 passageiros impedidos de embarcar .
Question #92 - Quem fundou a escola estóica?	
zenão de cítio	**pt/z/e/n/Zenão_de_Cítio_7891.html**: zenão de cítio . aos 42 anos , fundou a escola estóica , reunindo seus alunos sob os pórticos (em grego , " stoa ") de templos , mercados e ginásios .
stoa	**pt/z/e/n/Zenão_de_Cítio_7891.html**: zenão de cítio . aos 42 anos , fundou a escola estóica , reunindo seus alunos sob os pórticos (em grego , " stoa ") de templos , mercados e ginásios .
ginásios	**pt/z/e/n/Zenão_de_Cítio_7891.html**: zenão de cítio . aos 42 anos , fundou a escola estóica , reunindo seus alunos sob os pórticos (em grego , " stoa ") de templos , mercados e ginásios .
Question #93 - Quais são as regiões da Bélgica?	
flamenga	**pt/r/e/g/Região_flamenga.html**: a região flamenga , correntemente designada por flandres , é uma das três regiões autónomas da bélgica . juridicamente , todas as funções desta região administrativa são desempenhadas pela comunidade flamenga .
países baixos	**pt/h/i/s/História_dos_Países_Baixos_cba5.html**: história dos países baixos . a maior parte dos pequenos estados que existiam na região onde são actualmente a holanda e a bélgica foram finalmente unidos pelo duque da borgonha em 1 433 .
França	**PUBLICO-19951231-089**: revelam que o alentejo é a segunda região mais pobre da união europeia , seguida dos açores . • são desactivados os controlos fronteiriços entre a França , a alemanha , a bélgica ,
Question #94 - Qual é o 31º estado dos Estados Unidos?	
los angeles	**pt/l/o/s/Los_Angeles_983d.html**: los angeles . se o 31 º estado dos estados unidos .
califórnia	**pt/l/o/s/Los_Angeles_983d.html**: los angeles . . . . . deles , 1 485 576 nasceram na califórnia , 663 746 em outro estado americano , e 31 792 nasceram em um território dos estados unidos .
guerra civil americana	**pt/n/e/v/Nevada.html**: em 31 de outubro de 1 864 , o nevada foi elevada à categoria de estado dos estados unidos , durante a guerra civil americana .
Question #95 - E o 37º?	

nos jogos olímpicos de verão	**pt/e/s/g/Esgrima_nos_Jogos_Olímpicos_de_Verão_de_2004_0bf2.html**: esgrima nos jogos olímpicos de verão de 2 004 . 5 vladimir kalujny 11 15 hungria 37 - 31 estados unidos domonkos ferjancsik 5 -
países	**pt/l/i/s/Lista_de_países_por_PIB_nominal_87f5.html**: lista de países por pib nominal . 746 28 áfrica do sul 239 144 29 grécia 222 878 30 irlanda 199 722 31 irã 196 409 32 finlândia 193 491 33 portugal 183 436 34 argentina 181 662 35 hong kong 177 723 36 tailândia 168 774 37 emirados árabes unidos 133 768 38 venezuela 132 848 39 malásia 130 796 40 república tcheca 123 603
lista	**FSP940618-084**: do mundo se a lista da fifa definisse os 24 participantes da copa dos estados unidos . se esse fosse o critério , bulgária (29 º lugar) , grécia (31 º) , do grupo d , coréia do sul (37 º)
Question #96 - O que era a RSFSR?	
república oeste distante era localisada	**pt/r/e/p/República_Oeste_Distante_a43a.html**: a república oeste distante era localisada entre a rsfsr e o japão , por ser comandada pela rússia integrou - se a rsfsr .
likbez	**pt/l/i/k/Likbez.html**: likbez . . de acordo com o censo de 1 939 , os alfabetizados já eram 89 7 % (rsfsr , idades de 9 a 49 anos) .
sovnarkom da rsfsr sob o nome	**pt/g/o/s/Gosplan.html**: gosplan . o comitê foi criado em 22 de fevereiro de 1 921 por decreto do sovnarkom da rsfsr sob o nome de " comissão de planejamento estatal da rsfsr " .
Question #97 - Quantos atletas participaram nos Jogos Olímpicos de 1976?	
um atleta	**pt/c/a/r/Carlos_Lopes_b80c.html**: carlos lopes . viren era um atleta de excepção , e ganhou também o ouro nos 5 000 metros . era a primeira vez , desde há décadas , que portugal conquistava uma medalha olímpica , e a primeira vez no atletismo . palmarés 1 976 venceu o campeonato do mundo de
6 804 atletas	**pt/j/o/g/Jogos_Olímpicos_de_Verão_de_1976_f8f4.html**: jogos olímpicos de verão de 1 976 . os jogos olímpicos de montreal , no canadá , realizados entre 17 de julho e 1 e agosto de 1 976 , com a participação de 6 804 atletas de
Question #98 - Em que país se realizaram?	
portugal	**pt/c/a/r/Carlos_Lopes_b80c.html**: carlos lopes . viren era um atleta de excepção , e ganhou também o ouro nos 5 000 metros . era a primeira vez , desde há décadas , que portugal conquistava uma medalha olímpica , e a primeira vez no atletismo . palmarés 1 976 venceu o campeonato do mundo de

israel	**pt/e/h/u/Ehud_Barak_d32a.html**: ehud barak . atletas israelitas nos jogos olímpicos de munique no ano anterior . foram - lhe concedidos a medalha pelos " serviços distinguidos " e outras quatro condecorações pela bravura e eficácia operacional . sendo considerado o soldado mais condecorado da história de israel . em 1 976 fez o bacharelado em física e
suécia	**pt/j/o/g/Jogos_Paraolímpicos_9250.html**: os primeiros jogos para atletas com deficiências organizados à imagem dos jogos olímpicos realizaram - se em roma , em 1 960 , e ficaram conhecidos como jogos paraolímpicos . os primeiros jogos paraolímpicos de inverno realizaram - se em örnsköldsvik , na suécia , em 1 976 .
Question #99 - E em que cidade?	
los angeles	**pt/j/o/g/Jogos_da_Boa_Vontade_8f42.html**: jogos da boa vontade . pela primeira vez desde 1 976 , atletas americanos e soviéticos voltaram a se enfrentar numa grande competição . os estados unidos tinham boicotado os jogos olímpicos de 1 980 , em moscou , e os soviéticos fizeram o mesmo na olímpiada de 1 984 , em los angeles . para atrair atletas do atletismo ,
montreal	**pt/j/o/g/Jogos_Olímpicos_de_Verão_de_1976_f8f4.html**: jogos olímpicos de verão de 1 976 . os jogos olímpicos de montreal , no canadá , realizados entre 17 de julho e 1 e agosto de 1 976 , com a participação de 6 804 atletas de
tomar	**pt/t/i/r/Tiro_desportivo.html**: tiro desportivo . mais velho a se drogar nos jogos de 1 976 , o atleta paul cerutti foi desclassificado das provas de tiro por estar dopado . o mais curioso é que ele tinha terminado em 43 º lugar entre 44 competidores . cerutti entrou ainda para a história olímpica como o mais velho a tomar drogas .
Question #100 - O que é um berimbau?	
o berimbau é um instrumento de percussão usado tradicionalmente na capoeira , para marcar o ritmo da luta . no brasil é ainda conhecido pelos seguintes nomes : urucungo , urucurgo , orucungo , oricungo , uricungo , rucungo , ricungo , berimbau de barriga , gobo , marimbau , bucumbumba , bucumbunga , gunga , macungo , matungo , mutungo , aricongo , arco musical e rucumbo	**pt/b/e/r/Berimbau.html**: o berimbau é um instrumento de percussão usado tradicionalmente na capoeira , para marcar o ritmo da luta . no brasil é ainda conhecido pelos seguintes nomes : urucungo , urucurgo , orucungo , oricungo , uricungo , rucungo , ricungo , berimbau de barriga , gobo , marimbau , bucumbumba , bucumbunga , gunga , macungo , matungo , mutungo , aricongo , arco musical e rucumbo . no sul de moçambique , este instrumento tradicional tem o nome de xitende . o berimbau é constituido de um arco feito de uma vara de madeira de comprimento aproximado de 1 20 m e um fio de aço (arame) preso nas extremidades da vara .

Table C.7: IdSay Answers and Support: Part 1 of 2 (Questions 101-200)

A#: Answer	Support
Question #101 - Que países fazem fronteira com a Itália?	
eslovénia	**pt/p/r/o/Província_de_Trieste_8133.html**: fazendo fronteira a este com a eslovénia e a sudoeste com o golfo de trieste (no mar adriático) . ver também lista de comunas na província de trieste ligações externas site oficial categorias : ! esboços sobre geografia da itália | províncias da itália
luxemburgo	**pt/f/r/a/França.html**: a França funciona com um istmo que liga a península ibérica ao resto do continente , fazendo fronteira com a bélgica , luxemburgo , alemanha , suíça , itália , espanha , andorra e com o principado de mônaco .
suíça	**pt/f/r/a/França.html**: a França funciona com um istmo que liga a península ibérica ao resto do continente , fazendo fronteira com a bélgica , luxemburgo , alemanha , suíça , itália , espanha , andorra e com o principado de mônaco .
Question #102 - Como se chama o xadrez japonês?	
menos comum	**FSP950203-128**: o ton hoi serve os pratos comuns aos chineses paulistanos - como um leve frango - xadrez , um crocante rolinho primavera , um frango frito bem sequinho ou , menos comum , o saboroso kilza (pastel que os japoneses chamam guiosa , feito no vapor , cozido ou frito) .
calças	**FSP950208-118**: mesmo quando traz um look 60 que ele chama de " jeca future " : calças baixas , ternos com vinco ou calças de xadrez sintéticas misturados às t - shirts de desenhos animados japoneses .
desenhos animados	**FSP950208-118**: mesmo quando traz um look 60 que ele chama de " jeca future " : calças baixas , ternos com vinco ou calças de xadrez sintéticas misturados às t - shirts de desenhos animados japoneses .
Question #103 - Qual é a temperatura do zero absoluto?	
273 graus negativos	**PUBLICO-19940114-050**: existe em todos os materiais arrefecidos até perto do zero absoluto (273 graus negativos) , mas há décadas que os especialistas tentam obter materiais que sejam supercondutores a temperaturas mais altas .
0 01 ° c	**pt/k/e/l/Kelvin.html**: é definida por dois factos : zero kelvin é o zero absoluto (quando param os movimentos moleculares) , e um kelvin é a fracção 1 273 16 da temperatura termodinâmica do ponto triplo da água (0 01 ° c) .

269 ° c	**pt/n/u/v/Nuvem_de_Oort_8492.html**: a temperatura na nuvem de oort deve ser de - 269 ° c , ou seja 4 ° c acima do zero absoluto .
Question #104 - Quem era a deusa da sabedoria?	
saori kido	**pt/s/a/o/Saori_Kido_3b53.html**: saori kido . atena é a deusa da sabedoria , da guerra defensiva , da estratégia , da justiça e da esperança .
minerva	**pt/m/i/n/Minerva.html**: minerva . . . por ser também deusa da sabedoria .
mitologia grega	**pt/m/i/n/Minerva_McGonagall_79b5.html**: minerva mcgonagall . origem do nome minerva é o nome de uma deusa da mitologia romana que é a mesma que a deusa da sabedoria atena da mitologia grega .
Question #105 - Que rio banha Paris?	
piraí	**pt/r/i/o/Rio_Água_Branca_a352.html**: o rio piraí - mirim é um rio brasileiro que banha o estado do paraná .
arroio poço grande	**pt/a/r/r/Arroio_do_Ouro_9bd2.html**: o arroio poço grande é um rio brasileiro que banha o estado do paraná .
açu	**pt/r/i/o/Rio_Imbaú_e4fe.html**: o rio açu é um rio brasileiro que banha o estado do paraná .
Question #106 - Qual o comprimento do Spree?	
182 km	**pt/s/p/r/Spree.html**: do rio , os sórbios . geografia o spree tem um comprimento de cerca de 400 km , dos quais 182 km são navegáveis .
400 km	**pt/s/p/r/Spree.html**: do rio , os sórbios . geografia o spree tem um comprimento de cerca de 400 km , dos quais 182 km são navegáveis .
Question #107 - Qual é a capital do Cazaquistão?	
astana	**pt/a/s/t/Astana.html**: astana (em russo e cazaque : ??????) é a atual capital do cazaquistão .
ashkhabad	**pt/t/u/r/Turquemenistão.html**: turquemenistão . . capital : ashkhabad . história os primeiros habitantes do turquemenistão foram as tribos nômades turcas provenientes da região do atual cazaquistão desde o século x até o início do século xx .
moscovo	**PUBLICO-19951117-157**: moscovo com as capitais do « estrangeiro próximo » são , na maioria dos casos , de respeito mútuo e não ingerência exagerada nos seus assuntos internos . é opinião comum nos círculos do poder russo que é preferível , no próprio interesse de moscovo , que países como a ucrânia , a bielorrússia ou o cazaquistão se desenvolvam autonomamente

Question #108 - E a sua maior cidade?	
russo capital astana maior cidade almaty presidente nursultan nazarbayev	**pt/c/a/z/Cazaquistão.html**: sobre o presidente nazarbayev ????????? ???????????? ?????????? ????????? república do cazaquistão (detalhe) (detalhe) línguas oficiais cazaque e russo capital astana maior cidade almaty presidente nursultan nazarbayev primeiro -
línguas oficiais	**pt/c/a/z/Cazaquistão.html**: sobre o presidente nazarbayev ????????? ???????????? ?????????? ????????? república do cazaquistão (detalhe) (detalhe) línguas oficiais cazaque e russo capital astana maior cidade almaty presidente nursultan nazarbayev primeiro -
população	**pt/a/l/m/Almaty.html**: almaty (em cazaque : ??????) é a maior cidade do cazaquistão , com uma população de cerca de 1 185 900 (2 004) habitantes , ou seja , 8 % da população do país .
Question #109 - Quem é o actual presidente da Guatemala?	
óscar berger	**pt/ó/s/c/Óscar_Berger_e1f3.html**: óscar berger perdomo (nascido a 11 de agosto 1 946) é um político guatemalteco e atual presidente da guatemala .
prefeito da cidade da guatemala categorias	**pt/ó/s/c/Óscar_Berger_e1f3.html**: óscar berger . atual presidente da guatemala . é um advogado de profissão e entre 1 991 e 1 998 foi o prefeito da cidade da guatemala categorias : ! esboços de biografias | presidentes da guatemala
advogado de profissão	**pt/ó/s/c/Óscar_Berger_e1f3.html**: óscar berger . atual presidente da guatemala . é um advogado de profissão e entre 1 991 e 1 998 foi o prefeito da cidade da guatemala categorias : ! esboços de biografias | presidentes da guatemala
Question #110 - Qual era o cargo dele em 1991?	
partido social democrata	**FSP940815-038**: o partido social democrata indicou para o cargo sohei miyashita , que já chefiou a agência de defesa . eleição na guatemala tem 80 % de
nas eleições parlamentares	**FSP940815-038**: o partido social democrata indicou para o cargo sohei miyashita , que já chefiou a agência de defesa . eleição na guatemala tem 80 % de abstenção pode chegar a 80 % a abstenção nas eleições parlamentares da guatemala , encerradas na noite de ontem .
eleição	**FSP940815-038**: o partido social democrata indicou para o cargo sohei miyashita , que já chefiou a agência de defesa . eleição na guatemala tem 80 % de

Question #111 - Quantas faixas tem a bandeira dos Estados Unidos?	
13 faixas	**pt/b/a/n/Bandeira_dos_Estados_Unidos_da_América_b54f.html**: a bandeira dos estados unidos da américa consiste em 13 faixas horizontais , cujas cores são vermelho (que cobrem o topo e a parte de baixo da bandeira) alternando com branco .
cinco faixas	**pt/b/a/n/Bandeira_da_Índia_7f82.html**: bandeira da índia . usada nos estados unidos como símbolo para a índia por um curto período de tempo . bal gangadhar tilak e annie besant , principais representantes do movimento pela instalação de um governo próprio na índia (com o propósito de torná - la um país do commonwealth) adotaram uma nova bandeira que era composta de cinco faixas horizontais vermelhas
Question #112 - Quais as cores da bandeira da Hungria?	
áustria - hungria	**pt/á/u/s/Áustria-Hungria_8f5e.html**: áustria - hungria . armas bandeira da hungria coat of arms antes do compromisso de
igreja oficial igreja católica capital & amp	**pt/á/u/s/Áustria-Hungria_8f5e.html**: áustria - hungria . . armas bandeira da hungria coat of arms antes do compromisso de 1 867 bandeira do império habsburgo línguas oficiais alemão , húngaro igreja oficial igreja católica capital & maior cidade viena pop .
maior cidade	**pt/á/u/s/Áustria-Hungria_8f5e.html**: áustria - hungria . . armas bandeira da hungria coat of arms antes do compromisso de 1 867 bandeira do império habsburgo línguas oficiais alemão , húngaro igreja oficial igreja católica capital & maior cidade viena pop .
Question #113 - Quando ocorreu a batalha de Torres Vedras?	
22 de dezembro de 1 846	**pt/r/e/v/Revolução_da_Maria_da_Fonte_9220.html**: revolução da maria da fonte . com estas forças tentou avançar para sul , mas surpreendido pelas forças do marechal saldanha , retrocedeu sobre torres vedras . e foi naquela cidade que a 22 de dezembro de 1 846 , numa das batalhas decisivas da guerra , o brigadeiro josé lúcio travassos valdez , o 1 .
19 de agosto	**pt/b/a/t/Batalha_da_Roliça_2276.html**: batalha da roliça . na batalha de roliça , a 19 de agosto . o saldo do embate foi favorável aos britânicos , tendo o restante das tropas francesas se retirado para torres vedras , onde se uniu às tropas do general junot .

22 de dezembro . 1 758	**pt/2/2/_/22_de_Dezembro_12e3.html**: 22 de dezembro . 1 758 - fundação dos condados de schley , white e wilcox 1 761 - criação do ministério da fazenda 1 846 - na batalha de torres vedras josé travassos valdez é feito prisioneiro 1 904 -
Question #114 - Quem é o papa dos Infiéis?	
nicolau v	**pt/b/u/l/Bula.html**: bula . portugal para escravizar os infiéis da áfrica ocidental) romanus pontifex 1 455 - papa nicolau v (
áfrica ocidental	**pt/b/u/l/Bula.html**: bula . portugal para escravizar os infiéis da áfrica ocidental) romanus pontifex 1 455 - papa nicolau v (
ii	**pt/p/a/p/Papa_Urbano_II_e6e2.html**: " em 1 095 , o papa urbano ii convocou os cristãos a irem à terra santa expulsar os infiéis muçulmanos , nas chamadas cruzadas .
Question #115 - O que é VRML?	
vrml (virtual reality modelating language) , ou linguagem para modelagem de realidade virtual , é um padrão de aplicativos de realidade virtual utilizado na internet . por meio desta linguagem , escrita em modo texto , é possível criar objetos tridimensionais podendo definir cor , transparência , brilho , textura (associando - a a um bitmap)	**pt/v/r/m/VRML_a45e.html**: vrml (virtual reality modelating language) , ou linguagem para modelagem de realidade virtual , é um padrão de aplicativos de realidade virtual utilizado na internet . por meio desta linguagem , escrita em modo texto , é possível criar objetos tridimensionais podendo definir cor , transparência , brilho , textura (associando - a a um bitmap) . os objetos podem ser formas básicas , como esferas , cubos , ovóides , hexaedros , cones , cilindros , ou formas criadas pelo próprio programador , como as extrusões .
Question #116 - Onde está a Arca da Aliança?	
templo	**pt/h/i/s/História_Antiga_34b4.html**: história antiga . dentro do templo ficava a arca da aliança ,
querubim	**pt/q/u/e/Querubim.html**: querubim . um bebê alado que estava sobre o propiciatório da arca da aliança ,
bíblia	**pt/a/r/c/Arca_da_Aliança_fd7f.html**: construção a bíblia descreve a arca da aliança (
Question #117 - Como se chamava o Huambo durante a era colonial?	
próximos dias	**PUBLICO-19941023-047**: durante os próximos dias , no huambo (cidade que no tempo colonial se chamou nova lisboa)

cidade que no tempo	**PUBLICO-19941023-047**: durante os próximos dias , no huambo (cidade que no tempo colonial se chamou nova lisboa)
nova lisboa	**PUBLICO-19941023-047**: durante os próximos dias , no huambo (cidade que no tempo colonial se chamou nova lisboa)
Question #118 - Qual é a língua oficial do Egipto?	
árabe	**pt/l/i/s/Lista_de_estados_soberanos.html**: lista de estados soberanos . . oficial na língua do país egipto ou egito (bras .) - república árabe do egipto /
população total	**pt/d/e/m/Demografia_do_Egipto_9258.html**: demografia do egipto . idiomas a língua oficial do egito é árabe . o inglês e francês estão muito extendidos nas classes sociais mais altas . alfabetismo lêem e escrevem com mais de 15 anos 51 4 % da população total , sendo : homens : 63 6 % mulheres : 38 8 % . egipto história •
sudão	**pt/s/u/d/Sudão.html**: sudão . língua oficial árabe capital cartum presidente omar hasan ahmad al - bashir área - total - % água 10 . º maior 2 505 810 km ^{2} 5 % população - total - densidade 32 . º mais populoso 38 114 160 (est . julho de 2 003) 15 / km ^{2} independência - data do egipto
Question #119 - Quais os submarinos da Marinha Brasileira?	
história do brasil	**pt/h/i/s/História_do_Brasil_9420.html**: história do brasil . a decisão foi econômica : com a promessa dos eua em ajudar na construção de uma siderúgica - a csn - e após ataques submarinos a navios da marinha brasileira ,
após ataques submarinos a navios	**pt/h/i/s/História_do_Brasil_9420.html**: história do brasil . a decisão foi econômica : com a promessa dos eua em ajudar na construção de uma siderúgica - a csn - e após ataques submarinos a navios da marinha brasileira ,
força expedicionária	**pt/h/i/s/História_do_Brasil_9420.html**: história do brasil . . após ataques submarinos a navios da marinha brasileira , atribuídos a frota alemã , o brasil entrou na guerra em 1 942 ao lado dos aliados , enviando a força expedicionária brasileira (
Question #120 - Em que guerra combateu Joana de Arc?	
paz	**pt/j/o/a/Joana_d'Arc_9c58.html**: filipe iii , duque de borgonha . . . joana d ' arc lhe enviou uma carta no mesmo dia da consagração para lhe pedir paz .
cassandra	**FSP940623-073**: joana d ' arc , caterina sforza , catarina de médicis , isabel tudor , alessandra scala e cassandra fedele são algumas dessas mulheres .
Question #121 - Onde é que ela foi queimada?	
fogueira por bruxaria	**pt/1/4/3/1431.html**: 1 431 . joana d ' arc , de 19 anos , é queimada na fogueira por bruxaria .

aspectos controversos do catolicismo	**pt/a/s/p/Aspectos_controversos_do_Catolicismo_8b2d.html**: aspectos controversos do catolicismo . . casos como os das bruxas de salem , nos eua (país onde nunca houve a inquisição) e o de joana d ' arc heroína francesa queimada na inglaterra (outro país em que o tribunal do santo oficio jamais atuou) são exemplos de casos falsamente associados à inquisição .
história	**pt/j/o/a/Joana_d'Arc_9c58.html**: ser queimada viva . segundo a escritora irène kuhn , joana d ' arc foi esquecida pela história até ao século xix (
Question #122 - Quando?	
19 de abril . 1 839	**pt/1/9/_/19_de_Abril_f75e.html**: 19 de abril . 1 839 - a bélgica torna - se num reino através do tratado de londres 1 909 - joana de arc é canonizada .
1 909	**pt/1/9/_/19_de_Abril_f75e.html**: 19 de abril . 1 839 - a bélgica torna - se num reino através do tratado de londres 1 909 - joana de arc é canonizada .
Question #123 - Que idade tinha ela?	
dez anos	**pt/f/i/l/Filipe_III,_Duque_de_Borgonha_9d64.html**: filipe iii , duque de borgonha . à idade de dez anos . a 30 de maio de 1 431 , joana d ' arc , após ter sido julgada pela igreja , é queimada viva na praca da velha - marcha em rouen .
sete anos	**PUBLICO-19941203-102**: violência presenciadas em tenra idade , que nos perseguem até aos dias de hoje ? lembro - me de ter visto , aos sete anos , o notável filme de carl dreyer « joana d ' arc » ,
15 anos	**pt/g/u/e/Guerra_do_Contestado_0e08.html**: guerra do contestado . os fiéis que mudaram para caraguatá eram chefiadas por maria rosa , uma jovem com 15 anos de idade , considerada pelos historiadores como uma joana d ' arc do sertão ,
Question #124 - Desde quando está Fidel Castro no poder?	
1 969	**PUBLICO-19951007-046**: que se passa no seu próprio país » . a ideia é , portanto , que a informação produzida em cuba tenha um retorno , o que só pode desagradar a fidel castro . salvo ocasiões excepcionais , os órgãos de informação norte - americanos não operam na ilha desde 1 969 , quando os estados unidos ,
1 959	**FSP940825-042**: 125 quilômetros quadrados no extremo sul do território cubano . os estados unidos alugam esse terreno desde 1 934 , quando um acordo com o governo de cuba lhes deu tal direito pelo valor de us $ 4 085 anuais . desde que assumiu o poder em 1 959 , fidel castro só aceitou o pagamento do

1 964	**PUBLICO-19950607-066**: havana pode suavizar - se . cuba foi um dos fundadores da oea , em 1 948 , mas está suspensa desde 1 964 , por pressão dos estados unidos , que utilizaram essa sanção no seu braço de ferro com o regime de fidel castro .
Question #125 - Quando é que ele nasceu?	
13 de agosto de 1 926	**pt/b/i/r/Birán.html**: birán . este povoado tem uma inusual notoriedade por ser a localidade natal de fidel castro , que em 13 de agosto de 1 926 nasceu na propriedade rural de seu pai , o galego ángel castro argiz .
26 de julho	**pt/r/e/v/Revolução_Cubana_781d.html**: revolução cubana . . . . nasce o movimento revolucionário 26 de julho no mesmo ano , foi fundado o movimento revolucionário 26 de julho , constituído por fidel castro e
1 926	**pt/r/e/v/Revolução_Cubana_781d.html**: revolução cubana . . fileiras nasceu um movimento de novo tipo , encabeçado por fidel castro ruz (birán , 1 926) , um jovem advogado , formado pela faculdade de havana , cujas primeiras atividades políticas haviam se desenvolvido no meio universitário e às filas da ortodoxia . preconizando uma nova estratégia de luta armada contra a ditadura , fidel castro
Question #126 - Quem é o irmão dele?	
mary elizabeth donaldson	**pt/m/a/r/Mary_Elizabeth_Donaldson_f692.html**: mary elizabeth donaldson . ainda se sabe que mary se sentou entre o seu atual marido , príncipe frederik e o irmão dele , príncipe joachim . ela se relembra que eles começaram a conversar e simplesmente não pararam desde então .
ela se relembra que eles começaram	**pt/m/a/r/Mary_Elizabeth_Donaldson_f692.html**: mary elizabeth donaldson . ainda se sabe que mary se sentou entre o seu atual marido , príncipe frederik e o irmão dele , príncipe joachim . ela se relembra que eles começaram a conversar e simplesmente não pararam desde então .
princesa herdeira da dinamarca	**pt/m/a/r/Mary_Elizabeth,_Princesa_Herdeira_da_Dinamarca_bfbc.html**: mary elizabeth , princesa herdeira da dinamarca . ainda se sabe que mary se sentou entre o seu atual marido , príncipe frederik e o irmão dele , príncipe joachim . ela se relembra que eles começaram a conversar e simplesmente não pararam desde então .
Question #127 - O que são os forcados?	
forcarei é um município da espanha na província de pontevedra , comunidade autónoma da galiza , de área 168 70 km ^{2} com população de 4 616 habitantes (2 004) e densidade populacional de 27 36 hab / km ^{2}	**pt/f/o/r/Forcarei.html**: forcarei é um município da espanha na província de pontevedra , comunidade autónoma da galiza , de área 168 70 km ^{2} com população de 4 616 habitantes (2 004) e densidade populacional de 27 36 hab / km ^{2} . demografia variação demográfica do município entre 1 991 e 2 004 1 991 1 996 2 001 2 004 5 873 5 301 4 801 4 616 forcarei forcarei dados comunidade autónoma galiza província pontevedra área 168 70 km ^{2} população 4 616 hab .

Question #128 - Quando foi assinado o Tratado de Zamora?	
1 143	**pt/1/1/4/1143.html**: 1 143 . no tratado de zamora .
850	**PUBLICO-19941013-059**: nas comemorações do 850 º aniversário do tratado de zamora .
5 de outubro de 1 143	**pt/i/n/d/Independência_de_Portugal_f994.html**: só a 5 de outubro de 1 143 é reconhecida independência de portugal pelo rei afonso vii de castela , no tratado de zamora , assinando - se a paz definitiva .
Question #129 - O que é o fogo de São Telmo?	
fogo	**pt/f/o/g/Fogo_de_São_Telmo_6119.html**: fogo de são telmo
pedro gonzález telmo	**pt/p/e/d/Pedro_González_Telmo_b063.html**: pedro gonzález telmo . telmo . na iconogrtafia é representado vestido com o hábito branco e capa negra da ordem dominicana , levando na mão um círio azul , que representa o fogo de são telmo ,
efeitos da ionização	**pt/e/f/e/Efeitos_da_ionização.html**: efeitos da ionização . alo ver também ionização fogo de são telmo categorias : eletricidade | íons
Question #130 - O que é que é um brigadeiro?	
brigadeiro é uma patente militar de uso na aeronáutica ou força aérea , que designa a patente mais alta daquela força , equivalente a patente de general no exército	**pt/b/r/i/Brigadeiro.html**: brigadeiro é uma patente militar de uso na aeronáutica ou força aérea , que designa a patente mais alta daquela força , equivalente a patente de general no exército . categorias : !
Question #131 - Quem inventou o forno de microondas?	
percy spencer	**pt/p/e/r/Percy_Spencer_aacd.html**: percy spencer . . em inglês quem inventou o forno de microondas ?
inglês	**pt/p/e/r/Percy_Spencer_aacd.html**: percy spencer . . em inglês quem inventou o forno de microondas ?
industriais	**pt/c/u/l/Culinária.html**: e de alguns utensílios industriais , como as fritadeiras gigantes , devem ter sido inventadas as versões domésticas , mais pequenas ; já o fogão industrial é uma versão moderna e ampliada do fogão doméstico . o forno de microondas só foi possível com a revolução tecnológica . . . a restauração a culinária industrial a indústria alimentar passou
Question #132 - Qual a nacionalidade de Nicole Kidman?	
honolulu	**pt/n/i/c/Nicole_Kidman_310b.html**: com nicole kidman nicole kidman em cannes , em 2 001 nascimento 20 de junho de 1 967 honolulu , havaí nacionalidade norte -

nascimento	**pt/n/i/c/Nicole_Kidman_310b.html**: com nicole kidman nicole kidman em cannes , em 2 001 nascimento 20 de junho de 1 967 honolulu , havaí nacionalidade norte -
havaí nacionalidade norte	**pt/n/i/c/Nicole_Kidman_310b.html**: com nicole kidman nicole kidman em cannes , em 2 001 nascimento 20 de junho de 1 967 honolulu , havaí nacionalidade norte -
Question #133 - Quem patenteou o primeiro telégrafo sem fios?	
brasileiro roberto landell de moura patenteou em nova york	**FSP941122-066**: da reportagem local hoje faz 90 anos que o padre brasileiro roberto landell de moura patenteou em nova york o primeiro telégrafo sem fio .
levar adiante	**FSP941122-066**: moura patenteou em nova york o primeiro telégrafo sem fio . sem recursos e desestimulado pela ordem dos jesuítas , landell de moura não conseguiu levar adiante os seus projetos e a fama e fortuna provenientes do telégrafo sem fio ,
recursos	**FSP941122-066**: da reportagem local hoje faz 90 anos que o padre brasileiro roberto landell de moura patenteou em nova york o primeiro telégrafo sem fio . sem recursos e
Question #134 - Qual é a companhia francesa de caminhos-de-ferro ?	
korea train express	**pt/k/o/r/Korea_Train_Express_e1be.html**: korea train express . . . . . os carrís das linhas foram cosntruídos com a ajuda de técnicos da companhia francesa de caminhos - de - ferro sncf .
sncf	**PUBLICO-19950406-159**: caminhos - de - ferro da sncf (a companhia ferroviária francesa) , uma estação de táxis e um enorme parque de estacionamento .
comboio de alta velocidade	**pt/c/o/m/Comboio_de_alta_velocidade.html**: comboio de alta velocidade . a companhia francesa de caminhos - de - ferro , iniciando o plano de estudos a 1 966 e a construção em 1 976 .
Question #135 - O que é a Feplam?	
uma das pioneiras no ensino a distância no sul do país	**FSP940927-058**: a feplam é uma das pioneiras no ensino a distância no sul do país e
guaíba	**pt/g/u/a/Guaíba.html**: guaíba . porto alegre , março / abril / 97 ano ix - 23 , feplam , p .
porto alegre	**pt/g/u/a/Guaíba.html**: guaíba . porto alegre , março / abril / 97 ano ix - 23 , feplam , p .
Question #136 - Qual a dotação do Prémio Cervantes?	
18 mil contos	**PUBLICO-19941129-137**: o prémio cervantes , criado em 1 975 com uma dotação de 15 milhões de pesetas - - cerca de 18 mil contos - - foi atribuído no ano passado ao romancista miguel delibes .

Question #137 - Quem é que ganhou o prémio em 1994?	
ano passado ao romancista miguel	**PUBLICO-19941129-137**: o prémio cervantes , criado em 1 975 com uma dotação de 15 milhões de pesetas - - cerca de 18 mil contos - - foi atribuído no ano passado ao romancista miguel delibes .
contos	**PUBLICO-19941129-137**: o prémio cervantes , criado em 1 975 com uma dotação de 15 milhões de pesetas - - cerca de 18 mil contos - - foi atribuído no ano passado ao romancista miguel delibes .
pesetas	**PUBLICO-19941129-137**: o prémio cervantes , criado em 1 975 com uma dotação de 15 milhões de pesetas - - cerca de 18 mil contos - - foi atribuído no ano passado ao romancista miguel delibes .
Question #138 - Quem são os co-príncipes de Andorra?	
estado	**pt/g/o/v/Governo_de_Andorra_8c49.html**: governo de andorra . . . andorra) chefe de estado (co - príncipe episcopal)
principado	**pt/a/n/d/Andorra.html**: maior cidade andorra la vella língua oficial catalão ; espanhol e francês também é falado governo parlamentarismo co - principado - co - príncipe francês jacques chirac independência -
conselho geral dos vales	**pt/a/n/d/Andorra.html**: os chefes de estado , ou co - príncipes , são o presidente da república francesa e o bispo da comarca catalã de urgel . o chefe de governo é eleito pela maioria do conselho geral dos vales . os principais partidos políticos são o pla (partido liberal de andorra)
Question #139 - Que tipo de tecido é o damasco?	
religião	**pt/l/i/s/Lista_de_terremotos.html**: michel aflaq . notas biográfica educação francesa na sorbonne no anos 30 ele nasceu em damasco , numa família de religião cristã grega - ortodoxa , da classe média .
Question #140 - Quantos jogadores tem uma equipa de voleibol?	
dez jogadores	**PUBLICO-19950223-027**: no voleibol nacional , uma equipa faltou a um jogo . protagonista deste acto inédito , a académica de s . mamede , que deveria ter defrontado o nacional , no funchal , em jogo dos oitavos - de - final da taça de portugal . na origem desta ausência esteve a indisponibilidade de cinco dos dez jogadores do
Question #141 - Quando é que viveu Zenão de Eleia?	
NIL	
Question #142 - Qual é a área da Groenlândia?	
2 170 600 km 2	**FSP940620-108**: a oeste da groenlândia . ilha - a maior do mundo é a groenlândia , com área de 2 170 600 km 2 (6 270 vezes a ilha de são sebastião , litoral norte paulista , a maior do brasil) .

Question #143 - Quem foi a primeira mulher no espaço?	
valentina tereshkova	**pt/e/x/p/Exploração_espacial.html**: exploração espacial . . . a primeira mulher no espaço foi a russa valentina tereshkova (
vostok	**pt/s/o/y/Soyuz_T-7_1817.html**: soyuz t - 7 . svetlana savitskaya foi a primeira mulher no espaço desde valentina tereshkova (que vôou em 1 963 na vostok 6) .
programa espacial soviético considerou enviar	**pt/v/a/l/Valentina_Tereshkova_b8cd.html**: valentina tereshkova . no mesmo ano , o programa espacial soviético considerou enviar mulheres ao espaço , numa forma de colocar a primeira mulher no espaço e superar os estados unidos .
Question #144 - E a segunda?	
herói da união soviética	**pt/h/e/r/Herói_da_União_Soviética_d891.html**: herói da união soviética . caça soviética na segunda guerra mundial , a melhor ás feminina do mundo otto schmidt richard sorge valentina tereshkova - cosmonauta , primeira mulher no espaço anna yegorova -
guerra mundial	**pt/h/e/r/Herói_da_União_Soviética_d891.html**: herói da união soviética . caça soviética na segunda guerra mundial , a melhor ás feminina do mundo otto schmidt richard sorge valentina tereshkova - cosmonauta , primeira mulher no espaço anna yegorova -
caça	**pt/h/e/r/Herói_da_União_Soviética_d891.html**: herói da união soviética . caça soviética na segunda guerra mundial , a melhor ás feminina do mundo otto schmidt richard sorge valentina tereshkova - cosmonauta , primeira mulher no espaço anna yegorova -
Question #145 - Diga um jornal libanês.	
disse que sabia que iam assassiná	**PUBLICO-19940907-116**: entrevistado por um jornal libanês (em 1 979) , « carlos » , « o chacal » , disse que sabia que iam assassiná - lo um dia .
chacal	**PUBLICO-19940907-116**: entrevistado por um jornal libanês (em 1 979) , « carlos » , « o chacal » , disse que sabia que iam assassiná - lo um dia .
dia	**PUBLICO-19940907-116**: entrevistado por um jornal libanês (em 1 979) , « carlos » , « o chacal » , disse que sabia que iam assassiná - lo um dia .
Question #146 - Quantos refugiados haitianos estão na base de Guantanamo?	
dois mil refugiados	**PUBLICO-19940811-128**: são levados para a base de guantanamo , em cuba (onde se encontram mais de 15 mil) , para antigua , república dominicana e granada . estão também em curso negociações entre washington e o suriname para que este país da américa do sul aceite um total de dois mil refugiados haitianos .

16 mil refugiados	**PUBLICO-19940906-139**: a base militar dos eua em guantanamo , cuba , já abriga 16 mil refugiados haitianos . entidades de proteção a refugiados reclamaram que eles estão sendo mal tratados no local .
Question #147 - Quando foi fundado o Vasco da Gama?	
1 945	**pt/c/l/u/Clube_de_Futebol_Vasco_da_Gama_0bee.html**: história o clube foi fundado em 1 945 e o actual presidente é antónio galvão . ligas 2 005 2 006 - , o clube de futebol vasco da gama terminou a 1 ª divisão distrital da associação de
1	**pt/c/l/u/Clube_de_Futebol_Vasco_da_Gama_0bee.html**: história o clube foi fundado em 1 945 e o actual presidente é antónio galvão . ligas 2 005 2 006 - , o clube de futebol vasco da gama terminou a 1 ª divisão distrital da associação de
1 966	**pt/v/a/s/Vasco_da_Gama_Atlético_Clube_f3f9.html**: localização o vasco da gama atlético clube é um clube português , sedeado na cidade de sines , distrito de setúbal . história o clube foi fundado em 1 966 e
Question #148 - Por quem foi fundado?	
o condado	**pt/c/o/n/Condado_de_Klickitat_ac29.html**: condado de klickitat . o condado foi fundado em ?
esporte clube	**pt/e/s/p/Esporte_Clube_Mamoré_a129.html**: esporte clube mamoré . fundado em 13 de
história	**pt/h/i/s/História_da_Tunísia_cdea.html**: história da tunísia . haviam fundado ,
Question #149 - Quando nasceu Vasco da Gama?	
20 de dezembro de 1 995	**pt/a/f/o/Afonso_Maló_29d5.html**: afonso maló . o afonso nasceu no dia 20 de dezembro de 1 995 em lisboa . joga rugby no benfica e faz surf na lss . frequenta o colégio vasco da gama .
18 de maio de 1 978	**pt/h/e/l/Helton_da_Silva_Arruda_a74b.html**: helton da silva arruda , nasceu a 18 de maio de 1 978 no brasil , tendo passado pelo vasco da gama e união de leiria até assinar em 2 005 pelo fc porto .
1 959	**pt/a/s/s/Associação_Atlética_Portuguesa_(Santos)_e679.html**: associação atlética portuguesa (santos) . nascia assim , a " briosa " . história e pioneirismo a portuguesa é a dentetora da " fita azul " desde 1 959 - título criado para homenagear as equipes que faziam excursões ao exterior com sucesso . essa honra já foi do vasco da gama ,

Question #150 - Onde é que ele morreu?	
fausto dos santos	**pt/f/a/u/Fausto_dos_Santos_96f0.html**: fausto dos santos . . . morreu precocemente em razão da tuberculose . categorias : ! esboços sobre futebolistas | futebolistas do brasil | clube de regatas flamengo | atletas do club de regatas vasco da gama
flamengo	**PUBLICO-19950509-025**: brasil um homem morreu e três ficaram feridos , na sequência de incidentes entre adeptos do flamengo e do vasco da gama ,
nau	**pt/p/a/u/Paulo_da_Gama_34a0.html**: paulo da gama . angra 1 499) , irmão mais velho de vasco da gama , comandou a nau s . rafael quando acompanhava o seu mais novo irmão na rota marítima para a índia , mas veio a morrer no fim da viagem de
Question #151 - Em que distrito fica Sines?	
NIL	
Question #152 - Qual é a capital de Dublin?	
londres	**FSP940915-133**: free - lance para a folha comparada com outras capitais européias , como paris ou londres , dublin é uma cidade pequena .
belfast	**PUBLICO-19951129-169**: três dias que bill clinton hoje inicia à capital britânica , a belfast e a dublin .
atenas	**PUBLICO-19941207-008**: à semelhança do que já aconteceu em duas capitais europeias da cultura , atenas e dublin .
Question #153 - Em que ano é que Halle Berry venceu o Óscar?	
NIL	
Question #154 - Por que estados corre o Havel?	
saxônia	**pt/r/i/o/Rio_Havel_11cf.html**: rio havel . o havel é um rio que corre nos estados federais alemães de brandemburgo , berlim e saxônia - anhalt , alemanha .
berlim	**pt/r/i/o/Rio_Havel_11cf.html**: rio havel . o havel é um rio que corre nos estados federais alemães de brandemburgo , berlim e saxônia - anhalt , alemanha .
brandemburgo	**pt/r/i/o/Rio_Havel_11cf.html**: rio havel . o havel é um rio que corre nos estados federais alemães de brandemburgo , berlim e saxônia - anhalt , alemanha .
Question #155 - Diga um escritor irlandês.	
james joyce pelos críticos	**FSP951019-104**: folha - como escritor irlandês , o que acha de ser comparado a james joyce pelos críticos ? doyle - enquanto escrevo , não penso em mim como um escritor irlandês .

oliver goldsmith	**pt/o/l/i/Oliver_Goldsmith_eeac.html**: oliver goldsmith .) - 4 de abril de 1 774) foi um escritor irlandês .
penso	**FSP951019-104**: doyle - enquanto escrevo , não penso em mim como um escritor irlandês .
Question #156 - Quem foi Carl Barks?	
carl barks (27 de março de 1 901 - 25 de agosto de 2 000) foi um famoso ilustrador dos estúdios disney e criador de arte seqüencial , responsável pela invenção de patópolis e muitos de seus habitantes : tio patinhas (1 947) , gastão (1 948) , irmãos metralha (1 951) , professor pardal (1 952) , maga patalójika (1 961) e outros	**pt/c/a/r/Carl_Barks_0a4e.html**: carl barks (27 de março de 1 901 - 25 de agosto de 2 000) foi um famoso ilustrador dos estúdios disney e criador de arte seqüencial , responsável pela invenção de patópolis e muitos de seus habitantes : tio patinhas (1 947) , gastão (1 948) , irmãos metralha (1 951) , professor pardal (1 952) , maga patalójika (1 961) e outros . a qualidade de seus roteiros e desenhos o rendeu os apelidos o homem dos patos e o bom artista dos patos . o autor de quadrinhos will eisner o chamou de hans christian andersen dos quadrinhos .
Question #157 - Onde é que ele nasceu?	
ele tinha um irmão mais velho chamado	**pt/c/a/r/Carl_Barks_0a4e.html**: carl barks . biografia barks nasceu em merrill , oregon , filho de william barks e sua esposa arminta johnson . ele tinha um irmão mais velho chamado clyde . seu avô paterno se chamava david barks e seus avós maternos eram carl johnson e
avós maternos	**pt/c/a/r/Carl_Barks_0a4e.html**: carl barks . biografia barks nasceu em merrill , oregon , filho de william barks e sua esposa arminta johnson . ele tinha um irmão mais velho chamado clyde . seu avô paterno se chamava david barks e seus avós maternos eram carl johnson e
legou o pato donald	**FSP950722-096**: em especial carl barks " . barks foi o desenhista que legou o pato donald como o conhecemos e um dos maiores talentos que trabalharam com disney . quando nasceu ,
Question #158 - Quem eram os pais dele?	
tio patinhas	**pt/t/i/o/Tio_Patinhas_0ad9.html**: tio patinhas . mas geralmente não espera nenhuma gratidão deles . patinhas também expressou opinião de que só nos contos de fadas os maus se tornam bons , e que é velho demais para acreditar em contos de fadas . carl barks deu a patinhas uma ética definitiva consoante com a era em que construiu

expressou opinião de que só nos contos de fadas os maus	**pt/t/i/o/Tio_Patinhas_0ad9.html**: tio patinhas . mas geralmente não espera nenhuma gratidão deles . patinhas também expressou opinião de que só nos contos de fadas os maus se tornam bons , e que é velho demais para acreditar em contos de fadas . carl barks deu a patinhas uma ética definitiva consoante com a era em que construiu
tornam	**pt/t/i/o/Tio_Patinhas_0ad9.html**: tio patinhas . mas geralmente não espera nenhuma gratidão deles . patinhas também expressou opinião de que só nos contos de fadas os maus se tornam bons , e que é velho demais para acreditar em contos de fadas . carl barks deu a patinhas uma ética definitiva consoante com a era em que construiu
Question #159 - O que é um kilt?	
o kilt é o saiote pregueado , parcialmente trespassado , e quadriculado em cores correspondentes a cada clã ou família , e que faz parte do traje típico masculino da escócia	**pt/k/i/l/Kilt.html**: o kilt é o saiote pregueado , parcialmente trespassado , e quadriculado em cores correspondentes a cada clã ou família , e que faz parte do traje típico masculino da escócia . tradicionalmente ele era utilizado por guerreiros e batedores dos clãs , sendo que cada clã (ou clan) tinha um tartan diferente . categorias : !
Question #160 - Quem realizou «Os Pássaros»?	
ilhas selvagens	**pt/i/l/h/Ilhas_Selvagens_34a5.html**: ilhas selvagens . cientificas são realizadas anualmente nas ilhas . as selvagens têm 150 espécies de plantas , a maioria rasteiras . as ilhas mais ricas em flora são a ilha selvagem pequena e o ilhéu de fora porque nunca houve introdução de animais e plantas não indígenas . as ilhas tornaram - se conhecidas como um santuário para os pássaros :
one hundred and one dalmatians	**pt/o/n/e/One_Hundred_and_One_Dalmatians_fe8f.html**: one hundred and one dalmatians . foi um dos últimos desenhos animados realizados sob a supervisão de walt disney . no original em inglês , a voz do dálmata pongo é de rod taylor , ator australiano que trabalhou na adaptação de h . g . wells a máquina do tempo , de 1 961 , e em os pássaros ,
exposição individual no centro cultural português	**pt/g/r/a/Graça_Morais_915c.html**: graça morais . em maio de 1 978 , realiza uma exposição individual no centro cultural português em paris . " um elemento chave : o papel com que se decoravam as prateleiras da cozinha , cheio de histórias , o caçador , o cão , um leitão , a paisagem , os pássaros .

Question #161 - Quantos filmes realizou Jean Vigo?	
quatro filmes	**PUBLICO-19950206-099**: jean vigo apenas realizou quatro filmes mas o maremoto que constituiu a inventiva da sua obra do início dos anos 30 foi tão apaixonante como a sua trágica e
Question #162 - Diga um desses filmes.	
intervalos	**FSP940316-161**: o filme foi realizado durante os intervalos das filmagens de " hammett " . num desses intervalos , o cineasta rodou também " o filme de nick " , que registra a morte de nicholas ray , um dos ícones do cinema americano .
americano tom gunning	**FSP940405-136**: realizado em 1 913 pelo português radicado no brasil francisco santos . para amanhã , além da exibição desses mesmos filmes , está previsto um seminário internacional com o pesquisador norte - americano tom gunning .
morte de nicholas ray	**FSP940316-161**: o filme foi realizado durante os intervalos das filmagens de " hammett " . num desses intervalos , o cineasta rodou também " o filme de nick " , que registra a morte de nicholas ray , um dos ícones do cinema americano .
Question #163 - Qual o comprimento da Ponte do Øresund?	
18 quilómetros	**PUBLICO-19941028-134**: através do Øresund) , e outra entre as duas metades da dinamarca (separadas pelo storebælt) . ambas com 18 quilómetros , curiosamente também o comprimento da futura ponte de sacavém ao montijo .
Question #164 - Que companhia está baseada no Aeroporto Ben Gurion?	
NIL	
Question #165 - Que navio americano foi afundado em Pearl Harbor in 1941?	
couraçados	**pt/a/t/a/Ataque_a_Pearl_Harbor_e16e.html**: ataque a pearl harbor . . . . modo a este não afundar e bater no chão ao ser lançado do ar , podiam ser lançados mesmo até nas águas pouco profundas de pearl harbor ; facilitando assim , a utilização dos mesmos contra os couraçados norte - americanos ancorados .
Question #166 - E que navio japonês?	
couraçados	**pt/a/t/a/Ataques_aéreos_a_Darwin_ed53.html**: ataque a pearl harbor . . . . modo a este não afundar e bater no chão ao ser lançado do ar , podiam ser lançados mesmo até nas águas pouco profundas de pearl harbor ; facilitando assim , a utilização dos mesmos contra os couraçados norte - americanos ancorados .

Question #167 - O que é o Crescente Fértil?	
o crescente fértil é uma região do oriente médio compreendendo os atuais israel , cisjordânia e líbano bem como partes da jordânia , da síria , do iraque , do egito e do sudeste da turquia	**pt/c/r/e/Crescente_Fértil_01e7.html**: o crescente fértil é uma região do oriente médio compreendendo os atuais israel , cisjordânia e líbano bem como partes da jordânia , da síria , do iraque , do egito e do sudeste da turquia . o termo « crescente fértil » foi criado pelo arqueólogo james henry breasted , da universidade de chicago , em referência ao fato de o arco formado pelas diferentes zonas assemelhar - se a uma lua crescente . irrigada pelo jordão , pelo eufrates , pelo tigre e pelo nilo , a região cobre uma superfície de cerca de 400 000 a 500 000 km ^{2} e é povoada por 40 a 50 milhões de indivíduos .
Question #168 - Diga um clube de futebol de Campinas.	
são paulo	**pt/g/u/a/Guarani.html**: esportes guarani futebol clube , um clube de futebol de campinas , são paulo .
guarani	**pt/g/u/a/Guarani.html**: esportes guarani futebol clube , um clube de futebol de campinas , são paulo .
associação atlética ponte preta	**pt/a/s/s/Associação_Atlética_Ponte_Preta_bebd.html**: associação atlética ponte preta . futebol , situado no bairro da ponte preta , em campinas , estado de são paulo . o time foi fundado no dia 11 de agosto de 1 900 , sendo considerado o segundo mais antigo clube do brasil em atividade ininterrupta ,
Question #169 - E um de Belo Horizonte.	
minas gerais	**pt/c/r/u/Cruzeiro_Esporte_Clube_b654.html**: cruzeiro esporte clube é um clube de futebol brasileiro , com sede na cidade de belo horizonte , minas gerais .
brasileiro	**pt/c/r/u/Cruzeiro_Esporte_Clube_b654.html**: cruzeiro esporte clube é um clube de futebol brasileiro , com sede na cidade de belo horizonte , minas gerais .
sede na cidade	**pt/c/r/u/Cruzeiro_Esporte_Clube_b654.html**: cruzeiro esporte clube é um clube de futebol brasileiro , com sede na cidade de belo horizonte , minas gerais .
Question #170 - Qual a capital do Mato Grosso?	
cuiabá	**pt/c/u/i/Cuiabá.html**: cuiabá é a capital do estado brasileiro de mato grosso .
belo horizonte	**pt/m/i/n/Minas_Gerais_4450.html**: minas gerais brasão bandeira hino localização região sudeste capital belo horizonte estados limítrofes são paulo , rio de janeiro , espírito santo , bahia , mato grosso do sul ,

porto	**pt/p/o/r/Porto_(desambiguação).html**: porto (desambiguação) . capital do rio grande do sul porto alegre do norte - município do estado do mato grosso porto alegre do piauí -
Question #171 - Quem foi o oitavo marido de Elizabeth Taylor?	
casa com larry fortensky	**pt/6/_/d/6_de_Outubro_ac2c.html**: 6 de outubro . elizabeth taylor se casa com larry fortensky , seu oitavo marido 2 002 -
6 de outubro	**pt/6/_/d/6_de_Outubro_ac2c.html**: 6 de outubro . elizabeth taylor se casa com larry fortensky , seu oitavo marido 2 002 -
Question #172 - Quando é que eles se casaram?	
63	**FSP950901-056**: liz taylor anuncia seu sétimo divórcio a atriz elizabeth taylor , 63 , anunciou ontem que está se divorciando de seu sétimo marido , larry fortensky , 43 . eles estão casados há quatro anos .
43	**FSP950901-056**: liz taylor anuncia seu sétimo divórcio a atriz elizabeth taylor , 63 , anunciou ontem que está se divorciando de seu sétimo marido , larry fortensky , 43 . eles estão casados há quatro anos .
11	**FSP951110-147**: na vida real , rooney se casou muitas vezes ; tantas , talvez , quanto sua colega de profissão elizabeth taylor . elizabeth taylor filha de uma ex - atriz , elizabeth taylor ficou famosa com 11 anos ,
Question #173 - Qual é a nacionalidade dela?	
nbc	**PUBLICO-19940819-015**: elizabeth taylor processa nbc a actriz elizabeth taylor vai processar a cadeia de
richard burton	**PUBLICO-19940324-116**: mankiewicz com elizabeth taylor , richard burton e rex harrison i parte 130 min .
vai processar	**PUBLICO-19940819-015**: elizabeth taylor processa nbc a actriz elizabeth taylor vai processar a cadeia de
Question #174 - Quantos gêneros tem o alemão?	
três gêneros	**FSP940830-024**: declinação na língua alemã . o alemão conserva três gêneros : masculino , feminino e neutro ; dois números : singular e plural ; e quatro casos gramaticais : nominativo , acusativo , dativo e genitivo .
Question #175 - E quantos tem o romanche?	
1 552 ,	**pt/l/í/n/Língua_romanche.html**: língua romanche . . . história o primeiro registro escrito da língua romanche data de 1 552 , na forma de uma lição de catecismo chamada christiauna fuorma , rigistrada por jacob bifrun no dialeto engadino .

1 938 ,	**pt/l/í/n/Língua_romanche.html**: língua romanche . . . . . até 1 938 , o romanche não era considerado uma língua oficial da suíça , tendo seu status reconhecido apenas naquele ano .
Question #176 - Quanto tempo reinou Ramsés II?	
seis anos	**pt/t/a/u/Tausert.html**: tausert . o reinado de seti ii durou seis anos , sugerindo alguns autores que o rei foi perturbado por um usurpador , amenmosé , descendente de ramsés ii .
1 213 a . c .	**pt/r/a/m/Ramsés_II_3363.html**: ramsés ii foi o terceiro faraó da xix dinastia egípcia , uma das dinastias que compõem o império novo . reinou entre aproximadamente 1 279 e 1 213 a . c . . o seu reinado foi possivelmente o mais prestigioso da história egípcia tanto no aspecto econômico
66 anos	**pt/a/b/u/Abu_Simbel_68cb.html**: abu simbel . . ramsés ii iniciou o seu reinado em 1 290 a . c . e reinou durante 66 anos ,
Question #177 - Quando começou o seu reinado?	
1 284	**pt/a/b/u/Abu_Simbel_68cb.html**: abu simbel . a construção começou a cerca de 1 284 a . c . e terminou aproximadamente vinte anos mais tarde . ramsés ii iniciou o seu reinado em 1 290 a .
1 290	**pt/a/b/u/Abu_Simbel_68cb.html**: abu simbel . a construção começou a cerca de 1 284 a . c . e terminou aproximadamente vinte anos mais tarde . ramsés ii iniciou o seu reinado em 1 290 a .
vinte	**pt/a/b/u/Abu_Simbel_68cb.html**: abu simbel . a construção começou a cerca de 1 284 a . c . e terminou aproximadamente vinte anos mais tarde . ramsés ii iniciou o seu reinado em 1 290 a .
Question #178 - Ele ordenou a construção de que templos?	
abu simbel	**pt/a/n/t/Antigo_Egipto_b5aa.html**: antigo egipto . foi também ramsés ii que ordenou a construção dos templos de abu simbel .
Question #179 - Que se passou a 9 de Novembro de 1991?	
museu	**pt/l/i/s/Lista_de_jogos_do_Super_NES_9992.html**: • 13 de março de 1 991 : um quadro de renoir « jeunes femmes à la campagne » foi retirado do museu impressionista de bagnols - sur - cèze (gard) . • 18 novembro de
Question #180 - Quantos actos tem a ópera Verdi da Aida?	
quatro atos	**pt/a/i/d/Aida.html**: aida é uma ópera em quatro atos com música de giuseppe verdi e libreto de antonio ghislazoni , com estréia mundial na casa da ópera , cairo , aos 24 dezembro de 1 871 .

dois atos	**pt/a/i/d/Aida_(Musical)_5ff5.html**: aida (musical) . aida é um drama musical em dois atos baseado na ópera italiana hômonima de giuseppe verdi , que , por sua vez , é baseada numa história de auguste mariette . o musical foi produzido pela hyperion theatricals , uma filiada da disney theatrical ,
Question #181 - Quem escreveu o libretto dessa ópera?	
ópera	**FSP951106-132**: da ópera " aida " , de verdi .
otello	**pt/g/i/u/Giuseppe_Verdi_b105.html**: giuseppe verdi . . . após aida , verdi escreveu ainda as óperas otello e falstaff , baseadas em shakespeare além de algumas peças religiosas .
óperas	**PUBLICO-19940420-132**: o ciclo completa - se , depois , com três óperas de verdi : « aida » ,
Question #182 - Quando é que estreou a ópera?	
1 871	**pt/g/i/u/Giuseppe_Verdi_b105.html**: giuseppe verdi . . . ano da unificação e , posteriormente , senador . e continuou escrevendo óperas : em 1 871 estreou aida , em comemoração à abertura do canal de suez . após aida , verdi escreveu ainda as óperas otello e
1 861	**pt/g/i/u/Giuseppe_Verdi_b105.html**: giuseppe verdi . durante esse período , verdi era aclamado como um patriota , sendo eleito deputado em 1 861 , ano da unificação e , posteriormente , senador . e continuou escrevendo óperas : em 1 871 estreou aida , em comemoração à abertura do canal de suez .
maio de 1 987	**PUBLICO-19941122-106**: a ópera aida , de giuseppe verdi , que conta a história de amor entre um oficial egípcio e uma escrava etíope . quando a ópera foi pela primeira vez estreada , em maio de 1 987 , no templo faraónico de luxor , um edifício velho de 3 200 anos situado na parte leste da cidade ,
Question #183 - Quem se tornou lider do Partido Quebequense em 2005?	
quebec	**pt/q/u/e/Quebec.html**: quebec . em outubro de 1 968 , o partido quebequense foi fundado por rené lévesque .
andré boisclair	**pt/a/n/d/André_Boisclair_6943.html**: andré boisclair . ele é o atual líder do partido quebequense , o principal partido político separatista da província de quebec .
venceu as eleições	**pt/q/u/e/Quebec.html**: quebec . . . . . o partido quebequense venceu as eleições provinciais de 1 994 .
Question #184 - Qual é a maior cidade do Canadá?	
toronto	**pt/t/o/r/Toronto.html**: ca toronto é a maior cidade do canadá , e a capital da província de ontário .

montreal	**pt/m/o/n/Montreal.html**: montreal . muitos dos habitantes da cidade não querem que a cidade sedie outra olimpíada . o maior centro urbano do canadá e
nova iorque	**pt/t/o/r/Toronto.html**: toronto . . . . . valores do canadá , a segunda maior do continente americano (atrás apenas do new york stock exchange , localizado na cidade americana de nova iorque) e a sexta maior do mundo .
Question #185 - O que é o Gil Vicente FC?	
sc	**pt/l/i/g/Liga_de_Honra_1898.html**: liga de honra . praia cd feirense gondomar sc leixões sc futebol clube penafiel rio ave futebol clube gil vicente fc sc olhanense portimonense sporting clube cd santa clara varzim sc fc vizela olivais e
caso mateus	**pt/c/a/s/Caso_Mateus_53ef.html**: caso mateus . . . . . o gil vicente fc e
campeonato português	**pt/c/a/m/Campeonato_português_de_futebol_(1992-93).html**: campeonato português de futebol (1 992 93) . . . resultados fc porto sl benfica sporting cp boavista fc cs marítimo sc farense cf belenenses sc beira mar gil vicente fc fc paços ferreira vsc guimarães sc braga gd estoril -
Question #186 - Quem foi Gil Vicente?	
gil vicente (1 465 - 1 536 ?) é geralmente considerado o primeiro grande dramaturgo português , além de poeta de renome . há quem o identifique com o ourives , autor da custódia de belém , mestre da balança , e com o mestre de retórica do rei dom manuel	**pt/g/i/l/Gil_Vicente_f346.html**: gil vicente (1 465 - 1 536 ?) é geralmente considerado o primeiro grande dramaturgo português , além de poeta de renome . há quem o identifique com o ourives , autor da custódia de belém , mestre da balança , e com o mestre de retórica do rei dom manuel . enquanto homem de teatro , parece ter também desempenhado as tarefas de músico , actor e encenador . é frequentemente considerado , de uma forma geral , o pai do teatro português , ou mesmo do teatro ibérico já que também escreveu em castelhano - partilhando a paternidade da dramaturgia espanhola com juan del encina .
Question #187 - Quem foi o ”pai do teatro português”?	
corte surge também gil vicente	**pt/m/a/n/Manuel_I_de_Portugal_4cf8.html**: manuel i de portugal . na sua corte surge também gil vicente , o pai do teatro português e duarte pacheco pereira o geógrafo , autor do esmeraldo de situ orbis .
duarte pacheco pereira o geógrafo	**pt/m/a/n/Manuel_I_de_Portugal_4cf8.html**: manuel i de portugal . na sua corte surge também gil vicente , o pai do teatro português e duarte pacheco pereira o geógrafo , autor do esmeraldo de situ orbis .

dramaturgia espanhola com juan del encina	**pt/g/i/l/Gil_Vicente_f346.html**: gil vicente . é frequentemente considerado , de uma forma geral , o pai do teatro português , ou mesmo do teatro ibérico já que também escreveu em castelhano - partilhando a paternidade da dramaturgia espanhola com juan del encina .
Question #188 - Qual a área do Parque Estadual Guariba?	
72 296 331 hectares	**pt/l/i/s/Lista_de_parques_estaduais_do_Brasil_fb87.html**: lista de parques estaduais do brasil . . região norte acre parque estadual chandless amapá amazonas parque estadual guariba , com 72 296 331 hectares , criado em 2 005 , em manicoré parque estadual nhamundá parque estadual rio negro setor norte parque estadual rio negro setor sul parque estadual samaúma ,
51 hectares	**pt/l/i/s/Lista_de_parques_estaduais_do_Brasil_fb87.html**: lista de parques estaduais do brasil . . . . . amapá amazonas parque estadual guariba , com 72 296 331 hectares , criado em 2 005 , em manicoré parque estadual nhamundá parque estadual rio negro setor norte parque estadual rio negro setor sul parque estadual samaúma , com 51 hectares , criado em 2 003 , em manaus parque estadual serra do araça parque estadual sucunduri ,
Question #189 - Quando foi criado o parque?	
2 005	**pt/l/i/s/Lista_de_parques_estaduais_do_Brasil_fb87.html**: lista de parques estaduais do brasil . região norte acre parque estadual chandless amapá amazonas parque estadual guariba , com 72 296 331 hectares , criado em 2 005 ,
Question #190 - O que é a Torre do Tombo?	
torre do tombo é o nome do arquivo central do estado português desde a idade média . com mais de 600 anos , é uma das mais antigas instituições portuguesas ainda activas	**pt/t/o/r/Torre_do_Tombo_99b6.html**: torre do tombo é o nome do arquivo central do estado português desde a idade média . com mais de 600 anos , é uma das mais antigas instituições portuguesas ainda activas . ao longo do tempo , a conservação dos documentos foi prejudicada por um conjunto de circunstâncias : não apenas pelo terramoto de 1 755 , mas também as frequentes mudanças de local , incêndios , a transferência da corte para o rio de janeiro no brasil , o desvio de materiais aquando do domínio filipino e das invasões francesas etc .
Question #191 - Onde fica?	
goza de privilégios no que respeita a consultas	**PUBLICO-19940119-004**: « assim , ficamos a saber que um indivíduo , por ser director da torre do tombo , goza de privilégios no que respeita a consultas relativamente aos outros investigadores .

guarda do arquivo nacional	**PUBLICO-19940104-066**: mas seriam depositados à guarda do arquivo nacional da torre do tombo (antt) , responsabilizado pela sua organização ; que ficariam fechados ,
documento	**PUBLICO-19940303-013**: fica - se a saber por um documento interno da torre do tombo , em novembro , que não é o investigador que paga este trabalho , mas a câmara de oeiras .
Question #192 - Que país faz fronteira com Cuba?	
espanha	**PUBLICO-19940902-062**: o bloqueio a cuba gostaria de propor um exercício de imaginação . em 1 960 , o governo de espanha decidiu que portugal não deveria ser independente , por já ter sido possessão espanhola . fechou então as suas fronteiras aos portugueses ,
portugal	**PUBLICO-19941009-073**: se para analisar a situação na fronteira do kuwait . • a associação de amizade portugal - cuba realiza uma caravana lisboa - vigo , com produtos recolhidos em todo o país e que se destinam a cuba .
méxico	**pt/a/l/e/Alexander_von_Humboldt_682a.html**: alexander von humboldt . cuba e méxico (foi impedido de permanecer no brasil , pois os portugueses consideraram - no um possível espião alemão , após encontrarem - no em terras brasileiras perto da fronteira venezuelana) .
Question #193 - Qual é o comprimento do metro de Coimbra?	
4 metros	**FSP950815-033**: a vala tem 5 metros de comprimento por 4 metros de largura . a profundidade , disse o encarregado , é de até 3 metros . parte do terreno foi tomado por covas de crianças mortas . para a presidente do grupo tortura nunca mais , cecília coimbra ,
3 metros	**FSP950815-033**: a vala tem 5 metros de comprimento por 4 metros de largura . a profundidade , disse o encarregado , é de até 3 metros . parte do terreno foi tomado por covas de crianças mortas . para a presidente do grupo tortura nunca mais , cecília coimbra ,
2 602 metros	**pt/c/a/m/Campo_de_Provas_Brigadeiro_Velloso_7f14.html**: campo de provas brigadeiro velloso . 2 602 metros de comprimento . o campo de provas de cachimbo foi criado em 7 de março de 1 983 , sendo subortinado ao centro tecnológico aeroespacial (cta) . o nome da unidade foi alterado para campo de provas brigadeiro - do - ar haroldo coimbra velloso em 17 de

Question #194 - Quantas esposas tinha Ngungunhane?	
NIL	
Question #195 - Como é que se chamava o filho dele?	
quando o rei	**pt/n/g/u/Ngungunhane.html**: zixaxa casa e tem um filho , também chamado roberto zixaxa , fundando uma família que ainda está presente na sociedade angrense . quando o rei d . carlos i de portugal visita os açores em 1 901 , ngungunhane é levado a passear ao campo durante a permenência do monarca na
casa	**pt/n/g/u/Ngungunhane.html**: ngungunhane embriaga - se frequentemente , sendo várias vezes detido por desacatos praticados em tal estado . apenas molungo se retrai e evita falar português . zixaxa casa e tem um filho , também chamado roberto zixaxa ,
várias vezes	**pt/n/g/u/Ngungunhane.html**: ngungunhane embriaga - se frequentemente , sendo várias vezes detido por desacatos praticados em tal estado . apenas molungo se retrai e evita falar português . zixaxa casa e tem um filho , também chamado roberto zixaxa ,
Question #196 - Qual é a capital de Cuba?	
camagüey	**pt/c/a/m/Camagüey.html**: é capital da provincia de camagüey . categorias : ! artigos mínimos | cidades de cuba
guantánamo	**pt/g/u/a/Guantánamo.html**: guantánamo é a capital da província de guantánamo , situada no sudeste de cuba .
key west	**FSP951116-128**: com relação a cuba . o presidente harry s . truman foi um visitante regular de key west durante seu período na casa branca , e adquiriu na época uma propriedade que ainda é conhecida como a pequena casa branca de key west , atualmente aberta a visitação pública . " eu adoraria mudar a capital do país para
Question #197 - Quem criou o primeiro alfabeto?	
negociantes semíticos	**pt/l/í/b/Líbano.html**: líbano . é a pátria histórica dos fenícios , negociantes semíticos da antiguidade , cuja cultura marítima floresceu na região durante mais de 2 000 anos e que criaram o primeiro alfabeto , do qual sairam todos os demais , tanto semíticos como indo - europeus .
marítima floresceu na região durante	**pt/l/í/b/Líbano.html**: líbano . é a pátria histórica dos fenícios , negociantes semíticos da antiguidade , cuja cultura marítima floresceu na região durante mais de 2 000 anos e que criaram o primeiro alfabeto , do qual sairam todos os demais , tanto semíticos como indo - europeus .

líbano	**pt/l/í/b/Líbano.html**: líbano . é a pátria histórica dos fenícios , negociantes semíticos da antiguidade , cuja cultura marítima floresceu na região durante mais de 2 000 anos e que criaram o primeiro alfabeto , do qual sairam todos os demais , tanto semíticos como indo - europeus .
Question #198 - Quando é que Porto Rico se tornou um estados dos EUA?	
1 953	**PUBLICO-19941115-018**: dos eua , tornaram - se a primeira dupla a vencer por três vezes consecutivas a taça do mundo de golfe , uma competição que remonta a 1 953 e que este ano decorreu no percurso de dorado beach , em porto rico .
70	**pt/s/a/l/Salsa.html**: salsa . . americanas nos eua e porto rico , depois a cuba , venezuela , colômbia e outros países de língua espanhola . nomes como tito puente , celia cruz , johny pacheco se tornaram expoentes do gênero . o excessivo comercialismo em fins dos anos 70 converteu a salsa numa fórmula que apenas
Question #199 - Onde fica Livorno?	
comuna italiana da região	**pt/s/a/s/Sassetta.html**: sassetta é uma comuna italiana da região da toscana , província de livorno , com cerca de 548 habitantes .
toscana	**pt/h/i/s/História_da_Toscana_35e1.html**: história da toscana . livorno em 1 421 .
comunas	**pt/l/i/v/Livorno.html**: esboços sobre geografia da itália | comunas da toscana | comunas de livorno
Question #200 - O que são os iaques?	
os iaques (bos grunniens) , também conhecido como boi - cavalo , são bois selvagens asiáticos , encontrados no planalto tibetano , em altitudes que variam entre 4 500 m e 6 000 m	**pt/i/a/q/Iaque.html**: os iaques (bos grunniens) , também conhecido como boi - cavalo , são bois selvagens asiáticos , encontrados no planalto tibetano , em altitudes que variam entre 4 500 m e 6 000 m . possuem uma longa pelagem negra a marrom - escura e grandes chifres curvados para cima e para frente . foram domesticado em algumas regiões da ásia central . quando domesticados , fornecem lã , carne e leite , bem como sendo utilizados como animais de tração .

C.3 Official Results for IdSay

C.3.1 Summary

QA@Clef Download Interface

Menu

- Home
- User Login
- Registration Form
- Contact Us

User Menu

- Download
- Upload
- Analysis
- QA-WSD Pilot Task
- Logout

Analysis Result :

ACCURACY MEASURE OF ALL ANSWERS:
The file idsa081ptpt.xml contains a total of 463 answers
were assessed 463
- 65 Right
- 119 Wrong
- 8 ineXact
- 8 Unsupported

Accuracy calculated over the FIRST answer:
Overall accuracy = 65/200 = 32.500 %

Single Answers Returned = 59
- Correct Single Answers = 34
Multiple Answers Returned = 141
- Correct Multiple Answers = 51

Factoids:
The file contains a total of 162 factoids
- 47 Right
- 100 Wrong
- 7 ineXact
- 8 Unsupported

Accuracy calculated over factoids = 47/162 = 29.012 %

Lists:
The file contains a total of 10 lists
- 0 Right
- 9 Wrong
- 1 ineXact
- 0 Unsupported

Accuracy calculated over lists = 0/10 = 0.000 %

Definitions:
The file contains a total of 28 definitions
- 18 Right
- 10 Wrong
- 0 ineXact
- 0 Unsupported

Accuracy calculated over definitions = 18/28 = 64.286 %

NIL Answers Returned:
Total NIL Returned: 12
- 2 Right
- 10 Wrong
- 0 ineXact
- 0 Unsupported

Accuracy Over NIL Answers Returned = 2/12 = 16.667 %

Temporally Restricted Questions:
Total Temp. Restricted Questions: 16
- 3 Right
- 13 Wrong
- 0 ineXact
- 0 Unsupported

Accuracy Over Temp. Restricted Questions = 3/16 = 18.750 %

CONFIDENCE WEIGHTED SCORE:
Unable to be calculated.No scores in dataset.

Mean reciprocal rank SCORE:
The overall Mean reciprocal rank (MRR) over all assessed multple answers is : 40.167/141 = 0.28487

Mean reciprocal rank* (MMR*)SCORE:
The overall Mean reciprocal rank (MRR) over all assessed answers is : 74.167/200 = 0.37083

Figure C.1: IdSay Results - Summary

C.3.2 Details

QA@Clef Download Interface

Menu

- Home
- User Login
- Registration Form
- Contact Us

User Menu

- Download
- Upload
- Analysis
- QA-WSD Pilot Task
- Logout

Judgment Result : idsa081ptpt.xml

0001	língua	W
| 0001	mão	W
0002	hergé	R
| 0002	tibete	W
| 0002	dava forma	W
0003	1 929	R
| 0003	17 de novembro de 1 954	W
| 0003	10 de janeiro de 1 929	R
0004	sempre acompanhado	W
| 0004	jovem repórter que viaja pelo mundo solucionando mistérios	W
| 0004	milou	R
0005	NIL	W
0006	gres estação primeira de mangueira	W
| 0006	poucos todos os outros	W
| 0006	rol	W
0007	16 de julho de 1 990	W
| 0007	1 991	R
| 0007	20 por cento	W
0008	1 g	W
0009	itália	U
| 0009	rio rubicão	W
| 0009	direito romano da época	W
0010	quatro províncias	U
0011	pico de orizaba	R
| 0011	fiorde	W
| 0011	aimé bonpland	W
0012	monte fuji	R
| 0012	alpes japoneses	W
| 0012	sobre as ilhas	W
0013	puer aeternus	W
| 0013	guerra civil espanhola	W
| 0013	padrão do arquétipo	W
0014	2 117 m	R
| 0014	2 103 metros	R
| 0014	150 quilómetros	W
0015	richard bach	U
| 0015	romance	W
| 0015	admirável mundo novo	W
0016	menir - monumentos pré - históricos em pedras , cravadas verticalmente no solo (ortóstatos) , às vezes de tamanho bem elevado (megalito denominado menir) . a palavra menir foi adotada , através do francês , pelos arqueólogos do século xix com base nas palavras do bretão significando men = pedra e hir = longa (comparar com o gaélico : maen hir = pedra longa)	R
0017	20	W
| 0017	1 108	W
| 0017	24	W
0018	três ossos	W
| 0018	um osso	W
0019	3 500	W
| 0019	500	W
| 0019	1 941	W
0020	5 de fevereiro de 1 943	W
| 0020	1 955	W
| 0020	90	W
0021	1 955	R
0022	NIL	W
0023	fhc	W
| 0023	plano	W
| 0023	governo	W
0024	álvaro de campos (1 890 - 1 935) é um dos heterónimos mais conhecidos de fernando pessoa . nascido em tavira , teve a educação de liceu comum de sua época , posteriormente foi para a escócia estudar engenharia mecânica , e depois engenharia naval	R
0025	fernando pessoa	W
| 0025	alberto caeiro	W
| 0025	ricardo reis	W
0026	noite	W
0027	pierre athanase larousse (toucy , 23 de outubro de 1 817 — paris , 3 de janeiro de 1 875) foi	R

Figure C.2: IdSay Results - Details: Part 1 of 7

	um pedagogo , editor e enciclopedista francês	
0028	la brabançonne é o hino nacional da bélgica	R
| 0028	land of the free	W
| 0028	belga capital bruxelas	W
0029	família	W
| 0029	ramphastidae que vivem nas florestas da américa central	X
| 0029	outras presas	W
0030	NIL	W
0031	rio rubicão	W
| 0031	famosa frase	W
| 0031	césar	R
0032	rio rubicão	R
0033	nicolas sarkozy	R
0034	a cítara é um instrumento musical de várias cordas presas sobre um arco de madeira , com ou sem caixa de ressonância , que se tocavam com ambas as mãos	R
0035	mitsubishi	W
| 0035	principal caça da marinha japonesa durante toda a segunda guerra mundial	R
| 0035	corsair	W
0036	hélio	R
| 0036	superfluidez	W
| 0036	série química	W
0037	série química	R
| 0037	grupo	W
| 0037	enxofre	R
0038	apenas quando a sua órbita	W
| 0038	planeta anão	W
| 0038	nome também	W
0039	amiens	R
0040	1 833	W
| 0040	1 262	W
| 0040	mar	W
0041	ministro da defesa	W
| 0041	deputado josé maria benegas	W
| 0041	julián garcía vargas	W
0042	rocinante	R
| 0042	sancho pança	W
| 0042	diabo	W
0043	londres	W
| 0043	tóquio	W
| 0043	los angeles	W
0044	condado condado de galway	W
| 0044	norte	W
| 0044	república	W
0045	sintetizador	W
0046	concordata	W
0047	a saga do tio patinhas	X
| 0047	carl barks	W
| 0047	don rosa	W
0048	tio patinhas	W
| 0048	ducktales	W
| 0048	maga patalógika	W
0049	margarida (banda desenhada)	X
| 0049	superpato	W
| 0049	poucos sabem sua identidade	W
0050	chefe da segurança da loja	W
| 0050	brodka	W
| 0050	entende nada	W
0051	llanito	W
| 0051	culinária de espanha	W
| 0051	casa	W
0052	finlândia	W
| 0052	noruega	W
0053	açor	W
| 0053	pap	W
| 0053	arroz	W
0054	o feta é uma variedade de queijo típica da grécia , fabricado com leite de cabra ou de ovelha , exclusivamente . é um queijo branco , farelento e levemente salgado	R
0055	gouda	W
| 0055	grécia	R
0056	NIL	W
0057	13 anos	R
| 0057	dois anos	W
| 0057	14 anos	R
0058	toshiro mifune	W
| 0058	serviu ao exército	W

Figure C.3: IdSay Results - Details: Part 2 of 7

⌊ 0058	americanas	W
0059	15 repúblicas	R
⌊ 0059	seis repúblicas	W
0060	geórgia	W
⌊ 0060	rússia	R
0061	geórgia	U
⌊ 0061	tskhinvali	W
⌊ 0061	república	W
0062	34 km	R
0063	jogo de tabuleiro	W
⌊ 0063	considerações	W
⌊ 0063	jogar a todos	W
0064	ele	W
⌊ 0064	arnulfo de metz	R
⌊ 0064	dia em uma dessa datas	W
0065	ele	W
⌊ 0065	arnulfo de metz	W
⌊ 0065	dia em uma dessa datas	W
0066	chá com adição de rum	R
0067	135 155 cm	R
⌊ 0067	55 cm	W
0068	1 kg	R
0069	NIL	W
0070	24 províncias	R
0071	madrugada	W
0072	NIL	R
0073	roman abramovich	R
⌊ 0073	inglês de futebol	W
⌊ 0073	britânico mais rico	W
0074	300 000 habitantes	R
0075	dois anos	W
0076	engenharia civil	W
⌊ 0076	doutor em engenheria	W
⌊ 0076	geólogo	W
0077	91 estados	R
0078	roma	R
⌊ 0078	esta organização	W
⌊ 0078	bienal	W
0079	NIL	R
0080	instituto português do património arquitectónico	W
⌊ 0080	miguel ângelo lupi	W
⌊ 0080	lista de museus	W
0081	manuel de portugal	W
⌊ 0081	ii	W
⌊ 0081	príncipe real	W
0082	religião	W
⌊ 0082	moscóvia	W
0083	baleeiro	W
0084	iluminuras	W
⌊ 0084	ran	W
⌊ 0084	mundo moderno	W
0085	454 445 votos	W
⌊ 0085	11 168 228 votos	W
⌊ 0085	um voto	W
0086	2 003	U
⌊ 0086	janeiro de 2 003	R
⌊ 0086	8 de junho de 2 006	W
0087	carlos martel	R
⌊ 0087	luís iii de França	W
⌊ 0087	rei	W
0088	baden powell de aquino , violonista brasileiro , (varre - sai , rj , 6 de agosto de 1 937 - rio de janeiro , rj , 26 de setembro , de 2 000)	R
0089	janete clair	W
⌊ 0089	rudyard kipling	R
⌊ 0089	eugenia	W
0090	nos estados unidos	W
⌊ 0090	novo romance	W
⌊ 0090	mercy	W
0091	hokkaido	R
0092	zenão de cítio	R
⌊ 0092	stoa	W
⌊ 0092	ginásios	W
0093	flamenga	W
⌊ 0093	países baixos	W

Figure C.4: IdSay Results - Details: Part 3 of 7

⌊ 0093	frança	W
0094	los angeles	W
⌊ 0094	califórnia	U
⌊ 0094	guerra civil americana	W
0095	nos jogos olímpicos de verão	W
⌊ 0095	países	W
⌊ 0095	lista	W
0096	república oeste distante era localisada	W
⌊ 0096	likbez	W
⌊ 0096	sovnarkom da rsfsr sob o nome	W
0097	um atleta	W
⌊ 0097	6 804 atletas	R
0098	portugal	W
⌊ 0098	israel	W
⌊ 0098	suécia	W
0099	los angeles	W
⌊ 0099	montreal	R
⌊ 0099	tomar	W
0100	o berimbau é um instrumento de percussão usado tradicionalmente na capoeira , para marcar o ritmo da luta . no brasil é ainda conhecido pelos seguintes nomes : urucungo , urucurgo , orucungo , oricungo , uricungo , rucungo , ricungo , berimbau de barriga , gobo , marimbau , bucumbumba , bucumbunga , gunga , macungo , matungo , mutungo , aricongo , arco musical e rucumbo	R
0101	eslovénia	W
⌊ 0101	luxemburgo	W
⌊ 0101	suíça	W
0102	menos comum	W
⌊ 0102	calças	W
⌊ 0102	desenhos animados	W
0103	273 graus negativos	R
⌊ 0103	0 01 ° c	W
⌊ 0103	269 ° c	W
0104	saori kido	W
⌊ 0104	minerva	U
⌊ 0104	mitologia grega	W
0105	piraí	W
⌊ 0105	arroio poço grande	W
⌊ 0105	açu	W
0106	182 km	W
⌊ 0106	400 km	R
0107	astana	R
⌊ 0107	ashkhabad	W
⌊ 0107	moscovo	W
0108	russo capital astana maior cidade almaty presidente nursultan nazarbayev	X
⌊ 0108	línguas oficiais	W
⌊ 0108	população	W
0109	óscar berger	R
⌊ 0109	prefeito da cidade da guatemala categorias	X
⌊ 0109	advogado de profissão	W
0110	partido social democrata	W
⌊ 0110	nas eleições parlamentares	W
⌊ 0110	eleição	W
0111	13 faixas	R
⌊ 0111	cinco faixas	W
0112	áustria - hungria	W
⌊ 0112	igreja oficial igreja católica capital & amp	W
⌊ 0112	maior cidade	W
0113	22 de dezembro de 1 846	R
⌊ 0113	19 de agosto	W
⌊ 0113	22 de dezembro . 1 758	W
0114	nicolau v	W
⌊ 0114	África ocidental	W
⌊ 0114	ii	W
0115	vrml (virtual reality modelating language) , ou linguagem para modelagem de realidade virtual , é um padrão de aplicativos de realidade virtual utilizado na internet . por meio desta linguagem , escrita em modo texto , é possível criar objetos tridimensionais podendo definir cor , transparência , brilho , textura (associando - a a um bitmap)	R
0116	templo	W
⌊ 0116	querubim	W
⌊ 0116	bíblia	W
0117	próximos dias	W
⌊ 0117	cidade que no tempo	W
⌊ 0117	nova lisboa	R
0118	árabe	R
⌊ 0118	população total	W
⌊ 0118	sudão	W

Figure C.5: IdSay Results - Details: Part 4 of 7

0119	história do brasil	W
| 0119	após ataques submarinos a navios	W
| 0119	força expedicionária	W
0120	paz	W
| 0120	cassandra	W
0121	fogueira por bruxaria	W
| 0121	aspectos controversos do catolicismo	W
| 0121	história	W
0122	19 de abril . 1 839	W
| 0122	1 909	W
0123	dez anos	W
| 0123	sete anos	W
| 0123	15 anos	W
0124	1 969	W
| 0124	1 959	R
| 0124	1 964	W
0125	13 de agosto de 1 926	R
| 0125	26 de julho	W
| 0125	1 926	R
0126	mary elizabeth donaldson	W
| 0126	ela se relembra que eles começaram	W
| 0126	princesa herdeira da dinamarca	W
0127	forcarei é um município da espanha na província de pontevedra , comunidade autónoma da galiza , de área 168 70 km ² com população de 4 616 habitantes (2 004) e densidade populacional de 27 36 hab / km ²	W
0128	1 143	U
| 0128	850	W
| 0128	5 de outubro de 1 143	R
0129	fogo	W
| 0129	pedro gonzález telmo	W
| 0129	efeitos da ionização	W
0130	brigadeiro é uma patente militar de uso na aeronáutica ou força aérea , que designa a patente mais alta daquela força , equivalente a patente de general no exército	R
0131	percy spencer	U
| 0131	inglês	W
| 0131	industriais	W
0132	honolulu	W
| 0132	nascimento	W
| 0132	havaí nacionalidade norte	W
0133	brasileiro roberto landell de moura patenteou em nova york	X
| 0133	levar adiante	W
| 0133	recursos	W
0134	korea train express	W
| 0134	sncf	R
| 0134	comboio de alta velocidade	W
0135	uma das pioneiras no ensino a distância no sul do país	R
| 0135	guaíba	W
| 0135	porto alegre	W
0136	18 mil contos	R
0137	ano passado ao romancista miguel	W
| 0137	contos	W
| 0137	pesetas	W
0138	estado	W
| 0138	principado	W
| 0138	conselho geral dos vales	W
0139	religião	W
0140	dez jogadores	W
0141	NIL	W
0142	2 170 600 km 2	R
0143	valentina tereshkova	R
| 0143	vostok	W
| 0143	programa espacial soviético considerou enviar	W
0144	herói da união soviética	W
| 0144	guerra mundial	W
| 0144	caça	W
0145	disse que sabia que iam assassiná	W
| 0145	chacal	W
| 0145	dia	W
0146	dois mil refugiados	W
| 0146	16 mil refugiados	R
0147	1 945	R
| 0147	1	W
| 0147	1 966	R
0148	o condado	W
| 0148	esporte clube	W

Figure C.6: IdSay Results - Details: Part 5 of 7

⌊ 0148	história	W
0149	20 de dezembro de 1 995	W
⌊ 0149	18 de maio de 1 978	W
⌊ 0149	1 959	W
0150	fausto dos santos	W
⌊ 0150	flamengo	W
⌊ 0150	nau	W
0151	NIL	W
0152	londres	W
⌊ 0152	belfast	W
⌊ 0152	atenas	W
0153	NIL	W
0154	saxônia	W
⌊ 0154	berlim	X
⌊ 0154	brandemburgo	X
0155	james joyce pelos críticos	X
⌊ 0155	oliver goldsmith	R
⌊ 0155	penso	W
0156	carl barks (27 de março de 1 901 - 25 de agosto de 2 000) foi um famoso ilustrador dos estúdios disney e criador de arte seqüencial , responsável pela invenção de patópolis e muitos de seus habitantes : tio patinhas (1 947) , gastão (1 948) , irmãos metralha (1 951) , professor pardal (1 952) , maga patalójika (1 961) e outros	R
0157	ele tinha um irmão mais velho chamado	W
⌊ 0157	avós maternos	W
⌊ 0157	legou o pato donald	W
0158	tio patinhas	W
⌊ 0158	expressou opinião de que só nos contos de fadas os maus	W
⌊ 0158	tornam	W
0159	o kilt é o saiote pregueado , parcialmente trespassado , e quadriculado em cores correspondentes a cada clã ou família , e que faz parte do traje típico masculino da escócia	R
0160	ilhas selvagens	W
⌊ 0160	one hundred and one dalmatians	W
⌊ 0160	exposição individual no centro cultural português	W
0161	quatro filmes	R
0162	intervalos	W
⌊ 0162	americano tom gunning	W
⌊ 0162	morte de nicholas ray	W
0163	18 quilômetros	U
0164	NIL	W
0165	couraçados	W
0166	couraçados	W
0167	o crescente fértil é uma região do oriente médio compreendendo os atuais israel , cisjordânia e líbano bem como partes da jordânia , da síria , do iraque , do egito e do sudeste da turquia	R
0168	são paulo	W
⌊ 0168	guarani	R
⌊ 0168	associação atlética ponte preta	R
0169	minas gerais	W
⌊ 0169	brasileiro	W
⌊ 0169	sede na cidade	W
0170	cuiabá	R
⌊ 0170	belo horizonte	W
⌊ 0170	porto	W
0171	casa com larry fortensky	X
⌊ 0171	6 de outubro	W
0172	63	W
⌊ 0172	43	W
⌊ 0172	11	W
0173	nbc	W
⌊ 0173	richard burton	W
⌊ 0173	vai processar	W
0174	três gêneros	R
0175	1 552 ,	W
⌊ 0175	1 938 ,	W
0176	seis anos	W
⌊ 0176	1 213 a . c .	W
⌊ 0176	66 anos	R
0177	1 284	W
⌊ 0177	1 290	X
⌊ 0177	vinte	W
0178	abu simbel	R
0179	museu	W
0180	quatro atos	R
⌊ 0180	dois atos	X
0181	ópera	W
⌊ 0181	otello	W

Figure C.7: IdSay Results - Details: Part 6 of 7

| 0181	óperas	W
0182	1 871	R
| 0182	1 861	W
| 0182	maio de 1 987	W
0183	quebec	W
| 0183	andré boisclair	U
| 0183	venceu as eleições	W
0184	toronto	R
| 0184	montreal	W
| 0184	nova iorque	W
0185	sc	W
| 0185	caso mateus	W
| 0185	campeonato português	W
0186	gil vicente (1 465 - 1 536 ?) é geralmente considerado o primeiro grande dramaturgo português , além de poeta de renome . há quem o identifique com o ourives , autor da custódia de belém , mestre da balança , e com o mestre de retórica do rei dom manuel	R
0187	corte surge também gil vicente	X
| 0187	duarte pacheco pereira o geógrafo	W
| 0187	dramaturgia espanhola com juan del encina	W
0188	72 296 331 hectares	R
| 0188	51 hectares	W
0189	2 005	R
0190	torre do tombo é o nome do arquivo central do estado português desde a idade média . com mais de 600 anos , é uma das mais antigas instituições portuguesas ainda activas	R
0191	goza de privilégios no que respeita a consultas	W
| 0191	guarda do arquivo nacional	W
| 0191	documento	W
0192	espanha	W
| 0192	portugal	W
| 0192	méxico	W
0193	4 metros	W
| 0193	3 metros	W
| 0193	2 602 metros	W
0194	NIL	W
0195	quando o rei	W
| 0195	casa	W
| 0195	várias vezes	W
0196	camagüey	W
| 0196	guantánamo	W
| 0196	key west	W
0197	negociantes semíticos	X
| 0197	marítima floresceu na região durante	W
| 0197	líbano	X
0198	1 953	W
| 0198	70	W
0199	comuna italiana da região	W
| 0199	toscana	W
| 0199	comunas	W
0200	os iaques (bos grunniens) , também conhecido como boi - cavalo , são bois selvagens asiáticos , encontrados no planalto tibetano , em altitudes que variam entre 4 500 m e 6 000 m	R

Figure C.8: IdSay Results - Details: Part 7 of 7

D IdSay Results after Improvements

Table D.1: IdSay Answers and Support: Part 1 of 2 (Questions 1-100)

A#: Answer	Support
Question #1 - Que animal é o Cocas?	
falcão	**PUBLICO-19950930-097**: a saudade de um tempo em que o leite e os ovos vinham de animais e não de supermercados , os sapos , mesmo verdes , não se chamavam cocas - - e os livros podiam ter letras pequenas e quatrocentas páginas , como « memed , meu falcão » .
burro	**PUBLICO-19950911-137**: viegas bateu à esquerda e direita , do mrpp « do burro e do preguiça » ao pp do « sapo cocas da manuela moura guedes » .
anfíbio	**PUBLICO-19941030-100**: o grande porta estandarte dos marretas , o sapo cocas , tornou - se no primeiro anfíbio a discursar perante a mais famosa associação de alunos da universidade de oxford , a « oxford union » ,
Question #2 - Quem foi o criador de Tintin?	
hergé	**pt/2/2/_/22_de_Maio_4855.html**: 22 de maio . 1 907 * * hergé , criador de histórias em quadrinhos , como tintin .
kurt vonnegut	**pt/l/i/s/Lista_de_países_fictícios.html**: lista de países fictícios . no livro cama de gato (autoria de kurt vonnegut) san theodoros , nuevo rico , bordúria , sildávia e el - khemed , nos quadrinhos de tintin (autoria de hergé) santa fé ,
cama de gato	**pt/l/i/s/Lista_de_países_fictícios.html**: lista de países fictícios . no livro cama de gato (autoria de kurt vonnegut) san theodoros , nuevo rico , bordúria , sildávia e el - khemed , nos quadrinhos de tintin (autoria de hergé) santa fé ,
Question #3 - Quando é que ele foi criado?	
1 929	**FSP951214-158**: a coleção completa de tintin , o repórter criado pelo belga hergé em 1 929 , pode ser encontrada nas livrarias de pokhara a us $ 2,5 o exemplar da série .

10 de janeiro de 1 929	**pt/t/i/n/Tintin.html**: tintin . tintin , no original) , criado pelo quadrinista belga conhecido como hergé em 10 de janeiro de 1 929 .
10 de janeiro	**pt/1/0/_/10_de_Janeiro_ee4f.html**: 10 de janeiro . tintin , personagem criado pelo cartunista hergé .
Question #4 - Como se chama o cão dele?	
amigo capitão haddock	**FSP950903-176**: tintin é um jovem repórter que viaja pelo mundo solucionando mistérios , sempre acompanhado por seu cão , milou , e pelo amigo capitão haddock .
jovem repórter que viaja pelo mundo solucionando mistérios	**FSP950903-176**: tintin é um jovem repórter que viaja pelo mundo solucionando mistérios , sempre acompanhado por seu cão , milou , e pelo amigo capitão haddock .
sempre acompanhado	**FSP950903-176**: tintin é um jovem repórter que viaja pelo mundo solucionando mistérios , sempre acompanhado por seu cão , milou , e pelo amigo capitão haddock .
Question #5 - De que raça é o cão?	
deste	**PUBLICO-19940728-127**: autores dignos de crédito , como alexandra david - neel e fosco maraini [exploradores do tibete no começo deste século] , que passaram longas estadias no tibete . » o modo como hergé trabalha este material ressalta claramente em « tintin
pouco	**PUBLICO-19940629-127**: a conhecida aventura de tintin - - e o « piscar de olho » cúmplice aos personagens de hugo pratt (com corto maltese à cabeça) contribuem para dissipar um pouco a imagem vivamente « feminista » de uma autora de histórias aos quadradinhos inteligente , sensível ,
verá	**PUBLICO-19950118-132**: autor deste pequeno livrinho . a personalidade do professor e as suas numerosas descobertas fundamentais são passados em revista , sem esquecer os seus colegas e as peripécias em que , acidental ou deliberadamente , se verá envolvido ao longo da sua vida . graças a albert algoud , o leitor menos sistemático das aventuras de tintin
Question #6 - Diga uma escola de samba fundada nos anos 40.	
Question #7 - Em que ano houve um terramoto no Irão?	
julho de 1 990	**PUBLICO-19950322-131**: depois do terramoto que abalou o norte do irão em julho de 1 990 ,
1 991	**pt/a/b/b/Abbas_Kiarostami_1c17.html**: abbas kiarostami . 1 991) foi um brilhante retrato do trágico terramoto que assolou o irão em 1 991 .
16 de julho de 1 990	**PUBLICO-19950118-153**: irão : 35 a 36 mil mortos . 16 de julho de 1 990 - - 1 641 pessoas morrem , 969 desaparecem e 3 441 ficam feridas em luçon , principal ilha das filipinas , após um terramoto de 7,7 graus .

Question #8 - Quanto pesa um beija-flor?	
19 a 21 gramas	**pt/b/e/i/Beija-flor.html**: beija - flor . a maior espécie conhecida é o beija - flor - gigante da patagónia , que , mesmo assim é de tamanho diminuto em comparação com outras aves , com 19 a 21 gramas de peso .
2,5 g	**pt/e/s/t/Estrelinha.html**: estrelinha . flor de ampla distribuição na américa do sul , incluindo todo o brasil . tal espécie mede cerca de 8,6 cm de comprimento e 2,5 g , com cauda profundamente bifurcada e garganta vermelho - rosada ametística . também é conhecida pelos nomes de beija
menos de 1 g	**FSP951006-102**: entre os destaques da exposição , ovos de beija - flor (com apenas 11 milímetros de diâmetro e pesando menos de 1 g) e uma réplica perfeita de um ovo da extinta ave - elefante ,
Question #9 - Onde ficava a Gália Cisalpina?	
gália romana	**pt/g/á/l/Gália_romana.html**: gália romana . a cidadania romana foi estendida à gália transpadana por césar em 49 ac e toda a gália cisalpina foi incorporada à itália por augusto ,
itália	**pt/p/r/i/Primeira_Rebelião_Judaica_ec15.html**: primeira rebelião judaica . os legionários dessa unidade , homens da gália cisalpina , no norte da itália , eram sólidos e confiáveis .
ac	**pt/g/á/l/Gália_romana.html**: gália romana . a cidadania romana foi estendida à gália transpadana por césar em 49 ac e toda a gália cisalpina foi incorporada à itália por augusto ,
Question #10 - Quantas províncias tem a Catalunha?	
quatro províncias	**PUBLICO-19951119-035**: situada no nordeste , entre os pirenéus e o mediterrâneo , a catalunha tem 6,2 milhões de habitantes , ou seja , 16 por cento do total nacional , repartidos em quatro províncias (barcelona , gerona , lérida e tarragona) .
cinquenta províncias	**pt/e/s/p/Espanha.html**: espanha . país basco e catalunha) gozam da condição de " nacionalidade histórica " reconhecida na constituição , juntamente com um " estatuto de autonomia " o que reverte num maior poder e capacidade de decisão e soberania com respeito às outras comunidades . as comunidades dividem - se ainda em cinquenta províncias .
Question #11 - Qual é a montanha mais alta do México?	
pico de orizaba	**pt/p/i/c/Pico_de_Orizaba_8a80.html**: pico de orizaba . pico de orizaba pico de orizaba , a montanha mais alta do méxico

maior palácio	**pt/r/e/c/Recordes_mundiais.html**: recordes mundiais . alta montanha - russa : kingda ka ; maior palácio : palácio do parlamento , romênia , 350 000 m ^{2} ; maior palácio residencial : istana nurul iman , brunei , 200 000 m ^{2} . monumentos maior : grande pirâmide de cholula , méxico ,
terceiro vulcão mais alto no hemisfério ocidental	**pt/p/i/c/Pico_de_Orizaba_8a80.html**: pico de orizaba . o vulcão pico de orizaba ou citlaltépetl (do nahuatl citlalli = estrela , e tepetl = mountanha) localiza - se no méxico e é a montanha mais alta deste país , a terceira na américa do norte e o terceiro vulcão mais alto no hemisfério ocidental .
Question #12 - E do Japão?	
famoso monte fuji	**pt/j/a/p/Japão.html**: japão . a montanha mais alta do japão é o famoso monte fuji , com 3 776 m de altitude .
sudoeste de kyushu até perto de taiwan	**pt/j/a/p/Japão.html**: japão . japão inclui cerca de 3 000 ilhas menores , parte das quais constituem as ilhas ryukyu , que se estendem a sudoeste de kyushu até perto de taiwan . cerca de 73 % do país é montanhoso , com uma cordilheira a ocupar o centro de cada uma das ilhas principais . a montanha mais alta
ilhas	**pt/j/a/p/Japão.html**: japão . japão inclui cerca de 3 000 ilhas menores , parte das quais constituem as ilhas ryukyu , que se estendem a sudoeste de kyushu até perto de taiwan . cerca de 73 % do país é montanhoso , com uma cordilheira a ocupar o centro de cada uma das ilhas principais . a montanha mais alta
Question #13 - Onde fica Saint-Exupéry?	
pequeno príncipe	**FSP950324-128**: paulo coelho colunista da folha o autor de " o pequeno príncipe " , saint - exupéry , foi alguns anos comandante de um pequeno aeroporto ao norte da áfrica .
hoje	**PUBLICO-19940731-007**: mantendo a dúvida acesa . faz hoje cinquenta anos que saint - exupéry , que se considerava primeiro aviador e escritor em segundo lugar , desapareceu numa missão de reconhecimento sobre a França ocupada , algures entre a córsega e a riviera francesa .
nova iorque à terra do fogo	**PUBLICO-19940731-007**: exupéry escreveu « terre des hommes » , depois de quase ter morrido ao tentar fazer a ligação aérea de nova iorque à terra do fogo , em 1 938 . na sua última carta , onde dizia ter nascido « para ser jardineiro » , saint

Question #14 - Qual a altura do Kebnekaise?	
cerca de 150 quilómetros	**pt/k/e/b/Kebnekaise.html**: kebnekaise . o kebnekaise situa - se na lapónia , a cerca de 150 quilómetros (ca .
2 103 metros	**pt/k/e/b/Kebnekaise.html**: kebnekaise . o maciço do kebnekaise , que faz parte das montanhas escandinavas , tem dois picos , dos quais o mais a sul atinge 2 103 metros (ca .
2 117 m	**pt/k/i/r/Kiruna.html**: kiruna . o monte kebnekaise , no município de kiruna , é a montanha mais alta da suécia e tem 2 117 m de altitude .
Question #15 - Quem escreveu Fernão Capelo Gaivota?	
neil diamond	**PUBLICO-19950224-130**: tudo à sombra da banda sonora composta por neil diamond para o filme fernão capelo gaivota ... os « brancos » do carrefour são um êxito . e outras empresas de distribuição copiam - lhe a receita .
carrefour	**PUBLICO-19950224-130**: tudo à sombra da banda sonora composta por neil diamond para o filme fernão capelo gaivota ... os « brancos » do carrefour são um êxito . e outras empresas de distribuição copiam - lhe a receita .
Question #16 - O que é um menir?	
menir - monumentos pré - históricos em pedras , cravadas verticalmente no solo (ortóstatos) , às vezes de tamanho bem elevado (megalito denominado menir) . a palavra menir foi adotada , através do francês , pelos arqueólogos do século xix com base nas palavras do bretão significando men = pedra e hir = longa (comparar com o gaélico : maen hir = pedra longa) .	**pt/m/e/n/Menir.html**: menir - monumentos pré - históricos em pedras , cravadas verticalmente no solo (ortóstatos) , às vezes de tamanho bem elevado (megalito denominado menir) . a palavra menir foi adotada , através do francês , pelos arqueólogos do século xix com base nas palavras do bretão significando men = pedra e hir = longa (comparar com o gaélico : maen hir = pedra longa) . no bretão moderno usa - se a palavra peulvan .

Question #17 - Em que ano é que Ernie Els venceu o Dubai Open?	
junho	**PUBLICO-19950104-013**: ernie els , que da 20 [a] posição escalou até ao sexto lugar , depois de ter acumulado 1 108 pontos , mais quatro do que price . um ano inesquecível para o golfe africano , portanto . para além do seu triunfo no open dos eua , em junho , els venceu o dubai
Question #18 - Quantos ossos têm a face?	
três ossos	**pt/o/u/v/Ouvido_médio.html**: ouvido médio . face da pessoa . anatomia comparativa os mamíferos são os únicos que têm três ossos no ouvido . a bigorna e o estribo desenvolvem de ossos da mandíbula , e permitem a melhor detecção do som . alguns mamíferos - como os gatos - tem um ouvido médio maior localizado em um osso
um osso	**pt/f/í/b/Fíbula.html**: fíbula . a fíbula , chamada anteriormente de perônio , é um osso longo situado na face externa da perna , da qual constitui o esqueleto , junto com a tíbia .
14 ossos	**pt/c/a/b/Cabeça.html**: cabeça . parietal (2) temporal (2) frontal occipital esfenóide etmóide ossos da face a face é constituída por 14 ossos , sendo que seis são pares e dois são ímpares .
Question #19 - Quando começou o Neolítico?	
3 500 ac	**pt/d/u/n/Dundalk.html**: dundalk . geografia história começou a ser habitada no período neolítico (3 500 ac) e mais tarde pelos os celtas (500 d. c) antes da chegada dos normandos em 1 169 .
1 169	**pt/d/u/n/Dundalk.html**: dundalk . geografia história começou a ser habitada no período neolítico (3 500 ac) e mais tarde pelos os celtas (500 d. c) antes da chegada dos normandos em 1 169 .
1 912	**FSP940711-137**: começou a desmoronar e , pouco depois , leningrado foi desrebatizada , voltando a se chamar são petersburgo . nos bálcãs , o futuro voltou ao passado , primeiro a 1 941 e , agora , a 1 912 ou mesmo antes . no cáucaso e na áfrica central o que está ocorrendo é um renascimento do neolítico
Question #20 - Quando nasceu Thomas Mann?	
6 de junho de 1 875	**pt/t/h/o/Thomas_Mann_7f52.html**: thomas mann nent thomas mann (6 de junho de 1 875 - 12 de agosto de 1 955) foi um romancista alemão , considerado por alguns como um dos maiores romancistas do século xx , tendo recebido o prémio nobel da literatura em 1 929 .

século xx	**pt/t/h/o/Thomas_Mann_7f52.html**: thomas mann nent thomas mann (6 de junho de 1 875 - 12 de agosto de 1 955) foi um romancista alemão , considerado por alguns como um dos maiores romancistas do século xx , tendo recebido o prémio nobel da literatura em 1 929 .
12 de agosto de 1 955	**pt/t/h/o/Thomas_Mann_7f52.html**: thomas mann nent thomas mann (6 de junho de 1 875 - 12 de agosto de 1 955) foi um romancista alemão , considerado por alguns como um dos maiores romancistas do século xx , tendo recebido o prémio nobel da literatura em 1 929 .
Question #21 - E quando morreu?	
12 de agosto de 1 955	**pt/t/h/o/Thomas_Mann_7f52.html**: thomas mann nent thomas mann (6 de junho de 1 875 - 12 de agosto de 1 955) foi um romancista alemão , considerado por alguns como um dos maiores romancistas do século xx , tendo recebido o prémio nobel da literatura em 1 929 .
século xx	**pt/t/h/o/Thomas_Mann_7f52.html**: thomas mann nent thomas mann (6 de junho de 1 875 - 12 de agosto de 1 955) foi um romancista alemão , considerado por alguns como um dos maiores romancistas do século xx , tendo recebido o prémio nobel da literatura em 1 929 .
6 de junho de 1 875	**pt/t/h/o/Thomas_Mann_7f52.html**: thomas mann nent thomas mann (6 de junho de 1 875 - 12 de agosto de 1 955) foi um romancista alemão , considerado por alguns como um dos maiores romancistas do século xx , tendo recebido o prémio nobel da literatura em 1 929 .
Question #22 - A que partido pertence Zapatero?	
partido socialista operário espanhol	**pt/p/a/r/Partido_Socialista_Operário_Espanhol_b6e5.html**: partido socialista operário espanhol . com o apoio de outros partidos , zapatero foi eleito presidente de governo pelo congresso .
secretário geral	**pt/p/a/r/Partido_Socialista_Operário_Espanhol_b6e5.html**: partido socialista operário espanhol . no congresso do partido celebrado no verão de 2 000 , foi eleito secretário geral o então desconhecido josé luis rodríguez zapatero , em detrimento de outros candidatos do partido mais conhecidos .
sexo	**pt/f/e/l/Felipe_de_Bourbon_cc2f.html**: felipe de bourbon . espera - se que o actual governo espanhol , liderado por josé luis rodríguez zapatero , promova a alteração constitucional que permitiria a leonor no futuro ser rainha de espanha , mesmo que venha a ter irmãos mais jovens do sexo masculino .

Question #23 - Quem é FHC?	
fernando henrique cardoso (rio de janeiro , 18 de junho de 1 931) é um sociólogo , professor , e político brasileiro formado pela universidade de são paulo .	**pt/f/e/r/Fernando_Henrique_Cardoso_7443.html**: fernando henrique cardoso (rio de janeiro , 18 de junho de 1 931) é um sociólogo , professor , e político brasileiro formado pela universidade de são paulo . é também comumente conhecido por seu acrônimo fhc . foi senador por são paulo , ministro das relações exteriores e ministro da fazenda no governo de itamar franco . foi presidente do brasil por dois mandatos consecutivos , de 1 de janeiro de 1 995 a 31 de dezembro de 2 002 . foi ideólogo do mdb e teve uma participação decisiva nos bastidores das diretas - já e na eleição no colégio eleitoral de tancredo neves à presidência da república .
Question #24 - Quem foi Álvaro de Campos?	
álvaro de campos (1 890 - 1 935) é um dos heterónimos mais conhecidos de fernando pessoa . nascido em tavira , teve a educação de liceu comum de sua época , posteriormente foi para a escócia estudar engenharia mecânica , e depois engenharia naval .	**pt/á/l/v/Álvaro_de_Campos_9b10.html**: álvaro de campos (1 890 - 1 935) é um dos heterónimos mais conhecidos de fernando pessoa . nascido em tavira , teve a educação de liceu comum de sua época , posteriormente foi para a escócia estudar engenharia mecânica , e depois engenharia naval . em férias fez uma viagem ao oriente onde escreveu o opiário . entre todos os heterónimos , campos foi o único a manifestar fases poéticas diferentes ao longo de sua obra . era um engenheiro de educação inglesa e origem portuguesa , mas sempre com a sensação de ser um estrangeiro em qualquer parte do mundo .
Question #25 - Diga uma das suas obras.	
obras - primas	**PUBLICO-19950909-093**: de maistre escreveu na rússia « os serões de são petersburgo » , que alguns consideram uma das suas obras - primas .
considerar obra	**PUBLICO-19950909-093**: de maistre escreveu na rússia « os serões de são petersburgo » , que alguns consideram uma das suas obras - primas .
usos e costumes locais , tendo	**pt/b/a/i/Baião_(Portugal)_d72b.html**: baião (portugal) . nas suas paisagens , nos usos e costumes locais , tendo escrito , a propósito , uma das suas obras mais conhecidas : « a cidade e as serras » . com esta obra , imortalizou baião e toda a região circundante ,
Question #26 - Os tucanos são membros de que partido?	
partido democrático brasileiro) , era visto com olhares receosos pela esquerda conservadora do partido do trabalhadores	**pt/h/e/n/Henrique_Meirelles_5e4a.html**: henrique meirelles . na época da submissão do seu nome ao senado , o então tucano (membro do partido democrático brasileiro) , era visto com olhares receosos pela esquerda conservadora do partido do trabalhadores ,

henrique meirelles	**pt/h/e/n/Henrique_Meirelles_5e4a.html**: henrique meirelles . na época da submissão do seu nome ao senado , o então tucano (membro do partido democrático brasileiro) , era visto com olhares receosos pela esquerda conservadora do partido do trabalhadores ,
presidente lula	**pt/h/e/n/Henrique_Meirelles_5e4a.html**: henrique meirelles . partido dos trabalhadores (pt) do recém eleito presidente lula , fizeram dele nome altamente cotado a um dos mais altos cargos no conhecido sistema financeiro nacional (sfn) brasileiro , superado apenas pelo ministro da fazenda . na época da submissão do seu nome ao senado , o então tucano (membro
Question #27 - Quem foi Pierre Larousse?	
pierre athanase larousse (toucy , 23 de outubro de 1 817 - paris , 3 de janeiro de 1 875) foi um pedagogo , editor e enciclopedista francês .	**pt/p/i/e/Pierre_Larousse_94be.html**: pierre athanase larousse (toucy , 23 de outubro de 1 817 - paris , 3 de janeiro de 1 875) foi um pedagogo , editor e enciclopedista francês . sua sepultura se encontra no cemitério de montparnasse . ligações externas business week online , november 11,2002 , commentary : ' french publisher for sale : no foreigners , please ' (em inglês) categorias : !
Question #28 - O que é a Brabançonne?	
hino nacional da bélgica	**pt/h/i/n/Hino_nacional_da_Bélgica_4a7b.html**: hino nacional da bélgica . la brabançonne é o hino nacional da bélgica .
lista de hinos nacionais e regionais	**pt/l/i/s/Lista_de_hinos_nacionais_e_regionais.html**: lista de hinos nacionais e regionais . brabançonne
coelho com cerveja de leffe	**FSP950210-125**: escalopes de foie gras de pato com frutos silvestres , coelho com cerveja de leffe e ameixa , perdiz à brabançonne e queijo branco com mel e amêndoas .
Question #29 - Onde vivem as aves tucanos?	
emas	**pt/g/e/o/Geografia_do_Paraguai_e386.html**: geografia do paraguai . também existem muitas espécies de pássaros e aves tropicais , como o íbis , a garça , o tucano , pombos , perdizes , emas , seriemas e papagaios .
pica	**pt/i/t/a/Itajubá.html**: itajubá . algumas espécies de aves encontradas na região são : bem - te - vi , fogo - apagou , juriti , maritaca , pica - pauzinho , risadinha , rolinha , sabiá - amarelo , sanhaço , tiê - preto e o tucano .
papagaios	**pt/g/e/o/Geografia_do_Paraguai_e386.html**: geografia do paraguai . também existem muitas espécies de pássaros e aves tropicais , como o íbis , a garça , o tucano , pombos , perdizes , emas , seriemas e papagaios .

Question #30 - Em que estado brasileiro habitam os tucanos?	
novo presidente do brasil	**FSP950101-155**: como o novo presidente do brasil em meio a uma atmosfera de grande otimismo , que obscurece até mesmo a ascensão dos 27 novos governadores que também chegam ao poder ungidos pela maior eleição a que o país assistiu . esse otimismo contamina o próprio tucano .
alto rio	**pt/t/u/c/Tucanos.html**: tucanos . em etnologia , o termo tucanos , além de designar o grupos indígenas cujas línguas pertencem à família lingüística tucano , remete ainda ao grupo indígena que habita no noroeste do estado brasileiro do amazonas , mais precisamente as áreas indígenas alto rio negro , médio rio negro i , médio rio negro ii ,
henrique cardoso	**FSP941106-022**: economista que dirige programa de governo tucano deve ocupar secretaria do planejamento , que terá mais poder gabriela wolthers enviada especial ao paraguai o presidente eleito do brasil , fernando henrique cardoso ,
Question #31 - Quem disse "alea iacta est"?	
rio rubicão	**pt/r/i/o/Rio_Rubicão_fc5d.html**: rio rubicão . segundo suetônio , césar teria então proferido a famosa frase alea iacta est (" a sorte está lançada ") .
césar	**pt/r/i/o/Rio_Rubicão_fc5d.html**: rio rubicão . segundo suetônio , césar teria então proferido a famosa frase alea iacta est (" a sorte está lançada ") .
suetônio	**pt/r/i/o/Rio_Rubicão_fc5d.html**: rio rubicão . segundo suetônio , césar teria então proferido a famosa frase alea iacta est (" a sorte está lançada ") .
Question #32 - Ao atravessar que rio?	
rio rubicão	**pt/r/i/o/Rio_Rubicão_fc5d.html**: rio rubicão . segundo suetônio , césar teria então proferido a famosa frase alea iacta est (" a sorte está lançada ") . o mesmo autor também descreve como césar parecia indeciso ao se aproximar do rio e atribui a decisão de atravessar a uma aparição sobrenatural .
lei	**pt/r/i/o/Rio_Rubicão_fc5d.html**: rio rubicão . quando júlio césar atravessou o rubicão , em 49 ac , presumivelmente em 10 de janeiro do calendário romano , em perseguição a pompeu , violou a lei e tornou inevitável o conflito armado . segundo suetônio , césar teria então proferido a famosa frase alea iacta est
Question #33 - Que político é conhecido como Iznogoud?	
nicolas sarkozy	**pt/i/z/n/Iznogoud.html**: iznogoud . politicos , como nicolas sarkozy , são chamado " iznogoud " por sua ambicão e pequenha tamanho .
vizir	**pt/i/z/n/Iznogoud.html**: iznogoud . trata - se do grão - vizir iznogoud que ambiciona ser califa e elabora diversos planos para usurpar o trono do califa haroun el poussah (harun al sindar ,

Question #34 - O que é uma cítara?	
a cítara é um instrumento musical de várias cordas presas sobre um arco de madeira , com ou sem caixa de ressonância , que se tocavam com ambas as mãos .	**pt/c/í/t/Cítara.html**: a cítara é um instrumento musical de várias cordas presas sobre um arco de madeira , com ou sem caixa de ressonância , que se tocavam com ambas as mãos . a lenda diz que o imperador nero queimou roma tocando uma cítara . composta por onze cordas de ressonância e sete que são tocadas , é muito leve , feita geralmente com duas cabaças , uma para o corpo e uma acoplada no braço do instrumento para servir apenas como ressonância . as cordas são feitas de cobre ou bronze . é afinada em quintas , entre os tons dó , dó # e ré .
Question #35 - O que era o A6M Zero?	
o a 6 m zero foi o principal caça da marinha japonesa durante toda a segunda guerra mundial , e ganhou reputação de invencível no início da participação nipônica no conflito , com poder de manobra , alcance e razão de subida inigualáveis por qualquer caça ocidental , tanto de terra quanto embarcado .	**pt/m/i/t/Mitsubishi_A6M_Zero_8a24.html**: o a 6 m zero foi o principal caça da marinha japonesa durante toda a segunda guerra mundial , e ganhou reputação de invencível no início da participação nipônica no conflito , com poder de manobra , alcance e razão de subida inigualáveis por qualquer caça ocidental , tanto de terra quanto embarcado . porém , logo que modelos aperfeiçoados entraram em serviço , começou a ficar obsoleto , sendo , porém mantido na linha de frente , por falta de um substituto . categorias : !
Question #36 - Diga um gás nobre.	
hélio	**pt/s/u/p/Superfluidez.html**: superfluidez . modelo teórico por ser um gás nobre , o hélio exibe pouca interação intermolecular .
nível de energia	**pt/s/é/r/Série_química.html**: série química . o hélio é um gás nobre cuja configuração ocupa um único nível de energia : 1 s 2 .
série química	**pt/s/é/r/Série_química.html**: série química . o hélio é um gás nobre cuja configuração ocupa um único nível de energia : 1 s 2 .
Question #37 - E um não-metal	
classe , série química	**pt/b/r/o/Bromo.html**: bromo . símbolo , número bromo , br , 35 classe , série química não - metal , representativo (halogênio) grupo , período , bloco 17 (viia) , 4 , p densidade , dureza 3 119 kg / m 3 (300 k) ,
dureza 3 119 kg	**pt/b/r/o/Bromo.html**: bromo . símbolo , número bromo , br , 35 classe , série química não - metal , representativo (halogênio) grupo , período , bloco 17 (viia) , 4 , p densidade , dureza 3 119 kg / m 3 (300 k) ,

grupo	**pt/b/r/o/Bromo.html**: bromo . símbolo , número bromo , br , 35 classe , série química não - metal , representativo (halogênio) grupo , período , bloco 17 (viia) , 4 , p densidade , dureza 3 119 kg / m 3 (300 k) ,
Question #38 - Qual é o asteróide número 4?	
planeta anão ceres	**pt/c/i/n/Cintura_de_asteróides.html**: cintura de asteróides . asteróides , e estima - se que o número alcance os milhões . cerca de 220 deles são maiores que 100 km . a massa total da cintura , contando com o planeta anão ceres é estimada em 3,0 - 3,6 predefinição : e , o que é cerca de 4
lista de asteróides	**pt/l/i/s/Lista_de_asteróides.html**: lista de asteróides . os asteróides geralmente recebem nomes sistemáticos (como " 1 989 ac ") e , entretanto , um número (como 4 179) .
recebem nomes sistemáticos	**pt/l/i/s/Lista_de_asteróides.html**: lista de asteróides . os asteróides geralmente recebem nomes sistemáticos (como " 1 989 ac ") e , entretanto , um número (como 4 179) .
Question #39 - Qual a capital da Picardia?	
capital é a cidade de beauvais	**pt/o/i/s/Oise.html**: oise . oise é um departamento da França localizado na região picardia . sua capital é a cidade de beauvais .
amiens	**PUBLICO-19940506-109**: um pouco mais para sul , amiens capital da picardia foi a grande derrotada .
França	**pt/o/i/s/Oise.html**: oise . oise é um departamento da França localizado na região picardia . sua capital é a cidade de beauvais .
Question #40 - Quando reinou Isabel II de Castela?	
1 833	**pt/r/e/i/Reino_do_Algarve_afab.html**: reino do algarve . dado ter adquirido em 1 262 os restos do reino de niebla / algarve , situados já além do odiana - os demais reis de castela , e depois da espanha , até à subida ao trono da rainha isabel ii (1 833) ,
20 de janeiro	**pt/2/0/_/20_de_Janeiro_cb3b.html**: 20 de janeiro . 1 479 - fernando ii torna - se rei de aragão e passa a reinar em conjunto com a sua mulher isabel i de castela a maior parte da península ibérica .
1 479	**pt/2/0/_/20_de_Janeiro_cb3b.html**: 20 de janeiro . 1 479 - fernando ii torna - se rei de aragão e passa a reinar em conjunto com a sua mulher isabel i de castela a maior parte da península ibérica .
Question #41 - Quem é Narcís Serra?	
josé barrionuevo	**FSP951019-058**: os ex - ministros narcís serra (defesa) e josé barrionuevo (interior) também serão investigados .

felipe gonzález	**FSP950629-051**: felipe gonzález , aceitou a demissão de seu vice , narcís serra , e do ministro da defesa , julián garcía vargas .
governo o vice	**FSP950708-039**: o escândalo já tirou do governo o vice - primeiro - ministro , narcís serra , e o ministro da defesa , julián garcía vargas .
Question #42 - Como se chamava o cavalo do Dom Quixote?	
rocinante	**pt/d/o/m/Dom_Quixote_9ddd.html**: dom quixote . as figuras de dom quixote , de sancho pança e do cavalo de dom quixote , rocinante , depressa conquistaram a imaginação popular .
diabo autor : josé	**PUBLICO-19950102-077**: título : a cavalo no diabo autor : josé cardoso pires editor : dom quixote 206 pgs .
sancho pança	**pt/d/o/m/Dom_Quixote_9ddd.html**: dom quixote . as figuras de dom quixote , de sancho pança e do cavalo de dom quixote , rocinante , depressa conquistaram a imaginação popular .
Question #43 - Qual é a capital do estado de Nova York?	
eua	**FSP950608-131**: do enviado especial aos eua com o tema ‘ ‘ nova york - capital do mundo ” e o símbolo ‘ ‘ ny 95 ” ,
brasil	**FSP940203-030**: ” continuamos otimistas porque o brasil está encontrando seu caminho ” , disse jean vandewalle , da aliança capital de nova york .
londres	**FSP941227-029**: para 31 dias (capital de giro) : entre 67 % e 78 % ao ano . no exterior feriados em nova york e em londres .
Question #44 - Quais são as províncias da Irlanda?	
condado	**pt/c/o/n/Condado_de_Sligo_e4cf.html**: condado de sligo . sligo (sligeach em gaélico irlandês) é um condado na província de connacht , a oeste da irlanda .
república da irlanda	**pt/b/a/l/Ballinskelligs.html**: ballinskelligs . república da irlanda . ballinskelligs baile na sceilge província
norte	**PUBLICO-19941014-159**: irlanda do norte (que é uma província britânica) ,
Question #45 - Que instrumento tocava Ringo Starr?	
vocal	**pt/t/h/e/The_Beatles_3737.html**: the beatles . paul mccartney (baixo e vocal) , george harrison (guitarra e vocal) e ringo starr (bateria e vocal) , obtiveram notoriedade até hoje inédita para uma banda musical . atingiram
guitarra	**pt/h/e/r/Here_Comes_the_Sun_456a.html**: here comes the sun . outra composição dele , um solo de guitarra soa durante toda a música . é uma das músicas mais famosas e regravadas dos beatles , já ganhando até versão em orquestra de flautas . the beatles john lennon | paul mccartney | george harrison | ringo starr

flautas	**pt/h/e/r/Here_Comes_the_Sun_456a.html**: here comes the sun . outra composição dele , um solo de guitarra soa durante toda a música . é uma das músicas mais famosas e regravadas dos beatles , já ganhando até versão em orquestra de flautas . the beatles john lennon | paul mccartney | george harrison | ringo starr
Question #46 - Que papa sucedeu a Leão X?	
papa clemente vii	**pt/p/a/p/Papa_Clemente_VII_6fe1.html**: papa clemente vii . morto leão x , teve papel decisivo na escolha insperada do papa adriano vi , a quem sucederia no conclave de novembro de 1 523 .
adriano vi	**pt/p/a/p/Papa_Clemente_VII_6fe1.html**: papa clemente vii . morto leão x , teve papel decisivo na escolha insperada do papa adriano vi , a quem sucederia no conclave de novembro de 1 523 .
cardeal	**pt/p/a/p/Papa_Leão_X_a580.html**: papa leão x . pode - se considerar uma censura a leão x a eleição , para sucedê - lo , do cardeal de utrecht , adriano florenzo como papa adriano vi) , conhecido apenas por ter sido preceptor do jovem carlos v , mas que mantinha feroz polêmica com lutero e ,
Question #47 - Quem é o pato mais rico do mundo?	
tio patinhas	**pt/a/_/s/A_saga_do_Tio_Patinhas_b89c.html**: a saga do tio patinhas . que vão embora e deixam uma surpresa para patinhas : ele se tornara o pato mais rico do mundo .
carl barks	**pt/m/a/c/Mac_Mônei_b190.html**: mac mônei . mac mônei é supostamente o " segundo pato mais rico do mundo " , tal como surgiu nos quadrinhos de carl barks dos anos 50 e 60 ,
mac mônei	**pt/m/a/c/Mac_Mônei_b190.html**: mac mônei . mac mônei é supostamente o " segundo pato mais rico do mundo " , tal como surgiu nos quadrinhos de carl barks dos anos 50 e 60 ,
Question #48 - Quem são os sobrinhos do Pato Donald?	
tio patinhas	**pt/t/i/o/Tio_Patinhas_0ad9.html**: tio patinhas . este é um traço em comum com seu sobrinho pato donald .
huguinho , zezinho e luisinho	**pt/t/i/o/Tio_Patinhas_0ad9.html**: tio patinhas . na história , patinhas convida seu sobrinho pato donald e sobrinhos - netos huguinho , zezinho e luisinho para sua cabana nas montanhas ,
maga patalógika	**pt/m/a/g/Maga_Patalógika_d03d.html**: maga patalógika . maga patalójika também era personagem semi - regular no seriado de animação ducktales , opondo - se a pato donald e seus sobrinhos quando não estão com patinhas .

Question #49 - E a namorada dele?	
margarida (banda desenhada	**pt/m/a/r/Margarida_(banda_desenhada).html**: margarida (banda desenhada) . ela é a namorada do pato donald , embora o gastão também seja um pretendente .
pequena sereia	**FSP951006-101**: pato donald) e outros mais novos , como pocahontas e a pequena sereia .
melhor amigo é o pateta	**pt/m/i/c/Mickey_Mouse_6607.html**: mickey mouse . na banda desenhada actual , o seu melhor amigo é o pateta , a sua mascote é o pluto e a sua namorada é a minnie . ele costumava andar com o pato donald (ambos inclusive moram na mesma cidade) , mas os universos dos dois são separados .
Question #50 - Qual a profissão dele?	
era inspirado pelas próprias experiências	**pt/c/a/r/Carl_Barks_0a4e.html**: carl barks . donald vagueia de trabalho em trabalho , habitualmente sem sucesso , no que era inspirado pelas próprias experiências de carl . até quando tinha êxito , era apenas um sucesso temporário antes de outra decepção para o pato .
carl barks	**pt/c/a/r/Carl_Barks_0a4e.html**: carl barks . donald vagueia de trabalho em trabalho , habitualmente sem sucesso , no que era inspirado pelas próprias experiências de carl . até quando tinha êxito , era apenas um sucesso temporário antes de outra decepção para o pato .
sem sucesso	**pt/c/a/r/Carl_Barks_0a4e.html**: carl barks . donald vagueia de trabalho em trabalho , habitualmente sem sucesso , no que era inspirado pelas próprias experiências de carl . até quando tinha êxito , era apenas um sucesso temporário antes de outra decepção para o pato .
Question #51 - O que é a paella?	
a paelha (em castelhano e catalão paella) é o prato típico da costa mediterrânea espanhola confeccionado com arroz , mariscos , galinha e vegetais . como tempero é usado o açafrão e cozinhado numa frigideira de fundo ligeiramente abaulado , chamada paella .	**pt/p/a/e/Paelha.html**: a paelha (em castelhano e catalão paella) é o prato típico da costa mediterrânea espanhola confeccionado com arroz , mariscos , galinha e vegetais . como tempero é usado o açafrão e cozinhado numa frigideira de fundo ligeiramente abaulado , chamada paella . os ingredientes variam conforme a região de espanha . categorias : !

Question #52 - Que países abrange a Lapónia?	
finlândia , no norte do país	**pt/p/r/o/Província_da_Lapónia_c5e8.html**: província da lapónia . a lapónia é uma das províncias da finlândia , no norte do país . tem 98 946 km ^{2} e 187 777 habitantes (2 002) .
noruega e islândia , o país menos populoso da europa	**pt/d/e/m/Demografia_da_Finlândia_5b5f.html**: demografia da finlândia . a finlândia tem cinco milhões de habitantes e uma densidade populacional de 17 habitantes por quilômetro quadrado . isto faz da finlândia , depois da noruega e islândia , o país menos populoso da europa . na lapónia finlandesa , no ártico ,
capital é a cidade de murmansk	**pt/m/u/r/Murmansk_(oblast).html**: murmansk (oblast) . fica localizado na região noroeste do país . possui uma área de 144 900 km ^{2} e uma população de 892 534 , segundo o censo de 2 002 . a sua capital é a cidade de murmansk . geografia o oblast fica na península de kola e faz parte da lapónia ,
Question #53 - O que é a açorda?	
a açorda à alentejana é uma sopa típica do alentejo - ao contrário da maioria das sopas , esta não é cozinhada , mas basicamente pão em água quente temperada .	**pt/a/ç/o/Açorda_à_alentejana.html**: a açorda à alentejana é uma sopa típica do alentejo - ao contrário da maioria das sopas , esta não é cozinhada , mas basicamente pão em água quente temperada . (outra sopa que não é cozinhada é o gaspacho espanhol) .
Question #54 - O que é o feta?	
o feta é uma variedade de queijo típica da grécia , fabricado com leite de cabra ou de ovelha , exclusivamente . é um queijo branco , farelento e levemente salgado .	**pt/f/e/t/Feta.html**: o feta é uma variedade de queijo típica da grécia , fabricado com leite de cabra ou de ovelha , exclusivamente . é um queijo branco , farelento e levemente salgado . ligações externas queijo feta (em inglês) categorias : !
Question #55 - De que país é originário?	
grécia	**pt/f/e/t/Feta.html**: feta . o feta é uma variedade de queijo típica da grécia , fabricado com leite de cabra ou de ovelha , exclusivamente .
edam	**pt/q/u/e/Queijo.html**: queijo . [1] tipos de queijos appenzeller brie camembert cheddar chèvre cottage edam feta
portugueses	**PUBLICO-19950729-142**: assim , não sucederá aos queijos portugueses mais apetecidos - - serra , serpa e azeitão , entre outros - - o mesmo que aconteceu ao conhecido feta , um queijo grego de leite de ovelha ou de ovelha e cabra , que é actualmente produzido pelas indústrias de lacticínios dinamarquesa , alemã e francesa ,

Question #56 - Em que ano foi construída a sinagoga de Curação?	
Question #57 - Com que idade o Mequinho foi campeão brasileiro de xadrez?	
13 anos	**FSP940205-141**: a comparação com mequinho , campeão brasileiro aos 13 anos , é inevitável , mas giovanni procura fugir dela . ao contrário de mequinho , evita se dedicar ao xadrez em tempo integral .
dois anos	**FSP940205-144**: sua evolução foi semelhante às de fischer e do russo garry kasparov , campeão do mundo pela dissidente pca (associação de xadrez profissional) . mequinho foi campeão brasileiro aos 13 anos , em 1 965 , e campeão sul - americano dois anos depois .
14 anos	**pt/2/7/_/27_de_Dezembro_5328.html**: 27 de dezembro . mequinho - torna - se aos 14 anos campeão brasileiro de xadrez 1 978 - a espanha aprova a constituição democrática ,
Question #58 - Quem dirigiu o Japão durante a Segunda Guerra Mundial?	
toshiro mifune	**pt/t/o/s/Toshiro_Mifune_6f4c.html**: toshiro mifune . que o dirigiu dezesseis vezes - e em produções americanas . mifune trabalhou como fotógrafo em xangai e serviu ao exército do japão durante a segunda guerra mundial .
hirohito	**pt/7/_/d/7_de_Janeiro_0770.html**: 7 de janeiro . guitarrista e compositor brasileiro 1 989 - hirohito , imperador do japão , aos 87 anos , que conduziu o japão durante a segunda guerra mundial e foi o último monarca de seu país a ser reconhecido como uma divindade .
estados unidos da américa	**pt/b/a/l/Balão_bomba.html**: balão bomba . estima - se que cerca de 9 000 desses balões tenham sido lançados pelo japão durante a segunda guerra mundial , a partir de novembro de 1 944 , para atacar cidades , florestas e fazendas dos estados unidos da américa ,
Question #59 - Quantas repúblicas formavam a URSS?	
seis repúblicas	**pt/g/u/e/Guerra_da_Bósnia_96b6.html**: guerra da bósnia . nacionalismo com o fim dos regimes socialistas , a partir da desintegração da urss , emergem as diferenças étnicas , culturais e religiosas entre as seis repúblicas que formam a iugoslávia , impulsionando movimentos pela independência .
15 repúblicas	**FSP940911-062**: o balneário de stálin conflitos étnicos e caos econômico arrasam a geórgia do enviado especial à geórgia entre as 15 repúblicas que formavam a urss , a geórgia chamava a atenção por suas estações balneárias e por seus vinhos .

três repúblicas	**PUBLICO-19940131-030**: fazer . p. - - uma europa do atlântico até vladivostoque , no pacífico ? r. - - já está decidido que as repúblicas asiáticas da ex - urss não são europa . estão em causa as três repúblicas do cáucaso .
Question #60 - Em que país fica a Ossétia do Norte?	
república soviética da geórgia	**pt/o/s/s/Ossétia.html**: ossétia . além disso , a ossétia do sul , território ligado à ex - república soviética da geórgia , luta para ser anexada a ossétia do norte , a qual pertence à federação russa .
russos	**PUBLICO-19941214-050**: soldados e tanques russos continuavam a avançar sobre grozni , encontrando uma resistência cada vez maior . a não muitos quilómetros , na república da ossétia do norte ,
russa	**pt/o/s/s/Ossétia_do_Norte_46ca.html**: ossétia do norte . ossétia do norte região da federação russa . a sua capital é vladikavkaz .
Question #61 - E a Ossétia do Sul?	
subdivisões da geórgia	**pt/s/u/b/Subdivisões_da_Geórgia_5564.html**: subdivisões da geórgia . ossétia do sul ossétia do sul a república da ossétia do sul é uma república independente dentro da geórgia .
lista de hinos nacionais e regionais	**pt/l/i/s/Lista_de_hinos_nacionais_e_regionais.html**: lista de hinos nacionais e regionais . nigeria ' s call obey noruega ja , vi elsker dette landet (oui , nous aimons ce pays) nova zelândia god defend new zealand ossétia do sul hino de ossétia do
ossétia norte	**pt/o/s/s/Ossétia.html**: ossétia . além disso , a ossétia do sul , território ligado à ex - república soviética da geórgia , luta para ser anexada a ossétia do norte , a qual pertence à federação russa .
Question #62 - Qual a largura do Canal da Mancha no seu ponto mais estreito?	
34 km	**pt/c/a/l/Calais.html**: calais . calais está localizada no estreito de dover , no ponto mais estreito do canal da mancha com apenas 34 km de largura , sendo a cidade francesa mais próxima da inglaterra .
Question #63 - Quem criou Descobridores de Catan?	
português	**pt/d/e/s/Descobridores_de_Catan_69e1.html**: descobridores de catan . descobridores de catan : o jogo standart , settlers of catan (1 995) , é requerido para jogar a todos os mapas . foi traduzido para português e encontra - se dísponivel através da devir . expansões : devido à popularidade do jogo original uma nova expansão tem sido edita quase todos os anos ,

páginas	**pt/d/e/s/Descobridores_de_Catan_69e1.html**: descobridores de catan . por exemplo construir em novas ilhas ou construir rotas comerciais . vêr as páginas dedicadas a esses mapas para mais detalhes . variantes existe um artigo especifíco sobre as variantes ao jogo pode consulta - lo em descobridores de catan , variantes .
ilhas	**pt/d/e/s/Descobridores_de_Catan_69e1.html**: descobridores de catan . por exemplo construir em novas ilhas ou construir rotas comerciais . vêr as páginas dedicadas a esses mapas para mais detalhes . variantes existe um artigo especifíco sobre as variantes ao jogo pode consulta - lo em descobridores de catan , variantes .
Question #64 - Quem é o santo patrono dos cervejeiros?	
arnulfo de metz	**pt/a/r/n/Arnulfo_de_Metz_b694.html**: arnulfo de metz . arnulfo foi canonizado santo pela igreja católica romana e é conhecido como o santo patrono dos cervejeiros .
carlos magno	**pt/a/r/n/Arnulfo_de_Metz_b694.html**: arnulfo de metz . ele é freqüentemente confundido nas lendas com arnoldo de soissons , que é outro santo patrono dos cervejeiros . ancestralidade incerta enquanto arnulfo é reconhecido como um dos mais antigos ancestrais documentados de carlos magno , e através disso de muitas famílias reais européias modernas ,
freqüentemente	**pt/a/r/n/Arnulfo_de_Metz_b694.html**: arnulfo de metz . ele é freqüentemente confundido nas lendas com arnoldo de soissons , que é outro santo patrono dos cervejeiros . ancestralidade incerta enquanto arnulfo é reconhecido como um dos mais antigos ancestrais documentados de carlos magno , e através disso de muitas famílias reais européias modernas ,
Question #65 - E do pão?	
arnulfo de metz	**pt/a/r/n/Arnulfo_de_Metz_b694.html**: arnulfo de metz . arnulfo foi canonizado santo pela igreja católica romana e é conhecido como o santo patrono dos cervejeiros .
santo pela igreja católica romana	**pt/a/r/n/Arnulfo_de_Metz_b694.html**: arnulfo de metz . arnulfo foi canonizado santo pela igreja católica romana e é conhecido como o santo patrono dos cervejeiros .
lendas com arnoldo de soissons	**pt/a/r/n/Arnulfo_de_Metz_b694.html**: arnulfo de metz . ele é freqüentemente confundido nas lendas com arnoldo de soissons , que é outro santo patrono dos cervejeiros . ancestralidade incerta enquanto arnulfo é reconhecido como um dos mais antigos ancestrais documentados de carlos magno , e através disso de muitas famílias reais européias modernas ,

Question #66 - O que é o jagertee?	
chá com adição de rum	**pt/c/h/á/Chá.html**: chá . o jagertee é chá com adição de rum .
Question #67 - Qual a envergadura de um milhafre-preto?	
135 - 155 cm	**pt/m/i/l/Milhafre-preto.html**: milhafre - preto . o milhafre - preto mede cerca de 55 cm de comprimento e 135 - 155 cm de envergadura , para cerca de 1 kg de peso .
cerca de 55 cm	**pt/m/i/l/Milhafre-preto.html**: milhafre - preto . o milhafre - preto mede cerca de 55 cm de comprimento e 135 - 155 cm de envergadura , para cerca de 1 kg de peso .
Question #68 - Quanto é que ele pesa?	
cerca de 1 kg	**pt/m/i/l/Milhafre-preto.html**: milhafre - preto . o milhafre - preto mede cerca de 55 cm de comprimento e 135 - 155 cm de envergadura , para cerca de 1 kg de peso .
Question #69 - Que tipo de ave é?	
abelheiro	**PUBLICO-19951228-088**: preto merganso - de - poupa falcão - abelheiro peneireiro - cinzento milhafre
Question #70 - Quantas províncias tem a Ucrânia?	
24 províncias	**pt/u/c/r/Ucrânia.html**: ucrânia . subdivisões ucrânia é subdividida em 24 províncias (óblasts) e em uma república autônoma (crimeia) .
81 províncias	**pt/i/m/p/Império_Russo_a93b.html**: império russo . cáucaso , ucrânia , belarus , boa parte da polônia (antigo reino da polônia) , moldávia (bessarábia) e quase toda a ásia central . também contava com zonas de influência no irã , mongólia e norte da china . em 1 914 o império russo estava dividido em 81 províncias
Question #71 - Que partido foi fundado por Amílcar Cabral?	
partido africano da independência da guiné e cabo verde	**PUBLICO-19951217-030**: nascido do antigo partido africano da independência da guiné e cabo verde (paigc) , fundado por amílcar cabral em 1 956 , o paicv surgiu depois do golpe de estado que , em novembro de 1 980 ,
meses	**PUBLICO-19951124-065**: o mpd foi fundado em 1 990 nos primeiros meses da abertura política finalmente admitida pelo partido africano da independência de cabo verde (paicv) , do presidente aristides pereira e do primeiro - ministro pedro pires , dois discípulos de amílcar cabral .
joão bernardo vieira	**PUBLICO-19951217-030**: fundado por amílcar cabral em 1 956 , o paicv surgiu depois do golpe de estado que , em novembro de 1 980 , derrubou o presidente luís cabral e conduziu joão bernardo vieira ao poder em bissau .

Question #72 - Quantos filhos teve a rainha Cristina da Suécia?	
Question #73 - Quem é o dono do Chelsea?	
roman abramovich	**pt/r/o/m/Roman_Abramovich_efcb.html**: roman abramovich . roman abramovich é um investidor russo , dono do clube inglês de futebol chelsea .
clube inglês de futebol chelsea	**pt/r/o/m/Roman_Abramovich_efcb.html**: roman abramovich . roman abramovich é um investidor russo , dono do clube inglês de futebol chelsea .
britânico mais rico	**pt/r/o/m/Roman_Abramovich_efcb.html**: roman abramovich . dono do clube inglês de futebol chelsea . talvez por boris berezovski ser um asilado , e não um residente na grã - bretanha , seu antigo amigo e hoje rival roman abramovich é o residente britânico mais rico , dono de 13,1 bilhões de dólares ,
Question #74 - Quantos habitantes tinha Berlim em 1850?	
300 000 habitantes	**pt/b/e/r/Berlim.html**: berlim . em 1 850 berlim já tinha 300 000 habitantes .
Question #75 - Quantos tem hoje em dia?	
cinco anos	**PUBLICO-19941109-143**: há cinco anos , berlim era uma festa .
dois anos	**PUBLICO-19950804-145**: há dois anos fomos ao festival de berlim mas só fizemos um concerto .
três anos	**PUBLICO-19950617-127**: nos últimos três anos , em berlim , quase 20 pessoas foram mortas na guerra tipo mafia que « gangs » vietnamitas rivais travam entre si .
Question #76 - O que é o ICCROM?	
assembleia geral bienal cujos delegados	**PUBLICO-19951221-008**: financiado pelas contribuições anuais dos seus 91 estados membros , o iccrom é gerido por uma assembleia geral bienal cujos delegados examinam e aprovam o programa de funcionamento e respectivo orçamento .
centro internacional de estudos para a conservação	**PUBLICO-19951221-008**: presidente do instituto português de museus , foi eleita presidente do conselho do centro internacional de estudos para a conservação e restauração dos bens culturais (iccrom) , com sede em roma .
presidente do instituto português de museus	**PUBLICO-19951221-008**: presidente do instituto português de museus , foi eleita presidente do conselho do centro internacional de estudos para a conservação e restauração dos bens culturais (iccrom) , com sede em roma .
Question #77 - Quantos estados membros tinha em 1995?	
91 estados	**PUBLICO-19951221-008**: financiado pelas contribuições anuais dos seus 91 estados membros , o iccrom é gerido por uma assembleia geral bienal cujos delegados examinam e aprovam o programa de funcionamento e respectivo orçamento .

Question #78 - Onde tem a sua sede?	
restauração dos bens	**PUBLICO-19951221-008**: foi eleita presidente do conselho do centro internacional de estudos para a conservação e restauração dos bens culturais (iccrom) , com sede em roma .
centro internacional	**PUBLICO-19951221-008**: foi eleita presidente do conselho do centro internacional de estudos para a conservação e restauração dos bens culturais (iccrom) , com sede em roma .
roma	**PUBLICO-19951221-008**: foi eleita presidente do conselho do centro internacional de estudos para a conservação e restauração dos bens culturais (iccrom) , com sede em roma .
Question #79 - Quantas vezes ganhou Portugal a Taça Davis?	
Question #80 - O que é o IPM em Portugal?	
português de museus	**PUBLICO-19940122-011**: fundação oriente (fo) , instituto camões (ic) e instituto português de museus (ipm) - - , vão organizar em 1 997 a exposição « portugal e o mundo » .
direcção - geral dos edifícios	**pt/d/i/r/Direcção-Geral_dos_Edifícios_e_Monumentos_Nacionais_2f8d.html**: direcção - geral dos edifícios e monumentos nacionais . para uma extensa lista de património em portugal , baseada na lista do ippar , em julho de 2 005 . instituto português do património arquitectónico (ippar) instituto português de arqueologia (ipa) instituto português de museus (ipm
miguel ângelo lupi	**pt/m/i/g/Miguel_Ângelo_Lupi_bf9d.html**: miguel ângelo lupi . em 16 de dezembro de 1 883 . referências maria de aires silveira , cristina azevedo tavares , adelaide ginga tchen , miguel ângelo lupi , museu do chiado , lisboa , ipm , 2 002 (isbn 972 - 776 - 124 - 0) . ligações externas miguel lupi no portugal
Question #81 - Quem foi o último rei de Portugal?	
manuel ii de portugal	**pt/1/9/3/1932.html**: 1 932 . 2 de julho - manuel ii de portugal , último rei de portugal .
príncipe real de portugal	**pt/l/u/í/Luís_Filipe,_Príncipe_Real_de_Portugal_2c2f.html**: luís filipe , príncipe real de portugal . duque de beja , que assim subiu ao trono como d. manuel ii , e que viria a ser o último rei de portugal .
pedro iv	**pt/d/i/n/Dinastia_de_Bragança_a184.html**: dinastia de bragança . rei d. pedro iv , que reinou em portugal até ao fim da monarquia , em 1 910 . o último rei , d. manuel ii , faleceu em 1 932 , sem deixar filhos ,
Question #82 - Em que período foi ele rei?	
duque	**pt/p/e/d/Pedro_de_Portugal,_Duque_de_Coimbra_fdb2.html**: pedro de portugal , duque de coimbra . no entanto , o período da sua regência nunca foi esquecido e pedro foi citado muitas vezes pelo rei joão ii de portugal (seu neto) como sendo a sua maior influência .

história de portugal	**pt/h/i/s/História_de_Portugal_333c.html**: história de portugal . rei de castela , no tratado de zamora , assinando - se a paz definitiva . d. afonso henriques dirigiu - se ao papa inocêncio ii e declarou portugal tributário da santa sé , tendo reclamado para a nova monarquia a protecção pontifícia . durante o período que se segue ,
brasil	**pt/p/e/d/Pedro_I_do_Brasil_8e35.html**: pedro i do brasil . º rei de portugal (título herdado de seu pai , d. joão vi) , durante um período de sete dias (entre 26 de abril e 2 de maio de 1 826) ,
Question #83 - Em que barco ele embarcou para o exílio?	
navio	**PUBLICO-19940613-118**: depois de uma estadia de exílio do mar , no interior da ilha , no colégio e num emprego de escritório , o protagonista consegue finalmente concretizar o seu sonho secreto , embarcar num navio , o seu navio argo , o barco
iate	**pt/m/a/n/Manuel_II_de_Portugal_7504.html**: manuel ii de portugal . consumada a vitória republicana em lisboa e a adesão do resto do país ao novo regime , d. manuel ii decidiu - se pelo exílio , embarcando na ericeira no iate real amélia .
timoneiro	**PUBLICO-19941001-057**: tortura , isolamento , trabalhos forçados , arruinaram a saúde , mas não o espírito deste resistente , que ousou desafiar deng e que o « pequeno timoneiro » tomou de ponta , como exemplo para o resto da dissidência .
Question #84 - Diga uma batalha ocorrida durante a Guerra dos Cem Anos	
frança e a inglaterra	**pt/1/3/3/1337.html**: 1 337 . início da guerra dos cem anos entre a frança e a inglaterra .
durante a guerra	**pt/1/3/4/1346.html**: 1 346 . durante a guerra dos cem anos .
batalha de poitiers	**pt/1/3/5/1356.html**: 1 356 . 19 de setembro - batalha de poitiers , no âmbito da guerra dos cem anos .
Question #85 - Quantos votos teve o Lula nas eleições presidenciais de 2002?	
7 790 392 votos	**pt/p/a/r/Partido_da_Social_Democracia_Brasileira_4c41.html**: partido da social democracia brasileira . presidencial de 15 de novembro de 1 989 , tendo o senador mário covas , seu candidato , conquistado o quarto lugar , num total de 22 candidatos , obtendo 11,52 % dos votos válidos , correspondente a 7 790 392 votos , quando o segundo colocado , luís inácio lula

215 votos	**PUBLICO-19950822-043**: sucedendo assim a luis inácio lula da silva , que se afasta depois de duas derrotas sucessivas nas presidenciais . o congresso do pt , reunido domingo em guarapari (espírito santo , nordeste) , deu 215 votos a dirceu , representante da corrente « moderada » ,
650 mil votos	**PUBLICO-19950618-095**: recebendo para cima de 650 mil votos que o consagraram como o deputado mais votado da história política do brasil . em 1 989 e 199 lula concorreu às eleições presidenciais contra collor de melo e fernando henrique cardoso mas viria a sair derrotado ,
Question #86 - Quando é que ele tomou posse?	
janeiro de 2 003	**pt/g/i/l/Gilberto_Gil_205d.html**: gilberto gil . quando o presidente luís inácio lula da silva tomou posse em janeiro de 2 003 , escolheu gilberto gil para ser ministro da cultura do brasil ,
9 de agosto de 2 006	**pt/m/a/r/Maria_Thereza_Rocha_de_Assis_Moura_b17f.html**: maria thereza rocha de assis moura . luiz inácio lula da silva , no dia 8 de junho de 2 006 , a partir de lista tríplice enviada pelo superior tribunal de justiça , stj . ela tomou posse dia 9 de agosto de 2 006 , ocupando
8 de junho de 2 006	**pt/m/a/r/Maria_Thereza_Rocha_de_Assis_Moura_b17f.html**: maria thereza rocha de assis moura . luiz inácio lula da silva , no dia 8 de junho de 2 006 , a partir de lista tríplice enviada pelo superior tribunal de justiça , stj . ela tomou posse dia 9 de agosto de 2 006 , ocupando
Question #87 - Quem era o pai de Carlomano?	
carlomano , filho de carlos martel	**pt/c/a/r/Carlomano,_filho_de_Carlos_Martel_e27c.html**: carlomano , filho de carlos martel . com a morte de carlos em 741 , ele e seu meio irmão pepino o breve sucederam seu pai em seus cargos , pepino na nêustria e carlomano na austrásia .
rei	**pt/l/u/í/Luís_III_de_França_3a5f.html**: luís iii de frança . quando o seu pai morreu , em 879 , luís tornou - se rei , juntamente com o seu irmão carlomano .
pepino , o breve	**pt/p/e/p/Pepino,_o_Breve_4b6b.html**: pepino , o breve . assumindo o poder , pepino e carlomano , que não haviam sido testados em batalha na defesa de seus domínios como seu pai , instalaram o merovíngio childerico iii como rei ,
Question #88 - Quem foi Baden Powell de Aquino?	
baden powell de aquino , violonista brasileiro , (varre - sai , rj , 6 de agosto de 1 937 - rio de janeiro , rj , 26 de setembro , de 2 000) .	**pt/b/a/d/Baden_Powell_de_Aquino_e61d.html**: baden powell de aquino , violonista brasileiro , (varre - sai , rj , 6 de agosto de 1 937 - rio de janeiro , rj , 26 de setembro , de 2 000) . filho de dona adelina e do violinista lino de aquino , que deu - lhe esse nome por ser fã do criador do escotismo , general britânico robert stephenson smyth baden - powell . é pai do pianista e tecladista phillipe baden powell e do violonista louis marcel powell (ambos nascidos na frança) e primo do violonista joão de aquino .

Question #89 - Quem escreveu o Livro da Selva?	
rudyard kipling	**pt/m/o/g/Mogli.html**: mogli . mogli é uma personagem do conto o livro da selva de rudyard kipling .
the jungle book	**pt/j/a/s/Jason_Lee_Scott_fe6e.html**: jason lee scott . seu próximo papel estrelado foi em the jungle book - 1 994 (o livro da selva) .
mogli	**pt/m/o/g/Mogli.html**: mogli . mogli é uma personagem do conto o livro da selva de rudyard kipling .
Question #90 - Quem é a personagem principal do livro?	
mt	**PUBLICO-19940215-099**: ele detestava ser judeu » , escreve henry roth no seu novo romance « mercy of a rude stream (vol . 1 : a star shines over mt . morris park) » , ao falar do adolescente ira stigman , personagem principal do livro publicado recentemente nos estados unidos .
morris	**PUBLICO-19940215-099**: ele detestava ser judeu » , escreve henry roth no seu novo romance « mercy of a rude stream (vol . 1 : a star shines over mt . morris park) » , ao falar do adolescente ira stigman , personagem principal do livro publicado recentemente nos estados unidos .
mercy	**PUBLICO-19940215-099**: ele detestava ser judeu » , escreve henry roth no seu novo romance « mercy of a rude stream (vol . 1 : a star shines over mt . morris park) » , ao falar do adolescente ira stigman , personagem principal do livro publicado recentemente nos estados unidos .
Question #91 - Em que ilha fica Sapporo?	
hokkaido	**pt/s/a/p/Sapporo.html**: sapporo . sapporo (sapporo , ?? em japonês) é a quinta maior cidade do japão , na ilha de hokkaido , onde é a capital da província de hokkaido .
sapporo dome	**pt/s/a/p/Sapporo_Dome_a40d.html**: sapporo dome . o sapporo dome (em japonês : ????? , sapporo d?mu) é um estádio localizado em sapporo , na ilha de hokkaido , norte do japão .
osaka	**FSP940220-120**: é a cidade de osaka (região central) . em 1 980 , a organização tentou expandir seu território incluindo hokkaido , a ilha mais setentrional do japão . vestindo blazers brancos e camisas pólo pretas , quase 200 membros da yamaguchi - gumi voaram para sapporo , capital da ilha ,
Question #92 - Quem fundou a escola estóica?	
zenão de cítio	**pt/z/e/n/Zenão_de_Cítio_7891.html**: zenão de cítio . aos 42 anos , fundou a escola estóica , reunindo seus alunos sob os pórticos (em grego , " stoa ") de templos , mercados e ginásios .

stoa	**pt/z/e/n/Zenão_de_Cítio_7891.html**: zenão de cítio . aos 42 anos , fundou a escola estóica , reunindo seus alunos sob os pórticos (em grego , " stoa ") de templos , mercados e ginásios .
tripartição	**pt/z/e/n/Zenão_de_Cítio_7891.html**: zenão de cítio . fundou a escola estóica , reunindo seus alunos sob os pórticos (em grego , " stoa ") de templos , mercados e ginásios . zenão propôs uma tripartição na filosofia : lógica , física e ética . a lógica fornece um critério de verdade . a física constitui
Question #93 - Quais são as regiões da Bélgica?	
bélgica e alemanha , na região delimitada par venlo , colônia , aachen , maastricht e hasselt	**pt/l/í/n/Língua_limburguesa.html**: língua limburguesa . o limburguês é também falado em bélgica e alemanha , na região delimitada par venlo , colônia , aachen , maastricht e hasselt .
frança , nos anos 80	**PUBLICO-19941006-104**: esteve na bélgica e na região de annemasse , na frança , nos anos 80 .
Ås , condado de akershus , noruega Ås , comuna de krokom , suécia Ås , comuna de nora , suécia Ås , comuna de gislaved , suécia aš pode ser	**pt/a/s/_/AS_a2c2.html**: as . província de limburgo , bélgica Ås pode ser : Ås , condado de akershus , noruega Ås , comuna de krokom , suécia Ås , comuna de nora , suécia Ås , comuna de gislaved , suécia aš pode ser : aš , região de carlsbad (karlovy vary) ,
Question #94 - Qual é o 31º estado dos Estados Unidos?	
unidos da américa	**pt/c/r/o/Cronologia_da_descolonização_de_África_3454.html**: cronologia da descolonização de áfrica . estados unidos da américa) século xx 31 de maio de 1 910 - independência da áfrica do sul , como união sul - africana , na forma de domínio do império britânico 28 de fevereiro de 1 922 - independência do egipto (do reino unido
atletismo nos jogos olímpicos	**pt/a/t/l/Atletismo_nos_Jogos_Olímpicos_de_1936_81fa.html**: atletismo nos jogos olímpicos de 1 936 . unido 2 h 31 ' 23 bronze shoryu nan japão 2 h 31 ' 42 estafetas 4 x 100 metros masculinos medalhas nome país classificação ouro jesse owens ralph metcalfe foy draper franck clifford wykoff estados
esboços sobre geografia	**pt/c/o/n/Condado_de_Cannon_7cf4.html**: condado de cannon . 31 de janeiro de 1 836 website : [http : / /] categorias : ! esboços sobre geografia dos estados unidos da américa | condados do tennessee
Question #95 - E o 37º?	
esgrima nos jogos olímpicos de verão	**pt/e/s/g/Esgrima_nos_Jogos_Olímpicos_de_Verão_de_2004_0bf2.html**: esgrima nos jogos olímpicos de verão de 2 004 . 37 - 31 estados unidos

nova zelândia	**pt/c/o/p/Copa_do_Mundo_de_Rugby_de_1991_fcd5.html**: copa do mundo de rugby de 1 991 . gloucester 8 de outubro : inglaterra 36 - 6 itália twickenham , londres 11 de outubro : inglaterra 37 - 9 estados unidos twickenham , londres 13 de outubro : nova zelândia 31 - 21 itália welford road ,
itália	**pt/c/o/p/Copa_do_Mundo_de_Rugby_de_1991_fcd5.html**: copa do mundo de rugby de 1 991 . gloucester 8 de outubro : inglaterra 36 - 6 itália twickenham , londres 11 de outubro : inglaterra 37 - 9 estados unidos twickenham , londres 13 de outubro : nova zelândia 31 - 21 itália welford road ,
Question #96 - O que era a RSFSR?	
a república socialista federada soviética da rússia (rsfs da rússia , ou ?????????? ????????? ???????????? ???????????????? ?????????? , ????? em russo) foi a mais extensa e a mais povoada das repúblicas da união soviética .	**pt/r/e/p/República_Socialista_Federada_Soviética_da_Rússia_439e.html**: a república socialista federada soviética da rússia (rsfs da rússia , ou ?????????? ????????? ???????????? ???????????????? ?????????? , ????? em russo) foi a mais extensa e a mais povoada das repúblicas da união soviética . sua capital era moscou , que também era a capital da urss . história a rsfs da rússia se origina em 7 de novembro de 1 917 , na revolução russa . em 10 de julho de 1 918 a constituição soviética de 1 918 , na qual a rsfs seria oficialmente criada , é aprovada .
Question #97 - Quantos atletas participaram nos Jogos Olímpicos de 1976?	
dez atletas	**PUBLICO-19941128-033**: jogos olímpicos de sydney do ano 2 000 » . quanto à utilização de equipas de sete ou dez atletas , pugh mostrou maior simpatia pela segunda alternativa , por « se aproximar mais do râguebi de quinze » . o râguebi fez parte do programa olímpico
6 804 atletas	**pt/j/o/g/Jogos_Olímpicos_de_Verão_de_1976_f8f4.html**: jogos olímpicos de verão de 1 976 . os jogos olímpicos de montreal , no canadá , realizados entre 17 de julho e 1 e agosto de 1 976 , com a participação de 6 804 atletas de 92 nações competindo em 21 esportes ,
um atleta	**PUBLICO-19950815-007**: um bom resultado numa grande prova , campeonatos do mundo , da europa , jogos olímpicos , fica para sempre no currículo de um atleta .
Question #98 - Em que país se realizaram?	
jogos olímpicos de verão	**pt/f/u/t/Futebol_nos_Jogos_Olímpicos_de_Verão_de_1976_a526.html**: futebol nos jogos olímpicos de verão de 1 976 . jogos olímpicos de verão de 1 976 atletismo

sydney	**pt/f/u/t/Futebol_nos_Jogos_Olímpicos_de_Verão_de_1996_4e1a.html**: futebol nos jogos olímpicos de verão de 1 996 . 976 | moscou 1 980 | los angeles 1 984 seul 1 988 | barcelona 1 992 | atlanta 1 996 | sydney 2 000 | atenas 2 004 modalidades nos jogos olímpicos de verão de 1 996 atletismo
seul	**pt/f/u/t/Futebol_nos_Jogos_Olímpicos_de_Verão_de_1996_4e1a.html**: futebol nos jogos olímpicos de verão de 1 996 . 976 | moscou 1 980 | los angeles 1 984 seul 1 988 | barcelona 1 992 | atlanta 1 996 | sydney 2 000 | atenas 2 004 modalidades nos jogos olímpicos de verão de 1 996 atletismo
Question #99 - E em que cidade?	
los angeles 1 984 seul	**pt/f/u/t/Futebol_nos_Jogos_Olímpicos_de_Verão_de_1996_4e1a.html**: futebol nos jogos olímpicos de verão de 1 996 . 976 | moscou 1 980 | los angeles 1 984 seul 1 988 | barcelona 1 992 | atlanta 1 996 | sydney 2 000 | atenas 2 004 modalidades nos jogos olímpicos de verão de 1 996 atletismo
atenas	**pt/f/u/t/Futebol_nos_Jogos_Olímpicos_de_Verão_de_1996_4e1a.html**: futebol nos jogos olímpicos de verão de 1 996 . 976 | moscou 1 980 | los angeles 1 984 seul 1 988 | barcelona 1 992 | atlanta 1 996 | sydney 2 000 | atenas 2 004 modalidades nos jogos olímpicos de verão de 1 996 atletismo
sydney	**pt/f/u/t/Futebol_nos_Jogos_Olímpicos_de_Verão_de_1996_4e1a.html**: futebol nos jogos olímpicos de verão de 1 996 . 976 | moscou 1 980 | los angeles 1 984 seul 1 988 | barcelona 1 992 | atlanta 1 996 | sydney 2 000 | atenas 2 004 modalidades nos jogos olímpicos de verão de 1 996 atletismo
Question #100 - O que é um berimbau?	
o berimbau é um instrumento de percussão usado tradicionalmente na capoeira , para marcar o ritmo da luta . no brasil é ainda conhecido pelos seguintes nomes : urucungo , urucurgo , orucungo , oricungo , uricungo , rucungo , ricungo , berimbau de barriga , gobo , marimbau , bucumbumba , bucumbunga , gunga , macungo , matungo , mutungo , aricongo , arco musical e rucumbo .	**pt/b/e/r/Berimbau.html**: o berimbau é um instrumento de percussão usado tradicionalmente na capoeira , para marcar o ritmo da luta . no brasil é ainda conhecido pelos seguintes nomes : urucungo , urucurgo , orucungo , oricungo , uricungo , rucungo , ricungo , berimbau de barriga , gobo , marimbau , bucumbumba , bucumbunga , gunga , macungo , matungo , mutungo , aricongo , arco musical e rucumbo . no sul de moçambique , este instrumento tradicional tem o nome de xitende . o berimbau é constituido de um arco feito de uma vara de madeira de comprimento aproximado de 1,20 m e um fio de aço (arame) preso nas extremidades da vara .

Table D.2: IdSay Answers and Support: Part 1 of 2 (Questions 101-200)

A#: Answer	Support
Question #101 - Que países fazem fronteira com a Itália?	
bélgica , luxemburgo , alemanha , suíça , itália , espanha , andorra e com o principado de mônaco	**pt/f/r/a/França.html**: frança . a frança funciona com um istmo que liga a península ibérica ao resto do continente , fazendo fronteira com a bélgica , luxemburgo , alemanha , suíça , itália , espanha , andorra e com o principado de mônaco .
frança	**pt/f/r/a/França.html**: frança . a frança funciona com um istmo que liga a península ibérica ao resto do continente , fazendo fronteira com a bélgica , luxemburgo , alemanha , suíça , itália , espanha , andorra e com o principado de mônaco .
reino unido , passando por baixo do mar do norte	**pt/f/r/a/França.html**: frança . fazendo fronteira com a bélgica , luxemburgo , alemanha , suíça , itália , espanha , andorra e com o principado de mônaco . o eurotúnel liga a frança ao reino unido , passando por baixo do mar do norte .
Question #102 - Como se chama o xadrez japonês?	
livro chama a atenção em meio	**FSP940205-144**: no apartamento do pai de vescovi , no morumbi , um livro chama a atenção em meio à estante dedicada ao xadrez : " os meninos - prodígio do xadrez " , do espanhol pablo morán .
jogo de xadrez	**FSP950917-145**: no livro , arrabal acrescenta mais uma paixão ao delírio dos seus personagens : o jogo de xadrez .
fernando pessoa que não só diz	**PUBLICO-19950520-012**: chama - se fernando pessoa que não só diz a « ode marítima » de álvaro de campos , enquanto deambula no « foyer » entre os espectadores , como sobe depois ao palco (um estrado que é um tabuleiro de xadrez) .
Question #103 - Qual é a temperatura do zero absoluto?	
273 graus centígrados negativos	**PUBLICO-19950407-060**: em princípio , nenhum material apresenta resistência à passagem da corrente eléctrica à temperatura de zero graus kelvin - - o zero absoluto (273 graus centígrados negativos) - - ,
0 k	**pt/z/e/r/Zero_absoluto.html**: zero absoluto . o zero absoluto , ou zero kelvin (0 k) , corresponde à temperatura
zero graus kelvin	**PUBLICO-19950407-060**: em princípio , nenhum material apresenta resistência à passagem da corrente eléctrica à temperatura de zero graus kelvin - - o zero absoluto (273 graus centígrados negativos) - - ,

Question #104 - Quem era a deusa da sabedoria?	
minerva	**pt/m/i/n/Minerva.html**: minerva . deusa da sabedoria , das artes e da guerra , era filha de júpiter .
atena	**FSP950212-150**: outra personagem mitológica que ajudou nas curas e ressurreições de asclépio foi atena , a deusa da sabedoria .
deus	**PUBLICO-19950317-098**: indicar mais dois « 0 670 » inteiramente votados não a mercúrio , deus do marketing , mas a minerva , deusa da sabedoria .
Question #105 - Que rio banha Paris?	
sena	**pt/s/e/n/Sena.html**: sena . sena pode ser : o rio sena , que banha paris .
namorados	**pt/r/i/o/Rio_Sena_a4ae.html**: rio sena . em francês) é um rio francês que banha a capital , paris que vai desaguar no oceano atlântico . possui uma extensão de 776 km . ponte alexandre iii sobre o sena o rio sena é conhecido como o rio dos namorados .
ponte	**pt/r/i/o/Rio_Sena_a4ae.html**: rio sena . em francês) é um rio francês que banha a capital , paris que vai desaguar no oceano atlântico . possui uma extensão de 776 km . ponte alexandre iii sobre o sena o rio sena é conhecido como o rio dos namorados .
Question #106 - Qual o comprimento do Spree?	
cerca de 400 km	**pt/s/p/r/Spree.html**: spree . geografia o spree tem um comprimento de cerca de 400 km , dos quais 182 km são navegáveis .
182 km	**pt/s/p/r/Spree.html**: spree . geografia o spree tem um comprimento de cerca de 400 km , dos quais 182 km são navegáveis .
382 km	**pt/l/i/s/Lista_de_rios_da_Alemanha_13e8.html**: lista de rios da alemanha . spree - 382 km rio weser - 440 km categorias : listas de rios | rios da alemanha
Question #107 - Qual é a capital do Cazaquistão?	
pavlodar	**pt/p/a/v/Pavlodar.html**: pavlodar . pavlodar (???????? , em cazaque) é uma cidade do nordeste do cazaquistão . é a capital da província de pavlodar .
atyrau	**pt/a/t/y/Atyrau.html**: atyrau . atyrau (?????? , em cazaque e russo) é a capital da província de atyrau , no cazaquistão .
qaraghandy	**pt/q/a/r/Qaraghandy.html**: qaraghandy . qaraghandy (????????? , em cazaque) ou karaganda (????????? , em russo) é a capital da província de qaraghandy , no cazaquistão .
Question #108 - E a sua maior cidade?	
russo capital astana	**pt/c/a/z/Cazaquistão.html**: cazaquistão . dc histórico , e sobre o presidente nazarbayev ????????? ???????????? ?????????? ????????? república do cazaquistão (detalhe) (detalhe) línguas oficiais cazaque e russo capital astana maior cidade

almaty	**pt/a/l/m/Almaty.html**: almaty . almaty (em cazaque : ??????) é a maior cidade do cazaquistão , com uma população de cerca de 1 185 900 (2 004) habitantes , ou seja , 8 % da população do país .
presidente nazarbayev ????????? ???????????? ?????????? ?????????	**pt/c/a/z/Cazaquistão.html**: cazaquistão . dc histórico , e sobre o presidente nazarbayev ????????? ???????????? ?????????? ????????? república do cazaquistão (detalhe) (detalhe) línguas oficiais cazaque e russo capital astana maior cidade
Question #109 - Quem é o actual presidente da Guatemala?	
óscar berger	**pt/ó/s/c/Óscar_Berger_e1f3.html**: óscar berger . óscar berger perdomo (nascido a 11 de agosto 1 946) é um político guatemalteco e atual presidente da guatemala .
freddie mercury	**pt/1/9/4/1946.html**: 1 946 . ator 9 de julho - bon scott , vocalista de banda ac / dc 11 de agosto - óscar berger , actual presidente da guatemala 5 de setembro - freddie mercury , vocalista da banda britânica queen (m. 1 991) 15 de setembro oliver stone ,
oliver stone	**pt/1/9/4/1946.html**: 1 946 . ator 9 de julho - bon scott , vocalista de banda ac / dc 11 de agosto - óscar berger , actual presidente da guatemala 5 de setembro - freddie mercury , vocalista da banda britânica queen (m. 1 991) 15 de setembro oliver stone ,
Question #110 - Qual era o cargo dele em 1991?	
óscar berger	**pt/ó/s/c/Óscar_Berger_e1f3.html**: óscar berger . é um advogado de profissão e entre 1 991 e 1 998 foi o prefeito da cidade da guatemala categorias : !
prefeito da cidade da guatemala	**pt/ó/s/c/Óscar_Berger_e1f3.html**: óscar berger . é um advogado de profissão e entre 1 991 e 1 998 foi o prefeito da cidade da guatemala categorias : !
advogado de profissão	**pt/ó/s/c/Óscar_Berger_e1f3.html**: óscar berger . é um advogado de profissão e entre 1 991 e 1 998 foi o prefeito da cidade da guatemala categorias : !
Question #111 - Quantas faixas tem a bandeira dos Estados Unidos?	
13 faixas	**pt/b/a/n/Bandeira_dos_Estados_Unidos_da_América_b54f.html**: bandeira dos estados unidos da américa . a bandeira dos estados unidos da américa consiste em 13 faixas horizontais , cujas cores são vermelho (que cobrem o topo e a parte de baixo da bandeira) alternando com branco .

quatro faixas	**pt/o/f/f/Off_The_Wall_c775.html**: off the wall . compactos nos estados unidos , quatro faixas de off the wall se tornaram compacto e figuraram entre as dez mais da lista pop da billboard .
três faixas	**pt/b/e/a/Beautiful_Stranger_c2e6.html**: beautiful stranger . um " remix " de victor calderone foi também um grande sucessos nos estados unidos . o single só tem três faixas , a versão original e dois " remixes " .
Question #112 - Quais as cores da bandeira da Hungria?	
tricolor horizontal de vermelho , branco e verde	**pt/b/a/n/Bandeira_da_Hungria_c25a.html**: bandeira da hungria . a bandeira da hungria é uma tricolor horizontal de vermelho , branco e verde .
império habsburgo línguas oficiais alemão	**pt/á/u/s/Áustria-Hungria_8f5e.html**: áustria - hungria . bandeira da hungria coat of arms antes do compromisso de 1 867 bandeira do império habsburgo línguas oficiais alemão ,
igreja oficial igreja católica capital & maior cidade viena pop	**pt/á/u/s/Áustria-Hungria_8f5e.html**: áustria - hungria . hungria coat of arms antes do compromisso de 1 867 bandeira do império habsburgo línguas oficiais alemão , húngaro igreja oficial igreja católica capital & maior cidade viena pop .
Question #113 - Quando ocorreu a batalha de Torres Vedras?	
22 de dezembro de 1 846	**pt/j/o/s/José_Travassos_Valdez_b28d.html**: josé travassos valdez . na batalha de torres vedras em 22 de dezembro de 1 846 foi prisioneiro , sendo conduzido a lisboa , de onde passou a bordo de diferentes navios do estado ,
22 de dezembro	**pt/2/2/_/22_de_Dezembro_12e3.html**: 22 de dezembro . white e wilcox 1 761 - criação do ministério da fazenda 1 846 - na batalha de torres vedras josé travassos valdez é
1 761	**pt/2/2/_/22_de_Dezembro_12e3.html**: 22 de dezembro . white e wilcox 1 761 - criação do ministério da fazenda 1 846 - na batalha de torres vedras josé travassos valdez é
Question #114 - Quem é o papa dos Infiéis?	
papa nicolau v	**pt/1/4/5/1452.html**: 1 452 . do papa nicolau v (autorização a afonso v de portugal para escravizar os infiéis
urbano ii	**pt/p/a/p/Papa_Urbano_II_e6e2.html**: papa urbano ii . " em 1 095 , o papa urbano ii convocou os cristãos a irem à terra santa expulsar os infiéis muçulmanos , nas chamadas cruzadas .
portugal	**pt/1/4/5/1452.html**: 1 452 . do papa nicolau v (autorização a afonso v de portugal para escravizar os infiéis

Question #115 - O que é VRML?	
vrml (virtual reality modelating language) , ou linguagem para modelagem de realidade virtual , é um padrão de aplicativos de realidade virtual utilizado na internet . por meio desta linguagem , escrita em modo texto , é possível criar objetos tridimensionais podendo definir cor , transparência , brilho , textura (associando - a a um bitmap) .	**pt/v/r/m/VRML_a45e.html**: vrml (virtual reality modelating language) , ou linguagem para modelagem de realidade virtual , é um padrão de aplicativos de realidade virtual utilizado na internet . por meio desta linguagem , escrita em modo texto , é possível criar objetos tridimensionais podendo definir cor , transparência , brilho , textura (associando - a a um bitmap) . os objetos podem ser formas básicas , como esferas , cubos , ovóides , hexaedros , cones , cilindros , ou formas criadas pelo próprio programador , como as extrusões .
Question #116 - Onde está a Arca da Aliança?	
israel	**pt/s/a/m/Samuel_(Bíblia)_bb59.html**: samuel (bíblia) . a morte de eli os filisteus invadiram e venceram israel , e capturaram a arca da aliança .
santo	**pt/s/a/n/Santo_dos_santos.html**: santo dos santos . santo dos santos era uma sala do templo de salomão onde ficava guardada a arca da aliança .
jerusalém	**pt/a/r/o/Aron_Kodesh_07a8.html**: aron kodesh . " arca sagrada " , uma referência ao nome hebraico da arca da aliança , que guardava as tábuas dos mandamentos no templo de jerusalém .
Question #117 - Como se chamava o Huambo durante a era colonial?	
cidade que no tempo colonial se chamou nova lisboa	**PUBLICO-19941023-047**: durante os próximos dias , no huambo (cidade que no tempo colonial se chamou nova lisboa) ,
nova lisboa	**PUBLICO-19941023-047**: durante os próximos dias , no huambo (cidade que no tempo colonial se chamou nova lisboa) ,
durante os próximos	**PUBLICO-19941023-047**: durante os próximos dias , no huambo (cidade que no tempo colonial se chamou nova lisboa) ,
Question #118 - Qual é a língua oficial do Egipto?	
lista de estados soberanos	**pt/l/i/s/Lista_de_estados_soberanos.html**: lista de estados soberanos . oficial na língua do país egipto ou egito (bras .
extendidos nas classes sociais mais altas	**pt/d/e/m/Demografia_do_Egipto_9258.html**: demografia do egipto . idiomas a língua oficial do egito é árabe . o inglês e francês estão muito extendidos nas classes sociais mais altas . alfabetismo lêem e escrevem com mais de 15 anos 51,4 % da população total , sendo : homens : 63,6 % mulheres : 38,8 % . egipto

população total	**pt/d/e/m/Demografia_do_Egipto_9258.html**: demografia do egipto . idiomas a língua oficial do egito é árabe . o inglês e francês estão muito extendidos nas classes sociais mais altas . alfabetismo lêem e escrevem com mais de 15 anos 51,4 % da população total , sendo : homens : 63,6 % mulheres : 38,8 % . egipto
Question #119 - Quais os submarinos da Marinha Brasileira?	
alemã , o brasil entrou na guerra em 1 942 ao lado dos aliados , enviando a força expedicionária brasileira	**pt/h/i/s/História_do_Brasil_9420.html**: história do brasil . submarinos a navios da marinha brasileira , atribuídos a frota alemã , o brasil entrou na guerra em 1 942 ao lado dos aliados , enviando a força expedicionária brasileira (feb) à europa ,
projeto , construção e reparação " desses meios	**pt/s/_/t/S_Tikuna_badf.html**: s tikuna . é o quarto submarino da marinha brasileira construído dentro da estratégia de aquisição do domínio completo do ciclo " projeto , construção e reparação " desses meios .
presidente getúlio vargas	**pt/i/l/h/Ilha_da_Rita_faf5.html**: ilha da rita . marinha brasileira . em 1 940 se constrói ali uma pequena base naval de abastecimento , inaugurada com grande pompa pelo presidente getúlio vargas . haviam tanques de armazenamento de óleo , paióis de carvão e um pequeno alojamento . foi instalado um aqueduto submarino desde o continente ,
Question #120 - Em que guerra combateu Joana de Arc?	
maria rosa (contestado	**pt/m/a/r/Maria_Rosa_(Contestado)_860a.html**: maria rosa (contestado) . dizem os historiadores que , com apenas 15 anos , maria rosa lutou como homem nesta guerra . considerada como uma joana d ' arc do sertão , " combatia montada em um cavalo branco com arreios forrados de veludo , vestida de branco ,
gilles rer	**pt/g/i/l/Gilles_de_Rais_e536.html**: gilles de rais . joana d ' arc . deixou a vida militar e refugiou - se na bretanha francesa , mais precisamente no castelo de tiffauges , onde seus demônios e sentimentos mais perversos afloraram . a mente do ex - comandante ficara ainda mais confusa com as tragédias da guerra e
força	**pt/j/o/ã/João,_Duque_de_Bedford_a2a7.html**: joão , duque de bedford . bedford foi bem sucedido até aparecer joana de arc , que funcionou como força de união no lado francês .
Question #121 - Onde é que ela foi queimada?	
lídia laport	**FSP950126-109**: a lídia laport morre queimada na fogueira das vaidades da novela das oito , mas a história da vera fischer , que não é joana d ' arc , continua .

carneiros	**PUBLICO-19950130-009**: público - - o que é que sabia sobre joana d ' arc ? sandrine bonnaire - - quase nada , só sabia que ela guardava carneiros , que escutava vozes e que morreu queimada .
bruxas de salem	**pt/a/s/p/Aspectos_controversos_do_Catolicismo_8b2d.html**: aspectos controversos do catolicismo . casos como os das bruxas de salem , nos eua (país onde nunca houve a inquisição) e o de joana d ' arc heroína francesa queimada na inglaterra (outro país em que o tribunal do santo oficio jamais atuou) são exemplos de casos falsamente associados à inquisição .
Question #122 - Quando?	
1 de maio	**PUBLICO-19950426-049**: a 1 de maio , « dia de joana de arc » . « existe agora no nosso país uma quarta família política , tão importante como os socialistas , gaullistas ou centristas » , disse ele .
19 de abril	**pt/1/9/_/19_de_Abril_f75e.html**: 19 de abril . 1 839 - a bélgica torna - se num reino através do tratado de londres 1 909 - joana de arc é canonizada .
1 892	**PUBLICO-19941222-121**: joana de arc : « é preciso que o sangue de França seja o senhor » , quando tivermos adoptado o nosso programa que se resume numa palavra : a França aos franceses ! » . in « la livre parole » , 1 892 .
Question #123 - Que idade tinha ela?	
sete anos	**PUBLICO-19941203-102**: adultos , na nossa memória imagens de violência presenciadas em tenra idade , que nos perseguem até aos dias de hoje ? lembro - me de ter visto , aos sete anos , o notável filme de carl dreyer « joana d ' arc » ,
15 anos	**pt/g/u/e/Guerra_do_Contestado_0e08.html**: guerra do contestado . os fiéis que mudaram para caraguatá eram chefiadas por maria rosa , uma jovem com 15 anos de idade , considerada pelos historiadores como uma joana d ' arc do sertão , já que " combatia montada em um cavalo branco com arreios forrados de veludo , vestida de branco ,
dez anos	**pt/f/i/l/Filipe_III,_Duque_de_Borgonha_9d64.html**: filipe iii , duque de borgonha . catarina de valois , é consagrado rei da França na catedral notre - dame de paris , à idade de dez anos . a 30 de maio de 1 431 , joana d ' arc , após ter sido julgada pela igreja , é queimada viva na praca da velha - marcha em rouen .

Question #124 - Desde quando está Fidel Castro no poder?	
1 959	**FSP940429-020**: castrista : partidário do líder socialista cubano fidel castro , no poder desde 1 959 .
1 993	**PUBLICO-19941101-049**: fidel castro tenta recompor - se de duas más colheitas consecutivas : 4,2 milhões de toneladas de 1 992 - 93 e quatro milhões de toneladas em 1 993 - 1 994 . desde o colapso da antiga união soviética ,
26 de julho de 1 953	**pt/c/u/l/Cultura_de_Cuba_d602.html**: cultura de cuba . presidente desde então . 1 de maio dia do trabalhador día de los trabajadores dia internacional do trabalhador 26 de julho comemoração do assalto ao quartel moncada asalto al cuartel moncada na manhã de 26 de julho de 1 953 , cerca de 160 homens sob o comando de fidel castro
Question #125 - Quando é que ele nasceu?	
1 959	**FSP940926-124**: junto com fidel castro , foi o principal líder da revolução cubana que chegou ao poder em 1 959 .
1 980	**PUBLICO-19940618-076**: é um problema migratório » . fidel castro estará tão furioso como em 1 980 .
26 de julho	**pt/m/o/v/Movimento_26_de_Julho_0be5.html**: movimento 26 de julho . o movimento 26 de julho foi um movimento revolucionário cubano , fundado em 1 954 por fidel castro e seus companheiros , dentre eles che guevara , contra o ditador fulgencio batista .
Question #126 - Quem é o irmão dele?	
príncipe frederik	**pt/m/a/r/Mary_Elizabeth_Donaldson_f692.html**: mary elizabeth donaldson . ainda se sabe que mary se sentou entre o seu atual marido , príncipe frederik e o irmão dele , príncipe joachim . ela se relembra que eles começaram a conversar e simplesmente não pararam desde então .
mary elizabeth , princesa herdeira da dinamarca	**pt/m/a/r/Mary_Elizabeth,_Princesa_Herdeira_da_Dinamarca_bfbc.html**: mary elizabeth , princesa herdeira da dinamarca . ainda se sabe que mary se sentou entre o seu atual marido , príncipe frederik e o irmão dele , príncipe joachim . ela se relembra que eles começaram a conversar e simplesmente não pararam desde então .
dinamarca	**pt/m/a/r/Mary_Elizabeth,_Princesa_Herdeira_da_Dinamarca_bfbc.html**: mary elizabeth , princesa herdeira da dinamarca . ainda se sabe que mary se sentou entre o seu atual marido , príncipe frederik e o irmão dele , príncipe joachim . ela se relembra que eles começaram a conversar e simplesmente não pararam desde então .

Question #127 - O que são os forcados?	
os forcados são grupos amadores de vários homens que numa corrida de touros pegam o touro , ou seja , agarram o touro pela sua cara . quando se executa uma pega , oito homens entram na arena , o primeiro é o caras que agarra o touro pela cara ; os outros são os ajudas que ajudam o caras a parar o touro e há um último que segura no rabo do touro para que quando os outros o largarem este não fuja .	**pt/f/o/r/Forcado.html**: os forcados são grupos amadores de vários homens que numa corrida de touros pegam o touro , ou seja , agarram o touro pela sua cara . quando se executa uma pega , oito homens entram na arena , o primeiro é o caras que agarra o touro pela cara ; os outros são os ajudas que ajudam o caras a parar o touro e há um último que segura no rabo do touro para que quando os outros o largarem este não fuja .
garfo é um utensílio culinário utilizada pela civilização ocidental moderna para a alimentação . serve principalmente para segurar alimentos rígidos e levá - los à boca , mas também se usa na cozinha para segurar os produtos a cozinhar e para moer , por exemplo , batatas ou cenouras cozidas , em puré .	**pt/g/a/r/Garfo.html**: garfo é um utensílio culinário utilizada pela civilização ocidental moderna para a alimentação . serve principalmente para segurar alimentos rígidos e levá - los à boca , mas também se usa na cozinha para segurar os produtos a cozinhar e para moer , por exemplo , batatas ou cenouras cozidas , em puré . normalmente construído em metal (modernamente em aço inoxidável) , o garfo é composto por uma parte mais larga dividida em dentes - geralmente quatro - e uma pega . enquanto que o garfo de mesa tem dimensões próprias para levar os alimentos à boca , o garfo de cozinha ou o de trinchar são de maiores dimensões .
Question #128 - Quando foi assinado o Tratado de Zamora?	
1 143	**PUBLICO-19950809-024**: mais concretamente na igreja de almacave , que , há 852 anos , tiveram lugar as primeiras cortes nacionais , a culminar a celebração do tratado de zamora , em 1 143 .
5 de outubro de 1 143	**pt/i/n/d/Independência_de_Portugal_f994.html**: independência de portugal . só a 5 de outubro de 1 143 é reconhecida independência de portugal pelo rei afonso vii de castela , no tratado de zamora , assinando - se a paz definitiva .

850	**PUBLICO-19941013-059**: tem lugar um concerto de música renascentista , que se integra , igualmente , nas comemorações do 850 º aniversário do tratado de zamora . no bar terraço do centro cultural de belém , entre as 19 h e as 21 h ,
Question #129 - O que é o fogo de São Telmo?	
fogo	**pt/f/o/g/Fogo_de_São_Telmo_6119.html**: fogo de são telmo . fogo de são telmo
pedro gonzález telmo	**pt/p/e/d/Pedro_González_Telmo_b063.html**: pedro gonzález telmo . em portugal e no brasil , sob a invocação de são pedro gonzales (ou gonçalves) e são telmo existem numerosas igrejas e capelas , bem como confrarias , geralmente relacionadas com actividades de pescadores , marinheiros ou armadores .
ou directamente uma chama azul	**pt/p/e/d/Pedro_González_Telmo_b063.html**: pedro gonzález telmo . na iconogrtafia é representado vestido com o hábito branco e capa negra da ordem dominicana , levando na mão um círio azul , que representa o fogo de são telmo , ou directamente uma chama azul ; às vezes representa - se alimentando os pescadores .
Question #130 - O que é que é um brigadeiro?	
brigadeiro é uma patente militar de uso na aeronáutica ou força aérea , que designa a patente mais alta daquela força , equivalente a patente de general no exército .	**pt/b/r/i/Brigadeiro.html**: brigadeiro é uma patente militar de uso na aeronáutica ou força aérea , que designa a patente mais alta daquela força , equivalente a patente de general no exército . categorias : !
Question #131 - Quem inventou o forno de microondas?	
percy spencer	**pt/p/e/r/Percy_Spencer_aacd.html**: percy spencer . ligações externas hall da fama dos inventores - em inglês breve história do microondas - em inglês fatos fascinantes sobre o inventor do microondas - em inglês quem inventou o forno de microondas ?
americana ficou	**FSP950123-044**: entrou na onda bolando uma cesta que combina garrafas do seu riesling com potes de mostarda caseira . por conta da excentricidade americana ficou a idéia da fazenda wendell , que conseguiu entrar nos supermercados com uma nova variedade de milho de pipoca que se leva ao forno de microondas com sabugo e tudo .

sabugo	**FSP950123-044**: entrou na onda bolando uma cesta que combina garrafas do seu riesling com potes de mostarda caseira . por conta da excentricidade americana ficou a idéia da fazenda wendell , que conseguiu entrar nos supermercados com uma nova variedade de milho de pipoca que se leva ao forno de microondas com sabugo e tudo .
Question #132 - Qual a nacionalidade de Nicole Kidman?	
honolulu	**pt/n/i/c/Nicole_Kidman_310b.html**: nicole kidman . com nicole kidman nicole kidman em cannes , em 2 001 nascimento 20 de junho de 1 967 honolulu , havaí nacionalidade
havaí	**pt/n/i/c/Nicole_Kidman_310b.html**: nicole kidman . com nicole kidman nicole kidman em cannes , em 2 001 nascimento 20 de junho de 1 967 honolulu , havaí nacionalidade
nascimento	**pt/n/i/c/Nicole_Kidman_310b.html**: nicole kidman . com nicole kidman nicole kidman em cannes , em 2 001 nascimento 20 de junho de 1 967 honolulu , havaí nacionalidade
Question #133 - Quem patenteou o primeiro telégrafo sem fios?	
brasileiro roberto landell de moura	**FSP941122-066**: da reportagem local hoje faz 90 anos que o padre brasileiro roberto landell de moura patenteou em nova york o primeiro telégrafo sem fio .
ficaram para o italiano guglielmo marconi	**FSP941122-066**: telégrafo sem fio , base para o rádio de longa distância , ficaram para o italiano guglielmo marconi (1 874 - 1 937) . enquanto marconi se mudou da itália para a inglaterra , onde obteve apoio para desenvolver as pesquisas que lhe dariam
centro cultural light	**FSP950405-040**: destes países com empresas italianas similares . ontem ela esteve no centro cultural light inaugurando a exposição comemorativa do centenário da primeira experiência de transmissão à distância (telégrafo sem fio , usando ondas de rádio) ,
Question #134 - Qual é a companhia francesa de caminhos-de-ferro ?	
sncf	**PUBLICO-19950406-159**: estação de caminhos - de - ferro da sncf (a companhia ferroviária francesa) , uma estação de táxis e um enorme parque de estacionamento . para breve , está prevista uma linha de eléctricos rápidos .
parque de estacionamento	**PUBLICO-19950406-159**: estação de caminhos - de - ferro da sncf (a companhia ferroviária francesa) , uma estação de táxis e um enorme parque de estacionamento . para breve , está prevista uma linha de eléctricos rápidos .
partido único da revolução	**PUBLICO-19951017-049**: caminhos - de - ferro (capital francês) e das companhias de gaz (britânicas) . janeiro - - é institucionalizado o partido único da revolução , o justicialista .

Question #135 - O que é a Feplam?	
diz luís carlos soares	**FSP940927-058**: o aproveitamento cai muito " , diz luís carlos soares , da feplam (fundação educacional e cultural padre landell de moura) .
uma das pioneiras no ensino a distância no sul do país	**FSP940927-058**: a feplam é uma das pioneiras no ensino a distância no sul do país e não abre mão das reuniões periódicas entre seus alunos ,
p. 21 jornal correio do povo	**pt/g/u/a/Guaíba.html**: guaíba . porto alegre , março / abril / 97 ano ix - 23 , feplam , p. 21 jornal correio do povo - ano 1 926 arquivo do museu carlos nobre sant ' anna , carlos affonso .
Question #136 - Qual a dotação do Prémio Cervantes?	
cerca de 18 mil contos	**PUBLICO-19941129-137**: o prémio cervantes , criado em 1 975 com uma dotação de 15 milhões de pesetas - - cerca de 18 mil contos - - foi atribuído no ano passado ao romancista miguel delibes .
Question #137 - Quem é que ganhou o prémio em 1994?	
mario vargas llosa	**PUBLICO-19941204-064**: • o escritor peruano mario vargas llosa ganha o prémio cervantes 1 994 , o maior galardão literário espanhol .
camilo josé cela	**PUBLICO-19951214-005**: camilo josé cela ganha prémio literário ibero - americano prémio cervantes é uma « mierda » o escritor galego , camilo josé cela , prémio nobel da literatura em 1 989 , venceu ontem o prémio miguel cervantes ,
josé	**PUBLICO-19951214-005**: camilo josé cela ganha prémio literário ibero - americano prémio cervantes é uma « mierda » o escritor galego , camilo josé cela , prémio nobel da literatura em 1 989 , venceu ontem o prémio miguel cervantes ,
Question #138 - Quem são os co-príncipes de Andorra?	
governo parlamentarismo	**pt/a/n/d/Andorra.html**: andorra . andorra la vella língua oficial catalão ; espanhol e francês também é falado governo parlamentarismo co - principado - co - príncipe
presidente da república	**pt/a/n/d/Andorra.html**: andorra . os chefes de estado , ou co - príncipes , são o presidente da república francesa e o bispo da comarca catalã de urgel . o chefe de governo é eleito pela maioria do conselho geral dos vales . os principais partidos políticos são o pla (partido liberal de andorra) ,
jacques chirac	**pt/j/a/c/Jacques_Chirac_001d.html**: jacques chirac . como presidente , é também co - príncipe de andorra , por inerência (ex officio) .
Question #139 - Que tipo de tecido é o damasco?	
alto - relevo	**pt/d/a/m/Damasco_(tecido).html**: damasco (tecido) . em tapeçaria , o damasco é um tecido usualmente de seda (mas que também pode ser de lã , linho ou algodão) ornado em alto - relevo ,

grande família	**PUBLICO-19941018-057**: a sua principal queixa era a de sentirem desligados de uma grande família que já não existia em damasco .
algodão	**pt/d/a/m/Damasco_(tecido).html**: damasco (tecido) . em tapeçaria , o damasco é um tecido usualmente de seda (mas que também pode ser de lã , linho ou algodão) ornado em alto - relevo ,
Question #140 - Quantos jogadores tem uma equipa de voleibol?	
seis jogadores	**pt/v/o/l/Voleibol.html**: voleibol . voleibol é um desporto praticado numa quadra dividida em dois por uma rede , por duas equipas de seis jogadores cada .
três jogadores	**pt/s/e/p/Sepaktakraw.html**: sepaktakraw . combina a destreza e a habilidade do futebol , os fundamentos do voleibol e a agilidade das artes marciais . o esporte é jogado por duas equipes , com três jogadores cada , mais um reserva em uma quadra idêntica a de badminton .
um jogador	**pt/a/n/d/Andrija_Geric_4946.html**: andrija geric . novi sad , sérvia) é um jogador de voleibol sérvio . carreira geric despontou para o cenário internacional durante os jogos olímpicos de atlanta , quando o jovem time da então iugoslávia ,
Question #141 - Quando é que viveu Zenão de Eleia?	
século v ac	**pt/g/ö/d/Gödel,_Escher,_Bach_3d73.html**: gödel , escher , bach . zenão de eleia , inventor de paradoxos , viveu no século v ac .
1 991	**PUBLICO-19940824-098**: nome que , para uma figura que vive , em 1 991 , numa aldeia cinco quilómetros a sul de berlim , na ex - rda , só pode ser emblemático . zenão de eleia é o conhecido filósofo dos paradoxos e dos sofismas ,
Question #142 - Qual é a área da Groenlândia?	
2 170 600 km 2	**FSP940620-108**: ilha - a maior do mundo é a groenlândia , com área de 2 170 600 km 2 (6 270 vezes a ilha de são sebastião , litoral norte paulista , a maior do brasil) .
2 166 086 km ^{2}	**pt/r/e/c/Recordes_mundiais.html**: recordes mundiais . fenômenos físicos geografia ilhas maior : groenlândia , 2 166 086 km ^{2} ; mais remota habitada : tristão da cunha , 2 816 km . lagos mais profundo lago de água doce : lago baikal , 1 637 m ; maior lago de água doce por área : lago superior ,
Question #143 - Quem foi a primeira mulher no espaço?	
valentina tereshkova	**pt/e/x/p/Exploração_espacial.html**: exploração espacial . a primeira mulher no espaço foi a russa valentina tereshkova (1 937 -) ,

lista de astronautas	**pt/l/i/s/Lista_de_astronautas_(1961-1979).html**: lista de astronautas (1 961 - 1 979) . primeira mulher no espaço
svetlana savitskaya	**pt/v/o/s/Vostok_VI_6d36.html**: vostok vi . a cosmonauta foi valentina tereshkova , a primeira mulher no espaço , feito que se repetiria apenas 19 anos depois com svetlana savitskaya .
Question #144 - E a segunda?	
feminina do mundo otto schmidt richard sorge valentina tereshkova	**pt/h/e/r/Herói_da_União_Soviética_d891.html**: herói da união soviética . segunda guerra mundial , a melhor ás feminina do mundo otto schmidt richard sorge valentina tereshkova - cosmonauta , primeira mulher no espaço
herói da união	**pt/h/e/r/Herói_da_União_Soviética_d891.html**: herói da união soviética . segunda guerra mundial , a melhor ás feminina do mundo otto schmidt richard sorge valentina tereshkova - cosmonauta , primeira mulher no espaço
cosmonauta	**pt/h/e/r/Herói_da_União_Soviética_d891.html**: herói da união soviética . segunda guerra mundial , a melhor ás feminina do mundo otto schmidt richard sorge valentina tereshkova - cosmonauta , primeira mulher no espaço
Question #145 - Diga um jornal libanês.	
disse que sabia que iam assassiná	**PUBLICO-19940907-116**: entrevistado por um jornal libanês (em 1 979) , « carlos » , « o chacal » , disse que sabia que iam assassiná - lo um dia .
carlos	**PUBLICO-19940907-116**: entrevistado por um jornal libanês (em 1 979) , « carlos » , « o chacal » , disse que sabia que iam assassiná - lo um dia .
chacal	**PUBLICO-19940907-116**: entrevistado por um jornal libanês (em 1 979) , « carlos » , « o chacal » , disse que sabia que iam assassiná - lo um dia .
Question #146 - Quantos refugiados haitianos estão na base de Guantanamo?	
16 mil refugiados	**FSP940718-041**: a base militar dos eua em guantanamo , cuba , já abriga 16 mil refugiados haitianos .
dois mil refugiados	**PUBLICO-19940716-072**: base de guantanamo , em cuba (onde se encontram mais de 15 mil) , para antigua , república dominicana e granada . estão também em curso negociações entre washington e o suriname para que este país da américa do sul aceite um total de dois mil refugiados haitianos .
Question #147 - Quando foi fundado o Vasco da Gama?	
1 945	**pt/c/l/u/Clube_de_Futebol_Vasco_da_Gama_0bee.html**: clube de futebol vasco da gama . localização o clube de futebol vasco da gama é um clube de futebol português localizado na vila da vidigueira , distrito de beja . história o clube foi fundado em 1 945 e o actual presidente é antónio galvão .

19 de julho de 1 919	**pt/c/a/m/Campeonato_Catarinense_de_Futebol_80d4.html**: campeonato catarinense de futebol . vasco da gama de lages (lages) 1 videirense (videira) 1 xv de novembro (rio do sul) 1 clube fundado em 19 de julho de 1 919 como brasil football club .
21 de agosto	**pt/2/1/_/21_de_Agosto_6f67.html**: 21 de agosto . 1 898 - é fundado o club de regatas vasco da gama , na cidade do rio de janeiro .
Question #148 - Por quem foi fundado?	
a fundada é uma freguesia portuguesa do concelho de vila de rei , com 36,29 km ^{2} de área e 676 habitantes (2 001) . densidade : 18,6 hab / km ^{2} .	**pt/f/u/n/Fundada.html**: a fundada é uma freguesia portuguesa do concelho de vila de rei , com 36,29 km ^{2} de área e 676 habitantes (2 001) . densidade : 18,6 hab / km ^{2} . festas religiosas festa de santa margarida - realiza - se no quarto fim - de - semana de agosto .
esta página precisa ser reciclada . sinta - se livre para editá - la para que esta possa atingir um nível de qualidade superior . baseado é nome popular , no brasil , do cigarro cujo fumo é a maconha .	**pt/b/a/s/Baseado.html**: esta página precisa ser reciclada . sinta - se livre para editá - la para que esta possa atingir um nível de qualidade superior . baseado é nome popular , no brasil , do cigarro cujo fumo é a maconha . é geralmente confeccionado a partir de papéis de seda ou arroz . não é possível delimitar um padrão para seu formato . o baseado também é popularmente conhecido como fino , se for feito com pouca maconha , ou ainda bomba , braço - de - judas , bucha , tora ou vela , quando este é grosso por conter muita maconha , ou ponta , quando apenas uma fração final do baseado .
Question #149 - Quando nasceu Vasco da Gama?	
1 469	**pt/v/a/s/Vasco_da_Gama_3d28.html**: vasco da gama nent vasco da gama (sines , portugal , 1 469 - cochim , índia , 24 de dezembro de 1 524) foi um navegador português . filho do alcaide - mor de sines , estevão da gama , o rei d. manuel i (1
25 de abril	**PUBLICO-19941002-097**: acompanhado de vasco da gama fernandes , que viria a ser presidente da assembleia da república depois do 25 de abril . gama fernandes apresentou - se então , aos participantes ,
1 498	**PUBLICO-19950428-153**: conhecer « o outro » ainda é muito difícil . « em 1 498 , chegaram a moçambique os primeiros portugueses , chefiados por vasco da gama .
Question #150 - Onde é que ele morreu?	
vice - rei da índia	**PUBLICO-19950328-181**: na segunda armada foi o vasco da gama , que era vice - rei da índia , e foi nele que haveria de morrer o afonso de albuquerque .

campeonato carioca de futebol	**pt/c/a/m/Campeonato_Carioca_de_Futebol_Feminino_3c39.html**: campeonato carioca de futebol feminino . em 1 999) e o torneio início (com o vasco da gama vencendo em 1 999 e 2 000 , únicas edições do torneio) .
nacional brasil	**pt/o/s/m/Osmar_Fortes_Barcellos_7d60.html**: osmar fortes barcellos . vasco da gama grêmio sele (c) ção nacional brasil 23 categorias : !
Question #151 - Em que distrito fica Sines?	
distrito de setúbal	**pt/s/i/n/Sines.html**: sines . sines é uma cidade portuguesa no distrito de setúbal , região do alentejo e subregião do alentejo litoral , com cerca de 12 500 habitantes .
alentejo e subregião do alentejo litoral	**pt/s/i/n/Sines.html**: sines . sines é uma cidade portuguesa no distrito de setúbal , região do alentejo e subregião do alentejo litoral , com cerca de 12 500 habitantes .
santiago do cacém	**PUBLICO-19950406-096**: distrito de setúbal - - alcácer do sal , grândola , santiago do cacém e sines . uma nota da região de turismo a que a agência lusa teve
Question #152 - Qual é a capital de Dublin?	
lista de cidades na irlanda	**pt/l/i/s/Lista_de_cidades_na_Irlanda_2445.html**: lista de cidades na irlanda . esta é uma lista de cidades na irlanda : dublin - capital
paris ou londres	**FSP940915-133**: como paris ou londres , dublin é uma cidade pequena . suas vizinhas oferecem um número muito maior de atrações turísticas mas , em temos de atmosfera , a capital da irlanda é bem mais acolhedora -sua população soma 1 milhão de habitantes .
norte	**PUBLICO-19940212-055**: dominic mcglinchey tinha 40 anos e vivia em drogheda , 32 quilómetros a norte de dublin , a capital da república da irlanda .
Question #153 - Em que ano é que Halle Berry venceu o Óscar?	
Question #154 - Por que estados corre o Havel?	
brandemburgo , berlim e saxônia - anhalt , alemanha	**pt/r/i/o/Rio_Havel_11cf.html**: rio havel . o havel é um rio que corre nos estados federais alemães de brandemburgo , berlim e saxônia - anhalt , alemanha .
presidente	**PUBLICO-19951102-053**: foi definitivamente encerrada ao pública a cela onde o presidente havel passou , isolado , alguns dos longos anos de oposição ao regime comunista .
rios	**pt/l/i/s/Lista_de_rios_por_continente.html**: lista de rios por continente . veja lista de rios portugueses rio danúbio rio ebro rio elba rio guadalquivir rio havel

Question #155 - Diga um escritor irlandês.	
james joyce pelos críticos	**FSP951019-104**: folha - como escritor irlandês , o que acha de ser comparado a james joyce pelos críticos ? doyle - enquanto escrevo , não penso em mim como um escritor irlandês .
oscar wilde	**pt/o/s/c/Oscar_Wilde_3590.html**: oscar wilde . oscar fingal o ' flahertie wills wilde (dublin , 16 de outubro de 1 854 - paris , 30 de novembro de 1 900) foi um escritor irlandês .
principal obra no desenvolvimento	**pt/b/r/a/Bram_Stoker_987d.html**: bram stoker . clontarf , subúrbio de dublin , irlanda - 20 de abril de 1 912 , londres , inglaterra) foi um escritor irlandês bastante conhecido por ter sido o autor de drácula , a principal obra no desenvolvimento do mito literário moderno do vampiro .
Question #156 - Quem foi Carl Barks?	
carl barks (27 de março de 1 901 - 25 de agosto de 2 000) foi um famoso ilustrador dos estúdios disney e criador de arte seqüencial , responsável pela invenção de patópolis e muitos de seus habitantes : tio patinhas (1 947) , gastão (1 948) , irmãos metralha (1 951) , professor pardal (1 952) , maga patalójika (1 961) e outros .	**pt/c/a/r/Carl_Barks_0a4e.html**: carl barks (27 de março de 1 901 - 25 de agosto de 2 000) foi um famoso ilustrador dos estúdios disney e criador de arte seqüencial , responsável pela invenção de patópolis e muitos de seus habitantes : tio patinhas (1 947) , gastão (1 948) , irmãos metralha (1 951) , professor pardal (1 952) , maga patalójika (1 961) e outros . a qualidade de seus roteiros e desenhos o rendeu os apelidos o homem dos patos e o bom artista dos patos . o autor de quadrinhos will eisner o chamou de hans christian andersen dos quadrinhos .
Question #157 - Onde é que ele nasceu?	
patópolis	**pt/p/a/t/Patópolis.html**: patópolis . a primeira menção a patópolis foi feita numa história de carl barks na edição 49 de walt disney ' s comics and stories em 1 944 .
don	**pt/d/o/n/Don_Rosa_2c5d.html**: don rosa . conhecido pelo nome artístico don rosa , (louisville , 29 de junho de 1 951) é um cartunista norte - americano , considerado o sucessor de carl barks .
brasil	**pt/t/i/o/Tio_Patinhas_0ad9.html**: tio patinhas . tio por parte de mãe do personagem já existente pato donald , surgiu nos quadrinhos em dezembro de 1 947 em " christmas on bear mountain " (no brasil , natal nas montanhas) , história escrita e desenhada por carl barks .

Question #158 - Quem eram os pais dele?	
neville longbottom	**pt/n/e/v/Neville_Longbottom_dc85.html**: neville longbottom . pais dele ela acha que ele tem vergonha deles ; quando sua mãe dá de presente para ele uma embalagem de doce , ela diz para ele jogar aquilo no lixo ,
greaves	**pt/g/r/e/Greaves.html**: greaves . greaves nasceu em barbados nada tendo a ver com a guerra civil na inglaterra , ele pagou pelos pecados dos pais dele e era um escravo desde o momento do nascimento .
nada	**FSP950605-093**: os pais dele nada têm a ver com os meus .
Question #159 - O que é um kilt?	
o kilt é o saiote pregueado , parcialmente trespassado , e quadriculado em cores correspondentes a cada clã ou família , e que faz parte do traje típico masculino da escócia .	**pt/k/i/l/Kilt.html**: o kilt é o saiote pregueado , parcialmente trespassado , e quadriculado em cores correspondentes a cada clã ou família , e que faz parte do traje típico masculino da escócia . tradicionalmente ele era utilizado por guerreiros e batedores dos clãs , sendo que cada clã (ou clan) tinha um tartan diferente . categorias : !
Question #160 - Quem realizou «Os Pássaros»?	
alfred hitchcock	**PUBLICO-19940912-093**: é o seu primeiro filme , mas o seu papel mais importante - - exceptuando « miss daisy » - - foi em « os pássaros » , de alfred hitchcock .
eu	**PUBLICO-19950508-119**: / / eu odiava bonecas e / odiava jogos , os animais / eram inamistosos e os pássaros / levantavam voo e fugiam .
tippi hedren	**FSP940601-083**: a atriz tippi hedren parece que já esqueceu os ataques que recebeu em " os pássaros " , de hitchcock . tanto que agora é diretora de um refúgio de animais selvagens em acton , na califórnia .
Question #161 - Quantos filmes realizou Jean Vigo?	
um filme	**PUBLICO-19950607-157**: porque o jean vigo e varda fizeram ambos um filme sobre nice ; e eu fiz também « nice : à propos de jean vigo » . o meu não vai poder passar porque é em « double bande » .
100 filmes	**PUBLICO-19940520-064**: integrado no ciclo 100 dias 100 filmes , " l ' atalante " (a atalante) , de jean vigo , é o filme a que pode assistir , pelas 21 h 30 , no cinema tivoli .
37 filmes	**FSP950114-094**: seu endereço é / / www . maths . tcd . le / pub / films / movie - hypdocs . html . são 37 filmes de mais de 60 países , com obras de autores como jean vigo , ed wood , yasujiro ozu ,

Question #162 - Diga um desses filmes.	
avant - garde	**PUBLICO-19950607-157**: « entr ' acte » , de rené clair (França , 1 924) é um desses filmes « d ' avant - garde » .
usou o primeiro plano	**PUBLICO-19950607-157**: 1 924) é um desses filmes « d ' avant - garde » . o que é interessante aí é ver o esforço para descobrir o cinema . como o griffith , que usou o primeiro plano , ousou cortar as pernas , fazer uns « travellings » em balão ,
aí é ver	**PUBLICO-19950607-157**: 1 924) é um desses filmes « d ' avant - garde » . o que é interessante aí é ver o esforço para descobrir o cinema . como o griffith , que usou o primeiro plano , ousou cortar as pernas , fazer uns « travellings » em balão ,
Question #163 - Qual o comprimento da Ponte do Øresund?	
18 quilómetros	**PUBLICO-19941028-134**: Øresund) , e outra entre as duas metades da dinamarca (separadas pelo storebælt) . ambas com 18 quilómetros , curiosamente também o comprimento da futura ponte de sacavém ao montijo .
7 845 metros	**pt/p/o/n/Ponte_do_Øresund_986b.html**: ponte do Øresund . a ponte de Øresund (dinamarquês Øresund , sueco öresund) liga a ilha da zelândia (dinamarca) à suécia . com 7 845 metros de comprimento , é a maior ponte rodoviária e ferroviária (combinados) da europa .
800 metros 28 km	**pt/p/e/q/Pequeno_Belt_3df0.html**: pequeno belt . comprimento e sua larguara varia de 800 metros 28 km . sua profundidade máxima é de aproximadamente 75 m. várias pequenas ilhas localizam - se neste estreito . duas pontes atravessam o estreito , a velha ponte do pequeno belt e a nova ponte do pequeno belt . ver também grande belt öresund
Question #164 - Que companhia está baseada no Aeroporto Ben Gurion?	
el al	**pt/e/l/_/El_Al_26f7.html**: el al . fundada em 1 948 principais centros de operações aeroporto internacional ben gurion outros centros de operações programa de milhagem serviço vip aliança comercial frota 32 aviões destinos 43 companhia administradora sede tel aviv , israel pessoas importantes website oficial www .
lista	**pt/l/i/s/Lista_de_DDI_de_Israel_d9d6.html**: lista de ddi de israel . aeroporto ben gurion
programa de milhagem	**pt/e/l/_/El_Al_26f7.html**: el al . fundada em 1 948 principais centros de operações aeroporto internacional ben gurion outros centros de operações programa de milhagem serviço vip aliança comercial frota 32 aviões destinos 43 companhia administradora sede tel aviv , israel pessoas importantes website oficial www .

Question #165 - Que navio americano foi afundado em Pearl Harbor in 1941?	
uss arizona	**pt/a/r/i/Arizona_(desambiguação).html**: arizona (desambiguação) . uss arizona (bb - 39) , um navio de guerra americano , afundado em pearl harbor em 1 941 .
navio de guerra	**pt/a/r/i/Arizona_(desambiguação).html**: arizona (desambiguação) . uss arizona (bb - 39) , um navio de guerra americano , afundado em pearl harbor em 1 941 .
batalha naval	**pt/a/u/s/Austrália_nas_Grandes_Guerras_2683.html**: austrália nas grandes guerras . navio sydney e o cruzado alemão kormoran afundaram um ao outro numa batalha naval na costa oeste australiana . 645 australianos morreram no naufrágio . guerra no pacífico (1 942 - 1 945) após os ataques a pearl harbor e às ilhas aliadas da costa do pacífico em 8 de dezembro de 1 941 ,
Question #166 - E que navio japonês?	
uss arizona	**pt/a/r/i/Arizona_(desambiguação).html**: arizona (desambiguação) . uss arizona (bb - 39) , um navio de guerra americano , afundado em pearl harbor em 1 941 .
navio de guerra	**pt/a/r/i/Arizona_(desambiguação).html**: arizona (desambiguação) . uss arizona (bb - 39) , um navio de guerra americano , afundado em pearl harbor em 1 941 .
batalha naval	**pt/a/u/s/Austrália_nas_Grandes_Guerras_2683.html**: austrália nas grandes guerras . navio sydney e o cruzado alemão kormoran afundaram um ao outro numa batalha naval na costa oeste australiana . 645 australianos morreram no naufrágio . guerra no pacífico (1 942 - 1 945) após os ataques a pearl harbor e às ilhas aliadas da costa do pacífico em 8 de dezembro de 1 941 ,
Question #167 - O que é o Crescente Fértil?	
o crescente fértil é uma região do oriente médio compreendendo os atuais israel , cisjordânia e líbano bem como partes da jordânia , da síria , do iraque , do egito e do sudeste da turquia .	**pt/c/r/e/Crescente_Fértil_01e7.html**: o crescente fértil é uma região do oriente médio compreendendo os atuais israel , cisjordânia e líbano bem como partes da jordânia , da síria , do iraque , do egito e do sudeste da turquia . o termo « crescente fértil » foi criado pelo arqueólogo james henry breasted , da universidade de chicago , em referência ao fato de o arco formado pelas diferentes zonas assemelhar - se a uma lua crescente . irrigada pelo jordão , pelo eufrates , pelo tigre e pelo nilo , a região cobre uma superfície de cerca de 400 000 a 500 000 km ^{2} e é povoada por 40 a 50 milhões de indivíduos .

Question #168 - Diga um clube de futebol de Campinas.	
associação atlética ponte preta	**pt/a/s/s/Associação_Atlética_Ponte_Preta_bebd.html**: associação atlética ponte preta . é um time brasileiro de futebol , situado no bairro da ponte preta , em campinas , estado de são paulo . o time foi fundado no dia 11 de agosto de 1 900 , sendo considerado o segundo mais antigo clube do brasil em atividade ininterrupta ,
guarani futebol clube	**pt/g/u/a/Guarani.html**: guarani . esportes guarani futebol clube , um clube de futebol de campinas , são paulo .
esporte clube santo andré	**pt/e/s/p/Esporte_Clube_Santo_André_07a3.html**: esporte clube santo andré . fundação na data de fundação do santo andré , o que se tinha era a esperança de criar um clube que se rivalizasse com os principais expoentes do futebol paulista . campinas , ribeirão preto , são josé do rio preto e tantas outras cidades brilhavam com seus representantes .
Question #169 - E um de Belo Horizonte.	
cruzeiro esporte clube clube atlético mineiro américa	**pt/b/e/l/Belo_Horizonte_86b8.html**: belo horizonte . clubes de futebol cruzeiro esporte clube clube atlético mineiro américa futebol clube outras informações padroeira nossa senhora da boa viagem comarca belo horizonte domicílios 628 442 analfabetismo 4,6 % gini 0,62 ligações externas página da prefeitura belohorizonte .
nossa senhora da boa viagem	**pt/b/e/l/Belo_Horizonte_86b8.html**: belo horizonte . clubes de futebol cruzeiro esporte clube clube atlético mineiro américa futebol clube outras informações padroeira nossa senhora da boa viagem comarca belo horizonte domicílios 628 442 analfabetismo 4,6 % gini 0,62 ligações externas página da prefeitura belohorizonte .
minas gerais	**pt/c/r/u/Cruzeiro_Esporte_Clube_b654.html**: cruzeiro esporte clube . cruzeiro esporte clube é um clube de futebol brasileiro , com sede na cidade de belo horizonte , minas gerais .
Question #170 - Qual a capital do Mato Grosso?	
mato grosso do sul	**pt/m/a/t/Mato_Grosso_904c.html**: mato grosso . mato grosso brasão bandeira hino localização região centro - oeste capital cuiabá estados limítrofes pará , rondônia , amazonas , mato grosso do sul ,
região metropolitana municípios	**pt/i/v/i/Ivinhema.html**: ivinhema . mato grosso do sul microrregião iguatemi região metropolitana municípios limítrofes nova andradina , novo horizonte do sul , angélica , deodápolis e jateí distância até a capital

campo grande	**pt/m/a/t/Mato_Grosso_do_Sul_60f7.html**: mato grosso do sul . mato grosso do sul brasão bandeira hino localização região centro - oeste capital campo grande estados limítrofes go , mg , mt ,
Question #171 - Quem foi o oitavo marido de Elizabeth Taylor?	
josemaría escrivá	**pt/6/_/d/6_de_Outubro_ac2c.html**: 6 de outubro . lucas pires , apresenta sua demissão 1 991 - elizabeth taylor se casa com larry fortensky , seu oitavo marido 2 002 - canonização de josemaría escrivá ,
josé serra toda vez	**FSP940614-002**: josé serra toda vez que se anuncia mais um casamento de elizabeth taylor , os leitores que acompanham as notícias a respeito da atriz são tomados de um irresistível ceticismo sobre o futuro do novo compromisso . afinal , trata - se do sétimo ou oitavo .
larry fortensky	**pt/6/_/d/6_de_Outubro_ac2c.html**: 6 de outubro . lucas pires , apresenta sua demissão 1 991 - elizabeth taylor se casa com larry fortensky , seu oitavo marido 2 002 - canonização de josemaría escrivá ,
Question #172 - Quando é que eles se casaram?	
1 962	**PUBLICO-19940909-072**: « boulevard » do crepúsculo em 1 962 , enquanto elizabeth taylor entrava triunfalmente em roma na « cleópatra » de mankiewicz , soava em hollywood o toque de finados .
63	**FSP950919-111**: elizabeth taylor deixa hospital após taquicardia a atriz elizabeth taylor , 63 , deixou o santa monica hospital , na califórnia ,
1 956	**PUBLICO-19941212-088**: o famoso " silver cloud " , de 1 956 , que outrora pertenceu a bette davis e a mike todd , com quem a actriz elizabeth taylor esteve casada na década de 50 , foi adquirido pelo preço de 500 mil dólares (80 mil contos) .
Question #173 - Qual é a nacionalidade dela?	
televisão norte - americana nbc por produzir uma série	**PUBLICO-19940819-015**: elizabeth taylor processa nbc a actriz elizabeth taylor vai processar a cadeia de televisão norte - americana nbc por produzir uma série acerca da sua vida que ,
richard burton	**FSP950410-088**: filmada com richard burton e elizabeth taylor nos papéis principais .
montgomery clift	**FSP940703-186**: com montgomery clift , elizabeth taylor , shelley winters , raymond burr .
Question #174 - Quantos gêneros tem o alemão?	
dois gêneros	**FSP940903-076**: chailly -os dois gêneros me são gratificantes .
três gêneros	**pt/d/e/c/Declinação_na_língua_alemã.html**: declinação na língua alemã . o alemão conserva três gêneros : masculino , feminino e neutro ; dois números : singular e plural ; e quatro casos gramaticais : nominativo , acusativo , dativo e genitivo .

17 gêneros	**pt/b/a/l/Balanophoraceae.html**: balanophoraceae . balanophoraceae inclui 17 gêneros e aproximadamente 50 espécies .
Question #175 - E quantos tem o romanche?	
1 552 ,	**pt/l/í/n/Língua_romanche.html**: língua romanche . história o primeiro registro escrito da língua romanche data de 1 552 , na forma de uma lição de catecismo chamada christiauna fuorma , rigistrada por jacob bifrun no dialeto engadino .
1 938 ,	**pt/l/í/n/Língua_romanche.html**: língua romanche . até 1 938 , o romanche não era considerado uma língua oficial da suíça , tendo seu status reconhecido apenas naquele ano .
um)	**pt/s/u/í/Suíça.html**: suíça . omnes pro uno (português : um por todos , todos por um) línguas oficiais alemão , francês , italiano e romanche
Question #176 - Quanto tempo reinou Ramsés II?	
1 213 ac	**pt/r/a/m/Ramsés_II_3363.html**: ramsés ii . ramsés ii foi o terceiro faraó da xix dinastia egípcia , uma das dinastias que compõem o império novo . reinou entre aproximadamente 1 279 e 1 213 ac .
66 anos	**pt/a/b/u/Abu_Simbel_68cb.html**: abu simbel . ramsés ii iniciou o seu reinado em 1 290 ac e reinou durante 66 anos ,
1 290 ac	**pt/a/b/u/Abu_Simbel_68cb.html**: abu simbel . ramsés ii iniciou o seu reinado em 1 290 ac e reinou durante 66 anos ,
Question #177 - Quando começou o seu reinado?	
1 290 ac	**pt/a/b/u/Abu_Simbel_68cb.html**: abu simbel . a construção começou a cerca de 1 284 ac e terminou aproximadamente vinte anos mais tarde . ramsés ii iniciou o seu reinado em 1 290 ac e reinou durante 66 anos ,
século xiii ac	**pt/c/a/n/Canal_de_Suez_4928.html**: canal de suez . evidências indicam sua existência pelo menos no século xiii ac durante o reinado de ramsés ii (ver [1] , [2] , [3] , [4] inglês , [5] , em espanhol) .
1 284 ac	**pt/a/b/u/Abu_Simbel_68cb.html**: abu simbel . a construção começou a cerca de 1 284 ac e terminou aproximadamente vinte anos mais tarde . ramsés ii iniciou o seu reinado em 1 290 ac e reinou durante 66 anos ,
Question #178 - Ele ordenou a construção de que templos?	
abu simbel	**pt/a/n/t/Antigo_Egipto_b5aa.html**: antigo egipto . foi também ramsés ii que ordenou a construção dos templos de abu simbel .
Question #179 - Que se passou a 9 de Novembro de 1991?	
seki	**pt/l/i/s/Lista_de_asteróides_(5001-6000).html**: lista de asteróides (5 001 - 6 000) . h. mori 5 914 - 1 990 wk 20 de novembro de 1 990 rh mcnaught 5 915 yoshihiro 1 991 eu 9 de março de 1 991 t. seki 5 916 van der

normas morais das testemunhas de jeová	**pt/n/o/r/Normas_morais_das_Testemunhas_de_Jeová_1509.html**: normas morais das testemunhas de jeová . os valores parecem ter - se fixado por volta dos 40 mil desassociados e dissociados anuais , a nível mundial , conforme se pode observar na revista a sentinela , nas edições de 11 de novembro de 1 991 , pág . 9 , e 1 de abril de 1 994 ,
ultima	**pt/l/i/s/Lista_de_jogos_do_Super_NES_9992.html**: lista de jogos do super nes . setembro de 1 991 ultima : runes of virtue 2 fci novembro de 1 994 ultima vi
Question #180 - Quantos actos tem a ópera Verdi da Aida?	
quatro atos	**pt/a/i/d/Aida.html**: aida . aida é uma ópera em quatro atos com música de giuseppe verdi e libreto de antonio ghislazoni , com estréia mundial na casa da ópera , cairo , aos 24 dezembro de 1 871 .
dois atos	**pt/a/i/d/Aida_(Musical)_5ff5.html**: aida (musical) . aida é um drama musical em dois atos baseado na ópera italiana hômonima de giuseppe verdi , que , por sua vez , é baseada numa história de auguste mariette . o musical foi produzido pela hyperion theatricals , uma filiada da disney theatrical ,
Question #181 - Quem escreveu o libretto dessa ópera?	
giuseppe verdi	**pt/a/i/d/Aida.html**: aida . aida é uma ópera em quatro atos com música de giuseppe verdi e libreto de antonio ghislazoni , com estréia mundial na casa da ópera , cairo , aos 24 dezembro de 1 871 .
ghislanzoni tipo do enredo	**pt/a/i/d/Aida.html**: aida . da ópera em midi bem como uma breve descrição do argumento (em espanhol) aida nome em português (personagem - título) idioma original italiano compositor giuseppe verdi libretista antonio ghislanzoni tipo do enredo épico número de atos
verdi	**pt/a/i/d/Aida.html**: aida . aida é uma ópera em quatro atos com música de giuseppe verdi e libreto de antonio ghislazoni , com estréia mundial na casa da ópera , cairo , aos 24 dezembro de 1 871 .
Question #182 - Quando é que estreou a ópera?	
maio de 1 987	**PUBLICO-19941122-106**: aida , de giuseppe verdi , que conta a história de amor entre um oficial egípcio e uma escrava etíope . quando a ópera foi pela primeira vez estreada , em maio de 1 987 , no templo faraónico de luxor , um edifício velho de 3 200 anos situado na parte leste da cidade ,
1 861	**pt/g/i/u/Giuseppe_Verdi_b105.html**: giuseppe verdi . durante esse período , verdi era aclamado como um patriota , sendo eleito deputado em 1 861 , ano da unificação e , posteriormente , senador . e continuou escrevendo óperas : em 1 871 estreou aida , em comemoração à abertura do canal de suez .

28 de março , 1 880	**pt/l/a/_/La_Gioconda_(ópera)_1800.html**: la gioconda (ópera) . milão , 28 de março , 1 880) , assim como o maior sucesso na história da ópera italiana entre a aida (1 871) e otello (1 887) de verdi .
Question #183 - Quem se tornou lider do Partido Quebequense em 2005?	
andré boisclair	**pt/p/a/r/Partido_Quebequense_8b0b.html**: partido quebequense . premier 2 001 - 2 003 louise harel (2 005 , ínterim) andré boisclair (2 005 - tempos atuais) partido quebequense nome em inglês não há nome em francês parti québécois data de fundação 11 de outubro de 1 968 atual líder andré boisclair sede montreal ,
parti québécois	**pt/p/a/r/Partido_Quebequense_8b0b.html**: partido quebequense . o partido quebequense (em francês : parti québécois) é um partido político do quebec , canadá ,
governo	**pt/m/o/n/Montreal.html**: montreal . uma ilha , uma cidade em 2 001 , o partido quebequense , então no governo da província , fundiu as outras 27 cidades que ocupavam a ilha a montreal ,
Question #184 - Qual é a maior cidade do Canadá?	
toronto	**pt/c/a/n/Canadá.html**: canadá . montreal é o maior pólo ferroviário do país , seguido por calgary e toronto . toronto e montreal possuem modernos sistemas de metrô . rodovias pavimentadas conectam as principais cidades do canadá entre si .
montreal	**pt/c/a/n/Canadá.html**: canadá . montreal é o maior pólo ferroviário do país , seguido por calgary e toronto . toronto e montreal possuem modernos sistemas de metrô . rodovias pavimentadas conectam as principais cidades do canadá entre si .
província summerside quebec	**pt/l/i/s/Lista_de_cidades_do_Canadá_de9d.html**: lista de cidades do canadá . maior cidade da província summerside quebec ao contrário do restante do canadá , o quebec não possui cidades primárias ou secundárias (cities e towns) .
Question #185 - O que é o Gil Vicente FC?	
o gil vicente futebol clube é uma agremiação desportiva de barcelos . dedica - se ao futebol e a sua equipa principal disputa a liga de honra . história do clube desde que subiu pela última vez ao escalão máximo do futebol português em 1 990 , a equipa de barcelos tem se mantido mantido nela , apenas tendo descido uma vez em 1 997 e regressado em 1 999 à 1 ª liga .	**pt/g/i/l/Gil_Vicente_Futebol_Clube_e640.html**: o gil vicente futebol clube é uma agremiação desportiva de barcelos . dedica - se ao futebol e a sua equipa principal disputa a liga de honra . história do clube desde que subiu pela última vez ao escalão máximo do futebol português em 1 990 , a equipa de barcelos tem se mantido mantido nela , apenas tendo descido uma vez em 1 997 e regressado em 1 999 à 1 ª liga . o gil vicente nasceu em maio de 1 924 , após o nascimento de outros clubes na cidade de barcelos , tal como o barcellos sporting club e o união foot - ball club barcellense .

Question #186 - Quem foi Gil Vicente?	
gil vicente (1 465 - 1 536 ?) é geralmente considerado o primeiro grande dramaturgo português , além de poeta de renome . há quem o identifique com o ourives , autor da custódia de belém , mestre da balança , e com o mestre de retórica do rei dom manuel .	**pt/g/i/l/Gil_Vicente_f346.html**: gil vicente (1 465 - 1 536 ?) é geralmente considerado o primeiro grande dramaturgo português , além de poeta de renome . há quem o identifique com o ourives , autor da custódia de belém , mestre da balança , e com o mestre de retórica do rei dom manuel . enquanto homem de teatro , parece ter também desempenhado as tarefas de músico , actor e encenador . é frequentemente considerado , de uma forma geral , o pai do teatro português , ou mesmo do teatro ibérico já que também escreveu em castelhano - partilhando a paternidade da dramaturgia espanhola com juan del encina .
Question #187 - Quem foi o "pai do teatro português"?	
gil vicente	**pt/m/a/n/Manuel_I_de_Portugal_4cf8.html**: manuel i de portugal . na sua corte surge também gil vicente , o pai do teatro português e duarte pacheco pereira o geógrafo , autor do esmeraldo de situ orbis .
duarte pacheco pereira o geógrafo	**pt/m/a/n/Manuel_I_de_Portugal_4cf8.html**: manuel i de portugal . na sua corte surge também gil vicente , o pai do teatro português e duarte pacheco pereira o geógrafo , autor do esmeraldo de situ orbis .
manuel i de portugal	**pt/m/a/n/Manuel_I_de_Portugal_4cf8.html**: manuel i de portugal . na sua corte surge também gil vicente , o pai do teatro português e duarte pacheco pereira o geógrafo , autor do esmeraldo de situ orbis .
Question #188 - Qual a área do Parque Estadual Guariba?	
72 296,331 hectares	**pt/l/i/s/Lista_de_parques_estaduais_do_Brasil_fb87.html**: lista de parques estaduais do brasil . região norte acre parque estadual chandless amapá amazonas parque estadual guariba , com 72 296,331 hectares , criado em 2 005 ,
Question #189 - Quando foi criado o parque?	
2 005	**pt/l/i/s/Lista_de_parques_estaduais_do_Brasil_fb87.html**: lista de parques estaduais do brasil . região norte acre parque estadual chandless amapá amazonas parque estadual guariba , com 72 296,331 hectares , criado em 2 005 ,
Question #190 - O que é a Torre do Tombo?	
torre do tombo é o nome do arquivo central do estado português desde a idade média . com mais de 600 anos , é uma das mais antigas instituições portuguesas ainda activas .	**pt/t/o/r/Torre_do_Tombo_99b6.html**: torre do tombo é o nome do arquivo central do estado português desde a idade média . com mais de 600 anos , é uma das mais antigas instituições portuguesas ainda activas . ao longo do tempo , a conservação dos documentos foi prejudicada por um conjunto de circunstâncias : não apenas pelo terramoto de 1 755 , mas também as frequentes mudanças de local , incêndios , a transferência da corte para o rio de janeiro no brasil , o desvio de materiais aquando do domínio filipino e das invasões francesas etc .

Question #191 - Onde fica?	
arquivo nacional torre tombo	**PUBLICO-19940112-005**: e manuela mendonça , subdirectora dos arquivos nacionais / torre do tombo (antt) , por onde tem passado o processamento dos pagamentos . na segunda reunião , a 4 de março , manuela mendonça só pôde ficar 20 minutos ,
nacional	**PUBLICO-19951222-005**: torre do tombo e a rede nacional de arquivos foi ontem reafirmado pelo secretário de estado da cultura , rui vieira nery ,
pide	**PUBLICO-19940112-088**: torre do tombo (antt) , impedindo a consulta do arquivo da pide e seus quejandos desde que passaram a ficar à sua guarda .
Question #192 - Que país faz fronteira com Cuba?	
coréia do sul e dos eua se defrontam com soldados da coréia do norte em alerta	**FSP950219-080**: é uma das últimas fronteiras da guerra fria (a outra é guantánamo , em cuba) . ali , como em todos os 160 km da fronteira intercoreana , tropas da coréia do sul e dos eua se defrontam com soldados da coréia do norte em alerta permanente .
país	**PUBLICO-19941209-039**: nascido na localidade de ameneiro , na comarca de tuy , perto da fronteira portuguesa , como muitos outros galegos emigrou para cuba , país de onde regressou a espanha no final da ditadura de primo de rivera , mas a que ficou para sempre ligado .
eua	**FSP940828-053**: a decisão também se deve a motivos de segurança : há temores de que , para aumentar a pressão sobre os eua , o governo de cuba libere a sua fronteira com guantánamo , o que poderia levar milhares de pessoas à base .
Question #193 - Qual é o comprimento do metro de Coimbra?	
cerca de 100 metros	**PUBLICO-19950812-094**: coimbra . aproveitando o areal ali existente , pretende - se agora construir um dique em pedra , com cerca de 100 metros de comprimento , que permita a manutenção de um lençol de água com metro e meio de profundidade .
cerca de 36 metros	**PUBLICO-19940123-066**: generoso nos adjectivos , o autor do estudo preliminar da introdução do metro de superfície em coimbra descreve assim o comboio ligeiro urbano . um veículo que funciona a tracção eléctrica , tem cerca de 36 metros de comprimento e tanto pode circular numa rede urbana de eléctricos (com marcha à vista) ,

5 metros	**FSP950815-033**: a vala tem 5 metros de comprimento por 4 metros de largura . a profundidade , disse o encarregado , é de até 3 metros . parte do terreno foi tomado por covas de crianças mortas . para a presidente do grupo tortura nunca mais , cecília coimbra ,
Question #194 - Quantas esposas tinha Ngungunhane?	
Question #195 - Como é que se chamava o filho dele?	
rei d. carlos i de portugal visita os açores	**pt/n/g/u/Ngungunhane.html**: ngungunhane . zixaxa casa e tem um filho , também chamado roberto zixaxa , fundando uma família que ainda está presente na sociedade angrense . quando o rei d. carlos i de portugal visita os açores em 1 901 , ngungunhane
passando a designar	**pt/n/g/u/Ngungunhane.html**: ngungunhane . em mossurize , mundagaz ascendeu ao trono nguni , passando a designar - se por ngungunhane , gungunyane ou gungunhana , o filho de muzila e o leão de gaza . o imperador ngungunhane
ascendeu ao trono nguni	**pt/n/g/u/Ngungunhane.html**: ngungunhane . em mossurize , mundagaz ascendeu ao trono nguni , passando a designar - se por ngungunhane , gungunyane ou gungunhana , o filho de muzila e o leão de gaza . o imperador ngungunhane
Question #196 - Qual é a capital de Cuba?	
havana	**pt/c/u/b/Cuba.html**: cuba . a sua capital é havana (em espanhol , la habana) . a república de cuba é atualmente o único estado socialista das américas e um dos poucos do mundo .
santiago de cuba	**pt/p/r/o/Província_de_Santiago_de_Cuba_3271.html**: província de santiago de cuba . a capital da província é a cidade de santiago de cuba .
norte	**PUBLICO-19950211-067**: cuba substituir o investimento norte - americano pelo de outros países . a ideia é portanto afugentar os investidores , num momento em que havana se esforça por atrair , e com algum êxito , capital estrangeiro .
Question #197 - Quem criou o primeiro alfabeto?	
alfabeto fonético internacional	**pt/x/-/s/X-SAMPA_e84b.html**: x - sampa . alfabeto fonético internacional . o resultado é uma remodelagem do afi inspirada no sampa em ascii de 7 bits . características os símbolos afi que são letras minúsculas comuns são representadas do mesmo modo no x - sampa . x - sampa usa a contrabarra como modificador para criar um novo símbolo .

internacional	**pt/x/-/s/X-SAMPA_e84b.html**: x - sampa . alfabeto fonético internacional . o resultado é uma remodelagem do afi inspirada no sampa em ascii de 7 bits . características os símbolos afi que são letras minúsculas comuns são representadas do mesmo modo no x - sampa . x - sampa usa a contrabarra como modificador para criar um novo símbolo .
fenícia	**pt/h/i/s/História_Antiga_34b4.html**: história antiga . quanto à escrita , foram pioneiros na arte de escrever , pois sinais e marcas chegaram à fenícia , onde foram simplificados , resultando no alfabeto que temos nos dias de hoje .
Question #198 - Quando é que Porto Rico se tornou um estados dos EUA?	
Question #199 - Onde fica Livorno?	
comuna italiana da região	**pt/l/i/v/Livorno.html**: livorno . livorno é uma comuna italiana da região da toscana , província de livorno , com cerca de 148 143 habitantes .
província de livorno	**pt/l/i/v/Livorno.html**: livorno . livorno é uma comuna italiana da região da toscana , província de livorno , com cerca de 148 143 habitantes .
cagliari chievo empoli	**pt/s/e/r/Serie_A_(2005-06)_f5a2.html**: serie a (2 005 - 06) . ascoli cagliari chievo empoli fiorentina internazionale juventus lazio lecce livorno messina milan palermo parma reggina roma sampdoria siena treviso udinese série b a série b é organizada pela lega calcio e tem como principal patrocinador a tim .
Question #200 - O que são os iaques?	
os iaques (bos grunniens) , também conhecido como boi - cavalo , são bois selvagens asiáticos , encontrados no planalto tibetano , em altitudes que variam entre 4 500 m e 6 000 m. possuem uma longa pelagem negra a marrom - escura e grandes chifres curvados para cima e para frente .	**pt/i/a/q/Iaque.html**: os iaques (bos grunniens) , também conhecido como boi - cavalo , são bois selvagens asiáticos , encontrados no planalto tibetano , em altitudes que variam entre 4 500 m e 6 000 m. possuem uma longa pelagem negra a marrom - escura e grandes chifres curvados para cima e para frente . foram domesticado em algumas regiões da ásia central . quando domesticados , fornecem lã , carne e leite , bem como sendo utilizados como animais de tração .

E Case Study Data

E.1 Data Collection

The time intervals are written as minutes:seconds in the video stream (mm:ss).

Table E.1: Manual Transcripts: Part 1 of 5 (Transcripts 1-18)

T.#	Manual Transcript
1	Pela quarta vez na história o hino Português tocou nos Jogos Olímpicos.
2	Para já dou-lhe conta de um descarrilamento na linha do Tua que provocou um morto e 37 feridos.
3	Às dez e quarenta da manhã, o metro de Mirandela transportava 47 pessoas, quase todas turistas que aproveitavam o dia para apreciar o Tua.
4	Além do inquérito instaurado pelo governo, este acidente está também a ser investigado pelo Núcleo de Investigação Criminal da GNR e pela Policia Judiciária
5	Uma composição do metro de Mirandela descarrilou na Linha do Tua. O acidente fez um ferido. Este é o terceiro acidente do género no mesmo local no espaço de um ano e meio. Em cento e vinte anos de existência, a Linha do Tua nunca tinha registado acidentes graves.
6	A maior estrela dos jogos olimpicos de Pequim está de férias no Algarve. Michael Phelps depois das 8 medalhas o descanso em Portugal.
7	João Ubaldo Ribeiro foi distinguido com o prémio Camões 2008. O mais importante galardão atribuído a autores de Língua Portuguesa, distinguiu este ano um escritor Brasileiro, depois de em 2007 ter ido para as mãos do Português António Lobo Antunes.
8	Passavam 14 minutos das onze da noite quando foi dado o alarme: estava a arder o número 23 da Avenida da Liberdade, uma das principais avenidas da Capital.
9	A selecção nacional de futebol já está na Suíça. A equipa partiu a meio da tarde do aeroporto de Lisboa e aterrou há cerca de duas horas. Depois seguiu viagem até Neuchâtel, a primeira cidade onde vai permanecer durante o Europeu de Futebol.
10	Um ano após a abertura o Museu Berardo teve mais de meio milhão de visitantes.
11	Na primeira intervenção como líder parlamentar do PSD, Paulo Rangel disse mesmo que Sócrates desrespeitou o Parlamento, antecipando o debate da nação.
12	Entraram em vigor as alterações ao Código da estrada. A partir de hoje bastam três contra ordenações muito graves para haver cassação da carta.
13	Reunidos em Osaca, no Japão, os ministros das finanças do G-8 não chegaram a acordo, nem sobre as causas da crise nem sobre as soluções, mas defendem uma maior transparência no mercado petrolífero.
14	Cerca de cinco mil manifestantes, mas apenas quatro por fila, foi a condição das autoridades para autorizarem a marcha da contestação ao G-8, e sob fortes medidas de vigilância dos vinte mil policias mobilizados para Hokkaido, no Japão.
15	Foi ontem inaugurado em Porto Alegre o museu de Iberê Camargo. O projecto do arquitecto Siza Vieira que já ganhou uma medalha de ouro na Bienal de Veneza, e que no Brasil já é considerado um dos edifícios contemporâneos mais bonitos do País.
16	A sede da fundação José Saramago vai ser na Casa dos Bicos, em Lisboa.
17	Com um salto de 17 metros e 67 centímetros, Nelson Évora fez subir a bandeira nacional ao mastro mais alto de Pequim. E quando soou a Portuguêsa esta tarde, Nelson Évora, campeão olímpico do triplo salto, confessa que se arrepiou.
18	Ao quartel-general em Montreal, no Canadá, lugar onde nasceu o Cirque du Soleil há quase 30 anos, chegam pessoas muito diferentes, de diferentes lugares, para serem desafiadas.

Table E.2: Manual Transcripts: Part 2 of 5 (Transcripts 19-33)

T.#	Manual Transcript
19	(. . .) É um caso que vamos conhecer na segunda parte do Telejornal, onde vamos olhar também para a estreia da Liga Sagres, o campeonato nacional de futebol, que está de regresso à RTP. Até já.
20	O Benfica também não vai contar com o defesa Nelson, vendido hoje ao Sevilha por 5 milhões de euros. O lateral Português assinou um contrato válido para as próximas 5 temporadas
21	Jaime Silva diz que são inaceitáveis os confrontos na lota de Matosinhos entre pescadores e comerciantes
22	Após passar seis anos numa prisão judaica voltou ao Líbano. Nassim Nasser tinha à sua espera uma recepção de herói. Em 2002 foi acusado em Israel de espiar a favor do Hezbollah. Perdeu cidadania Israelita que tinha graças à sua mãe, uma judia casada com um muçulmano Libanês. Foi agora entregue à Cruz Vermelha e deportado.
23	O preço do petróleo bateu mais um recorde. Em Nova Iorque o barril de light atingiu novo máximo histórico, nos cento e quarenta e cinco dólares e oitenta e cinco cêntimos. Em Londres o barril de brent atingiu também novo recorde a negociar perto dos cento e quarenta e sete dólares.
24	Dirigiu-se ao parlamento pela primeira vez como líder da bancada laranja para repetir palavras de Manuela Ferreira Leite. Paulo Rangel criticou o programa de obras públicas do governo.
25	A cantora Brasileira Rosa Passos está novamente em Portugal para uma digressão a começar esta sexta-feira. Rosa Passos actua sexta-feira no Centro Cultural de Belém, em Lisboa, para a apresentação do seu mais recente álbum, Romance.
26	Rosa Passos nasceu em Abril de 1952. Estreou-se em 1979 com o álbum Recriação, . . .
27	Carlos do Carmo, o grande senhor do fado Português revisita em Faro a sua carreira de mais de 40 anos.
28	Michael Phelps soma já 13 medalhas de ouro conquistadas em Jogos Olímpicos, é também o nadador mais bem pago da história, mas afinal quanto vale em euros a estrela dos Jogos Olímpicos de Pequim?
29	Phelps tornou-se milionário e atleta profissional da natação poucos meses antes de fazer 16 anos.
30	A provincia de Xichuan engalanou-se para receber a tocha olímpica. Mas a festa não durou muito. Horas depois um forte tremor de terra voltou a sacudir a região. O abalo, de 6.0 na escala de Richter provocou pelo menos 1 morto e 23 feridos. O epicentro foi a 1250 km da capital Chinesa...
31	Nelson Évora concretizou o sonho olímpico: é ele o novo campeão olímpico do triplo salto, conquistou a medalha de ouro para Portugal. Conseguiu o lugar ao quarto ensaio quando saltou 17 metros e 67 centímetros no estádio "Ninho de Pássaro".
32	Em 1969 Fontes Rocha e Fernando Alvim deram vida à formiga cantada por Amália, uma adaptação de Alain Oulman do poema de O'Neil velha fábula em Bossa Nova.
33	À Ossétia do Norte chegou entretanto o alto comissário da ONU para os refúgiados. António Guterres foi coordenar a ajuda aos milhares de deslocados que fugiram da Ossétia do Sul.

Table E.3: Manual Transcripts: Part 3 of 5 (Transcripts 34-45)

T.#	Manual Transcript
34	A cientista Portuguesa Elvira Fortunato ganhou o maior prémio alguma vez atribuído pelo European Research Council, considerado uma espécie de prémios Nobel na área da engenharia. A investigadora liderou uma equipa da Universidade Nova que conseguiu produzir transistores com papel, que não usam o silicio como isolador de curto circuitos.
35	O jamaicano Usain Bolt é o novo campeão olímpico dos 100 metros. Foi uma prova espantosa: ele tornou-se o primeiro atleta a correr a distância abaixo dos 9 segundos e 70 centésimos.
36	Quanto ao assalto de ontem na auto-estrada do sul, sabe-se que o explosivo utilizado para rebentar as portas da carrinha de valores foi um explosivo militar, de utilização relativamente restrita. Trata-se do C 4, um explosivo utilizado habitualmente pelas autoridades para fazerem demolições e arrombarem portas.
37	...é extremamente seguro, pode-se manusear, transportar com elevados padrões de segurança. Para se iniciar este tipo de explosivo é preciso que lhe criemos um estímulo energético elevado ...
38	O C-4 é como, imaginemos, a plasticina usada nas escolas, pelos garotos, pelos miúdos. Portanto, no fundo, é isso. É uma barra de plasticina ...
39	Manuela Ferreira Leite não vai participar na festa do Pombal, que decorre, do Pontal aliás, que decorre esta noite no Algarve. O orador principal é Angelo Correia.
40	Na cimeira internacional sobre a SIDA no México dezenas de manifestantes exigiram que os países apoiem mais estes doentes. As Nações Unidas calculam que a doença possa fazer mais de 50 milhões de crianças órfas só em África.
41	Por aqui diz-se que quem beber a água do Rio Bengo fica eternamente ligado a Angola. A crer na lenda, e aplicando aos negócios, isso significa para os empresários Portugueses uma vantagem, que não é mágica, mas sim histórica, face aos concorrentes Chineses, Sul Africanos ou Americanos; E se nos negócios os números dizem quase tudo, os Portugueses esperam que as palavras "Língua" e "Cultura" digam o resto.
42	Para o encontro frente às ilhas faroé, Queiróz chamou três guarda-redes: Quim do Benfica, Eduardo do Braga é chamado pela primeira vez à selecção nacional, e Daniel Fernandes do Bochum
43	O primeiro a fazer um comício no Pontal foi Sá Carneiro no verão quente de 75.
44	Michael Phelps voltou a deslumbrar. O norte-americano venceu os 100 metros mariposa e igualou Mark Spitz ao conquistar a sétima medalha de ouro nos Jogos Olímpicos
45	João Ubaldo Ribeiro nasceu em 1941, trabalhou como jornalista antes de se dedicar, por inteiro, e em exclusivo, à escrita

Table E.4: Manual Transcripts: Part 4 of 5 (Transcripts 46-58)

T.#	Manual Transcript
46	A Câmara está a finalizar um contrato de financiamento por parte do IHRU, o Instituto da Habitação e da Reabilitação Urbana com o apoio do BEI, do Banco Europeu do Investimento, com vista a retomar as empreitadas de reabilitação urbana que foram iniciadas no inicio dos anos 2000, e que não puderam prosseguir por falta de capacidade da Câmara Municipal.
47	José Sócrates debateu com o presidente Angolano José Eduardo dos Santos os laços comerciais entre os dois países. À saída, o primeiro ministro anunciou o reforço das linha de crédito para as empresas Portuguesas que invistam em Angola.
48	Enquanto as operações de rescaldo decorriam, chegava à Avenida da Liberdade o Presidente da Junta de freguesia mais rica do País: o abastado dono do número 21 quis em testamento que este edificio fosse para sempre propriedade das gentes da sua terra, Galveias.
49	Os 15 inquilinos da Junta de Galveias não podiam estar mais revoltados com o incêndio que lhes roubou o tecto.
50	A Brasileira EMBRAER é a terceira maior empresa mundial no fabrico de aviões, e o maior accionista da Portuguesa OGMA com 65 por cento do capital. O Estado Português detém os restantes 35 por cento.
51	As chamas obrigaram ao corte da Avenida da Liberdade às primeiras horas da madrugada. O incêndio aconteceu no último quarteirão antes da Praça dos Restauradores. Começou no número 23 da Avenida, um prédio de 6 andares, devoluto. As chamas saltaram os telhados, e depressa se propagaram ao edificio do lado, o número 21; aqui moravam 15 pessoas. As chamas não ficaram pela Avenida da Liberdade, elas foram projectadas para a rua de trás, atingindo o número 6 da Rua da Glória, onde viviam sete pessoas. Todos os moradores afirmam que o prédio devoluto há muito que devia estar isolado e limpo.
52	Os aviões da EMBRAER são de tamanha qualidade que até o governo está comprando dois aviões
53	Boa Noite, a EMBRAER uma das maiores construtoras mundiais de aviões vai investir cerca de 150 milhões de euros em duas fábricas na cidade de Évora. O investimento deverá permitir a criação de pelo menos 500 postos de trabalho directos e mais de 1000 indirectos
54	Para Jorge Nuno Pinto da Costa, presidente do Futebol Clube do Porto, o concelho de justiça da Federação não tem credibilidade.
55	Começo hoje mais uma edição do festival andanças. Até domingo em São Pedro do Sul há danças populares e música de todo o mundo.
56	A verdade é que a campanha de Obama está a revolucionar a forma de comunicar política. Através da internet construiu uma rede de 2 milhões de voluntários e recebeu o número recorde de 240 milhões de dólares de pequenos doadores.
57	Foi há precisamente 50 anos, em Julho de 1958, que João Gilberto gravou "chega de saudade". O tema marcou o início da bossa nova, movimento que cruzou o samba com as harmonias do jazz, e tornou a música Brasileira conhecida mundialmente.
58	Jerónimo de Sousa declarou aberta a Festa do Avante, a trigésima segunda edição da Festa ...

Table E.5: Manual Transcripts: Part 5 of 5 (Transcripts 59-68)

T.#	Manual Transcript
59	Barack Obama terminou a visita ao Médio Oriente e à Europa. A última paragem foi em Londres.
60	Começou à chuva a Festa do Avante. As portas da Quinta da Atalaia abriram às seis da tarde, para três dias de música, teatro e debates políticos.
61	Da cidade de Évora vão saír para o Brasil componentes para os aviões da EMBRAER, os melhores segundo Lula da Silva
62	Há cada vez mais casos de jovens com cancro cutâneo, e é também entre os jovens que se regista o maior número de queimaduras solares. Para chegar a mensagem principalmente aos mais novos, a Associação Portuguêsa de Cancro Cutâneo, levou figuras públicas à praia da Falésia, em Vilamoura no Algarve.
63	(. . .) com Manuel Salgado a dizer que existem em Lisboa quatro mil seiscentos e dois prédios devolutos, como este 23 da Avenida da Liberdade (. . .)
64	A primeira linha de crédito é uma linha de ajuda, de cem milhões de euros. Depois há uma linha garantia COSEC, de trezentos, que passa de 300 milhões para 500 milhões de euros porque estava praticamente esgotada, e há finalmente uma linha de crédito comercial, da Caixa Geral de Depósitos, de 500 milhões de euros.
65	Quando Emir Kusturica se junta à Smoking Orchestra, a festa é garantida. É o que vai acontecer na praia do Tonel, em Sagres, amanhã à noite, no encerramento do Super Bock Surf Fest
66	É a grande surpresa do torneio olímpico de ténis: Roger Federer, a quatro dias de perder o estatuto de número um mundial para Rafael Nadal, foi derrotado por James Blake, norte-americano, sétimo no ranking ATP . . .
67	Roger Federer foi eliminado pelo norte-americano James Blake
68	Estamos em Baião, no Vale do Tâmega, um dos concelhos mais pobres do País. Há dois dias a inauguração de um hotel deu um novo ânimo à região: criaram-se 35 postos de trabalho.

Table E.6: Transcript Document and Speaker Information: Part 1 of 2 (Transcripts 1-40)

T.#	Document	Time Interval	Speaker	Gender	Environment
1	2008_08_22-19_59_02-Telejornal-1.avi	00:07 to 00:12	JAC	M	A
2	2008_08_22-19_59_02-Telejornal-1.avi	01:24 to 01:30	JAC	M	A
3	2008_08_22-19_59_02-Telejornal-1.avi	01:50 to 01:59	SF	F	P + N
4	2008_08_22-19_59_02-Telejornal-1.avi	04:01 to 04:10	SF	F	P + N
5	2008_06_06-19_59_01-Telejornal-1.avi	15:48 to 16:05	JRS	M	A + J
6	2008_08_22-19_59_02-Telejornal-1.avi	28:23 to 28:31	JAC	M	A + J
7	2008_07_26-21_59_02-Jornal2-2.avi	15:41 to 15:57	CE	F	A
8	2008_07_07-19_59_01-Telejornal-1.avi	00:24 to 00:31	JAC	M	A + N
9	2008_06_01-21_59_02-Jornal2-2.avi	00:47 to 00:57	CC	F	A
10	2008_07_03-21_59_01-Jornal2-2.avi	15:26 to 15:30	TN	F	P
11	2008_07_03-21_59_01-Jornal2-2.avi	06:44 to 06:53	CC	F	A
12	2008_07_06-19_59_02-Telejornal-1.avi	15:14 to 15:23	JRS	M	A
13	2008_06_14-21_59_01-Jornal2-2.avi	03:08 to 03:20	AMF	F	A
14	2008_07_06-19_59_02-Telejornal-1.avi	48:05 to 48:19	FF	F	P + N
15	2008_06_01-21_59_02-Jornal2-2.avi	21:20 to 21:33	CC	F	A
16	2008_07_17-19_59_01-Telejornal-1.avi	58:53 to 58:58	JAC	M	A
17	2008_08_22-19_59_02-Telejornal-1.avi	28:35 to 28:50	JAC	M	A
18	2008_08_22-19_59_02-Telejornal-1.avi	42:27 to 42:40	TN	F	P + N
19	2008_08_22-19_59_02-Telejornal-1.avi	43:54 to 44:04	JAC	M	A + J
20	2008_08_22-19_59_02-Telejornal-1.avi	56:32 to 56:42	JAC	M	A + J
21	2008_06_01-21_59_02-Jornal2-2.avi	00:27 to 00:33	CC	F	A + J
22	2008_06_01-21_59_02-Jornal2-2.avi	19:01 to 19:23	AC	M	P + M
23	2008_07_03-21_59_01-Jornal2-2.avi	00:44 to 01:02	CC	F	A
24	2008_07_03-21_59_01-Jornal2-2.avi	06:55 to 07:06	DS	F	P + N
25	2008_07_03-21_59_01-Jornal2-2.avi	29:52 to 30:05	CC	F	A
26	2008_07_03-21_59_01-Jornal2-2.avi	30:16 to 30:23	CC	F	A + M
27	2008_08_14-21_59_01-Jornal2-2.avi	40:11 to 40:19	AMF	F	A + M
28	2008_08_16-19_59_02-Telejornal-1.avi	33:44 to 33:59	JS	F	A
29	2008_08_16-21_59_02-Jornal2-2.avi	30:21 to 30:29	LS	F	P + M
30	2008_08_05-21_59_02-Jornal2-2.avi	30:50 to 31:40	AC	M	P + N
31	2008_08_21-19_59_01-Telejornal-1.avi	00:07 to 00:28	JAC	M	A + N
32	2008_07_26-21_59_02-Jornal2-2.avi	32:25 to 33:23	ALR	F	P + M
33	2008_08_21-19_59_01-Telejornal-1.avi	50:44 to 50:44	AC	M	P + N
34	2008_07_26-21_59_02-Jornal2-2.avi	17:10 to 17:30	CE	F	A
35	2008_08_16-19_59_02-Telejornal-1.avi	26:48 to 27:01	JS	F	A
36	2008_08_21-19_59_01-Telejornal-1.avi	39:50 to 40:08	JAC	M	A
37	2008_08_21-19_59_01-Telejornal-1.avi	40:25 to 40:39	GS	M	S
38	2008_08_21-19_59_01-Telejornal-1.avi	40:45 to 40:55	VF	M	S + T
39	2008_08_14-21_59_01-Jornal2-2.avi	28:10 to 28:20	AMF	F	A
40	2008_08_05-21_59_02-Jornal2-2.avi	27:16 to 27:29	PG	F	A

Table E.7: Transcript Document and Speaker Information: Part 2 of 2 (Transcripts 41-68)

T.#	Document	Time Interval	Speaker	Gender	Environment
41	2008_07_17-21_59_02-Jornal2-2.avi	05:18 to 05:41	ILS	F	P + N
42	2008_08_14-21_59_01-Jornal2-2.avi	35:17 to 35:28	EP	M	P + N
43	2008_08_14-21_59_01-Jornal2-2.avi	29:34 to 29:39	LB	F	P + N
44	2008_08_16-19_59_02-Telejornal-1.avi	30:41 to 30:51	JS	F	A
45	2008_07_26-21_59_02-Jornal2-2.avi	15:58 to 16:06	CE	F	A + N
46	2008_07_07-19_59_01-Telejornal-1.avi	07:48 to 08:10	MS	M	S
47	2008_07_17-21_59_02-Jornal2-2.avi	03:32 to 03:45	ILS	F	P + N
48	2008_07_07-19_59_01-Telejornal-1.avi	02:47 to 03:00	SF	F	P + N
49	2008_07_07-19_59_01-Telejornal-1.avi	03:20 to 03:25	SF	F	P + N
50	2008_07_26-21_59_02-Jornal2-2.avi	02:25 to 03:36	JB	M	P + N
51	2008_07_07-19_59_01-Telejornal-1.avi	03:54 to 04:35	JAC	M	A
52	2008_07_26-21_59_02-Jornal2-2.avi	01:54 to 02:00	LS	M	S
53	2008_07_26-21_59_02-Jornal2-2.avi	00:42 to 00:57	CE	F	A
54	2008_07_06-19_59_02-Telejornal-1.avi	10:55 to 11:02	PS	M	P + N
55	2008_08_05-21_59_02-Jornal2-2.avi	30:21 to 30:29	PG	F	A
56	2008_08_14-21_59_01-Jornal2-2.avi	24:28 to 24:37	VG	M	P + N
57	2008_07_26-21_59_02-Jornal2-2.avi	29:40 to 29:57	CE	F	A
58	2008_09_05-21_59_01-Jornal2-2.avi	15:05 to 15:10	ASF	M	O
59	2008_07_26-21_59_02-Jornal2-2.avi	11:50 to 11:56	CE	F	A
60	2008_09_05-21_59_01-Jornal2-2.avi	14:12 to 14:22	AR	F	A
61	2008_07_26-21_59_02-Jornal2-2.avi	01:45 to 01:53	JB	M	P + N
62	2008_07_19-19_59_02-Telejornal-1.avi	04:27 to 04:44	JS	F	A
63	2008_07_07-19_59_01-Telejornal-1.avi	07:26 to 07:37	AS	F	O
64	2008_07_17-21_59_02-Jornal2-2.avi	03:46 04:03	JS	M	S
65	2008_08_14-21_59_01-Jornal2-2.avi	40:34 to 40:48	AMF	F	A + M
66	2008_08_14-21_59_01-Jornal2-2.avi	33:19 to 33:33	PS	M	P + N
67	2008_08_14-21_59_01-Jornal2-2.avi	33:07 to 33:11	AMF	F	A
68	2008_07_17-21_59_02-Jornal2-2.avi	14:30 to 14:43	SMS	F	P

Table E.8: Document and Transcript Chronological Order: Part 1 of 2

Document	Time Interval	T.#	Document	Time Interval	T.#
2008_06_01-21_59_02-Jornal2-2.avi	00:27 to 00:33	21	2008_07_17-21_59_02-Jornal2-2.avi	03:32 to 03:45	47
2008_06_01-21_59_02-Jornal2-2.avi	00:47 to 00:57	9	2008_07_17-21_59_02-Jornal2-2.avi	03:46 to 04:03	64
2008_06_01-21_59_02-Jornal2-2.avi	19:01 to 19:23	22	2008_07_17-21_59_02-Jornal2-2.avi	05:18 to 05:41	41
2008_06_01-21_59_02-Jornal2-2.avi	21:20 to 21:33	15	2008_07_17-21_59_02-Jornal2-2.avi	14:30 to 14:43	68
2008_06_06-19_59_01-Telejornal-1.avi	15:48 to 16:05	5	2008_07_19-19_59_02-Telejornal-1.avi	04:27 to 04:44	62
2008_06_14-21_59_01-Jornal2-2.avi	03:08 to 03:20	13	2008_07_26-21_59_02-Jornal2-2.avi	00:42 to 00:57	53
2008_07_03-21_59_01-Jornal2-2.avi	00:44 to 01:02	23	2008_07_26-21_59_02-Jornal2-2.avi	01:45 to 01:53	61
2008_07_03-21_59_01-Jornal2-2.avi	06:44 to 06:53	11	2008_07_26-21_59_02-Jornal2-2.avi	01:54 to 02:00	52
2008_07_03-21_59_01-Jornal2-2.avi	06:55 to 07:06	24	2008_07_26-21_59_02-Jornal2-2.avi	02:25 to 03:36	50
2008_07_03-21_59_01-Jornal2-2.avi	15:26 to 15:30	10	2008_07_26-21_59_02-Jornal2-2.avi	11:50 to 11:56	59
2008_07_03-21_59_01-Jornal2-2.avi	29:52 to 30:05	25	2008_07_26-21_59_02-Jornal2-2.avi	15:41 to 15:57	7
2008_07_03-21_59_01-Jornal2-2.avi	30:16 to 30:23	26	2008_07_26-21_59_02-Jornal2-2.avi	15:58 to 16:06	45
2008_07_06-19_59_02-Telejornal-1.avi	10:55 to 11:02	54	2008_07_26-21_59_02-Jornal2-2.avi	17:10 to 17:30	34
2008_07_06-19_59_02-Telejornal-1.avi	15:14 to 15:23	12	2008_07_26-21_59_02-Jornal2-2.avi	29:40 to 29:57	57
2008_07_06-19_59_02-Telejornal-1.avi	48:05 to 48:19	14	2008_07_26-21_59_02-Jornal2-2.avi	32:25 to 33:23	32
2008_07_07-19_59_01-Telejornal-1.avi	00:24 to 00:31	8	2008_08_05-21_59_02-Jornal2-2.avi	27:16 to 27:29	40
2008_07_07-19_59_01-Telejornal-1.avi	02:47 to 03:00	48	2008_08_05-21_59_02-Jornal2-2.avi	30:21 to 30:29	55
2008_07_07-19_59_01-Telejornal-1.avi	03:20 to 03:25	49	2008_08_05-21_59_02-Jornal2-2.avi	30:50 to 31:40	30
2008_07_07-19_59_01-Telejornal-1.avi	03:54 to 04:35	51	2008_08_14-21_59_01-Jornal2-2.avi	24:28 to 24:37	56
2008_07_07-19_59_01-Telejornal-1.avi	07:26 to 07:37	63	2008_08_14-21_59_01-Jornal2-2.avi	28:10 to 28:20	39
2008_07_07-19_59_01-Telejornal-1.avi	07:48 to 08:10	46	2008_08_14-21_59_01-Jornal2-2.avi	29:34 to 29:39	43

Table E.9: Document and Transcript Chronological Order: Part 2 of 2

Document	Time Interval	T.#	Document	Time Interval	T.#
2008_08_14-21_59_01-Jornal2-2.avi	33:07 to 33:11	67	2008_08_21-19_59_01-Telejornal-1.avi	50:44 to 50:44	33
2008_08_14-21_59_01-Jornal2-2.avi	33:19 to 33:33	66	2008_08_22-19_59_02-Telejornal-1.avi	00:07 to 00:12	1
2008_08_14-21_59_01-Jornal2-2.avi	35:17 to 35:28	42	2008_08_22-19_59_02-Telejornal-1.avi	01:24 to 01:30	2
2008_08_14-21_59_01-Jornal2-2.avi	40:11 to 40:19	27	2008_08_22-19_59_02-Telejornal-1.avi	01:50 to 01:59	3
2008_08_14-21_59_01-Jornal2-2.avi	40:34 to 40:48	65	2008_08_22-19_59_02-Telejornal-1.avi	04:01 to 04:10	4
2008_08_16-19_59_02-Telejornal-1.avi	26:48 to 27:01	35	2008_08_22-19_59_02-Telejornal-1.avi	28:23 to 28:31	6
2008_08_16-19_59_02-Telejornal-1.avi	30:41 to 30:51	44	2008_08_22-19_59_02-Telejornal-1.avi	28:35 to 28:50	17
2008_08_16-19_59_02-Telejornal-1.avi	33:44 to 33:59	28	2008_08_22-19_59_02-Telejornal-1.avi	42:27 to 42:40	18
2008_08_16-21_59_02-Jornal2-2.avi	30:21 to 30:29	29	2008_08_22-19_59_02-Telejornal-1.avi	43:54 to 44:04	19
2008_08_21-19_59_01-Telejornal-1.avi	00:07 to 00:28	31	2008_08_22-19_59_02-Telejornal-1.avi	56:32 to 56:42	20
2008_08_21-19_59_01-Telejornal-1.avi	39:50 to 40:08	36	2008_09_05-21_59_01-Jornal2-2.avi	14:12 to 14:22	60
2008_08_21-19_59_01-Telejornal-1.avi	40:25 to 40:39	37	2008_09_05-21_59_01-Jornal2-2.avi	15:05 to 15:10	58
2008_08_21-19_59_01-Telejornal-1.avi	40:45 to 40:55	38			

E.2 Question Set

Table E.10: Questions: Part 1 of 3 (Questions 1-35)

Q.#	C.#	Question
1	2600	Quantas vezes tocou o hino Português nos Jogos Olímpicos?
2	2601	Onde houve um descarrilamento?
3	2601	Quantos feridos provocou o descarrilamento?
4	2601	Quantas pessoas transportava o combóio?
5	2601	Além do governo, quem está a investigar o acidente?
6	2602	Quantos anos tem a Linha do Tua?
7	2603	Quantas medalhas de ouro ganhou Michael Phelps em Pequim?
8	2603	Para que país foi ele passar férias?
9	2604	Que prémio ganhou João Ubaldo Ribeiro em 2008?
10	2605	O que é a Avenida da Liberdade?
11	2606	Qual a primeira cidade onde vai permanecer a selecção nacional durante o europeu de futebol?
12	2607	Qual o número de visitantes do Museu Berardo, um ano após a sua abertura?
13	2608	Quem é Paulo Rangel?
14	2609	Quantas contra ordenações muito graves são necessárias para haver cassação da carta?
15	2610	Onde estiveram reunidos os ministros das finanças do G-8?
16	2611	Quantos policias foram mobilizados para a Marcha de Contestação do G-8?
17	2611	E quantos manifestantes foram autorizados?
18	2612	Quem é a maior estrela dos Jogos Olímpicos de Pequim?
19	2613	De quem é o projecto do museu Iberê Camargo?
20	2613	Em que cidade fica?
21	2614	Onde fica a fundação José Saramago?
22	2615	Quem é o campeão olímpico do triplo salto?
23	2616	Onde nasceu o Cirque du Soleil?
24	2617	O que é a Liga Sagres?
25	2618	Qual o montante que o Benfica vai receber pela venda do lateral Nelson ao Sevilha?
26	2619	Onde ocorreram confrontos entre pescadores e comerciantes?
27	2620	Quem espiou a favor do Hezzbollah?
28	2621	Qual o comprimento do salto de Nelson Évora em Pequim?
29	2622	Quantos dólares custa o barril de petróleo em Nova Iorque?
30	2622	Quantos dólares custa em Londres?
31	2623	Quando é que Paulo Rangel se dirigiu ao Parlamento pela primeira vez como Líder da Bancada Laranja?
32	2623	Qual o programa que criticou na sua intervenção?
33	2624	Quem é Rosa Passos?
34	2624	Quando nasceu?
35	2625	Diga o nome de um álbum de Rosa Passos.

Table E.11: Questions: Part 2 of 3 (Questions 36-70)

Q.#	C.#	Question
36	2626	Quantos anos tem a carreira de Carlos do Carmo?
37	2627	Quem é o nadador mais bem pago da história?
38	2628	Com que idade se tornou Phelps milionário?
39	2629	A que distância da capital chinesa se localizou o epicentro do abalo na provincia de Xichuan?
40	2630	Que medalha conquistou Nelson Évora para Portugal no triplo salto?
41	2631	Em que ensaio conseguiu Nelson Évora o lugar que o tornaria campeão olímpico?
42	2632	Quando ganhou António Lobo Antunes o Prémio Camões?
43	2633	Quem deu vida à "Formiga" cantada por Amália?
44	2633	Em que ano?
45	2634	Quem é António Guterres?
46	2635	Com que material conseguiu Elvira Fortunato produzir transístores?
47	2636	Qual a nacionalidade do novo campeão olímpico dos 100 metros?
48	2637	O que é o C 4?
49	2637	O que é necessário para iniciar este tipo de explosivo?
50	2637	Com que material usado nas escolas é que ele se parece?
51	2638	Em que festa não vai Manuela Ferreira Leite participar?
52	2639	Que prémio ganhou Elvira Fortunato?
53	2640	Quantos milhões de crianças podem ficar órfãs em África, segundo as Nações Unidas?
54	2641	Quais os concorrentes dos empresários Portugueses em Angola?
55	2642	Qual o guarda-redes do Benfica para o encontro frente às Ilhas Faroé?
56	2642	Qual o do Braga?
57	2643	Quem é Carlos do Carmo?
58	2644	Quem foi o primeiro a fazer um comício no Pontal?
59	2645	Quem igualou o recorde de medalhas de Mark Spitz?
60	2646	Quando nasceu João Ubaldo Ribeiro?
61	2647	Qual o banco que vai apoiar o financiamento das empreitadas de reabilitação urbana?
62	2648	Em que país fica o Rio Bengo?
63	2649	Quem é o Presidente Angolano?
64	2650	Quem é o dono do "número vinte e um" da Avenida da Liberdade?
65	2651	Quantos inquilinos tem o prédio da Junta de Galveias?
66	2652	O que é a EMBRAER?
67	2652	Qual a sua nacionalidade?
68	2653	Quantas pessoas moravam no prédio que foi atingido pelas chamas na Rua da Glória?
69	2654	Quantos aviões da EMBRAER está comprando o governo do Brasil?
70	2655	Diga o nome de uma construtora de aviões.

Table E.12: Questions: Part 3 of 3 (Questions 71-100)

Q.#	C.#	Question
71	2656	Quem é Jorge Nuno Pinto da Costa?
72	2657	Onde se realiza o festival Andanças?
73	2658	Em que estádio conquistou Nelson Évora a medalha de ouro olimpica?
74	2659	Quantos milhões de voluntários tem a rede que a campanha de Obama construiu através da internet?
75	2659	E quantos milhões de dólares recebeu a campanha de pequenos doadores por esta via?
76	2660	Que movimento iniciou João Gilberto?
77	2661	Quem gravou, há 50 anos, o tema "Chega de Saudade"?
78	2661	Em que data?
79	2662	Onde foi realizada a cimeira internacional sobre a SIDA?
80	2663	Quem declarou aberta a trigésima segunda Festa do Avante?
81	2664	Qual a última paragem da visita de Barack Obama à Europa?
82	2665	Que festa se realiza na Quinta da Atalaia?
83	2666	Segundo Lula da Silva quais os melhores aviões?
84	2667	Qual a participação do Estado Português na OGMA?
85	2668	Quem tem cada vez mais casos de cancro cutâneo?
86	2668	Quem regista maior número de queimaduras solares?
87	2669	Quantos prédios devolutos existem em Lisboa?
88	2670	Quem anunciou o reforço das linhas de crédito para as empresas Portuguesas que invistam em Angola?
89	2671	Quantos milhões de euros tem a linha de crédito de ajuda?
90	2672	Em que praia vai actuar Emir Kusturica amanhã à noite?
91	2673	Qual a posição de James Blake no ranking ATP?
92	2674	Que norte-americano eliminou Roger Federer do torneio olímpico de ténis?
93	2675	Quem é o orador principal da Festa do Pontal?
94	2676	Que regiões visitou Barack Obama em 2008?
95	2677	O que é a Bossa Nova?
96	2678	O que é o Prémio Camões?
97	2679	Qual a empresa maior accionista da OGMA?
98	2680	Em que cidade é que a EMBRAER vai investir em duas fábricas?
99	2680	Quantos postos de trabalho directos serão criados com o investimento?
100	2681	Quantos postos de trabalho foram criados com a inauguração de um hotel em Baião?

Table E.13: Manual Transcripts, Questions and Answers: Part 1 of 5 (Questions 1-25)

Q.#	C.#	T.#	Question	Answer
1	2600	1	Quantas vezes tocou o hino Português nos Jogos Olímpicos?	4
2	2601	2	Onde houve um descarrilamento?	Linha do Tua
3	2601	2	Quantos feridos provocou o descarrilamento?	37 feridos
4	2601	3	Quantas pessoas transportava o combóio?	47 pessoas
5	2601	4	Além do governo, quem está a investigar o acidente?	Núcleo de Investigação Criminal da GNR e Policia Judiciária
6	2602	5	Quantos anos tem a Linha do Tua?	120
7	2603	6	Quantas medalhas de ouro ganhou Michael Phelps em Pequim?	8
8	2603	6	Para que país foi ele passar férias?	Algarve, Portugal
9	2604	7	Que prémio ganhou João Ubaldo Ribeiro em 2008?	Prémio Camões
10	2605	8	O que é a Avenida da Liberdade?	Uma das principais avenidas da Capital
11	2606	9	Qual a primeira cidade onde vai permanecer a selecção nacional durante o europeu de futebol?	Neuchâtel
12	2607	10	Qual o número de visitantes do Museu Berardo, um ano após a sua abertura?	Mais de meio milhão de visitantes
13	2608	11	Quem é Paulo Rangel?	Líder parlamentar do PSD
14	2609	12	Quantas contra ordenações muito graves são necessárias para haver cassação da carta?	Três
15	2610	13	Onde estiveram reunidos os ministros das finanças do G-8?	Osaca, no Japão
16	2611	14	Quantos policias foram mobilizados para a Marcha de Contestação do G-8?	20 mil
17	2611	14	E quantos manifestantes foram autorizados?	Cerca de cinco mil
18	2612	6	Quem é a maior estrela dos Jogos Olímpicos de Pequim?	Michael Phelps
19	2613	15	De quem é o projecto do museu Iberê Camargo?	Siza Vieira
20	2613	15	Em que cidade fica?	Porto Alegre
21	2614	16	Onde fica a fundação José Saramago?	Lisboa
22	2615	17	Quem é o campeão olímpico do triplo salto?	Nelson Évora
23	2616	18	Onde nasceu o Cirque du Soleil?	Em Montreal no Canadá
24	2617	19	O que é a Liga Sagres?	Campeonato nacional de futebol
25	2618	20	Qual o montante que o Benfica vai receber pela venda do lateral Nelson ao Sevilha?	5 milhões de euros

Table E.14: Manual Transcripts, Questions and Answers: Part 2 of 5 (Questions 26-47)

Q.#	C.#	T.#	Question	Answer
26	2619	21	Onde ocorreram confrontos entre pescadores e comerciantes?	Na lota de Matosinhos
27	2620	22	Quem espiou a favor do Hezzbollah?	Nassim Nasser
28	2621	17	Qual o comprimento do salto de Nelson Évora em Pequim?	17 metros e 67 centímetros
29	2622	23	Quantos dólares custa o barril de petróleo em Nova Iorque?	Cento e quarenta e cinco dólares e oitenta e cinco cêntimos.
30	2622	23	Quantos dólares custa em Londres?	(perto dos) Cento e quarenta e sete dólares
31	2623	24	Quando é que Paulo Rangel se dirigiu ao Parlamento pela primeira vez como Líder da Bancada Laranja?	03 de Julho de 2008
32	2623	24	Qual o programa que criticou na sua intervenção?	Programa de obras públicas do governo
33	2624	25	Quem é Rosa Passos?	Cantora brasileira
34	2624	26	Quando nasceu?	Abril de 1952
35	2625	25	Diga o nome de um álbum de Rosa Passos.	Romance; Recriação
36	2626	27	Quantos anos tem a carreira de Carlos do Carmo?	Mais de 40 anos
37	2627	28	Quem é o nadador mais bem pago da história?	Michael Phelps
38	2628	29	Com que idade se tornou Phelps milionário?	(poucos meses antes de fazer) 16 anos
39	2629	30	A que distância da capital chinesa se localizou o epicentro do abalo na provincia de Xichuan?	1250 km
40	2630	31	Que medalha conquistou Nelson Évora para Portugal no triplo salto?	Medalha de ouro
41	2631	31	Em que ensaio conseguiu Nelson Évora o lugar que o tornaria campeão olímpico?	Quarto ensaio
42	2632	7	Quando ganhou António Lobo Antunes o Prémio Camões?	2007
43	2633	32	Quem deu vida à "Formiga" cantada por Amália?	Fontes Rocha e Fernando Alvim
44	2633	32	Em que ano?	1969
45	2634	33	Quem é António Guterres?	Alto comissário da ONU para os refugiados
46	2635	34	Com que material conseguiu Elvira Fortunato produzir transístores?	Papel
47	2636	35	Qual a nacionalidade do novo campeão olímpico dos 100 metros?	Jamaicana/o

Table E.15: Manual Transcripts, Questions and Answers: Part 3 of 5 (Questions 48-65)

Q.#	C.#	T.#	Question	Answer
48	2637	36	O que é o C 4?	C 4 é o explosivo utilizado habitualmente pelas autoridades para fazerem demolições e arrombarem portas
49	2637	37	O que é necessário para iniciar este tipo de explosivo?	Um estímulo energético elevado
50	2637	38	Com que material usado nas escolas é que ele se parece?	Plasticina
51	2638	39	Em que festa não vai Manuela Ferreira Leite participar?	Festa do Pontal
52	2639	34	Que prémio ganhou Elvira Fortunato?	Prémio do European Research Council, considerado uma espécie de prémios Nobel na área da engenharia.
53	2640	40	Quantos milhões de crianças podem ficar órfãs em África, segundo as Nações Unidas?	Mais de 50 milhões
54	2641	41	Quais os concorrentes dos empresários Portugueses em Angola?	Chineses, Sul Africanos ou Americanos.
55	2642	42	Qual o guarda-redes do Benfica para o encontro frente às Ilhas Faroé?	Quim
56	2642	42	Qual o do Braga?	Eduardo
57	2643	27	Quem é Carlos do Carmo?	O grande senhor do fado Português
58	2644	43	Quem foi o primeiro a fazer um comício no Pontal?	Sá Carneiro
59	2645	44	Quem igualou o recorde de medalhas de Mark Spitz?	Michael Phelps
60	2646	45	Quando nasceu João Ubaldo Ribeiro?	1941
61	2647	46	Qual o banco que vai apoiar o financiamento das empreitadas de reabilitação urbana?	BEI, Banco Europeu do Investimento
62	2648	41	Em que país fica o Rio Bengo?	Angola
63	2649	47	Quem é o Presidente Angolano?	José Eduardo dos Santos
64	2650	48	Quem é o dono do ”número vinte e um” da Avenida da Liberdade?	Galveias
65	2651	49	Quantos inquilinos tem o prédio da Junta de Galveias?	15

Table E.16: Manual Transcripts, Questions and Answers: Part 4 of 5 (Questions 66-88)

Q.#	C.#	T.#	Question	Answer
66	2652	50	O que é a EMBRAER?	A terceira maior empresa mundial no fabrico de aviões
67	2652	50	Qual a sua nacionalidade?	Brasileira
68	2653	51	Quantas pessoas moravam no prédio que foi atingido pelas chamas na Rua da Glória?	7
69	2654	52	Quantos aviões da EMBRAER está comprando o governo do Brasil?	Dois
70	2655	53	Diga o nome de uma construtora de aviões.	EMBRAER
71	2656	54	Quem é Jorge Nuno Pinto da Costa?	Presidente do Futebol Clube do Porto
72	2657	55	Onde se realiza o festival Andanças?	São Pedro do Sul
73	2658	31	Em que estádio conquistou Nelson Évora a medalha de ouro olimpica?	Ninho de Pássaro
74	2659	56	Quantos milhões de voluntários tem a rede que a campanha de Obama construiu através da internet?	2 milhões
75	2659	56	E quantos milhões de dólares recebeu a campanha de pequenos doadores por esta via?	240 milhões de dólares
76	2660	57	Que movimento iniciou João Gilberto?	Bossa Nova
77	2661	57	Quem gravou, há 50 anos, o tema "Chega de Saudade"?	João Gilberto
78	2661	57	Em que data?	Julho de 1958
79	2662	40	Onde foi realizada a cimeira internacional sobre a SIDA?	México
80	2663	58	Quem declarou aberta a trigésima segunda Festa do Avante?	Jerónimo de Sousa
81	2664	59	Qual a última paragem da visita de Barack Obama à Europa?	Londres
82	2665	60	Que festa se realiza na Quinta da Atalaia?	Festa do Avante
83	2666	61	Segundo Lula da Silva quais os melhores aviões?	Os da EMBRAER
84	2667	50	Qual a participação do Estado Português na OGMA?	0,35
85	2668	62	Quem tem cada vez mais casos de cancro cutâneo?	Jovens
86	2668	62	Quem regista maior número de queimaduras solares?	Jovens
87	2669	63	Quantos prédios devolutos existem em Lisboa?	4602
88	2670	47	Quem anunciou o reforço das linhas de crédito para as empresas Portuguesas que invistam em Angola?	O primeiro ministro

Table E.17: Manual Transcripts, Questions and Answers: Part 5 of 5 (Questions 89-100)

Q.#	C.#	T.#	Question	Answer
89	2671	64	Quantos milhões de euros tem a linha de crédito de ajuda?	Cem milhões de euros
90	2672	65	Em que praia vai actuar Emir Kusturica amanhã à noite?	Praia do Tonel, em Sagres
91	2673	66	Qual a posição de James Blake no ranking ATP?	Sétimo
92	2674	67	Que norte-americano eliminou Roger Federer do torneio olímpico de ténis?	James Blake
93	2675	39	Quem é o orador principal da Festa do Pontal?	Ângelo Correia
94	2676	59	Que regiões visitou Barack Obama em 2008?	Médio oriente e Europa
95	2677	57	O que é a Bossa Nova?	Movimento que cruzou o samba com as harmonias do jazz e tornou a música Brasileira conhecida mundialmente.
96	2678	7	O que é o Prémio Camões?	O mais importante galardão atribuído a autores de Língua Portuguesa
97	2679	50	Qual a empresa maior accionista da OGMA?	A Brasileira EMBRAER
98	2680	53	Em que cidade é que a EMBRAER vai investir em duas fábricas?	Évora
99	2680	53	Quantos postos de trabalho directos serão criados com o investimento?	(pelo menos) 500
100	2681	68	Quantos postos de trabalho foram criados com a inauguração de um hotel em Baião?	35

E.3 Answer Set

E.3.1 Answer Assessment per Question

The following 3 Tables (Table E.18 to Table E.20) present the Assessment that we made, manually for the answers of the system to the 8 situations that were tested. The assessement values used were the same ones used at QACLEF (and correspondent to the ones used in QAST), and were:

- 1 - 1st Answer Right
- 2 - 2nd Answer Right
- 3 - 3rd Answer Right
- -2 - Passage Right, Inexact Extraction
- -1 - Unsupported
- -3 - NIL Answer
- 0 - Wrong

Table E.18: Assessment Value per Question: Part 1 of 3 (Questions 1-35)
A - Transcriptspv (full stops and commas)
B - Transandwikipv (full stops and commas, Wikipedia)
C - Transcripts (only commas)
D - Transandwiki (only commas, Wikipedia)

		2008				2010			
Question #	Cluster #	A	B	C	D	A	B	C	D
1	2600	1	1	1	1	1	1	1	1
2	2601	1	2	0	2	1	-2	-3	1
3	2601	2	2	2	2	2	2	2	2
4	2601	-3	-3	0	0	-3	-3	0	0
5	2601	0	0	0	0	0	0	-3	0
6	2602	-2	-2	-2	-2	-2	-2	-2	-2
7	2603	1	1	1	1	1	2	1	1
8	2603	-3	0	-3	0	-3	0	-3	0
9	2604	-3	1	-3	1	-3	1	-3	1
10	2605	0	1	0	1	0	1	0	1
11	2606	-3	-2	-3	-2	-3	0	-3	0
12	2607	0	-2	0	0	0	-2	0	0
13	2608	1	0	2	0	3	0	3	0
14	2609	-2	-2	-2	-2	-2	-2	-2	-2
15	2610	1	0	1	0	0	0	1	0
16	2611	1	1	1	1	1	1	1	1
17	2611	1	1	1	1	1	1	1	1
18	2612	-3	0	-3	2	-3	-2	-3	-2
19	2613	-3	1	-3	1	-3	1	-3	1
20	2613	-3	1	-3	1	-3	0	-3	0
21	2614	3	1	-2	1	3	-2	0	1
22	2615	-2	-2	0	1	0	-2	0	-2
23	2616	-3	0	-3	0	-3	0	-3	0
24	2617	0	0	0	0	0	0	0	1
25	2618	-3	-3	-3	-3	-3	-3	-3	-3
26	2619	-2	1	-2	1	2	1	2	1
27	2620	-3	-2	-3	-2	-3	-2	-3	-2
28	2621	-3	-3	-3	-3	-3	-3	-3	-3
29	2622	-3	-2	-3	-2	-3	-2	-3	-2
30	2622	-3	-2	-3	-2	-3	-2	-3	-2
31	2623	-3	-3	-3	-3	-3	-3	-3	-3
32	2623	0	0	0	0	0	0	0	0
33	2624	0	0	0	0	0	0	-2	0
34	2624	0	0	0	0	0	0	0	0
35	2625	0	3	0	0	0	0	0	0

Table E.19: Assessment Value per Question: Part 2 of 3 (Questions 36-70)
A - Transcriptspv (full stops and commas)
B - Transandwikipv (full stops and commas, Wikipedia)
C - Transcripts (only commas)
D - Transandwiki (only commas, Wikipedia)

		2008				2010			
Question #	Cluster #	A	B	C	D	A	B	C	D
36	2626	-3	0	-3	0	-3	0	-3	0
37	2627	-3	1	-3	2	0	0	-3	0
38	2628	1	1	1	1	-3	0	-3	0
39	2329	-2	-2	-2	-2	0	0	0	0
40	2630	-3	-2	-3	-2	-3	0	-3	0
41	2631	-3	0	-3	0	-3	-3	-3	0
42	2632	1	1	1	1	1	1	1	1
43	2633	0	-2	0	-2	0	-2	0	-2
44	2633	1	1	2	2	-3	0	-3	0
45	2634	-2	1	-2	1	0	1	-2	1
46	2635	-3	0	-3	0	-3	-2	-3	-2
47	2636	0	0	0	0	0	0	0	0
48	2637	0	0	0	0	0	0	-3	0
49	2637	-3	0	-3	-2	-3	-3	-3	-3
50	2637	-3	0	-3	0	-3	0	-3	0
51	2638	-3	0	-3	0	-3	-2	-3	-2
52	2639	-3	-2	-3	0	-3	-2	-3	-2
53	2640	-3	-2	-3	-2	-3	-2	-3	-2
54	2641	0	0	0	0	0	0	0	0
55	2642	0	1	0	0	-2	-2	-2	-2
56	2642	0	-2	0	0	-2	-2	-2	-2
57	2643	0	1	0	1	0	1	-2	1
58	2644	0	-2	0	-2	0	-2	0	-2
59	2645	1	-2	1	-2	1	-2	1	-2
60	2646	-2	1	-2	1	-2	1	-2	1
61	2647	-2	3	-2	-2	-2	3	-2	3
62	2648	-3	1	-3	1	-3	1	-3	1
63	2649	0	-2	0	-2	-2	1	-2	-2
64	2650	-3	3	-3	3	-3	-2	-3	-2
65	2651	1	1	1	1	1	1	1	1
66	2652	1	1	-2	1	3	1	3	1
67	2652	2	-2	1	-2	3	0	3	0
68	2653	1	1	1	1	1	1	1	1
69	2654	-3	-3	-3	-3	-3	-3	-3	-3
70	2655	0	0	0	0	-2	0	-2	0

Table E.20: Assessment Value per Question: Part 3 of 3 (Questions 71-100)
A - Transcriptspv (full stops and commas)
B - Transandwikipv (full stops and commas, Wikipedia)
C - Transcripts (only commas)
D - Transandwiki (only commas, Wikipedia)

		2008				2010			
Question #	Cluster #	A	B	C	D	A	B	C	D
71	2656	-3	1	-3	1	-3	1	-3	1
72	2657	0	-2	0	-2	-2	-2	-2	0
73	2658	-3	0	-3	0	-3	-1	-3	-1
74	2659	-3	0	-3	0	-3	-2	-3	-2
75	2659	0	0	0	0	-2	-2	-2	-2
76	2660	-3	1	-3	1	-3	1	-3	1
77	2661	-3	-2	0	-2	-3	-2	0	-2
78	2661	-3	0	-3	0	-3	-2	-3	-2
79	2662	-2	-2	0	-2	0	-2	0	-2
80	2663	0	-2	0	-2	-2	3	-2	-2
81	2664	2	2	0	0	-3	-3	-3	-3
82	2665	-3	1	-3	1	-3	1	-3	1
83	2666	3	3	0	0	-2	-2	-2	-2
84	2667	0	0	0	0	-2	-2	-2	-2
85	2668	0	3	0	-2	-2	1	0	-2
86	2668	0	3	0	3	0	3	0	3
87	2669	-2	-2	-2	-2	-2	-2	-2	-2
88	2670	0	3	0	3	0	3	0	3
89	2671	-3	1	-3	0	-3	0	-3	0
90	2672	-3	-2	-3	-2	-3	-2	-3	-2
91	2673	-2	-2	-2	-2	-3	-3	-3	-3
92	2674	-3	0	-3	0	-3	0	-3	0
93	2675	-2	1	-2	1	-2	1	-2	1
94	2676	-3	1	-3	2	-3	-1	-3	2
95	2677	0	1	0	1	-2	1	0	1
96	2678	0	1	0	1	-2	1	-2	1
97	2679	-2	3	-2	-2	-2	-2	-2	-2
98	2680	-3	2	-3	1	-3	-2	-3	-2
99	2680	1	-3	1	-3	1	-3	1	1
100	2681	-3	-2	-3	-2	-3	-2	-3	-2

E.3.2 Detailed Answers per Question

Table E.21: Detailed Answers and Support

A - Transcriptspv (full stops and commas)

B - Transandwikipv (full stops and commas, Wikipedia)

C - Transcripts (only commas)

D - Transandwiki (only commas, Wikipedia)

Test	A#: Answer	Support
Question #1 - Quantas vezes tocou o hino Português nos Jogos Olímpicos?		
A 2008 B 2008	1: quarta vez	**2008_08_22-19_59_02-Telejornal-1 bloco 1**: boa noite pela quarta vez na história ou hino português tocou nos jogos olímpicos .
C 2008 D 2008	1: quarta vez	**2008_08_22-19_59_02-Telejornal-1 bloco 1**: boa noite pela quarta vez na história ou hino português tocou nos jogos olímpicos , foi o momento da consagração de nelson évora ,
A 2010 B 2010	1: quarta vez	**2008_08_22-19_59_02-Telejornal-1 bloco 1**: boa noite pela quarta vez na história o hino português tocou nos jogos olímpicos .
C 2010 D 2010	1: quarta vez	**2008_08_22-19_59_02-Telejornal-1 bloco 1**: boa noite pela quarta vez na história o hino português tocou nos jogos olímpicos , foi o momento da consagração de nelson évora ,
Question #2 - Onde houve um descarrilamento?		
A 2008	1: tua	**2008_08_27-21_59_01-Jornal2-2 bloco 2**: o descarrilamento no tua .
B 2008	1: barragem	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: tem os seus defensores . o maquinista da composição acidentada chegou a falar da coincidência dos descarrilamento os quatro em dezoito meses , e curiosamente aconteceram quando a notícia da barragem , começou a circular .
B 2008	2: mirandela	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: estando agora a gnr a investigar este incidente . o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua , na passada sexta - feira ,
B 2008	3: vou	**2008_06_12-21_59_01-Jornal2-2 bloco 3**: modo a frente vou não sou . a descarrilamento , de três , e que a campanha agressiva favor não teve papel importante .
C 2008	1: automotora	**2008_08_27-21_59_01-Jornal2-2 bloco 2**: afasta a existência de qualquer problema com automotora e com a linha , não foram detectadas causas pudessem ter provocado , o descarrilamento no tua , a doença o mistério sobre os acidentes da linha do tua , não é da da da da automotora não é da linha ,
C 2008	2: avença	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: afasta a existência de qualquer problema com a automotora e com a linha , não foram detectadas causas pudessem ter provocado , o descarrilamento no tua , a avença o mistério sobre os acidentes da linha do tua , não é da da da da automotora não é da linha ,

D 2008	1: divino espírito santo	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: descarrilamento no tua , a avença o mistério sobre os acidentes da linha do tua , não é da da da da automotora não é da linha , não é do motorista , até um casino costumo dizer uma expressão popular sobre por obra e graça do divino espírito santo , mas tem de haver
D 2008	2: mirandela	**2008_08_22-21_59_02-Jornal2-2 bloco 2**: boa noite um descarrilamento na linha do tua provocou um morto , e trinta e sete feridos , dois deles estão hospitalizados em estado grave , o acidente envolveu uma composição do metro de mirandela ,
D 2008	3: barragem	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: tem os seus defensores , o maquinista da composição acidentada chegou a falar da coincidência dos descarrilamento os quatro em dezoito meses , e curiosamente aconteceram quando a notícia da barragem , começou a circular ,
A 2010	1: tua	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: e com linha não foram detectadas causas pudessem ter provocado o descarrilamento no tua .
A 2010	2: detectadas	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: e com linha não foram detectadas causas pudessem ter provocado o descarrilamento no tua .
B 2010	1: superfície de mirandela	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: estando agora a gnr investigar este incidente . o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua na passada sexta - feira não se tratou de um mero acidente .
B 2010	2: divino espírito santo	**2008_08_27-21_59_01-Jornal2-2 bloco 2**: descarrilamento no tua . avença o mistério sobre os acidentes da linha do tua . não é da da da da automotora moeda linha no é do motorista . até um casino costumo dizer numa expressão popular só se for por obra e graça do divino espírito santo . mas tem
B 2010	3: barragem	**2008_08_27-21_59_01-Jornal2-2 bloco 2**: tem os seus defensores . o maquinista da composição acidentada chegou a falar da coincidência dos descarrilamento os quatro em dezoito meses e curiosamente aconteceram quando a notícia da barragem começou a circular .
C 2010	1: NIL	
D 2010	1: mirandela	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: e um do metro de mirandela afasta a existência de qualquer problema com auto motora , e com linha não foram detectadas causas pudessem ter provocado o descarrilamento no tua , avença o mistério sobre os acidentes da linha do tua ,
D 2010	2: divino espírito santo	**2008_08_27-21_59_01-Jornal2-2 bloco 2**: descarrilamento no tua , avença o mistério sobre os acidentes da linha do tua , não é da da da da automotora moeda linha no é do motorista , até um casino costumo dizer numa expressão popular só se for por obra e graça do divino espírito santo , mas tem

D 2010	3: barragem	**2008_08_27-21_59_01-Jornal2-2 bloco 2**: tem os seus defensores , o maquinista da composição acidentada chegou a falar da coincidência dos descarrilamento os quatro em dezoito meses e curiosamente aconteceram quando a notícia da barragem começou a circular ,
Question #3 - Quantos feridos provocou o descarrilamento?		
A 2008	1: sete feridos	**2008_08_22-21_59_02-Jornal2-2 bloco 2**: boa noite um descarrilamento na linha do tua provocou um morto , e trinta e sete feridos .
A 2008	2: dois feridos	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua , na passada sexta - feira , não se tratou de um mero acidente . o desastre provocou um morto e dois feridos .
A 2008	3: um ferido	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: será . o mesmo fez o deputado socialista manuel alegre , de eça de vulto . está a decorrer um inquérito para apurar as causas do terceiro descarrilamento na linha do tua . o acidente desta manhã fez um ferido ligeiro .
B 2008	1: sete feridos	**2008_08_22-21_59_02-Jornal2-2 bloco 2**: boa noite um descarrilamento na linha do tua provocou um morto , e trinta e sete feridos . dois deles estão hospitalizados em estado grave .
B 2008	2: dois feridos	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: estando agora a gnr a investigar este incidente . o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua , na passada sexta - feira , não se tratou de um mero acidente . o desastre provocou um morto e dois feridos .
B 2008	3: um ferido	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: será . o mesmo fez o deputado socialista manuel alegre , de eça de vulto . está a decorrer um inquérito para apurar as causas do terceiro descarrilamento na linha do tua . o acidente desta manhã fez um ferido ligeiro .
C 2008	1: sete feridos	**2008_08_22-21_59_02-Jornal2-2 bloco 2**: boa noite um descarrilamento na linha do tua provocou um morto , e trinta e sete feridos , dois deles estão hospitalizados em estado grave , o acidente envolveu uma composição do metro de mirandela ,
C 2008	2: dois feridos	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: descarrilamento na linha do tua , na passada sexta - feira , não se tratou de um mero acidente , o desastre provocou um morto e dois feridos , um deles ainda em estado grave , foi o quarto acidente em apenas ano e meio ,

C 2008	3: um ferido	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: será , o mesmo fez o deputado socialista manuel alegre , de eça de vulto , está a decorrer um inquérito para apurar as causas do terceiro descarrilamento na linha do tua , o acidente desta manhã fez um ferido ligeiro ,
D 2008	1: sete feridos	**2008_08_22-21_59_02-Jornal2-2 bloco 2**: boa noite um descarrilamento na linha do tua provocou um morto , e trinta e sete feridos , dois deles estão hospitalizados em estado grave , o acidente envolveu uma composição do metro de mirandela ,
D 2008	2: dois feridos	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: descarrilamento na linha do tua , na passada sexta - feira , não se tratou de um mero acidente , o desastre provocou um morto e dois feridos , um deles ainda em estado grave , foi o quarto acidente em apenas ano e meio ,
D 2008	3: um ferido	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: será , o mesmo fez o deputado socialista manuel alegre , de eça de vulto , está a decorrer um inquérito para apurar as causas do terceiro descarrilamento na linha do tua , o acidente desta manhã fez um ferido ligeiro ,
A 2010	1: sete feridos	**2008_08_22-19_59_02-Telejornal-1 bloco 1**: para já dou conta de um descarrilamento na linha do toa que provocou um morto e trinta e sete feridos .
A 2010	2: dois feridos	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua na passada sexta - feira não se tratou de um mero acidente . o desastre provocou um morto e dois feridos .
A 2010	3: um ferido	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: é esteves . o mesmo fez o deputado socialista manuel alegre . a essa futuro . por um lado , o vereador barbosa e nove . está a decorrer um inquérito para apurar as causas do terceiro descarrilamento na linha do tua . o acidente desta manhã fez um ferido ligeiro .
B 2010	1: sete feridos	**2008_08_22-21_59_02-Jornal2-2 bloco 1**: se um descarrilamento na linha do tua provocou um morto e trinta e sete feridos . dois deles estão hospitalizados em estado grave .
B 2010	2: dois feridos	**2008_08_24-21_09_01-Telejornal-1 bloco 3**: estando agora a gnr aim esticar o incidente . o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua na passada sexta - feira não se tratou de um mero acidente . o desastre provocou um morto e dois feridos
B 2010	3: um ferido	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: é esteves . o mesmo fez o deputado socialista manuel alegre . a essa futuro . por um lado , o vereador barbosa e nove . está a decorrer um inquérito para apurar as causas do terceiro descarrilamento na linha do tua . o acidente desta manhã fez um ferido ligeiro .

C 2010	1: sete feridos	**2008_08_22-19_59_02-Telejornal-1 bloco 1**: para já dou conta de um descarrilamento na linha do toa que provocou um morto e trinta e sete feridos , dois destes feridos estão hospitalizados em estado grave ainda ,
C 2010	2: dois feridos	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: descarrilamento na linha do tua na passada sexta - feira não se tratou de um mero acidente , o desastre provocou um morto e dois feridos , um deles ainda em estado grave , foi o quarto acidente em apenas ano e meio ,
C 2010	3: um ferido	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: é esteves , o mesmo fez o deputado socialista manuel alegre , a essa futuro , por um lado , o vereador barbosa e nove , está a decorrer um inquérito para apurar as causas do terceiro descarrilamento na linha do tua , o acidente desta manhã fez um ferido ligeiro ,
D 2010	1: sete feridos	**2008_08_22-21_59_02-Jornal2-2 bloco 1**: boa noite , se um descarrilamento na linha do tua provocou um morto e trinta e sete feridos , dois deles estão hospitalizados em estado grave ,
D 2010	2: dois feridos	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: descarrilamento na linha do tua na passada sexta - feira não se tratou de um mero acidente , o desastre provocou um morto e dois feridos , um deles ainda em estado grave , foi o quarto acidente em apenas ano e meio ,
D 2010	3: um ferido	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: é esteves , o mesmo fez o deputado socialista manuel alegre , a essa futuro , por um lado , o vereador barbosa e nove , está a decorrer um inquérito para apurar as causas do terceiro descarrilamento na linha do tua , o acidente desta manhã fez um ferido ligeiro ,
Question #4 - Quantas pessoas transportava o comboio?		
A 2008 B 2008	1: NIL	
C 2008 D 2008	1: quarenta pessoas	**2008_08_10-19_59_02-Telejornal-1 bloco 2**: quarenta pessoas continuam desaparecidas , há também notícia do descarrilamento de um comboio ,
A 2010 B 2010	1: NIL	
C 2010 D 2010	1: quarenta pessoas	**2008_08_10-19_59_02-Telejornal-1 bloco 2**: foi no ano de natação porque chuvas torrenciais fizeram setenta e oito mortos no norte do vietname quarenta pessoas continuam desaparecidas , há também notícia do descarrilamento de um comboio , mas os mais de mil passageiros sobretudo turistas foram resgatados com vida ,

Question #5 - Além do governo, quem está a investigar o acidente?		
A 2008	1: tua	**2008_08_27-21_59_01-Jornal2-2 bloco 2**: o descarrilamento no tua .
B 2008	1: tua	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: estando agora a gnr a investigar este incidente . o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua , na passada sexta - feira , não se tratou de um mero acidente .
B 2008	2: maquinista	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: estando agora a gnr a investigar este incidente . o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua , na passada sexta - feira , não se tratou de um mero acidente .
B 2008	3: ainda	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: investigar este incidente . o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua , na passada sexta - feira , não se tratou de um mero acidente . o desastre provocou um morto e dois feridos . um deles ainda em estado grave . foi o quarto acidente
C 2008	1: automotora	**2008_08_27-21_59_01-Jornal2-2 bloco 2**: afasta a existência de qualquer problema com automotora e com a linha , não foram detectadas causas pudessem ter provocado , o descarrilamento no tua , a doença o mistério sobre os acidentes da linha do tua , não é da da da da automotora não é da linha ,
C 2008	2: avença	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: afasta a existência de qualquer problema com a automotora e com a linha , não foram detectadas causas pudessem ter provocado , o descarrilamento no tua , a avença o mistério sobre os acidentes da linha do tua , não é da da da da automotora não é da linha ,
D 2008	1: tua	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: estando agora a gnr a investigar este incidente , o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua , na passada sexta - feira , não se tratou de um mero acidente , o desastre provocou um morto e dois feridos ,
D 2008	2: socialista manuel alegre	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: o mesmo fez o deputado socialista manuel alegre , de eça de vulto , está a decorrer um inquérito para apurar as causas do terceiro descarrilamento na linha do tua , o acidente desta manhã fez um ferido ligeiro ,
D 2008	3: maquinista	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: estando agora a gnr a investigar este incidente , o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua , na passada sexta - feira , não se tratou de um mero acidente , o desastre provocou um morto e dois feridos ,

A 2010	1: tua	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: e com linha não foram detectadas causas pudessem ter provocado o descarrilamento no tua .
A 2010	2: detectadas	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: e com linha não foram detectadas causas pudessem ter provocado o descarrilamento no tua .
B 2010	1: tua	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: estando agora a gnr investigar este incidente . o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua na passada sexta - feira não se tratou de um mero acidente .
B 2010	2: maquinista	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: estando agora a gnr investigar este incidente . o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua na passada sexta - feira não se tratou de um mero acidente .
B 2010	3: ainda	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: investigar este incidente . o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua na passada sexta - feira não se tratou de um mero acidente . o desastre provocou um morto e dois feridos . um deles ainda em estado grave . foi o quarto acidente
C 2010	1: NIL	
D 2010	1: tua	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: estando agora a gnr investigar este incidente , o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua na passada sexta - feira não se tratou de um mero acidente , o desastre provocou um morto e dois feridos ,
D 2010	2: maquinista	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: estando agora a gnr investigar este incidente , o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua na passada sexta - feira não se tratou de um mero acidente , o desastre provocou um morto e dois feridos ,
D 2010	3: ainda	**2008_08_24-21_59_01-Jornal2-2 bloco 2**: investigar este incidente , o maquinista do metro de superfície de mirandela acredita que o descarrilamento na linha do tua na passada sexta - feira não se tratou de um mero acidente , o desastre provocou um morto e dois feridos , um deles ainda em estado grave , foi o quarto acidente
Question #6 - Quantos anos tem a Linha do Tua?		
A 2008	1: vinte anos	**2008_06_06-19_59_01-Telejornal-1 bloco 1**: em cento e vinte anos de existência de a linha do tua nunca tinha registado acidentes graves do passado .
A 2008	2: dezoito meses	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: linha do tua . tem os seus defensores . o maquinista da composição acidentada chegou a falar da coincidência dos descarrilamento os quatro em dezoito meses , e curiosamente aconteceram quando a notícia da barragem , começou a circular . a vista alegre ou o material óptimo material num , eu tinha . o estado do

A 2008	3: trinta dias	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: tua na semana passada não conseguiu encontrar uma explicação , não foram detectados problemas nem na composição na linha . o ministério das obras públicas quer conclusões no prazo de trinta dias . o ministro mário lino não ficou satisfeito com a conclusão do relatório preliminar , exige respostas .
B 2008	1: vinte anos	**2008_06_06-19_59_01-Telejornal-1 bloco 1**: em cento e vinte anos de existência de a linha do tua nunca tinha registado acidentes graves do passado .
B 2008	2: vinte cinco anos	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: a linha do tua . o supremo tribunal de justiça rejeitou o recurso do antigo cabo da gnr de santa comba dão . antónio costa vai cumprir vinte cinco anos pelo homicídio de três raparigas .
B 2008	3: um ano	**2008_06_06-19_59_01-Telejornal-1 bloco 1**: uma composição do metro de mirandela descarrilou na linha do tua . o acidente fez um ferido . este é o terceiro acidente do género no mesmo local no espaço de um ano e meio . em cento e vinte anos
C 2008	1: dezoito meses	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: a tese de sabotagem alegadamente associada à construção de uma barragem que vem um dar grande parte da linha do tua , tem os seus defensores , o maquinista da composição acidentada chegou a falar da coincidência dos descarrilamento os quatro em dezoito meses ,
C 2008	2: vinte anos	**2008_06_06-19_59_01-Telejornal-1 bloco 1**: em cento e vinte anos de existência de a linha do tua nunca tinha registado acidentes graves do passado , uma composição descarrilou , uma carruagem caiu por uma ravina de sessenta metros novamente , matando ,
C 2008	3: vinte cinco anos	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: numa altura em que muitos interesses movem , no sentido de encerrar , a linha do tua , o supremo tribunal de justiça rejeitou o recurso do antigo cabo da gnr de santa comba dão , antónio costa vai cumprir vinte cinco anos pelo homicídio de três raparigas ,
D 2008	1: vinte anos	**2008_06_06-19_59_01-Telejornal-1 bloco 1**: em cento e vinte anos de existência de a linha do tua nunca tinha registado acidentes graves do passado , uma composição descarrilou , uma carruagem caiu por uma ravina de sessenta metros novamente , matando , três pessoas ,
D 2008	2: vinte cinco anos	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: no sentido de encerrar , a linha do tua , o supremo tribunal de justiça rejeitou o recurso do antigo cabo da gnr de santa comba dão , antónio costa vai cumprir vinte cinco anos pelo homicídio de três raparigas ,

D 2008	3: um ano	**2008_06_06-19_59_01-Telejornal-1 bloco 1**: uma composição do metro de mirandela descarrilou na linha do tua , o acidente fez um ferido , este é o terceiro acidente do género no mesmo local no espaço de um ano e meio , em cento e vinte anos
A 2010	1: vinte anos	**2008_06_06-19_59_01-Telejornal-1 bloco 1**: em cento e vinte anos de existência de a linha do tua nunca tinha registado acidentes graves do ano passado .
A 2010	2: trinta dias	**2008_08_27-19_59_02-Telejornal-1 bloco 2**: tua na semana passada não conseguiu encontrar uma explicação não foram detectados problemas nem na composição nem na linha . o ministério das obras públicas quer conclusões no prazo de trinta dias . o ministro mário lino não ficou satisfeito com a conclusão do
B 2010	1: vinte anos	**2008_06_06-19_59_01-Telejornal-1 bloco 1**: em cento e vinte anos de existência de a linha do tua nunca tinha registado acidentes graves do ano passado .
B 2010	2: sete anos	**2008_08_22-19_59_02-Telejornal-1 bloco 1**: e sete anos , perante uma notícia de um trágica que se repete pela quarta vez no último ano e meio a secretaria de estado dos transportes que ainda ontem fez este percurso ferroviário e dirigiu - se ao fim de um declive e anunciou o encerramento temporário da linha do tua ,
B 2010	3: vinte cinco anos	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: no sentido de encerrar a linha do tua . o supremo tribunal de justiça rejeitou o recurso do antigo cabo da gnr de santa comba dão . antónio costa vai cumprir vinte cinco anos pelo homicídio de três raparigas .
C 2010	1: vinte anos	**2008_06_06-19_59_01-Telejornal-1 bloco 1**: em cento e vinte anos de existência de a linha do tua nunca tinha registado acidentes graves do ano passado , uma composição descarrilou , uma carruagem caiu por uma ravina de sessenta metros de prova e um ,
C 2010	2: vinte cinco anos	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: numa altura em que muitos interesses movem , no sentido de encerrar a linha do tua , o supremo tribunal de justiça rejeitou o recurso do antigo cabo da gnr de santa comba dão , antónio costa vai cumprir vinte cinco anos pelo homicídio de três raparigas ,
C 2010	3: um ano	**2008_06_06-19_59_01-Telejornal-1 bloco 1**: lamento , novo lar ao álcool no sangue , uma composição do metro de mirandela descarrilou na linha do tua , o acidente fez um ferido , este é o terceiro acidente do género no mesmo local no espaço de um ano e meio ,
D 2010	1: vinte anos	**2008_06_06-19_59_01-Telejornal-1 bloco 1**: em cento e vinte anos de existência de a linha do tua nunca tinha registado acidentes graves do ano passado , uma composição descarrilou , uma carruagem caiu por uma ravina de sessenta metros de prova e um ,

D 2010	2: vinte cinco anos	**2008_06_06-21_59_02-Jornal2-2 bloco 3**: numa altura em que muitos interesses movem , no sentido de encerrar a linha do tua , o supremo tribunal de justiça rejeitou o recurso do antigo cabo da gnr de santa comba dão , antónio costa vai cumprir vinte cinco anos pelo homicídio de três raparigas ,
D 2010	3: sete anos	**2008_08_22-19_59_02-Telejornal-1 bloco 1**: e sete anos , perante uma notícia de um trágica que se repete pela quarta vez no último ano e meio a secretaria de estado dos transportes que ainda ontem fez este percurso ferroviário e dirigiu - se ao fim de um declive e anunciou o encerramento temporário da linha do tua ,
Question #7 - Quantas medalhas de ouro ganhou Michael Phelps em Pequim?		
A 2008	1: cinco medalhas	**2008_08_13-21_59_02-Jornal2-2 bloco 2**: o nadador conquista em pequim cinco medalhas de ouro a somar às seis , conseguiu em atenas ninguém até agora , tinha ganho ou onze medalhas de ouro . michael phelps
A 2008	2: onze medalhas	**2008_08_13-21_59_02-Jornal2-2 bloco 2**: o nadador conquista em pequim cinco medalhas de ouro a somar às seis , conseguiu em atenas ninguém até agora , tinha ganho ou onze medalhas de ouro . michael phelps
A 2008	3: segunda medalha	**2008_08_11-21_59_01-Jornal2-2 bloco 2**: michael phelps conquistou a segunda medalha de ouro em pequim depois de ter sido o primeiro nos quatrocentos metros livres .
B 2008	1: cinco medalhas	**2008_08_13-21_59_02-Jornal2-2 bloco 2**: o nadador conquista em pequim cinco medalhas de ouro a somar às seis , conseguiu em atenas ninguém até agora , tinha ganho ou onze medalhas de ouro . michael phelps
B 2008	2: onze medalhas	**2008_08_13-21_59_02-Jornal2-2 bloco 2**: o nadador conquista em pequim cinco medalhas de ouro a somar às seis , conseguiu em atenas ninguém até agora , tinha ganho ou onze medalhas de ouro . michael phelps
B 2008	3: segunda medalha	**2008_08_11-21_59_01-Jornal2-2 bloco 2**: michael phelps conquistou a segunda medalha de ouro em pequim depois de ter sido o primeiro nos quatrocentos metros livres .
C 2008	1: cinco medalhas	**2008_08_13-21_59_02-Jornal2-2 bloco 2**: o nadador conquista em pequim cinco medalhas de ouro a somar às seis , conseguiu em atenas ninguém até agora , tinha ganho ou onze medalhas de ouro , michael phelps
C 2008	2: onze medalhas	**2008_08_13-21_59_02-Jornal2-2 bloco 2**: o nadador conquista em pequim cinco medalhas de ouro a somar às seis , conseguiu em atenas ninguém até agora , tinha ganho ou onze medalhas de ouro , michael phelps
C 2008	3: segunda medalha	**2008_08_11-21_59_01-Jornal2-2 bloco 2**: para lutar pela medalha de ouro nos cem metros , com lucídio ribeiro , manuel fernandes silva , rtp um , vinte , michael phelps conquistou a segunda medalha de ouro em pequim depois de ter sido o primeiro nos quatrocentos metros livres ,

D 2008	1: cinco medalhas	**2008_08_13-21_59_02-Jornal2-2 bloco 2**: o nadador conquista em pequim cinco medalhas de ouro a somar às seis , conseguiu em atenas ninguém até agora , tinha ganho ou onze medalhas de ouro , michael phelps
D 2008	2: onze medalhas	**2008_08_13-21_59_02-Jornal2-2 bloco 2**: o nadador conquista em pequim cinco medalhas de ouro a somar às seis , conseguiu em atenas ninguém até agora , tinha ganho ou onze medalhas de ouro , michael phelps
D 2008	3: segunda medalha	**2008_08_11-21_59_01-Jornal2-2 bloco 2**: para lutar pela medalha de ouro nos cem metros , com lucídio ribeiro , manuel fernandes silva , rtp um , vinte , michael phelps conquistou a segunda medalha de ouro em pequim depois de ter sido o primeiro nos quatrocentos metros livres ,
A 2010	1: oito medalhas	**2008_08_25-21_59_01-Jornal2-2 bloco 1**: michael phelps conquistou em pequim oito medalhas de ouro nunca ninguém tinha conseguido antes .
A 2010	2: cinco medalhas	**2008_08_13-21_59_02-Jornal2-2 bloco 2**: nos olímpicos michael of health já entrou na história das olimpíadas o nadador conquistou em pequim cinco medalhas de ouro a somar às seis que conseguiu em atenas ninguém até agora , tinha ganho onze medalhas de ouro . michael phelps
A 2010	3: onze medalhas	**2008_08_13-21_59_02-Jornal2-2 bloco 2**: nos olímpicos michael of health já entrou na história das olimpíadas o nadador conquistou em pequim cinco medalhas de ouro a somar às seis que conseguiu em atenas ninguém até agora , tinha ganho onze medalhas de ouro . michael phelps
B 2010	1: cinco medalhas	**2008_08_13-19_59_02-Telejornal-1 bloco 3**: michael phelps já nos habitou nadar sempre à frente desta linha verde linha que marca o anterior recorde do mundo . em pequim fel que já conquistou cinco medalhas
B 2010	2: segunda medalha	**2008_08_11-21_59_01-Jornal2-2 bloco 1**: michael phelps conquistou a segunda medalha de ouro em pequim depois de ter sido primeiro nos quatrocentos metros livres .
B 2010	3: oito medalhas	**2008_08_25-21_59_01-Jornal2-2 bloco 1**: michael phelps conquistou em pequim oito medalhas de ouro nunca ninguém tinha conseguido antes .
C 2010	1: oito medalhas	**2008_08_25-21_59_01-Jornal2-2 bloco 1**: michael phelps conquistou em pequim oito medalhas de ouro nunca ninguém tinha conseguido antes , tornou - se na primeira grande figura dos jogos ,
C 2010	2: cinco medalhas	**2008_08_13-19_59_02-Telejornal-1 bloco 3**: pequim cinco medalhas de ouro a que se juntam as seis conseguidas há quatro anos em atenas algo que nunca ninguém conseguiu , em jogos olímpicos , michael phelps

C 2010	3: segunda medalha	**2008_08_11-21_59_01-Jornal2-2 bloco 1**: para lutar pela medalha de ouro nos cem metros , com lucídio ribeiro manuel fernandes silva , rtp , vinte , michael phelps conquistou a segunda medalha de ouro em pequim depois de ter sido primeiro nos quatrocentos metros livres ,
D 2010	1: cinco medalhas	**2008_08_13-19_59_02-Telejornal-1 bloco 3**: pequim cinco medalhas de ouro a que se juntam as seis conseguidas há quatro anos em atenas algo que nunca ninguém conseguiu , em jogos olímpicos , michael phelps
D 2010	2: segunda medalha	**2008_08_11-21_59_01-Jornal2-2 bloco 1**: para lutar pela medalha de ouro nos cem metros , com lucídio ribeiro manuel fernandes silva , rtp , vinte , michael phelps conquistou a segunda medalha de ouro em pequim depois de ter sido primeiro nos quatrocentos metros livres ,
D 2010	3: oito medalhas	**2008_08_25-21_59_01-Jornal2-2 bloco 1**: michael phelps conquistou em pequim oito medalhas de ouro nunca ninguém tinha conseguido antes , tornou - se na primeira grande figura dos jogos ,
Question #8 - Para que país foi ele passar férias?		
A 2008	1: NIL	
B 2008	1: jogos olímpicos	**2008_08_23-19_59_02-Telejornal-1 bloco 3**: michael phelps foi o primeiro grande destaque da vigésima nona edição dos jogos olímpicos da era moderna . o nadador norte - americano cumpriu o objectivo provada para pequim conquistar oito medalhas de ouro .
B 2008	2: norte - americano	**2008_08_23-19_59_02-Telejornal-1 bloco 3**: michael phelps foi o primeiro grande destaque da vigésima nona edição dos jogos olímpicos da era moderna . o nadador norte - americano cumpriu o objectivo provada para pequim conquistar oito medalhas de ouro .
B 2008	3: linha verde	**2008_08_13-21_59_02-Jornal2-2 bloco 2**: michael phelps já nos habituou a nadar sempre à frente desta linha verde linha que marca o anterior recorde do mundo . em pequim phelps já conquistou cinco medalhas de ouro ,
C 2008	1: NIL	
D 2008	1: jogos olímpicos	**2008_08_13-19_59_02-Telejornal-1 bloco 3**: o nadador norte - americano já conquistou em pequim , cinco medalhas de ouro , a que se juntam as seis conseguidas há quatro anos em atenas , algo que nunca ninguém conseguiu , em jogos olímpicos , michael phelps
D 2008	2: linha verde	**2008_08_13-19_59_02-Telejornal-1 bloco 3**: algo que nunca ninguém conseguiu , em jogos olímpicos , michael phelps já nos habituou a nadar sempre à frente desta linha verde da linha que marca o anterior recorde do mundo , em pequim phelps já conquistou cinco medalhas de ouro ,

D 2008	3: norte - americano	**2008_08_13-19_59_02-Telejornal-1 bloco 3**: o nadador norte - americano já conquistou em pequim , cinco medalhas de ouro , a que se juntam as seis conseguidas há quatro anos em atenas , algo que nunca ninguém conseguiu , em jogos olímpicos , michael phelps
A 2010	1: NIL	
B 2010	1: norte - americano	**2008_08_22-19_59_02-Telejornal-1 bloco 3**: pequim e para onde vou depois de entrar para a história . algarve o nadador norte - americano está de férias num hotel algarvio onde já jogou golfe . saiu à noite e nadou claro na piscina e sem bater recordes . das medalhas chega o descanso . o campeão da china para o algarve . michael phelps
B 2010	2: china	**2008_08_22-19_59_02-Telejornal-1 bloco 3**: pequim e para onde vou depois de entrar para a história . algarve o nadador norte - americano está de férias num hotel algarvio onde já jogou golfe . saiu à noite e nadou claro na piscina e sem bater recordes . das medalhas chega o descanso . o campeão da china para o algarve . michael phelps
C 2010	1: NIL	
D 2010	1: norte - americano	**2008_08_22-19_59_02-Telejornal-1 bloco 3**: pequim e para onde vou depois de entrar para a história , algarve o nadador norte - americano está de férias num hotel algarvio onde já jogou golfe , saiu à noite e nadou claro na piscina e sem bater recordes , das medalhas chega o descanso , o campeão da china para o algarve , michael phelps
D 2010	2: china	**2008_08_22-19_59_02-Telejornal-1 bloco 3**: pequim e para onde vou depois de entrar para a história , algarve o nadador norte - americano está de férias num hotel algarvio onde já jogou golfe , saiu à noite e nadou claro na piscina e sem bater recordes , das medalhas chega o descanso , o campeão da china para o algarve , michael phelps
Question #9 - Que prémio ganhou João Ubaldo Ribeiro em 2008?		
A 2008	1: NIL	
B 2008	1: prémio camões	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa .
B 2008	2: antônio cândido	**pt/a/n/t/Antônio_Carlos_Jobim_dd0e.html**: antônio carlos jobim . por pressão junto ao congresso nacional de uma comissão de notáveis , formada por chico buarque , oscar niemeyer , joão ubaldo ribeiro , antônio cândido , antônio houaiss e edu lobo , criada e pessoalmente coordenada pelo crítico ricardo cravo albin .

C 2008	1: NIL	
D 2008	1: prémio camões	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: ubaldo ribeiro , é o prémio camões dois mil e oito o mais importante prémio atribuído a um escritor de língua portuguesa , o júri reunido em lisboa escolheu por maioria o escritor brasileiro nascido na bahia , joão
D 2008	2: antônio cândido	**pt/a/n/t/Antônio_Carlos_Jobim_dd0e.html**: antônio carlos jobim . por pressão junto ao congresso nacional de uma comissão de notáveis , formada por chico buarque , oscar niemeyer , joão ubaldo ribeiro , antônio cândido , antônio houaiss e edu lobo , criada e pessoalmente coordenada pelo crítico ricardo cravo albin .
A 2010	1: NIL	
B 2010	1: prémio camões	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa .
B 2010	2: antônio cândido	**pt/a/n/t/Antônio_Carlos_Jobim_dd0e.html**: antônio carlos jobim . por pressão junto ao congresso nacional de uma comissão de notáveis , formada por chico buarque , oscar niemeyer , joão ubaldo ribeiro , antônio cândido , antônio houaiss e edu lobo , criada e pessoalmente coordenada pelo crítico ricardo cravo albin .
C 2010	1: NIL	
D 2010	1: prémio camões	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: ubaldo ribeiro , é o prémio camões dois mil e oito o mais importante prémio atribuído a um escritor de língua portuguesa , o júri reunido em lisboa escolheu por maioria o escritor brasileiro nascido na baía , joão
D 2010	2: antônio cândido	**pt/a/n/t/Antônio_Carlos_Jobim_dd0e.html**: antônio carlos jobim . por pressão junto ao congresso nacional de uma comissão de notáveis , formada por chico buarque , oscar niemeyer , joão ubaldo ribeiro , antônio cândido , antônio houaiss e edu lobo , criada e pessoalmente coordenada pelo crítico ricardo cravo albin .
Question #10 - O que é a Avenida da Liberdade?		
A 2008	1: elementos do departamento de reabilitação urbana	**2008_07_10-19_59_01-Telejornal-1 bloco 3**: os serviços da câmara de lisboa já concluíram a vistoria aos edifícios atingidos pelo incêndio da avenida da liberdade , elementos do departamento de reabilitação urbana , terminaram as vistorias necessárias para o apuramento das condições de segurança do edifício atingido pelo incêndio .

A 2008	2: onde eram interpor uma providência cautelar os quer impedir qualquer obra no prédio abandonado onde começou o fogo	**2008_07_08-19_59_01-Telejornal-1 bloco 2**: os moradores de um dos prédios atingidos no incêndio na avenida da liberdade , onde eram interpor uma providência cautelar os quer impedir qualquer obra no prédio abandonado onde começou o fogo .
A 2008	3: venezuela	**2008_07_24-21_59_02-Jornal2-2 bloco 2**: o presidente da venezuela passou quase avenida da liberdade para homenagear simón bolívar .
B 2008	1: depois do terramoto de 1 755 o marquês de pombal criou o passeio público na área ocupada pela parte inferior da avenida da liberdade e praça dos restauradores . apesar do nome , era rodeado por muros e portões por onde só passavam os membros da alta sociedade .	**pt/a/v/e/Avenida_da_Liberdade_680e.html**: depois do terramoto de 1 755 o marquês de pombal criou o passeio público na área ocupada pela parte inferior da avenida da liberdade e praça dos restauradores . apesar do nome , era rodeado por muros e portões por onde só passavam os membros da alta sociedade . em 1 821 , quando os liberais subiram ao poder , os muros foram derrubados e o passeio foi aberto a todos . a avenida que hoje se pode ver foi construída em 1 879 - 82 no estilo dos campos elísios em paris . a grande avenida arborizada tornou - se num centro de cortejos , festividades e manifestações .
C 2008	1: ninguém fica indiferente contestado	**2008_07_24-21_59_02-Jornal2-2 bloco 2**: numa cerimónia na avenida da liberdade em lisboa , sócrates acompanhava o presidente da venezuela , e teve uma recepção aposta , ninguém fica indiferente contestado por exiguidade por outro lado homólogo do ouro negro da casa av brasília em lisboa aclamado como um ídolo ,
C 2008	2: venezuela	**2008_07_24-21_59_02-Jornal2-2 bloco 2**: aplausos para os jovens , conversam campos para sócrates , como , o presidente da venezuela passou quase avenida da liberdade para homenagear simón bolívar , o herói da independência sul - americana ,
C 2008	3: é sempre deles são propriedade pública alguns de propriedade municipal	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: prédios devolutos com este vinte e três de à avenida da liberdade , é sempre deles são propriedade pública alguns de propriedade municipal , e que não a prioridade da autarquia para recuperar , esses prédios devolutos para isso mesmo , foi negociado , o empréstimo ,

D 2008	1: depois do terramoto de 1 755 o marquês de pombal criou o passeio público na área ocupada pela parte inferior da avenida da liberdade e praça dos restauradores . apesar do nome , era rodeado por muros e portões por onde só passavam os membros da alta sociedade .	**pt/a/v/e/Avenida_da_Liberdade_680e.html**: depois do terramoto de 1 755 o marquês de pombal criou o passeio público na área ocupada pela parte inferior da avenida da liberdade e praça dos restauradores . apesar do nome , era rodeado por muros e portões por onde só passavam os membros da alta sociedade . em 1 821 , quando os liberais subiram ao poder , os muros foram derrubados e o passeio foi aberto a todos . a avenida que hoje se pode ver foi construída em 1 879 - 82 no estilo dos campos elísios em paris . a grande avenida arborizada tornou - se num centro de cortejos , festividades e manifestações .
A 2010	1: venezuela	**2008_07_24-19_59_02-Telejornal-1 bloco 1**: numa cerimónia na avenida da liberdade josé sócrates acompanhava o presidente da venezuela e teve a recepção oposta .
A 2010	2: acompanhava	**2008_07_24-19_59_02-Telejornal-1 bloco 1**: numa cerimónia na avenida da liberdade josé sócrates acompanhava o presidente da venezuela e teve a recepção oposta .
A 2010	3: chamas obrigaram ao corte	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: as chamas obrigaram ao corte da avenida da liberdade às primeiras horas da madrugada .
B 2010	1: depois do terramoto de 1 755 o marquês de pombal criou o passeio público na área ocupada pela parte inferior da avenida da liberdade e praça dos restauradores . apesar do nome , era rodeado por muros e portões por onde só passavam os membros da alta sociedade .	**pt/a/v/e/Avenida_da_Liberdade_680e.html**: depois do terramoto de 1 755 o marquês de pombal criou o passeio público na área ocupada pela parte inferior da avenida da liberdade e praça dos restauradores . apesar do nome , era rodeado por muros e portões por onde só passavam os membros da alta sociedade . em 1 821 , quando os liberais subiram ao poder , os muros foram derrubados e o passeio foi aberto a todos . a avenida que hoje se pode ver foi construída em 1 879 - 82 no estilo dos campos elísios em paris . a grande avenida arborizada tornou - se num centro de cortejos , festividades e manifestações .

C 2010	1: é cortado	**2008_07_07-19_59_01-Telejornal-1 bloco 3**: da liberdade em lisboa , o incêndio começa no segundo andar depressa seu astra a todo o prédio , em estado volume todos os seis pisos , o trânsito na avenida da liberdade , é cortado ,
C 2010	2: acompanhava presidente da venezuela	**2008_07_24-21_59_02-Jornal2-2 bloco 2**: avenida da liberdade em lisboa sócrates acompanhava presidente da venezuela e teve uma recepção aposta , ninguém fica indiferente contestado por onze admirado por outro jogo da
C 2010	3: chamas obrigaram ao corte	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: as chamas obrigaram ao corte da avenida da liberdade às primeiras horas da madrugada , o incêndio aconteceu no último quarteirão antes da praça dos restauradores começou , no número vinte e três da avenida um prédio de seis andares , devoluto ,
D 2010	1: depois do terramoto de 1 755 o marquês de pombal criou o passeio público na área ocupada pela parte inferior da avenida da liberdade e praça dos restauradores . apesar do nome , era rodeado por muros e portões por onde só passavam os membros da alta sociedade .	**pt/a/v/e/Avenida_da_Liberdade_680e.html**: depois do terramoto de 1 755 o marquês de pombal criou o passeio público na área ocupada pela parte inferior da avenida da liberdade e praça dos restauradores . apesar do nome , era rodeado por muros e portões por onde só passavam os membros da alta sociedade . em 1 821 , quando os liberais subiram ao poder , os muros foram derrubados e o passeio foi aberto a todos . a avenida que hoje se pode ver foi construída em 1 879 - 82 no estilo dos campos elísios em paris . a grande avenida arborizada tornou - se num centro de cortejos , festividades e manifestações .
Question #11 - Qual a primeira cidade onde vai permanecer a selecção nacional durante o europeu de futebol?		
A 2008	1: NIL	
B 2008	1: suíça	**2008_06_01-21_59_02-Jornal2-2 bloco 1**: a selecção nacional de futebol já está na suíça . a equipa partiu a meio da tarde do aeroporto de lisboa e aterrou há cerca de duas horas depois e a viagem até nos à tel . a primeira cidade onde vai permanecer durante o europeu de futebol .
B 2008	2: lisboa	**2008_06_01-21_59_02-Jornal2-2 bloco 1**: a selecção nacional de futebol já está na suíça . a equipa partiu a meio da tarde do aeroporto de lisboa e aterrou há cerca de duas horas depois e a viagem até nos à tel . a primeira cidade onde vai permanecer durante o europeu de futebol .

B 2008	3: aterrou	**2008_06_01-21_59_02-Jornal2-2 bloco 1**: a selecção nacional de futebol já está na suíça . a equipa partiu a meio da tarde do aeroporto de lisboa e aterrou há cerca de duas horas depois e a viagem até nos à tel . a primeira cidade onde vai permanecer durante o europeu de futebol .
C 2008	1: NIL	
D 2008	1: suíça	**2008_06_01-21_59_02-Jornal2-2 bloco 1**: boa noite , a selecção nacional de futebol já está na suíça , a equipa partiu a meio da tarde do aeroporto de lisboa e aterrou há cerca de duas horas depois e a viagem até nos à tel , a primeira cidade onde vai permanecer durante o europeu de futebol ,
D 2008	2: noite	**2008_06_01-21_59_02-Jornal2-2 bloco 1**: boa noite , a selecção nacional de futebol já está na suíça , a equipa partiu a meio da tarde do aeroporto de lisboa e aterrou há cerca de duas horas depois e a viagem até nos à tel , a primeira cidade onde vai permanecer durante o europeu de futebol ,
D 2008	3: lisboa	**2008_06_01-21_59_02-Jornal2-2 bloco 1**: boa noite , a selecção nacional de futebol já está na suíça , a equipa partiu a meio da tarde do aeroporto de lisboa e aterrou há cerca de duas horas depois e a viagem até nos à tel , a primeira cidade onde vai permanecer durante o europeu de futebol ,
A 2010	1: NIL	
B 2010	1: suíça	**2008_06_01-21_59_02-Jornal2-2 bloco 1**: a selecção nacional de futebol já está na suíça a equipa partiu a meio da tarde do aeroporto de lisboa e aterrou há cerca de duas horas depois seguiu viagem até nos à tel a primeira cidade onde vai permanecer durante o europeu de futebol .
B 2010	2: lisboa	**2008_06_01-21_59_02-Jornal2-2 bloco 1**: a selecção nacional de futebol já está na suíça a equipa partiu a meio da tarde do aeroporto de lisboa e aterrou há cerca de duas horas depois seguiu viagem até nos à tel a primeira cidade onde vai permanecer durante o europeu de futebol .
B 2010	3: aterrou	**2008_06_01-21_59_02-Jornal2-2 bloco 1**: a selecção nacional de futebol já está na suíça a equipa partiu a meio da tarde do aeroporto de lisboa e aterrou há cerca de duas horas depois seguiu viagem até nos à tel a primeira cidade onde vai permanecer durante o europeu de futebol .
C 2010	1: NIL	
D 2010	1: noite	**2008_06_01-21_59_02-Jornal2-2 bloco 1**: boa noite , a selecção nacional de futebol já está na suíça a equipa partiu a meio da tarde do aeroporto de lisboa e aterrou há cerca de duas horas depois seguiu viagem até nos à tel a primeira cidade onde vai permanecer durante o europeu de futebol ,

D 2010	2: suíça	**2008_06_01-21_59_02-Jornal2-2 bloco 1**: boa noite , a selecção nacional de futebol já está na suíça a equipa partiu a meio da tarde do aeroporto de lisboa e aterrou há cerca de duas horas depois seguiu viagem até nos à tel a primeira cidade onde vai permanecer durante o europeu de futebol ,
D 2010	3: lisboa	**2008_06_01-21_59_02-Jornal2-2 bloco 1**: boa noite , a selecção nacional de futebol já está na suíça a equipa partiu a meio da tarde do aeroporto de lisboa e aterrou há cerca de duas horas depois seguiu viagem até nos à tel a primeira cidade onde vai permanecer durante o europeu de futebol ,
Question #12 - Qual o número de visitantes do Museu Berardo um ano após a sua abertura?		
A 2008	1: renovação de certo modo contratos	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: adianta que para pessoal não docente está autorizada a renovação de certo modo contratos de passagem os cortes de milícias hutus funcionou . esta noite um museu berardo comemora um ano de existência , com mais de quinhentos mil visitantes após a abertura .
A 2008	2: passa	**2008_07_03-19_59_01-Telejornal-1 bloco 6**: após a abertura , é o museu com mais visitantes em portugal . ela quase passa despercebida . mas ela está sempre à espreita . eles estão todos os dias . no museu berardo .
A 2008	3: espreita	**2008_07_03-19_59_01-Telejornal-1 bloco 6**: após a abertura , é o museu com mais visitantes em portugal . ela quase passa despercebida . mas ela está sempre à espreita . eles estão todos os dias . no museu berardo .
B 2008	1: ano após	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: um ano após a abertura o museu berardo teve mais de meio milhão de visitantes
B 2008	2: está	**2008_07_03-19_59_01-Telejornal-1 bloco 6**: museu berardo está a comemorar um ano de existência com mais de quinhentos mil visitantes após a abertura , é o museu com mais visitantes em portugal .
B 2008	3: meio	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: um ano após a abertura o museu berardo teve mais de meio milhão de visitantes
C 2008	1: parlamento em silêncio	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: e um é como se falasse com eles , parlamento em silêncio , um ano após a abertura o museu berardo teve mais de meio milhão de visitantes
C 2008	2: passa	**2008_07_03-19_59_01-Telejornal-1 bloco 6**: museu berardo está a comemorar um ano de existência com mais de quinhentos mil visitantes após a abertura , é o museu com mais visitantes em portugal , ela quase passa despercebida , mas ela está sempre à espreita , eles estão todos os dias , no museu berardo ,

C 2008	3: arte contemporânea	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: um ano após a abertura o museu berardo teve mais de meio milhão de visitantes são pessoas como a margarida e o gonçalo que vão continuar a receber quem passar pela maior colecção de arte contemporânea do país , e sigo agora ao encontro da jornalista diana palma de barco está precisamente no museu berardo ,
D 2008	1: mas ela está sempre à espreita , eles estão todos os dias	**2008_07_03-19_59_01-Telejornal-1 bloco 6**: museu berardo está a comemorar um ano de existência com mais de quinhentos mil visitantes após a abertura , é o museu com mais visitantes em portugal , ela quase passa despercebida , mas ela está sempre à espreita , eles estão todos os dias , no museu berardo ,
D 2008	2: maior colecção de arte contemporânea do país	**2008_07_03-19_59_01-Telejornal-1 bloco 6**: parlamento em silêncio , um ano após a abertura o museu berardo teve mais de meio milhão de visitantes são pessoas como a margarida e o gonçalo que vão continuar a receber quem passar pela maior colecção de arte contemporânea do país ,
D 2008	3: eles estão todos os dias	**2008_07_03-19_59_01-Telejornal-1 bloco 6**: museu berardo está a comemorar um ano de existência com mais de quinhentos mil visitantes após a abertura , é o museu com mais visitantes em portugal , ela quase passa despercebida , mas ela está sempre à espreita , eles estão todos os dias , no museu berardo ,
A 2010	1: passa despercebida	**2008_07_03-19_59_01-Telejornal-1 bloco 7**: após a abertura . é o museu com mais visitantes em portugal . ela quase passa despercebida está sempre à espreita . eles estão todos os dias no museu berardo .
A 2010	2: conhecer	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: após a abertura . é um museu com mais visitantes em portugal a jornalista teresa nicolau também fez uma visita mas desta vez , para conhecer quem por lá trabalha . ela quase passa despercebida está assim brás bem . eles estão todos os dias no museu berardo .
A 2010	3: teresa	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: após a abertura . é um museu com mais visitantes em portugal a jornalista teresa nicolau também fez uma visita mas desta vez , para conhecer quem por lá trabalha . ela quase passa despercebida está assim brás bem . eles estão todos os dias no museu berardo .
B 2010	1: ano após	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: um ano após a abertura o museu berardo teve mais de meio milhão de visitantes
B 2010	2: também fez uma visita mas desta	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: abertura . é um museu com mais visitantes em portugal a jornalista teresa nicolau também fez uma visita mas desta vez , para conhecer quem por lá trabalha . ela quase passa despercebida está assim brás bem . eles estão todos os dias no museu berardo .

B 2010	3: ela quase passa despercebida está assim brás bem	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: abertura . é um museu com mais visitantes em portugal a jornalista teresa nicolau também fez uma visita mas desta vez , para conhecer quem por lá trabalha . ela quase passa despercebida está assim brás bem . eles estão todos os dias no museu berardo .
C 2010	1: parlamento em silêncio	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: e é como se falasse com eles no parlamento em silêncio , um ano após a abertura o museu berardo teve mais de meio milhão de visitantes
C 2010	2: passa	**2008_07_03-19_59_01-Telejornal-1 bloco 7**: o museu berardo está a comemorar um ano de existência com mais de quinhentos mil visitantes após a abertura , é o museu com mais visitantes em portugal , ela quase passa despercebida está sempre à espreita , eles estão todos os dias no museu berardo ,
C 2010	3: teresa	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: esta noite um museu berardo comemora um ano de existência com mais de quinhentos mil visitantes após a abertura , é um museu com mais visitantes em portugal a jornalista teresa nicolau também fez uma visita mas desta vez ,
D 2010	1: nicolau também fez uma visita mas desta vez	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: esta noite um museu berardo comemora um ano de existência com mais de quinhentos mil visitantes após a abertura , é um museu com mais visitantes em portugal a jornalista teresa nicolau também fez uma visita mas desta vez ,
D 2010	2: ano após	**2008_07_03-21_59_01-Jornal2-2 bloco 3**: e é como se falasse com eles no parlamento em silêncio , um ano após a abertura o museu berardo teve mais de meio milhão de visitantes
D 2010	3: eles estão todos os dias	**2008_07_03-19_59_01-Telejornal-1 bloco 7**: o museu berardo está a comemorar um ano de existência com mais de quinhentos mil visitantes após a abertura , é o museu com mais visitantes em portugal , ela quase passa despercebida está sempre à espreita , eles estão todos os dias no museu berardo ,
Question #13 - Quem é Paulo Rangel?		
A 2008	1: o rosto que vai liderar os deputados do psd	**2008_06_25-19_59_02-Telejornal-1 bloco 1**: paulo rangel é o rosto que vai liderar os deputados do psd . a partir de amanhã sucedendo a santana lopes . durante quase todo o debate desta tarde o rangel
A 2008	2: josé sócrates	**2008_07_09-21_59_02-Jornal2-2 bloco 2**: a expectativa é grande , até porque pela primeira vez , paulo rangel e josé sócrates , vão estar frente a frente .
A 2008	3: ministro ganho para já na experiência	**2008_07_09-21_59_02-Jornal2-2 bloco 2**: o primeiro - ministro ganho para já na experiência , paulo rangel se está no parlamento há três anos , mas muitos factores que podem influenciar o rumo do debate .

B 2008	1: pedro paulo rangel	**pt/p/e/d/Pedro_Paulo_Rangel_0994.html**: pedro paulo rangel . pedro paulo rangel
B 2008	2: manuela ferreira leite	**2008_06_25-21_59_02-Jornal2-2 bloco 2**: o dito por manuela ferreira leite paulo rangel , é amanhã eleito novo líder parlamentar do psd .
B 2008	3: josé sócrates	**2008_07_09-19_59_01-Telejornal-1 bloco 3**: a expectativa é grande , até porque pela primeira vez , paulo rangel e josé sócrates , vão estar frente a frente .
C 2008	1: como é o caso de luís campos ferreira	**2008_06_25-19_59_02-Telejornal-1 bloco 1**: por perto estão outros nomes que fazem parte da nova equipa paulo rangel , como é o caso de luís campos ferreira , agustin bem comum ou regina bastos , na primeira fila ,
C 2008	2: o novo líder da bancada parlamentar do psd	**2008_06_26-21_59_02-Jornal2-2 bloco 1**: e votaram contra as restantes ao lado do ps que votou contra todas as propostas comunistas escritas , paulo rangel é o novo líder da bancada parlamentar do psd ,
C 2008	3: josé sócrates	**2008_07_09-19_59_01-Telejornal-1 bloco 3**: com visões opostas sobre o estado do país , um ano das eleições , a expectativa é grande , até porque pela primeira vez , paulo rangel e josé sócrates , vão estar frente a frente , uma estreia que os politólogos antevê mais difícil , para o novo líder parlamentar do psd ,
D 2008	1: josé sócrates	**2008_06_21-21_59_01-Jornal2-2 bloco 1**: paulo rangel a confrontar josé sócrates num ano decisivo , antes das eleições , paulo rangel já tem o apoio de manuela ferreira leite , mas resta - lhe a conquistar ,
D 2008	2: pedro paulo rangel	**pt/p/e/d/Pedro_Paulo_Rangel_0994.html**: pedro paulo rangel . pedro paulo rangel
D 2008	3: manuela ferreira leite	**2008_06_21-21_59_01-Jornal2-2 bloco 1**: paulo rangel a confrontar josé sócrates num ano decisivo , antes das eleições , paulo rangel já tem o apoio de manuela ferreira leite , mas resta - lhe a conquistar ,
A 2010	1: excelente datava lisa soeiro esta forma como manuela ferreira leite não tem assento no parlamento	**2008_06_21-19_59_01-Telejornal-1 bloco 1**: excelente datava lisa soeiro esta forma como manuela ferreira leite não tem assento no parlamento será paulo rangel a confrontar josé sócrates num ano decisivo antes das eleições paulo rangel já tem o apoio de manuela ferreira leite ,
A 2010	2: deputado pela primeira vez nesta legislatura	**2008_06_21-19_59_01-Telejornal-1 bloco 1**: paulo rangel é deputado pela primeira vez nesta legislatura , mas já fez um discurso que ficou na memória parlamentar .
A 2010	3: o rosto que vai liderar os deputados do psd	**2008_06_25-19_59_02-Telejornal-1 bloco 1**: paulo rangel é o rosto que vai liderar os deputados do psd , a partir de amanhã sucedendo a santana lopes .

B 2010	1: pedro paulo rangel	**pt/p/e/d/Pedro_Paulo_Rangel_0994.html**: pedro paulo rangel . pedro paulo rangel
B 2010	2: manuela ferreira leite	**2008_06_25-21_59_02-Jornal2-2 bloco 2**: escolhido por manuela ferreira leite paulo rangel .
B 2010	3: josé sócrates vão estar frente a frente	**2008_07_09-21_59_02-Jornal2-2 bloco 2**: a expectativa é grande até porque pela primeira vez , paulo rangel e josé sócrates vão estar frente a frente .
C 2010	1: josé sócrates vão estar frente a frente	**2008_07_09-19_59_01-Telejornal-1 bloco 3**: com visões opostas sobre o estado do país a um ano das eleições , a expectativa é grande até porque pela primeira vez , paulo rangel e josé sócrates vão estar frente a frente ,
C 2010	2: deputado pela primeira vez nesta legislatura	**2008_06_21-19_59_01-Telejornal-1 bloco 1**: pedro bright anos e daniela santiago rtp , paulo rangel é deputado pela primeira vez nesta legislatura , mas já fez um discurso que ficou na memória parlamentar ,
C 2010	3: o rosto que vai liderar os deputados do psd	**2008_06_25-19_59_02-Telejornal-1 bloco 1**: mas já não passa despercebido , paulo rangel é o rosto que vai liderar os deputados do psd , a partir de amanhã sucedendo a santana lopes , durante quase todo o debate desta tarde , rangel
D 2010	1: pedro paulo rangel	**pt/p/e/d/Pedro_Paulo_Rangel_0994.html**: pedro paulo rangel . pedro paulo rangel
D 2010	2: manuela ferreira leite	**2008_06_26-21_59_02-Jornal2-2 bloco 1**: paulo rangel é o novo líder da bancada parlamentar do psd , o deputado conta com o apoio de manuela ferreira leite e foi o único candidato nestas eleições ,
D 2010	3: josé sócrates	**2008_07_09-19_59_01-Telejornal-1 bloco 3**: o debate do estado da nação vai opor amanhã o governo e oposição pela primeira vez paulo rangel vai confrontar directamente josé sócrates , os politólogos prevêem um debate difícil ,
Question #14 - Quantas contra ordenações muito graves são necessárias para haver cassação da carta?		
A 2008 B 2008	1: três contra	**2008_07_06-21_59_01-Jornal2-2 bloco 2**: a partir de agora ao fim de três contra - ordenações muito graves . à cassação da carta .
B 2008	2: 192 contra	**pt/j/o/s/José_Dirceu_77c3.html**: josé dirceu . o placar da votação foi de 293 votos a favor da cassação e 192 contra .
B 2008	3: 190 contra	**pt/l/i/s/Lista_de_autoridades_absolvidas_após_o_escândalo_do_mensalão.html**: lista de autoridades absolvidas após o escândalo do mensalão . na contagem dos votos , dos 443 deputados que compareceram 228 votaram a favor da cassação , 190 contra , 19 abstenções , 5 brancos e 1 nulo .
C 2008 D 2008	1: três contra	**2008_07_06-19_59_02-Telejornal-1 bloco 3**: não causou qualquer vítima , entraram em vigor as alterações ao código da estrada , a partir de hoje bastam três contra - ordenações , muito graves , pra ver cassação da carta , joaquim génese de patrulhar as estradas os vinte e oito anos ,

D 2008	2: 192 contra	**pt/j/o/s/José_Dirceu_77c3.html**: josé dirceu . o placar da votação foi de 293 votos a favor da cassação e 192 contra .
D 2008	3: 190 contra	**pt/l/i/s/Lista_de_autoridades_absolvidas_após_o_escândalo_do_mensalão.html**: lista de autoridades absolvidas após o escândalo do mensalão . na contagem dos votos , dos 443 deputados que compareceram 228 votaram a favor da cassação , 190 contra , 19 abstenções , 5 brancos e 1 nulo .
A 2010 B 2010	1: três contra	**2008_07_06-21_59_01-Jornal2-2 bloco 2**: a partir de agora ao fim de três contra - ordenações muito graves . à cassação da carta .
B 2010	2: 192 contra	**pt/j/o/s/José_Dirceu_77c3.html**: josé dirceu . o placar da votação foi de 293 votos a favor da cassação e 192 contra .
B 2010	3: 190 contra	**pt/l/i/s/Lista_de_autoridades_absolvidas_após_o_escândalo_do_mensalão.html**: lista de autoridades absolvidas após o escândalo do mensalão . na contagem dos votos , dos 443 deputados que compareceram 228 votaram a favor da cassação , 190 contra , 19 abstenções , 5 brancos e 1 nulo .
C 2010 D 2010	1: três contra	**2008_07_05-19_59_01-Telejornal-1 bloco 2**: a , vai ser mais fácil ficar sem carta de condução bastam três contra - ordenações muito graves num espaço de cinco anos , entra amanhã em vigor as alterações ao código da estrada que facilitou o processo de cassação da carta ,
D 2010	2: 192 contra	**pt/j/o/s/José_Dirceu_77c3.html**: josé dirceu . o placar da votação foi de 293 votos a favor da cassação e 192 contra .
D 2010	3: 190 contra	**pt/l/i/s/Lista_de_autoridades_absolvidas_após_o_escândalo_do_mensalão.html**: lista de autoridades absolvidas após o escândalo do mensalão . na contagem dos votos , dos 443 deputados que compareceram 228 votaram a favor da cassação , 190 contra , 19 abstenções , 5 brancos e 1 nulo .
Question #15 - Onde estiveram reunidos os ministros das finanças do G-8?		
A 2008	1: osaca no japão	**2008_06_14-21_59_01-Jornal2-2 bloco 1**: que aumentem a produção para travar a escalada dos preços , reunidos em osaca no japão . os ministros das finanças do g oito . não chegaram a acordo , nem sobre as causas da crise , nem sobre as soluções .
A 2008	2: fiabilidade	**2008_06_14-21_59_01-Jornal2-2 bloco 1**: os ministros das finanças do g oito reunidos no japão . dizem que o mercado petrolífero funcionaria melhor . se existisse maior transparência e fiabilidade das informações de mercado nomeadamente existências e dimensão dos fluxos financeiros .

A 2008	3: fluxos	**2008_06_14-21_59_01-Jornal2-2 bloco 1**: os ministros das finanças do g oito reunidos no japão . dizem que o mercado petrolífero funcionaria melhor . se existisse maior transparência e fiabilidade das informações de mercado nomeadamente existências e dimensão dos fluxos financeiros .
B 2008	1: união europeia	**pt/c/o/m/Comissão_Europeia_7dbe.html**: comissão europeia . g - 8 em são petersburgo . ver também história da união europeia cronologia da união europeia personagens chave da união europeia política da união europeia ligações externas portal oficial em português . categorias : comissão europeia | instituições da união europeia | órgãos administrativos da união europeia
B 2008	2: está actualmente em fase	**pt/c/o/m/Comboio_de_alta_velocidade.html**: comboio de alta velocidade . agora g - 8) , esta tecnologia está actualmente em fase de testes , estando previsto que entre em serviço no secto seul - gwangju em 2 007 . o comboio proposto circularia mais rápido que o tgv ,
B 2008	3: política de marrocos	**pt/p/o/l/Política_de_Marrocos_ded7.html**: política de marrocos . g - 8 na geórgia . o fórum deverá reunir autoridades do meio político , personalidades da sociedade civil e homens de negócios de cerca de 25 países para examinar como promover as reformas democráticas na áfrica do norte , no oriente próximo ,
C 2008	1: osaca no japão	**2008_06_14-21_59_01-Jornal2-2 bloco 1**: reunidos em osaca no japão , os ministros das finanças do g oito , não chegaram a acordo , nem sobre as causas da crise , nem sobre as soluções , mas defendeu uma maior transparência no mercado petrolífero ,
C 2008	2: fiabilidade	**2008_06_14-21_59_01-Jornal2-2 bloco 1**: os ministros das finanças do g oito reunidos no japão , dizem que o mercado petrolífero funcionaria melhor , se existisse maior transparência e fiabilidade das informações de mercado nomeadamente existências e dimensão dos fluxos financeiros , os oito
C 2008	3: economizar	**2008_06_04-21_59_01-Jornal2-2 bloco 1**: mas o ministro das finanças do japão já veio dizer que os representantes do g oito não podem sozinhos , tomar medidas radicais de mim conta o que está na base de que se o movimento considera que os países consumidores deve economizar energia ,
D 2008	1: união europeia	**pt/c/o/m/Comissão_Europeia_7dbe.html**: comissão europeia . g - 8 em são petersburgo . ver também história da união europeia cronologia da união europeia personagens chave da união europeia política da união europeia ligações externas portal oficial em português . categorias : comissão europeia | instituições da união europeia | órgãos administrativos da união europeia

D 2008	2: está actualmente em fase	**pt/c/o/m/Comboio_de_alta_velocidade.html**: comboio de alta velocidade . agora g - 8) , esta tecnologia está actualmente em fase de testes , estando previsto que entre em serviço no secto seul - gwangju em 2 007 . o comboio proposto circularia mais rápido que o tgv ,
D 2008	3: política de marrocos	**pt/p/o/l/Política_de_Marrocos_ded7.html**: política de marrocos . g - 8 na geórgia . o fórum deverá reunir autoridades do meio político , personalidades da sociedade civil e homens de negócios de cerca de 25 países para examinar como promover as reformas democráticas na áfrica do norte , no oriente próximo ,
A 2010	1: fluxos	**2008_06_14-21_59_01-Jornal2-2 bloco 1**: os ministros das finanças do g oito reunidos no japão . dizem que o mercado petrolífero funcionaria melhor . se existisse maior transparência e fiabilidade das informações do mercado nomeadamente existências e dimensão dos fluxos financeiros .
A 2010	2: fiabilidade	**2008_06_14-21_59_01-Jornal2-2 bloco 1**: os ministros das finanças do g oito reunidos no japão . dizem que o mercado petrolífero funcionaria melhor . se existisse maior transparência e fiabilidade das informações do mercado nomeadamente existências e dimensão dos fluxos financeiros .
A 2010	3: itálico	**2008_06_14-21_59_01-Jornal2-2 bloco 1**: finanças do g oito reunidos no japão . dizem que o mercado petrolífero funcionaria melhor . se existisse maior transparência e fiabilidade das informações do mercado nomeadamente existências e dimensão dos fluxos financeiros . os oito países mais industrializados estados unidos frança reino unido japão ao manhã itálico canadá rússia . defende o diálogo e
B 2010	1: união europeia	**pt/c/o/m/Comissão_Europeia_7dbe.html**: comissão europeia . g - 8 em são petersburgo . ver também história da união europeia cronologia da união europeia personagens chave da união europeia política da união europeia ligações externas portal oficial em português . categorias : comissão europeia | instituições da união europeia | órgãos administrativos da união europeia
B 2010	2: está actualmente em fase	**pt/c/o/m/Comboio_de_alta_velocidade.html**: comboio de alta velocidade . agora g - 8) , esta tecnologia está actualmente em fase de testes , estando previsto que entre em serviço no secto seul - gwangju em 2 007 . o comboio proposto circularia mais rápido que o tgv ,
B 2010	3: política de marrocos	**pt/p/o/l/Política_de_Marrocos_ded7.html**: política de marrocos . g - 8 na geórgia . o fórum deverá reunir autoridades do meio político , personalidades da sociedade civil e homens de negócios de cerca de 25 países para examinar como promover as reformas democráticas na áfrica do norte , no oriente próximo ,

C 2010	1: fiabilidade	**2008_06_14-21_59_01-Jornal2-2 bloco 1**: os ministros das finanças do g oito reunidos no japão , dizem que o mercado petrolífero funcionaria melhor , se existisse maior transparência e fiabilidade das informações do mercado nomeadamente existências e dimensão dos fluxos financeiros ,
C 2010	2: osaka	**2008_06_14-21_59_01-Jornal2-2 bloco 1**: reunidos em osaka no japão , os ministros das finanças do g oito , não chegaram a acordo nem sobre as causas da crise nem sobre as soluções ,
C 2010	3: fluxos	**2008_06_14-21_59_01-Jornal2-2 bloco 1**: os ministros das finanças do g oito reunidos no japão , dizem que o mercado petrolífero funcionaria melhor , se existisse maior transparência e fiabilidade das informações do mercado nomeadamente existências e dimensão dos fluxos financeiros ,
D 2010	1: união europeia	**pt/c/o/m/Comissão_Europeia_7dbe.html**: comissão europeia . g - 8 em são petersburgo . ver também história da união europeia cronologia da união europeia personagens chave da união europeia política da união europeia ligações externas portal oficial em português . categorias : comissão europeia | instituições da união europeia | órgãos administrativos da união europeia
D 2010	2: está actualmente em fase	**pt/c/o/m/Comboio_de_alta_velocidade.html**: comboio de alta velocidade . agora g - 8) , esta tecnologia está actualmente em fase de testes , estando previsto que entre em serviço no secto seul - gwangju em 2 007 . o comboio proposto circularia mais rápido que o tgv ,
D 2010	3: política de marrocos	**pt/p/o/l/Política_de_Marrocos_ded7.html**: política de marrocos . g - 8 na geórgia . o fórum deverá reunir autoridades do meio político , personalidades da sociedade civil e homens de negócios de cerca de 25 países para examinar como promover as reformas democráticas na áfrica do norte , no oriente próximo ,
Question #16 - Quantos polícias foram mobilizados para a Marcha de Contestação do G-8?		
A 2008 B 2008	1: vinte mil polícias	**2008_07_06-21_59_01-Jornal2-2 bloco 2**: g oito será um sucesso . embora admita que a economia norte - americana está em enfrentar dificuldades . cerca de cinco mil manifestantes mas apenas quatro por fila . foi a condição das autoridades para autorizarem a marcha da contestação g oito , e sob fortes medidas de vigilância dos vinte mil polícias mobilizados para hokkaido
C 2008 D 2008	1: vinte mil polícias	**2008_07_06-21_59_01-Jornal2-2 bloco 2**: g oito será um sucesso , embora admita que a economia norte - americana está em enfrentar dificuldades , cerca de cinco mil manifestantes mas apenas quatro por fila , foi a condição das autoridades para autorizarem a marcha da contestação g oito , e sob fortes medidas de vigilância dos vinte mil polícias mobilizados para hokkaido

A 2010 B 2010	1: vinte mil polícias	**2008_07_06-19_59_02-Telejornal-1 bloco 5**: do g oito serão sucesso . embora admita que a economia americana está a enfrentar dificuldades . cerca de cinco mil manifestantes , mas apenas quatro por fila . foi a condição das autoridades para autorizar em marcha da contestação g oito e sob fortes medidas de vigilância dos vinte mil polícias mobilizados
C 2010 D 2010	1: vinte mil polícias	**2008_07_06-19_59_02-Telejornal-1 bloco 5**: do g oito serão sucesso , embora admita que a economia americana está a enfrentar dificuldades , cerca de cinco mil manifestantes , mas apenas quatro por fila , foi a condição das autoridades para autorizar em marcha da contestação g oito e sob fortes medidas de vigilância dos vinte mil polícias mobilizados
Question #17 - Quantos manifestantes foram autorizados?		
A 2008 B 2008	1: cerca de cinco mil manifestantes	**2008_07_06-21_59_01-Jornal2-2 bloco 2**: g oito será um sucesso . embora admita que a economia norte - americana está em enfrentar dificuldades . cerca de cinco mil manifestantes mas apenas quatro por fila . foi a condição das autoridades para autorizarem a marcha da contestação g oito , e sob fortes medidas de vigilância dos vinte mil polícias mobilizados para hokkaido
C 2008	1: cerca de cinco mil manifestantes	**2008_07_06-19_59_02-Telejornal-1 bloco 5**: george bush prometeu que a cimeira do g oito serão um sucesso , embora admita que a economia americana está a enfrentar dificuldades , cerca de cinco mil manifestantes mas apenas quatro por fila , foi a condição das autoridades para autorizarem a marcha da contestação g oito ,
D 2008	1: cerca de cinco mil manifestantes	**2008_07_06-21_59_01-Jornal2-2 bloco 2**: g oito será um sucesso , embora admita que a economia norte - americana está em enfrentar dificuldades , cerca de cinco mil manifestantes mas apenas quatro por fila , foi a condição das autoridades para autorizarem a marcha da contestação g oito , e sob fortes medidas de vigilância dos vinte mil polícias mobilizados para hokkaido
A 2010	1: cerca de cinco mil manifestantes	**2008_07_06-21_59_01-Jornal2-2 bloco 2**: do g oito será um sucesso . embora admita que a economia norte - americana está em enfrentar dificuldades . cerca de cinco mil manifestantes , mas apenas quatro por fila . foi a condição das autoridades para autorizar em marcha da contestação g oito
B 2010	1: cerca de cinco mil manifestantes	**2008_07_06-19_59_02-Telejornal-1 bloco 5**: do g oito serão sucesso . embora admita que a economia americana está a enfrentar dificuldades . cerca de cinco mil manifestantes , mas apenas quatro por fila . foi a condição das autoridades para autorizar em marcha da contestação g oito e sob fortes medidas de vigilância dos vinte mil polícias mobilizados

C 2010	1: cerca de cinco mil manifestantes	**2008_07_06-21_59_01-Jornal2-2 bloco 2**: do g oito será um sucesso , embora admita que a economia norte - americana está em enfrentar dificuldades , cerca de cinco mil manifestantes , mas apenas quatro por fila , foi a condição das autoridades para autorizar em marcha da contestação
D 2010	1: cerca de cinco mil manifestantes	**2008_07_06-19_59_02-Telejornal-1 bloco 5**: do g oito serão sucesso , embora admita que a economia americana está a enfrentar dificuldades , cerca de cinco mil manifestantes , mas apenas quatro por fila , foi a condição das autoridades para autorizar em marcha da contestação g oito e sob fortes medidas de vigilância dos vinte mil polícias mobilizados
Question #18 - Quem é a maior estrela dos Jogos Olímpicos de Pequim?		
A 2008	1: NIL	
B 2008	1: vieira da silva	**2008_08_09-19_59_01-Telejornal-1 bloco 3**: a primeira participação do jovem , estrela estavam ontem muito . nervosa . mais do que o costume matar . um . isso pesou um bocadinho desenrolar da prova outono . manuel vieira da silva deu o tiro de partida a participação portuguesa nos jogos olímpicos de pequim .
B 2008	2: dos	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: a estrela dos jogos olímpicos de pequim .
B 2008	3: portuguesa	**2008_08_09-19_59_01-Telejornal-1 bloco 3**: a primeira participação do jovem , estrela estavam ontem muito . nervosa . mais do que o costume matar . um . isso pesou um bocadinho desenrolar da prova outono . manuel vieira da silva deu o tiro de partida a participação portuguesa nos jogos olímpicos de pequim .
C 2008	1: NIL	
D 2008	1: vieira da silva	**2008_08_09-19_59_01-Telejornal-1 bloco 3**: a primeira participação do jovem , estrela estavam ontem muito , nervosa , mais do que o costume matar , um , isso pesou um bocadinho desenrolar da prova outono , manuel vieira da silva deu o tiro de partida a participação portuguesa nos jogos olímpicos de pequim ,
D 2008	2: field	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: a estrela dos jogos olímpicos de pequim , é a imagem da vitória , tal como outras estrelas do desporto mundial michael field representam um mundo de possibilidades para os especialistas em marketing , de um sonho para qualquer marca , câmara dá corpo ao ,
D 2008	3: corpo	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: a estrela dos jogos olímpicos de pequim , é a imagem da vitória , tal como outras estrelas do desporto mundial michael field representam um mundo de possibilidades para os especialistas em marketing , de um sonho para qualquer marca , câmara dá corpo ao ,

A 2010	1: NIL	
B 2010	1: field	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: jogos olímpicos de pequim . é a imagem da vitória . tal como outras estrelas do desporto mundial michael field representam um mundo de possibilidades para os especialistas em marketing e um sonho para qualquer marca . quando o mar bate o pé ao . um mais o véu de
B 2010	2: marca	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: jogos olímpicos de pequim . é a imagem da vitória . tal como outras estrelas do desporto mundial michael field representam um mundo de possibilidades para os especialistas em marketing e um sonho para qualquer marca . quando o mar bate o pé ao . um mais o véu de
B 2010	3: mas afinal	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: mas afinal quanto vale em euros a estrela dos jogos olímpicos de pequim .
C 2010	1: NIL	
D 2010	1: field	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: jogos olímpicos de pequim , é a imagem da vitória , tal como outras estrelas do desporto mundial michael field representam um mundo de possibilidades para os especialistas em marketing e um sonho para qualquer marca , quando o mar bate o pé ao ,
D 2010	2: marca	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: jogos olímpicos de pequim , é a imagem da vitória , tal como outras estrelas do desporto mundial michael field representam um mundo de possibilidades para os especialistas em marketing e um sonho para qualquer marca , quando o mar bate o pé ao ,
D 2010	3: mas afinal	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: mas afinal quanto vale em euros a estrela dos jogos olímpicos de pequim , é a imagem da vitória ,
Question #19 - De quem é o projecto do museu Iberê Camargo?		
A 2008	1: NIL	
B 2008	1: siza vieira	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: frente inaugurado em portalegre o museu de iberê camargo um projecto do arquitecto siza vieira , já ganho uma medalha de ouro na bienal de veneza em que no brasil ,
B 2008	2: fundação	**pt/f/u/n/Fundação_Iberê_Camargo_fd11.html**: fundação iberê camargo . ligações externas fundação iberê camargo projeto da nova sede foto do museu
B 2008	3: medalha	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: frente inaugurado em portalegre o museu de iberê camargo um projecto do arquitecto siza vieira , já ganho uma medalha de ouro na bienal de veneza em que no brasil ,
C 2008	1: NIL	
D 2008	1: siza vieira	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: frente inaugurado em portalegre o museu de iberê camargo um projecto do arquitecto siza vieira , já ganho uma medalha de ouro na bienal de veneza em que no brasil ,

D 2008	2: fundação	**pt/f/u/n/Fundação_Iberê_Camargo_fd11.html**: fundação iberê camargo . ligações externas fundação iberê camargo projeto da nova sede foto do museu
D 2008	3: medalha	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: frente inaugurado em portalegre o museu de iberê camargo um projecto do arquitecto siza vieira , já ganho uma medalha de ouro na bienal de veneza em que no brasil ,
A 2010	1: NIL	
B 2010	1: álvaro siza	**pt/á/l/v/Álvaro_Siza_Vieira_8f93.html**: álvaro siza vieira . os planos horizontais , a clareza das formas , o requinte do espaço . criando marcos arquitectónicos na história da arquitectura portuguesa como a casa de chá , as piscinas de matosinhos , o museu serralves , a igreja de marco de canavezes , ou mais recentemente , o museu para a fundação iberê camargo
B 2010	2: fundação	**pt/f/u/n/Fundação_Iberê_Camargo_fd11.html**: fundação iberê camargo . ligações externas fundação iberê camargo projeto da nova sede foto do museu
B 2010	3: sede foto	**pt/f/u/n/Fundação_Iberê_Camargo_fd11.html**: fundação iberê camargo . ligações externas fundação iberê camargo projeto da nova sede foto do museu
C 2010	1: NIL	
D 2010	1: álvaro siza vieira	**pt/á/l/v/Álvaro_Siza_Vieira_8f93.html**: álvaro siza vieira . os planos horizontais , a clareza das formas , o requinte do espaço . criando marcos arquitectónicos na história da arquitectura portuguesa como a casa de chá , as piscinas de matosinhos , o museu serralves , a igreja de marco de canavezes , ou mais recentemente , o museu para a fundação iberê camargo
D 2010	2: fundação	**pt/f/u/n/Fundação_Iberê_Camargo_fd11.html**: fundação iberê camargo . ligações externas fundação iberê camargo projeto da nova sede foto do museu
D 2010	3: sede foto	**pt/f/u/n/Fundação_Iberê_Camargo_fd11.html**: fundação iberê camargo . ligações externas fundação iberê camargo projeto da nova sede foto do museu
Question #20 - Em que cidade fica?		
A 2008	1: NIL	
B 2008	1: portalegre	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: frente inaugurado em portalegre o museu de iberê camargo um projecto do arquitecto siza vieira , já ganho uma medalha de ouro na bienal de veneza em que no brasil , já é considerado um dos edifícios contemporâneos mais bonitos do país .
B 2008	2: porto	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: oito anos depois do sonho o museu de iberê camargo em porto alegre , e está pronto e a funcionar . é o segundo projecto de siza vieira na américa latina , e levou .

B 2008	3: brasil	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: frente inaugurado em portalegre o museu de iberê camargo um projecto do arquitecto siza vieira , já ganho uma medalha de ouro na bienal de veneza em que no brasil , já é considerado um dos edifícios contemporâneos mais bonitos do país .
C 2008	1: NIL	
D 2008	1: portalegre	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: frente inaugurado em portalegre o museu de iberê camargo um projecto do arquitecto siza vieira , já ganho uma medalha de ouro na bienal de veneza em que no brasil , já é considerado um dos edifícios contemporâneos mais bonitos do país ,
D 2008	2: brasil	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: frente inaugurado em portalegre o museu de iberê camargo um projecto do arquitecto siza vieira , já ganho uma medalha de ouro na bienal de veneza em que no brasil , já é considerado um dos edifícios contemporâneos mais bonitos do país ,
D 2008	3: porto	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: projecto do arquitecto siza vieira , já ganho uma medalha de ouro na bienal de veneza em que no brasil , já é considerado um dos edifícios contemporâneos mais bonitos do país , oito anos depois do sonho o museu de iberê camargo em porto alegre , e está
A 2010	1: NIL	
B 2010	1: brasileiros	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: desenho de um dos mais reconhecidos pintor expressionista brasileiros do século vinte . iberê camargo era gaúcho e a capital do estado escolheu um português para conceber esta casa que lhe vai perpetuar o nome ele e naturalmente também org eterno . o museu já foi
B 2010	2: capital	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: desenho de um dos mais reconhecidos pintor expressionista brasileiros do século vinte . iberê camargo era gaúcho e a capital do estado escolheu um português para conceber esta casa que lhe vai perpetuar o nome ele e naturalmente também org eterno . o museu já foi
B 2010	3: marco de canavezes	**pt/á/l/v/Álvaro_Siza_Vieira_8f93.html**: álvaro siza vieira . os planos horizontais , a clareza das formas , o requinte do espaço . criando marcos arquitectónicos na história da arquitectura portuguesa como a casa de chá , as piscinas de matosinhos , o museu serralves , a igreja de marco de canavezes , ou mais recentemente , o museu para a fundação iberê camargo
C 2010	1: NIL	
D 2010	1: brasileiros	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: desenho de um dos mais reconhecidos pintor expressionista brasileiros do século vinte , iberê camargo era gaúcho e a capital do estado escolheu um português para conceber esta casa que lhe vai perpetuar o nome ele e naturalmente também org eterno , o museu já foi

D 2010	2: capital	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: desenho de um dos mais reconhecidos pintor expressionista brasileiros do século vinte , iberê camargo era gaúcho e a capital do estado escolheu um português para conceber esta casa que lhe vai perpetuar o nome ele e naturalmente também org eterno , o museu já foi
D 2010	3: marco de canavezes	**pt/á/l/v/Álvaro_Siza_Vieira_8f93.html**: álvaro siza vieira . os planos horizontais , a clareza das formas , o requinte do espaço . criando marcos arquitectónicos na história da arquitectura portuguesa como a casa de chá , as piscinas de matosinhos , o museu serralves , a igreja de marco de canavezes , ou mais recentemente , o museu para a fundação iberê camargo
Question #21 - Onde fica a fundação José Saramago?		
A 2008	1: bicos	**2008_07_17-21_59_02-Jornal2-2 bloco 4**: a sede da fundação josé saramago vai ser na casa dos bicos em lisboa .
A 2008	2: gratuitamente	**2008_07_17-21_59_02-Jornal2-2 bloco 4**: foi hoje assinado , entre a câmara municipal de lisboa e a fundação , ou acordo prevê a cedência do espaço gratuitamente , durante essa anos , já a partir do próximo ano . josé saramago presente na cerimónia .
A 2008	3: azinhaga	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: josé saramago e a mulher pilar del rio estiveram ontem na aldeia natal do escritor azinhaga no alentejo os prémio nobel da literatura inaugurou a sede local da fundação que tem o seu nome .
B 2008	1: lisboa	**2008_07_17-19_59_01-Telejornal-1 bloco 9**: na sede da fundação josé saramago vai ser na casa dos bicos em lisboa . o protocolo de cedência do histórico edifício localizado no campo das cebolas .
B 2008	2: anos	**2008_07_17-21_59_02-Jornal2-2 bloco 4**: protocol cedência do histórico edifício do campo das cebolas , foi hoje assinado , entre a câmara municipal de lisboa e a fundação , ou acordo prevê a cedência do espaço gratuitamente , durante essa anos , já a partir do próximo ano . josé saramago presente na cerimónia .
B 2008	3: câmara	**2008_07_17-21_59_02-Jornal2-2 bloco 4**: a sede da fundação josé saramago vai ser na casa dos bicos em lisboa . protocol cedência do histórico edifício do campo das cebolas , foi hoje assinado , entre a câmara municipal de lisboa e a fundação ,
C 2008	1: bicos	**2008_07_17-19_59_01-Telejornal-1 bloco 9**: na sede da fundação josé saramago vai ser na casa dos bicos em lisboa , o protocolo de cedência do histórico edifício localizado no campo das cebolas , foi assinado entre a câmara municipal e a fundação ,

C 2008	2: gratuitamente	**2008_07_17-21_59_02-Jornal2-2 bloco 4**: a sede da fundação josé saramago vai ser na casa dos bicos em lisboa , protocol cedência do histórico edifício do campo das cebolas , foi hoje assinado , entre a câmara municipal de lisboa e a fundação , ou acordo prevê a cedência do espaço gratuitamente ,
C 2008	3: cedência	**2008_07_17-19_59_01-Telejornal-1 bloco 9**: na sede da fundação josé saramago vai ser na casa dos bicos em lisboa , o protocolo de cedência do histórico edifício localizado no campo das cebolas , foi assinado entre a câmara municipal e a fundação ,
D 2008	1: lisboa	**2008_07_17-19_59_01-Telejornal-1 bloco 9**: na sede da fundação josé saramago vai ser na casa dos bicos em lisboa , o protocolo de cedência do histórico edifício localizado no campo das cebolas , foi assinado entre a câmara municipal e a fundação ,
D 2008	2: aldeia	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: vinte camas doses , josé saramago e a mulher pilar del rio estiveram ontem na aldeia natal do escritor azinhaga no alentejo os prémio nobel da literatura inaugurou a sede local da fundação que tem o seu nome , há festa na aldeia ,
D 2008	3: câmara	**2008_07_17-19_59_01-Telejornal-1 bloco 9**: na sede da fundação josé saramago vai ser na casa dos bicos em lisboa , o protocolo de cedência do histórico edifício localizado no campo das cebolas , foi assinado entre a câmara municipal e a fundação ,
A 2010	1: bicos	**2008_07_17-19_59_01-Telejornal-1 bloco 8**: a sede da fundação josé saramago vai ser na casa dos bicos em lisboa .
A 2010	2: gratuitamente	**2008_07_17-21_59_02-Jornal2-2 bloco 4**: fundação . o acordo prevê a cedência do espaço gratuitamente . perante dessa anos . já a partir do próximo ano . josé saramago presente na cerimónia mostrou - se honrado com esta solução .
A 2010	3: zequinha agha	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: o prémio nobel veio também zequinha agha para inaugurar a sede local da fundação josé saramago que tem a bedeteca computadores .
B 2010	1: bicos em lisboa	**2008_07_17-21_59_02-Jornal2-2 bloco 4**: a sede da fundação josé saramago vai ser na casa dos bicos em lisboa .
B 2010	2: anos	**2008_07_17-19_59_01-Telejornal-1 bloco 8**: foi assinado entre a câmara municipal e a fundação . o acordo prevê a cedência do espaço gratuitamente durante dez anos . já a partir de dois mil e nove . josé saramago presente na cerimónia mostrou - se .

B 2010	3: local	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: o prémio nobel veio também zequinha agha para inaugurar a sede local da fundação josé saramago que tem a bedeteca computadores . e um piso em que está
C 2010	1: cedência do espaço gratuitamente	**2008_07_17-21_59_02-Jornal2-2 bloco 4**: fundação , o acordo prevê a cedência do espaço gratuitamente , perante dessa anos , já a partir do próximo ano , josé saramago presente na cerimónia mostrou - se honrado com esta solução , a casa dos bicos ,
C 2010	2: bicos	**2008_07_17-19_59_01-Telejornal-1 bloco 8**: a sede da fundação josé saramago vai ser na casa dos bicos em lisboa , o protocolo de cedência do histórico edifício localizado no campo das cebolas ,
C 2010	3: honrado	**2008_07_17-21_59_02-Jornal2-2 bloco 4**: fundação , o acordo prevê a cedência do espaço gratuitamente , perante dessa anos , já a partir do próximo ano , josé saramago presente na cerimónia mostrou - se honrado com esta solução , a casa dos bicos ,
D 2010	1: lisboa	**2008_07_17-19_59_01-Telejornal-1 bloco 8**: a sede da fundação josé saramago vai ser na casa dos bicos em lisboa , o protocolo de cedência do histórico edifício localizado no campo das cebolas ,
D 2010	2: aldeia	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: quim quer nas doses , josé saramago e a mulher pilar del rio estiveram ontem na aldeia natal do escritor azinhaga no alentejo , prémio nobel da literatura inaugurou a sede local da fundação que tem o seu nove , há festa na aldeia ,
D 2010	3: anos	**2008_07_17-19_59_01-Telejornal-1 bloco 8**: foi assinado entre a câmara municipal e a fundação , o acordo prevê a cedência do espaço gratuitamente durante dez anos , já a partir de dois mil e nove , josé saramago presente na cerimónia mostrou - se , honrado com esta solução ,
Question #22 - Quem é o campeão olímpico do triplo salto?		
A 2008	1: nelson évora que concretizou	**2008_08_21-19_59_01-Telejornal-1 bloco 1**: boa noite nelson évora que concretizou o sonho olímpico é ele o novo campeão olímpico do triplo salto de conquistou a medalha de ouro para portugal .
A 2008	2: ocupados	**2008_09_03-19_59_02-Telejornal-1 bloco 4**: o campeão olímpico do triplo salto , todos os quartos estão ocupados menos este .
A 2008	3: gratuitamente tremendo	**2008_08_23-21_59_01-Jornal2-2 bloco 1**: senti gratuitamente tremendo dizer que o pódio era possível mas que ourém mim de maneira muito foi , nelson évora campeão olímpico do triplo salto e vanessa fernandes prata no triatlo , conseguirão os maiores feitos .

B 2008	1: vanessa fernandes prata no triatlo	**2008_08_23-19_59_02-Telejornal-1 bloco 3**: nelson évora campeão olímpico do triplo salto e vanessa fernandes prata no triatlo , conseguirão os maiores feitos . mas há ainda há assinalar , os sete diplomas olímpico conquistado , em cinco modalidades ,
B 2008	2: olímpico	**2008_09_03-19_59_02-Telejornal-1 bloco 4**: o campeão olímpico do triplo salto , todos os quartos estão ocupados menos este .
B 2008	3: portuguesa esta tarde nélson	**2008_08_22-19_59_02-Telejornal-1 bloco 3**: um salto de dezassete metros e sessenta e sete cento entre os nelson évora fez subir a bandeira nacional . ao mastro mais alto de pequim . enquanto sou portuguesa esta tarde nélson évora campeão olímpico do triplo salto .
C 2008	1: partida	**2008_07_12-21_59_02-Jornal2-2 bloco 3**: o que é que o campeão do mundo do triplo salto foi o ponto de partida para pequim , jogos olímpicos o que vou querer estar mais forte mais rápido , sector muito bons ,
C 2008	2: seguido	**2008_08_21-19_59_01-Telejornal-1 bloco 1**: numa das bancadas mesmo em frente à caixa de saltos notava - se a presença portuguesa , quase toda a missão olímpica foi apoiar , o campeão do mundo do triplo - salto , cada salto era seguido com atenção e ansiedade ,
C 2008	3: rápido	**2008_07_12-21_59_02-Jornal2-2 bloco 3**: o que é que o campeão do mundo do triplo salto foi o ponto de partida para pequim , jogos olímpicos o que vou querer estar mais forte mais rápido , sector muito bons ,
D 2008	1: nelson évora campeão olímpico	**2008_08_23-19_59_02-Telejornal-1 bloco 3**: nelson évora campeão olímpico do triplo salto e vanessa fernandes prata no triatlo , conseguirão os maiores feitos , mas há ainda há assinalar , os sete diplomas olímpico conquistado , em cinco modalidades ,
D 2008	2: portuguesa esta tarde nélson	**2008_08_22-19_59_02-Telejornal-1 bloco 3**: ao mastro mais alto de pequim , enquanto sou portuguesa esta tarde nélson évora campeão olímpico do triplo salto , confessa , que se arrepiou , foi às vinte e uma horas e trinta e oito minutos que nelson évora hugo operado análise ,
D 2008	3: vanessa fernandes prata no triatlo	**2008_08_23-19_59_02-Telejornal-1 bloco 3**: nelson évora campeão olímpico do triplo salto e vanessa fernandes prata no triatlo , conseguirão os maiores feitos , mas há ainda há assinalar , os sete diplomas olímpico conquistado , em cinco modalidades ,
A 2010	1: imanente	**2008_08_23-19_59_02-Telejornal-1 bloco 3**: era possível mas cor imanente primeira muito foi nelson évora campeão olímpico do triplo salto .
A 2010	2: tivoli	**2008_08_23-21_59_01-Jornal2-2 bloco 1**: sem tivoli coisa não terminem dizer que o pódio era possível mas cor imanente primeira muito foi nelson évora campeão olímpico do triplo salto .

A 2010	3: mágicas	**2008_08_22-21_59_02-Jornal2-2 bloco 2**: depois ouviu as palavras mágicas campeão olímpico . ouro olímpico para o detentor do título mundial do triplo salto .
B 2010	1: sete centeno outros nélson	**2008_08_22-19_59_02-Telejornal-1 bloco 3**: com um salto de dezassete metros e sessenta e sete centeno outros nélson évora fez subir a bandeira nacional ao mastro mais alto de pequim e quando somou portuguesa esta tarde nélson évora campeão olímpico do triplo salto .
B 2010	2: olímpico	**2008_09_03-19_59_02-Telejornal-1 bloco 4**: o campeão olímpico do triplo salto
B 2010	3: mas cor imanente	**2008_08_23-19_59_02-Telejornal-1 bloco 3**: era possível mas cor imanente primeira muito foi nelson évora campeão olímpico do triplo salto .
C 2010	1: tivoli	**2008_08_23-21_59_01-Jornal2-2 bloco 1**: perante tudo aconteceu , sem tivoli coisa não terminem dizer que o pódio era possível mas cor imanente primeira muito foi nelson évora campeão olímpico do triplo salto , e vanessa fernandes prata no triatlo conseguirão os maiores feitos ,
C 2010	2: imanente	**2008_08_23-21_59_01-Jornal2-2 bloco 1**: perante tudo aconteceu , sem tivoli coisa não terminem dizer que o pódio era possível mas cor imanente primeira muito foi nelson évora campeão olímpico do triplo salto , e vanessa fernandes prata no triatlo conseguirão os maiores feitos ,
C 2010	3: rendido	**2008_08_22-21_59_02-Jornal2-2 bloco 2**: primeiro pódio os adversários que bateu na final , depois ouviu as palavras mágicas campeão olímpico , ouro olímpico para o detentor do título mundial do triplo salto , o estádio inteiro rendido à classe o atleta português , o reconhecimento , impressionante ,
D 2010	1: sete centeno outros nélson	**2008_08_22-19_59_02-Telejornal-1 bloco 3**: com um salto de dezassete metros e sessenta e sete centeno outros nélson évora fez subir a bandeira nacional ao mastro mais alto de pequim e quando somou portuguesa esta tarde nélson évora campeão olímpico do triplo salto , confessa que se arrepiou ,
D 2010	2: vanessa fernandes prata no triatlo	**2008_08_23-21_59_01-Jornal2-2 bloco 1**: perante tudo aconteceu , sem tivoli coisa não terminem dizer que o pódio era possível mas cor imanente primeira muito foi nelson évora campeão olímpico do triplo salto , e vanessa fernandes prata no triatlo conseguirão os maiores feitos ,
D 2010	3: olímpico	**2008_09_03-19_59_02-Telejornal-1 bloco 4**: ninguém responde e as que encontramos um quarto desarrumado onde dormiu , o campeão olímpico do triplo salto
Question #23 - Onde nasceu o Cirque du Soleil?		
A 2008	1: NIL	
B 2008	1: bellagio em las vegas	**pt/o/_/(/O_(Cirque_du_Soleil)_879d.html**: o (cirque du soleil) . o é um espetáculo do cirque du soleil . ele é apresentado permanentemente no bellagio em las vegas , nevada , estados unidos .

B 2008	2: alegría	**pt/c/i/r/Cirque_du_Soleil_fe99.html**: cirque du soleil . será realizada no brasil . alegría quidam dralion varekai corteo produções fixas / localização la nouba - walt disney world o - espetáculo subaquático em las vegas mystère - las vegas love com trilha sonora dos beatles delirium espetáculo musical - las vegas zumanity - las vegas ligações externas cirque du soleil
B 2008	3: saltimbanco	**pt/s/a/l/Saltimbanco_(Cirque_du_Soleil)_45f3.html**: saltimbanco (cirque du soleil) . saltimbanco (cirque du soleil) esta página precisa ser reciclada .
C 2008	1: NIL	
D 2008	1: bellagio em las vegas	**pt/o/_/(/O_(Cirque_du_Soleil)_879d.html**: o (cirque du soleil) . o é um espetáculo do cirque du soleil . ele é apresentado permanentemente no bellagio em las vegas , nevada , estados unidos .
D 2008	2: alegría	**pt/c/i/r/Cirque_du_Soleil_fe99.html**: cirque du soleil . será realizada no brasil . alegría quidam dralion varekai corteo produções fixas / localização la nouba - walt disney world o - espetáculo subaquático em las vegas mystère - las vegas love com trilha sonora dos beatles delirium espetáculo musical - las vegas zumanity - las vegas ligações externas cirque du soleil
D 2008	3: saltimbanco	**pt/s/a/l/Saltimbanco_(Cirque_du_Soleil)_45f3.html**: saltimbanco (cirque du soleil) . saltimbanco (cirque du soleil) esta página precisa ser reciclada .
A 2010	1: NIL	
B 2010	1: bellagio em las vegas	**pt/o/_/(/O_(Cirque_du_Soleil)_879d.html**: o (cirque du soleil) . o é um espetáculo do cirque du soleil . ele é apresentado permanentemente no bellagio em las vegas , nevada , estados unidos .
B 2010	2: alegría	**pt/c/i/r/Cirque_du_Soleil_fe99.html**: cirque du soleil . será realizada no brasil . alegría quidam dralion varekai corteo produções fixas / localização la nouba - walt disney world o - espetáculo subaquático em las vegas mystère - las vegas love com trilha sonora dos beatles delirium espetáculo musical - las vegas zumanity - las vegas ligações externas cirque du soleil
B 2010	3: saltimbanco	**pt/s/a/l/Saltimbanco_(Cirque_du_Soleil)_45f3.html**: saltimbanco (cirque du soleil) . saltimbanco (cirque du soleil) esta página precisa ser reciclada .
C 2010	1: NIL	
D 2010	1: bellagio em las vegas	**pt/o/_/(/O_(Cirque_du_Soleil)_879d.html**: o (cirque du soleil) . o é um espetáculo do cirque du soleil . ele é apresentado permanentemente no bellagio em las vegas , nevada , estados unidos .
D 2010	2: alegría	**pt/c/i/r/Cirque_du_Soleil_fe99.html**: cirque du soleil . será realizada no brasil . alegría quidam dralion varekai corteo produções fixas / localização la nouba - walt disney world o - espetáculo subaquático em las vegas mystère - las vegas love com trilha sonora dos beatles delirium espetáculo musical - las vegas zumanity - las vegas ligações externas cirque du soleil

D 2010	3: saltimbanco	**pt/s/a/l/Saltimbanco_(Cirque_du_Soleil)_45f3.html**: saltimbanco (cirque du soleil) . saltimbanco (cirque du soleil) esta página precisa ser reciclada .
Question #24 - O que é a Liga Sagres?		
A 2008	1: jornada	**2008_08_27-19_59_02-Telejornal-1 bloco 5**: o clássico réu principal de jogo da segunda jornada da liga sagres .
A 2008	2: clássico	**2008_08_27-19_59_02-Telejornal-1 bloco 5**: o clássico réu principal de jogo da segunda jornada da liga sagres .
A 2008	3: campeões	**2008_07_29-19_59_01-Telejornal-1 bloco 5**: por isso os árbitros pedem celeridade na resolução destes casos do futebol português . o vitória de guimarães ainda quer entrar directamente na liga dos campeões .
B 2008	1: réu principal de jogo da segunda jornada	**2008_08_27-19_59_02-Telejornal-1 bloco 5**: o clássico réu principal de jogo da segunda jornada da liga sagres .
B 2008	2: vamos ao futebol no jogo grande	**2008_08_30-21_59_01-Jornal2-2 bloco 1**: a terminar vamos ao futebol no jogo grande da segunda jornada da liga sagres .
B 2008	3: principal	**2008_08_27-19_59_02-Telejornal-1 bloco 5**: o clássico réu principal de jogo da segunda jornada da liga sagres .
C 2008	1: jornada	**2008_08_30-21_59_01-Jornal2-2 bloco 1**: se nada , de encerramento , a terminar vamos ao futebol no jogo grande da segunda jornada da liga sagres , o benfica recebeu o porto no estádio da luz ,
C 2008	2: estádio da luz	**2008_08_30-21_59_01-Jornal2-2 bloco 1**: se nada , de encerramento , a terminar vamos ao futebol no jogo grande da segunda jornada da liga sagres , o benfica recebeu o porto no estádio da luz ,
C 2008	3: clássico	**2008_08_27-19_59_02-Telejornal-1 bloco 5**: futebol clube do porto olhou benfica defrontam - se neste próximo fim - de - semana , o clássico réu principal de jogo da segunda jornada da liga sagres , na equipa campeã nacional mariano gonzález está em risco ,
D 2008	1: benfica recebeu o porto no estádio da luz	**2008_08_30-21_59_01-Jornal2-2 bloco 1**: se nada , de encerramento , a terminar vamos ao futebol no jogo grande da segunda jornada da liga sagres , o benfica recebeu o porto no estádio da luz ,
D 2008	2: réu principal de jogo	**2008_08_27-19_59_02-Telejornal-1 bloco 5**: futebol clube do porto olhou benfica defrontam - se neste próximo fim - de - semana , o clássico réu principal de jogo da segunda jornada da liga sagres , na equipa campeã nacional mariano gonzález está em risco ,

D 2008	3: grande da segunda jornada	**2008_08_30-21_59_01-Jornal2-2 bloco 1**: se nada , de encerramento , a terminar vamos ao futebol no jogo grande da segunda jornada da liga sagres , o benfica recebeu o porto no estádio da luz ,
A 2010	1: patrocínio de no continente	**2008_08_24-21_09_01-Telejornal-1 bloco 2**: r a liga sagres tem o patrocínio de no continente .
A 2010	2: jornada	**2008_08_27-19_59_02-Telejornal-1 bloco 5**: o clássico é o principal de jogo da segunda jornada da liga sagres .
A 2010	3: clássico	**2008_08_27-19_59_02-Telejornal-1 bloco 5**: o clássico é o principal de jogo da segunda jornada da liga sagres .
B 2010	1: principal de jogo da segunda jornada	**2008_08_27-19_59_02-Telejornal-1 bloco 5**: o clássico é o principal de jogo da segunda jornada da liga sagres .
B 2010	2: parte do telejornal onde vamos olhar também para a estreia	**2008_08_22-19_59_02-Telejornal-1 bloco 3**: fomos bater na segunda parte do telejornal onde vamos olhar também para a estreia da liga sagres .
B 2010	3: vamos ao futebol no jogo grande	**2008_08_30-21_59_01-Jornal2-2 bloco 1**: a terminar vamos ao futebol no jogo grande da segunda jornada da liga sagres .
C 2010	1: patrocínio de no continente	**2008_08_24-21_09_01-Telejornal-1 bloco 2**: r a liga sagres tem o patrocínio de no continente , a direita a noruega e a sua vida ,
C 2010	2: estádio da luz	**2008_08_30-21_59_01-Jornal2-2 bloco 1**: amanhã à tarde um desvio com todos os participantes , pressionada de , encerramento , a terminar vamos ao futebol no jogo grande da segunda jornada da liga sagres , o benfica recebeu porto no estádio da luz , o jogo está nos minutos de desconto resultado está num impacte de um a um ,
C 2010	3: direita a noruega	**2008_08_24-21_09_01-Telejornal-1 bloco 2**: r a liga sagres tem o patrocínio de no continente , a direita a noruega e a sua vida ,
D 2010	1: campeã nacional mariano gonzalez está em risco	**2008_08_27-19_59_02-Telejornal-1 bloco 5**: o clássico é o principal de jogo da segunda jornada da liga sagres , na equipa campeã nacional mariano gonzalez está em risco ,
D 2010	2: principal de jogo da segunda jornada	**2008_08_27-19_59_02-Telejornal-1 bloco 5**: o clássico é o principal de jogo da segunda jornada da liga sagres , na equipa campeã nacional mariano gonzalez está em risco ,
D 2010	3: campeonato nacional de futebol	**2008_08_22-19_59_02-Telejornal-1 bloco 3**: com também tinham hiv , fomos bater na segunda parte do telejornal onde vamos olhar também para a estreia da liga sagres , o campeonato nacional de futebol que está de regresso à rtp , tejo ,

Question #25 - Qual o montante que o Benfica vai receber pela venda do lateral Nelson ao Sevilha?		
A 2008 B 2008 C 2008 D 2008 A 2010 B 2010 C 2010 D 2010	1: NIL	
Question #26 - Onde ocorreram confrontos entre pescadores e comerciantes?		
A 2008	1: lota	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: jaime silva já considerou inaceitáveis os confrontos que se registaram ontem , na lota de matosinhos , entre pescadores e comerciantes , e avisou que o diálogo entre o governo e os pescadores não é incondicional .
A 2008	2: incondicional	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: jaime silva já considerou inaceitáveis os confrontos que se registaram ontem , na lota de matosinhos , entre pescadores e comerciantes , e avisou que o diálogo entre o governo e os pescadores não é incondicional .
B 2008	1: lota de matosinhos	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: jaime silva já considerou inaceitáveis os confrontos que se registaram ontem , na lota de matosinhos , entre pescadores e comerciantes , e avisou que o diálogo entre o governo e os pescadores não é incondicional .
B 2008	2: governo	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: jaime silva já considerou inaceitáveis os confrontos que se registaram ontem , na lota de matosinhos , entre pescadores e comerciantes , e avisou que o diálogo entre o governo e os pescadores não é incondicional .
B 2008	3: silva	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: jaime silva já considerou inaceitáveis os confrontos que se registaram ontem , na lota de matosinhos , entre pescadores e comerciantes , e avisou que o diálogo entre o governo e os pescadores não é incondicional .
C 2008	1: lota	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: jaime silva já considerou inaceitáveis os confrontos que se registaram ontem , na lota de matosinhos , entre pescadores e comerciantes , e avisou que o diálogo entre o governo e os pescadores não é incondicional , ele é ministro tinham encontro mas apenas com caçadores ,

C 2008	2: caçadores	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: jaime silva já considerou inaceitáveis os confrontos que se registaram ontem , na lota de matosinhos , entre pescadores e comerciantes , e avisou que o diálogo entre o governo e os pescadores não é incondicional , ele é ministro tinham encontro mas apenas com caçadores ,
C 2008	3: frustrada	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: confrontos que se registaram ontem , na lota de matosinhos , entre pescadores e comerciantes , e avisou que o diálogo entre o governo e os pescadores não é incondicional , ele é ministro tinham encontro mas apenas com caçadores , a expectativa dos pescadores saiu por isso , frustrada ,
D 2008	1: lota de matosinhos	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: jaime silva já considerou inaceitáveis os confrontos que se registaram ontem , na lota de matosinhos , entre pescadores e comerciantes , e avisou que o diálogo entre o governo e os pescadores não é incondicional , ele é ministro tinham encontro mas apenas com caçadores ,
D 2008	2: governo	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: jaime silva já considerou inaceitáveis os confrontos que se registaram ontem , na lota de matosinhos , entre pescadores e comerciantes , e avisou que o diálogo entre o governo e os pescadores não é incondicional , ele é ministro tinham encontro mas apenas com caçadores ,
D 2008	3: ministro	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: jaime silva já considerou inaceitáveis os confrontos que se registaram ontem , na lota de matosinhos , entre pescadores e comerciantes , e avisou que o diálogo entre o governo e os pescadores não é incondicional , ele é ministro tinham encontro mas apenas com caçadores ,
A 2010	1: incondicional	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: confrontos que se registaram ontem . na lota de matosinhos entre pescadores e comerciantes e avisou que o diálogo entre o governo e os pescadores não é incondicional .
A 2010	2: lota	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: confrontos que se registaram ontem . na lota de matosinhos entre pescadores e comerciantes e avisou que o diálogo entre o governo e os pescadores não é incondicional .
B 2010	1: lota de matosinhos	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: confrontos que se registaram ontem . na lota de matosinhos entre pescadores e comerciantes e avisou que o diálogo entre o governo e os pescadores não é incondicional .
B 2010	2: governo	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: confrontos que se registaram ontem . na lota de matosinhos entre pescadores e comerciantes e avisou que o diálogo entre o governo e os pescadores não é incondicional .
B 2010	3: ministro	**2008_06_01-21_59_02-Jornal2-2 bloco 1**: ministro da agricultura e pescas diz que são inaceitáveis os confrontos que se registaram ontem na lota de matosinhos entre pescadores e comerciantes jaime silva avisa que o diálogo entre o governo e os pescadores , não é incondicional .

C 2010	1: incondicional	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: confrontos que se registaram ontem , na lota de matosinhos entre pescadores e comerciantes e avisou que o diálogo entre o governo e os pescadores não é incondicional , é loulé ministro tinham encontro ,
C 2010	2: lota	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: confrontos que se registaram ontem , na lota de matosinhos entre pescadores e comerciantes e avisou que o diálogo entre o governo e os pescadores não é incondicional , é loulé ministro tinham encontro ,
D 2010	1: lota de matosinhos	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: confrontos que se registaram ontem , na lota de matosinhos entre pescadores e comerciantes e avisou que o diálogo entre o governo e os pescadores não é incondicional , é loulé ministro tinham encontro ,
D 2010	2: governo	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: confrontos que se registaram ontem , na lota de matosinhos entre pescadores e comerciantes e avisou que o diálogo entre o governo e os pescadores não é incondicional , é loulé ministro tinham encontro ,
D 2010	3: loulé ministro	**2008_06_01-19_59_01-Telejornal-1 bloco 2**: confrontos que se registaram ontem , na lota de matosinhos entre pescadores e comerciantes e avisou que o diálogo entre o governo e os pescadores não é incondicional , é loulé ministro tinham encontro ,
Question #27 - Quem espiou a favor do Hezbollah?		
A 2008	1: NIL	
B 2008	1: acusado	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: em dois mil e dois foi acusado em gelo espiar a favor do hezbollah .
C 2008	1: NIL	
D 2008	1: acusado	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: medicina ser tinha à sua espera uma recepção de herói , em dois mil e dois foi acusado em gelo espiar a favor do hezbollah , perdeu a cidadania israelita que tinha graças a sua mãe , uma judia casada com um muçulmano libanês ,
D 2008	2: graças a sua mãe	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: medicina ser tinha à sua espera uma recepção de herói , em dois mil e dois foi acusado em gelo espiar a favor do hezbollah , perdeu a cidadania israelita que tinha graças a sua mãe , uma judia casada com um muçulmano libanês ,
D 2008	3: judeus	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: israel de portou para o líbano eis o cidadão seu que em dois mil e dois foi acusado de espiar para o hezbollah , em troca , o grupo shihita devolveu os restos mortais de soldados judeus mortos em dois mil e seis , na guerra com israel ,
A 2010	1: NIL	
B 2010	1: acusado	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: em dois mil e dois foi acusado em gelo espiar a favor do hezbollah .

C 2010	1: NIL	
D 2010	1: acusado	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: medicina ser tinha à sua espera uma recepção de herói , em dois mil e dois foi acusado em gelo espiar a favor do hezbollah , perdeu a cidadania israelita que tinha graças a sua mãe , uma judia casada com um muçulmano libanês ,
D 2010	2: graças a sua mãe	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: medicina ser tinha à sua espera uma recepção de herói , em dois mil e dois foi acusado em gelo espiar a favor do hezbollah , perdeu a cidadania israelita que tinha graças a sua mãe , uma judia casada com um muçulmano libanês ,
D 2010	3: judeus	**2008_06_01-21_59_02-Jornal2-2 bloco 2**: israel de portou para o líbano ex cidadão seu que em dois mil e dois foi acusado de espiar para o hezbollah , em troca o grupo shihita devolveu os restos mortais de soldados judeus mortos em dois mil e seis , na guerra com israel ,
Question #28 - Qual o comprimento do salto de Nelson Évora em Pequim?		
A 2008 B 2008 C 2008 D 2008 A 2010 B 2010 C 2010 D 2010	1: NIL	
Question #29 - Quantos dólares custa o barril de petróleo em Nova Iorque?		
A 2008	1: NIL	
B 2008	1: três dólares	**2008_07_24-19_59_02-Telejornal-1 bloco 2**: em nova iorque o barril de light baixou , para os cento e vinte e três dólares e meio .
B 2008	2: sete dólares	**2008_09_04-21_59_02-Jornal2-2 bloco 1**: em nova iorque o barril de light negociou ligeiramente acima dos cento e sete dólares .
B 2008	3: nove dólares	**2008_09_09-19_59_01-Telejornal-1 bloco 6**: num mínimo de noventa e nove dólares e trinta cêntimos por barril . já light crude transaccionado nova iorque .
C 2008	1: NIL	
D 2008	1: três dólares	**2008_07_24-19_59_02-Telejornal-1 bloco 2**: em nova iorque o barril de light baixou , para os cento e vinte e três dólares e meio , em londres o barril
D 2008	2: sete dólares	**2008_06_12-19_59_01-Telejornal-1 bloco 1**: os combustíveis em alta num dia em que o petróleo nos mercados de nova iorque e londres , quase como os cento e trinta e sete dólares , o barril , combustíveis mais caros mas a correr ,

D 2008	3: mais de quatro dólares	**2008_09_01-21_59_01-Jornal2-2 bloco 2**: não irá afectar a produção do golfo do méxico , nova iorque o barril de light perdeu mais de quatro dólares , e meio e negociou abaixo dos cento e onze dólares em londres ,
A 2010	1: NIL	
B 2010	1: três dólares	**2008_07_24-19_59_02-Telejornal-1 bloco 2**: em nova iorque o barril de light baixou para os cento e vinte e três dólares e meio .
B 2010	2: mais de três dólares	**2008_06_30-21_59_02-Jornal2-2 bloco 2**: em nova iorque o barril de light subiu mais de três dólares e atingiu um novo máximo histórico .
B 2010	3: sete dólares	**2008_09_04-21_59_02-Jornal2-2 bloco 1**: em nova iorque o barril de light negociou ligeiramente acima dos cento e sete dólares .
C 2010	1: NIL	
D 2010	1: três dólares	**2008_07_24-19_59_02-Telejornal-1 bloco 2**: em nova iorque o barril de light baixou para os cento e vinte e três dólares e meio , em londres o barril
D 2010	2: sete dólares	**2008_06_06-19_59_01-Telejornal-1 bloco 1**: nova iorque o barril de light subiu mais de nove dólares e atingiu novo máximo histórico a negociar nos cento e trinta e sete dólares e setenta cêntimos , em londres o barril
D 2010	3: mais de três dólares	**2008_08_19-21_59_02-Jornal2-2 bloco 1**: os preços do petróleo voltaram a subir nos mercados internacionais em nova iorque o barril de light ganho mais de três dólares , negociou próximo dos cento e dezasseis dólares ,
Question #30 - Quantos dólares custa em Londres?		
A 2008	1: NIL	
B 2008	1: quatro dólares	**2008_06_30-19_59_01-Telejornal-1 bloco 1**: em londres o barril de brent , subiu também para um novo máximo a rondar os cento e quarenta e quatro dólares .
B 2008	2: três dólares	**2008_08_19-19_59_02-Telejornal-1 bloco 4**: em londres o brent de referência para o mercado português . seguiu também quase três dólares para perto dos cento e quinze dólares por barril .
B 2008	3: sete dólares	**2008_09_11-19_59_02-Telejornal-1 bloco 2**: o crude em londres chegou hoje a cotar abaixo dos noventa e sete dólares por barril .
C 2008	1: NIL	
D 2008	1: sete dólares	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: nos cento e quarenta e cinco dólares e oitenta e cinco cêntimos , em londres o barril de brent atingiu também novo recorde , em negociar perto dos cento e quarenta e sete dólares ,
D 2008	2: quatro dólares	**2008_06_19-19_59_01-Telejornal-1 bloco 3**: barril de light que recuou cerca de três dólares , e negociou abaixo dos cento e trinta e quatro dólares , em londres o brent perdeu cerca de dois dólares e meio ,

D 2008	3: três dólares	**2008_07_24-19_59_02-Telejornal-1 bloco 2**: em nova iorque o barril de light baixou , para os cento e vinte e três dólares e meio , em londres o barril de brent de referência para o mercado português caiu para os cento e vinte e quatro dólares ,
A 2010	1: NIL	
B 2010	1: quatro dólares	**2008_07_02-21_59_02-Jornal2-2 bloco 1**: londres o barril de brent atingiu mesmo um novo máximo histórico a negociar acima dos cento e quarenta e quatro dólares .
B 2010	2: três dólares	**2008_07_24-19_59_02-Telejornal-1 bloco 2**: em nova iorque o barril de light baixou para os cento e vinte e três dólares e meio . em londres
B 2010	3: sete dólares	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: em londres o barril de brent atingiu também novo record a negociar perto dos cento e quarenta e sete dólares .
C 2010	1: NIL	
D 2010	1: sete dólares	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: em londres o barril de brent atingiu também novo record a negociar perto dos cento e quarenta e sete dólares ,
D 2010	2: quatro dólares	**2008_06_19-19_59_01-Telejornal-1 bloco 3**: barril de light que recuou cerca de três dólares e negociou abaixo dos cento e trinta e quatro dólares , em londres
D 2010	3: três dólares	**2008_07_24-19_59_02-Telejornal-1 bloco 2**: em nova iorque o barril de light baixou para os cento e vinte e três dólares e meio , em londres
Question #31 - Quando é que Paulo Rangel se dirigiu ao Parlamento pela primeira vez como Líder da Bancada Laranja?		
A 2008 B 2008 C 2008 D 2008 A 2010 B 2010 C 2010 D 2010	1: NIL	
Question #32 - Qual o programa que criticou na sua intervenção?		
A 2008	1: palavras de manuela	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: a primeira intervenção como líder parlamentar do psd . paulo rangel disse mesmo que sócrates desrespeitou o parlamento . antecipando o debate da nação . dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja , para repetir palavras de manuela ferreira leite . paulo rangel , criticou o programa

A 2008	2: nação	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: a primeira intervenção como líder parlamentar do psd . paulo rangel disse mesmo que sócrates desrespeitou o parlamento . antecipando o debate da nação . dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja , para repetir palavras de manuela ferreira leite . paulo rangel , criticou o programa
A 2008	3: desrespeitou	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: a primeira intervenção como líder parlamentar do psd . paulo rangel disse mesmo que sócrates desrespeitou o parlamento . antecipando o debate da nação . dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja , para repetir palavras de manuela ferreira leite . paulo rangel , criticou o programa
B 2008	1: palavras de manuela ferreira leite	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: a primeira intervenção como líder parlamentar do psd . paulo rangel disse mesmo que sócrates desrespeitou o parlamento . antecipando o debate da nação . dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja , para repetir palavras de manuela ferreira leite . paulo rangel , criticou o programa
B 2008	2: palavras de manuela ferreira leite	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: a primeira intervenção como líder parlamentar do psd . paulo rangel disse mesmo que sócrates desrespeitou o parlamento . antecipando o debate da nação . dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja , para repetir palavras de manuela ferreira leite . paulo rangel , criticou o programa
B 2008	3: sócrates desrespeitou	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: a primeira intervenção como líder parlamentar do psd . paulo rangel disse mesmo que sócrates desrespeitou o parlamento . antecipando o debate da nação . dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja , para repetir palavras de manuela ferreira leite . paulo rangel , criticou o programa
C 2008	1: palavras de manuela	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: a primeira intervenção como líder parlamentar do psd , paulo rangel disse mesmo que sócrates desrespeitou o parlamento , antecipando o debate da nação , dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja , para repetir palavras de manuela ferreira leite , paulo rangel , criticou o programa
C 2008	2: nação	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: a primeira intervenção como líder parlamentar do psd , paulo rangel disse mesmo que sócrates desrespeitou o parlamento , antecipando o debate da nação , dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja , para repetir palavras de manuela ferreira leite , paulo rangel , criticou o programa

C 2008	3: desrespeitou	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: a primeira intervenção como líder parlamentar do psd , paulo rangel disse mesmo que sócrates desrespeitou o parlamento , antecipando o debate da nação , dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja , para repetir palavras de manuela ferreira leite , paulo rangel , criticou o programa
D 2008	1: palavras de manuela ferreira leite	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: a primeira intervenção como líder parlamentar do psd , paulo rangel disse mesmo que sócrates desrespeitou o parlamento , antecipando o debate da nação , dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja , para repetir palavras de manuela ferreira leite , paulo rangel , criticou o programa
D 2008	2: sócrates desrespeitou	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: a primeira intervenção como líder parlamentar do psd , paulo rangel disse mesmo que sócrates desrespeitou o parlamento , antecipando o debate da nação , dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja , para repetir palavras de manuela ferreira leite , paulo rangel , criticou o programa
D 2008	3: debate da nação	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: a primeira intervenção como líder parlamentar do psd , paulo rangel disse mesmo que sócrates desrespeitou o parlamento , antecipando o debate da nação , dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja , para repetir palavras de manuela ferreira leite , paulo rangel , criticou o programa
A 2010	1: palavras de manuela	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: na primeira intervenção como líder parlamentar do psd paulo rangel disse mesmo que sócrates desrespeitou o parlamento . antecipando o debate da nação . dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja para repetir palavras de manuela ferreira leite . paulo rangel . criticou o programa
A 2010	2: nação	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: na primeira intervenção como líder parlamentar do psd paulo rangel disse mesmo que sócrates desrespeitou o parlamento . antecipando o debate da nação . dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja para repetir palavras de manuela ferreira leite . paulo rangel . criticou o programa
A 2010	3: desrespeitou	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: na primeira intervenção como líder parlamentar do psd paulo rangel disse mesmo que sócrates desrespeitou o parlamento . antecipando o debate da nação . dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja para repetir palavras de manuela ferreira leite . paulo rangel . criticou o programa

B 2010	1: palavras de manuela ferreira leite	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: na primeira intervenção como líder parlamentar do psd paulo rangel disse mesmo que sócrates desrespeitou o parlamento . antecipando o debate da nação . dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja para repetir palavras de manuela ferreira leite . paulo rangel . criticou o programa
B 2010	2: sócrates desrespeitou	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: na primeira intervenção como líder parlamentar do psd paulo rangel disse mesmo que sócrates desrespeitou o parlamento . antecipando o debate da nação . dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja para repetir palavras de manuela ferreira leite . paulo rangel . criticou o programa
B 2010	3: debate da nação	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: na primeira intervenção como líder parlamentar do psd paulo rangel disse mesmo que sócrates desrespeitou o parlamento . antecipando o debate da nação . dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja para repetir palavras de manuela ferreira leite . paulo rangel . criticou o programa
C 2010	1: palavras de manuela	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: na primeira intervenção como líder parlamentar do psd paulo rangel disse mesmo que sócrates desrespeitou o parlamento , antecipando o debate da nação , dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja para repetir palavras de manuela ferreira leite , paulo rangel , criticou o programa
C 2010	2: nação	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: na primeira intervenção como líder parlamentar do psd paulo rangel disse mesmo que sócrates desrespeitou o parlamento , antecipando o debate da nação , dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja para repetir palavras de manuela ferreira leite , paulo rangel , criticou o programa
C 2010	3: desrespeitou	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: na primeira intervenção como líder parlamentar do psd paulo rangel disse mesmo que sócrates desrespeitou o parlamento , antecipando o debate da nação , dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja para repetir palavras de manuela ferreira leite , paulo rangel , criticou o programa
D 2010	1: palavras de manuela ferreira leite	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: na primeira intervenção como líder parlamentar do psd paulo rangel disse mesmo que sócrates desrespeitou o parlamento , antecipando o debate da nação , dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja para repetir palavras de manuela ferreira leite , paulo rangel , criticou o programa

D 2010	2: sócrates desrespeitou	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: na primeira intervenção como líder parlamentar do psd paulo rangel disse mesmo que sócrates desrespeitou o parlamento , antecipando o debate da nação , dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja para repetir palavras de manuela ferreira leite , paulo rangel , criticou o programa
D 2010	3: debate da nação	**2008_07_03-21_59_01-Jornal2-2 bloco 2**: na primeira intervenção como líder parlamentar do psd paulo rangel disse mesmo que sócrates desrespeitou o parlamento , antecipando o debate da nação , dirigiu - se ao parlamento pela primeira vez como líder da bancada laranja para repetir palavras de manuela ferreira leite , paulo rangel , criticou o programa
Question #33 - Quem é Rosa Passos?		
A 2008	1: começou	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: rosa passos têm outros trabalhos começou muito nova .
A 2008	2: perth	**2008_06_10-21_59_01-Jornal2-2 bloco 2**: michael gore e rosa veloso a rtp madrid . em frança os camionistas também estão em protesto contra o aumento dos combustíveis . na cidade de perth isso junto à fronteira espanhola foram montados bloqueios de estrada . os piquetes permitiram a passagem de veículos ligeiros ,
A 2008	3: gesta	**2008_06_09-19_59_01-Telejornal-1 bloco 3**: excelente a mais consagrada rosa portugal . la sobre a. civilização pela sua cultura . e sobretudo pela sua grande simplicidade . a gesta prémio esta distinção como também uma decisão , a toda a rtp internacional , obviamente individualmente só muito feliz ,
B 2008	1: centro cultural de belém joana	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: rosa passos actua sexta - feira no centro cultural de belém joana para a apresentação do seu mais recente álbum romance .
B 2008	2: frank sinatra	**pt/b/o/l/Bolero.html**: bolero . e interpretado , entre outros , por : beatles , plácido domingo , diane krall , joão gilberto , cesária évora , rosa passos , frank sinatra , etc ...
B 2008	3: edu ribeiro	**pt/e/d/u/Edu_Ribeiro_0afc.html**: edu ribeiro . onde teve a oportunidade de trabalhar com vários artistas , entre eles yamandú costa , chico pinheiro , johny alf , rosa passos , arismar do espírito santo , joyce , dori caymmi , léa freire , bocato , hamilton de hollanda , toquinho , paulo moura , dominguinhos ,
C 2008	1: o benfica	**2008_07_15-19_59_01-Telejornal-1 bloco 4**: rosa , o benfica e o vitória de guimarães criticaram a atitude da uefa neste recurso para o tribunal arbitral do desporto , os dois clubes reagiram para já através da internet e prometem novas tomadas de posição , para quando conhecerem os fundamentos desta decisão ,

C 2008	2: perth	**2008_06_10-21_59_01-Jornal2-2 bloco 2**: michael gore e rosa veloso a rtp madrid , em frança os camionistas também estão em protesto contra o aumento dos combustíveis , na cidade de perth isso junto à fronteira espanhola foram montados bloqueios de estrada , os piquetes permitiram a passagem de veículos ligeiros ,
C 2008	3: gesta	**2008_06_09-19_59_01-Telejornal-1 bloco 3**: excelente a mais consagrada rosa portugal , la sobre a , civilização pela sua cultura , e sobretudo pela sua grande simplicidade , a gesta prémio esta distinção como também uma decisão , a toda a rtp internacional , obviamente individualmente só muito feliz ,
D 2008	1: centro cultural de belém joana	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: outros convites profissionais , socialmente , muito , a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira , rosa passos actua sexta - feira no centro cultural de belém joana para a apresentação do seu mais recente álbum romance ,
D 2008	2: novamente em portugal para uma digressão	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: são várias as razões apontadas para esta alegada demissão vão de questões ou seja orçamentais , outros convites profissionais , socialmente , muito , a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira ,
D 2008	3: cesária évora	**pt/b/o/l/Bolero.html**: bolero . e interpretado , entre outros , por : beatles , plácido domingo , diane krall , joão gilberto , cesária évora , rosa passos , frank sinatra , etc ...
A 2010	1: começou	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: rosa passos têm outros trabalhos começou muito nova ,
A 2010	2: na rosa ganha	**2008_06_06-19_59_01-Telejornal-1 bloco 5**: espanha e na rosa ganha . passada a rainha .
A 2010	3: rainha	**2008_06_06-19_59_01-Telejornal-1 bloco 5**: espanha e na rosa ganha . passada a rainha .
B 2010	1: cesária évora	**pt/b/o/l/Bolero.html**: bolero . e interpretado , entre outros , por : beatles , plácido domingo , diane krall , joão gilberto , cesária évora , rosa passos , frank sinatra , etc ...
B 2010	2: edu ribeiro	**pt/e/d/u/Edu_Ribeiro_0afc.html**: edu ribeiro . onde teve a oportunidade de trabalhar com vários artistas , entre eles yamandú costa , chico pinheiro , johny alf , rosa passos , arismar do espírito santo , joyce , dori caymmi , léa freire , bocato , hamilton de hollanda , toquinho , paulo moura , dominguinhos ,

B 2010	3: frank sinatra	**pt/b/o/l/Bolero.html**: bolero . e interpretado , entre outros , por : beatles , plácido domingo , diane krall , joão gilberto , cesária évora , rosa passos , frank sinatra , etc ...
C 2010	1: a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: muito , a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira , rosa passe actua sexta - feira no centro cultural de belém lisboa , para apresentação do seu mais recente álbum romance ,
C 2010	2: perth	**2008_06_10-21_59_01-Jornal2-2 bloco 2**: michael gore e rosa veloso rtp madrid , em frança os camionistas também estão em protesto contra o aumento dos combustíveis , a cidade de perth isso junto à fronteira espanhola foram montados bloqueios de estrada , os piquetes permitiram a passagem de veículos ligeiros ,
C 2010	3: vinculação posat	**2008_06_06-19_59_01-Telejornal-1 bloco 5**: com as falhas na real de os torres aura deu - se na saúde , grave é intermédia para vinculação posat , espanha e na rosa ganha , passada a rainha , foi sempre avesso agora ser hasta dão trabalho a guerra ,
D 2010	1: joão gilberto	**pt/b/o/l/Bolero.html**: bolero . e interpretado , entre outros , por : beatles , plácido domingo , diane krall , joão gilberto , cesária évora , rosa passos , frank sinatra , etc ...
D 2010	2: cesária évora	**pt/b/o/l/Bolero.html**: bolero . e interpretado , entre outros , por : beatles , plácido domingo , diane krall , joão gilberto , cesária évora , rosa passos , frank sinatra , etc ...
D 2010	3: plácido domingo	**pt/b/o/l/Bolero.html**: bolero . e interpretado , entre outros , por : beatles , plácido domingo , diane krall , joão gilberto , cesária évora , rosa passos , frank sinatra , etc ...
Question #34 - Quando nasceu?		
A 2008	1: vinte e um	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: de um de fim de um fundo de vinte e um de rosa passos é a minha convidado no jornal dois muito boa noite no belíssimo romance cem anos .
A 2008	2: sexta	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: rosa passos actua sexta - feira no centro cultural de belém joana para a apresentação do seu mais recente álbum romance . esta digressão vai passar do funchal aveiro caldas da rainha e coimbra . de onde viaja até espanha para actuar na carreira na no festival musicas de é

B 2008	1: vinte e um	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: de um de fim de um fundo de vinte e um de rosa passos é a minha convidado no jornal dois muito boa noite no belíssimo romance cem anos .
B 2008	2: 1 941	**pt/b/o/l/Bolero.html**: bolero . célebre bolero mexicano o mais célebre bolero mexicano é sem dúvida besame mucho , composto por consuelo velásquez em 1 941 , e interpretado , entre outros , por : beatles , plácido domingo , diane krall , joão gilberto , cesária évora , rosa passos , frank sinatra , etc ...
B 2008	3: 1 995	**pt/e/d/u/Edu_Ribeiro_0afc.html**: edu ribeiro . foi um dos selecionados para o projeto de música instrumental o som da demo , promovido pelo sesc / sp em 1 995 . em 1 996 mudou - se para são paulo , onde teve a oportunidade de trabalhar com vários artistas , entre eles yamandú costa , chico pinheiro , johny alf , rosa passos
C 2008	1: vinte e um	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: de um de fim de um fundo de vinte e um de rosa passos é a minha convidado no jornal dois muito boa noite no belíssimo romance cem anos , ou que o romance , é este , o uso que a declaração ,
C 2008	2: um , um	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: de angola não obterão um , um , perfume es um a um , de um de fim de um fundo de vinte e um de rosa passos é a minha convidado no jornal dois muito boa noite no belíssimo romance cem anos , ou que o romance , é este ,
C 2008	3: sexta	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: são várias as razões apontadas para esta alegada demissão vão de questões ou seja orçamentais , outros convites profissionais , socialmente , muito , a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira ,
D 2008	1: vinte e um	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: de um de fim de um fundo de vinte e um de rosa passos é a minha convidado no jornal dois muito boa noite no belíssimo romance cem anos , ou que o romance , é este , o uso que a declaração ,
D 2008	2: sexta	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: são várias as razões apontadas para esta alegada demissão vão de questões ou seja orçamentais , outros convites profissionais , socialmente , muito , a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira ,
D 2008	3: um , um	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: de angola não obterão um , um , perfume es um a um , de um de fim de um fundo de vinte e um de rosa passos é a minha convidado no jornal dois muito boa noite no belíssimo romance cem anos , ou que o romance , é este ,

A 2010	1: sexta	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: são várias as razões apontadas para esta alegada demissão vão questões úlcera orçamentais outros convites profissionais socialmente . muito . a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira .
B 2010	1: 1 941	**pt/b/o/l/Bolero.html**: bolero . célebre bolero mexicano o mais célebre bolero mexicano é sem dúvida besame mucho , composto por consuelo velásquez em 1 941 , e interpretado , entre outros , por : beatles , plácido domingo , diane krall , joão gilberto , cesária évora , rosa passos , frank sinatra , etc ...
B 2010	2: 1 995	**pt/e/d/u/Edu_Ribeiro_0afc.html**: edu ribeiro . foi um dos selecionados para o projeto de música instrumental o som da demo , promovido pelo sesc / sp em 1 995 . em 1 996 mudou - se para são paulo , onde teve a oportunidade de trabalhar com vários artistas , entre eles yamandú costa , chico pinheiro , johny alf , rosa passos
B 2010	3: sexta	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: são várias as razões apontadas para esta alegada demissão vão questões úlcera orçamentais outros convites profissionais socialmente . muito . a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira .
C 2010	1: sexta	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: nem diogo infante na câmara de lisboa falam sobre o assunto , são várias as razões apontadas para esta alegada demissão vão questões úlcera orçamentais outros convites profissionais socialmente , muito , a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira ,
D 2010	1: 1 941	**pt/b/o/l/Bolero.html**: bolero . célebre bolero mexicano o mais célebre bolero mexicano é sem dúvida besame mucho , composto por consuelo velásquez em 1 941 , e interpretado , entre outros , por : beatles , plácido domingo , diane krall , joão gilberto , cesária évora , rosa passos , frank sinatra , etc ...
D 2010	2: 1 995	**pt/e/d/u/Edu_Ribeiro_0afc.html**: edu ribeiro . foi um dos selecionados para o projeto de música instrumental o som da demo , promovido pelo sesc / sp em 1 995 . em 1 996 mudou - se para são paulo , onde teve a oportunidade de trabalhar com vários artistas , entre eles yamandú costa , chico pinheiro , johny alf , rosa passos
D 2010	3: sexta	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: nem diogo infante na câmara de lisboa falam sobre o assunto , são várias as razões apontadas para esta alegada demissão vão questões úlcera orçamentais outros convites profissionais socialmente , muito , a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira ,

Question #35 - Diga o nome de um álbum de Rosa Passos.		
A 2008	1: cultural de belém joana	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: rosa passos actua sexta - feira no centro cultural de belém joana para a apresentação do seu mais recente álbum romance .
A 2008	2: cinema de rodrigo leão	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: no álbum cinema de rodrigo leão . e no domingo um , um . de um a um . de angola não obterão um . um , perfume es um a um . de um de fim de um fundo de vinte e um de rosa passos
A 2008	3: angola	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: no álbum cinema de rodrigo leão . e no domingo um , um . de um a um . de angola não obterão um . um , perfume es um a um . de um de fim de um fundo de vinte e um de rosa passos
B 2008	1: feira no centro cultural de belém joana	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: rosa passos actua sexta - feira no centro cultural de belém joana para a apresentação do seu mais recente álbum romance .
B 2008	2: cinema de rodrigo leão	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: no álbum cinema de rodrigo leão . e no domingo um , um . de um a um . de angola não obterão um . um , perfume es um a um . de um de fim de um fundo de vinte e um de rosa passos
B 2008	3: romance	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: rosa passos actua sexta - feira no centro cultural de belém joana para a apresentação do seu mais recente álbum romance .
C 2008	1: cultural de belém joana	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: rosa passos actua sexta - feira no centro cultural de belém joana para a apresentação do seu mais recente álbum romance , esta digressão vai passar do funchal aveiro caldas da rainha e coimbra ,
C 2008	2: cinema de rodrigo leão	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: no álbum cinema de rodrigo leão , e no domingo um , um , de um a um , de angola não obterão um , um , perfume es um a um , de um de fim de um fundo de vinte e um de rosa passos
C 2008	3: angola	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: no álbum cinema de rodrigo leão , e no domingo um , um , de um a um , de angola não obterão um , um , perfume es um a um , de um de fim de um fundo de vinte e um de rosa passos
D 2008	1: esta digressão vai passar do funchal aveiro caldas da rainha	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: rosa passos actua sexta - feira no centro cultural de belém joana para a apresentação do seu mais recente álbum romance , esta digressão vai passar do funchal aveiro caldas da rainha e coimbra ,
D 2008	2: feira no centro cultural de belém joana	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: rosa passos actua sexta - feira no centro cultural de belém joana para a apresentação do seu mais recente álbum romance , esta digressão vai passar do funchal aveiro caldas da rainha e coimbra ,

D 2008	3: aveiro caldas da rainha	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: rosa passos actua sexta - feira no centro cultural de belém joana para a apresentação do seu mais recente álbum romance , esta digressão vai passar do funchal aveiro caldas da rainha e coimbra ,
A 2010	1: cantora brasileira	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira .
A 2010	2: começar	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira .
A 2010	3: começou	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: rosa passos têm outros trabalhos começou muito nova ,
B 2010	1: joão donato	**pt/j/o/ã/João_Donato_b958.html**: joão donato . rosa passos
B 2010	2: outros trabalhos começou muito nova	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: rosa passos têm outros trabalhos começou muito nova ,
B 2010	3: portugal para uma digressão a começar esta	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira .
C 2010	1: peixes um tó rocha	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: é todo momento maduro do trabalho monte um bom momento como a com peixes um tó rocha aqui é esse era muito soft , rosa passos têm outros trabalhos começou muito nova ,
C 2010	2: cultural de belém	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: são várias as razões apontadas para esta alegada demissão vão questões úlcera orçamentais outros convites profissionais socialmente , muito , a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira , rosa passe actua sexta - feira no centro cultural de belém lisboa ,
C 2010	3: cantora brasileira	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: são várias as razões apontadas para esta alegada demissão vão questões úlcera orçamentais outros convites profissionais socialmente , muito , a cantora brasileira rosa passos está novamente em portugal para uma digressão a começar esta sexta - feira , rosa passe actua sexta - feira no centro cultural de belém lisboa ,
D 2010	1: todo momento maduro do trabalho	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: é todo momento maduro do trabalho monte um bom momento como a com peixes um tó rocha aqui é esse era muito soft , rosa passos têm outros trabalhos começou muito nova ,
D 2010	2: joão donato	**pt/j/o/ã/João_Donato_b958.html**: joão donato . rosa passos
D 2010	3: tó rocha aqui	**2008_07_03-21_59_01-Jornal2-2 bloco 4**: é todo momento maduro do trabalho monte um bom momento como a com peixes um tó rocha aqui é esse era muito soft , rosa passos têm outros trabalhos começou muito nova ,

Question #36 - Quantos anos tem a carreira de Carlos do Carmo?		
A 2008	1: NIL	
B 2008	1: 26 semanas	**pt/r/i/b/Ribeiro_Cardoso_c876.html**: ribeiro cardoso . foi também co - autor , com josé jorge letria e carlos do carmo , do programa televisivo carlos do carmo , que a rtpi emitiu durante 26 semanas , de outubro de 97 a abril de 98 .
B 2008	2: 50 anos	**pt/j/o/ã/João_Braga_8fd0.html**: joão braga . carlos do carmo . canta , juntamente com teresa tarouca e antónio de mello - corrêa , na festa dos 50 anos de toureio de mestre joão branco núncio , no convento de são bento de castres (évora) . com a abertura , em abril de 1 966 , da " taverna do
C 2008	1: NIL	
D 2008	1: 26 semanas	**pt/r/i/b/Ribeiro_Cardoso_c876.html**: ribeiro cardoso . foi também co - autor , com josé jorge letria e carlos do carmo , do programa televisivo carlos do carmo , que a rtpi emitiu durante 26 semanas , de outubro de 97 a abril de 98 .
D 2008	2: 50 anos	**pt/j/o/ã/João_Braga_8fd0.html**: joão braga . carlos do carmo . canta , juntamente com teresa tarouca e antónio de mello - corrêa , na festa dos 50 anos de toureio de mestre joão branco núncio , no convento de são bento de castres (évora) . com a abertura , em abril de 1 966 , da " taverna do
A 2010	1: NIL	
B 2010	1: 26 semanas	**pt/r/i/b/Ribeiro_Cardoso_c876.html**: ribeiro cardoso . foi também co - autor , com josé jorge letria e carlos do carmo , do programa televisivo carlos do carmo , que a rtpi emitiu durante 26 semanas , de outubro de 97 a abril de 98 .
B 2010	2: 50 anos	**pt/j/o/ã/João_Braga_8fd0.html**: joão braga . carlos do carmo . canta , juntamente com teresa tarouca e antónio de mello - corrêa , na festa dos 50 anos de toureio de mestre joão branco núncio , no convento de são bento de castres (évora) . com a abertura , em abril de 1 966 , da " taverna do
C 2010	1: NIL	
D 2010	1: 26 semanas	**pt/r/i/b/Ribeiro_Cardoso_c876.html**: ribeiro cardoso . foi também co - autor , com josé jorge letria e carlos do carmo , do programa televisivo carlos do carmo , que a rtpi emitiu durante 26 semanas , de outubro de 97 a abril de 98 .
D 2010	2: 50 anos	**pt/j/o/ã/João_Braga_8fd0.html**: joão braga . carlos do carmo . canta , juntamente com teresa tarouca e antónio de mello - corrêa , na festa dos 50 anos de toureio de mestre joão branco núncio , no convento de são bento de castres (évora) . com a abertura , em abril de 1 966 , da " taverna do

Question #37 - Quem é o nadador mais bem pago da história?		
A 2008	1: NIL	
B 2008	1: celso java	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: celso java tinham um outro recorde que já lhe é um nadador mais bem pago da história .
B 2008	2: fazer	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: telles já batiam um outro recorde que já lhe é um nadador mais bem pago da história . phelps tornou - se milionário de atleta profissional de natação , poucos meses antes de fazer dezasseis anos .
B 2008	3: profissional	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: telles já batiam um outro recorde que já lhe é um nadador mais bem pago da história . phelps tornou - se milionário de atleta profissional de natação , poucos meses antes de fazer dezasseis anos .
C 2008	1: NIL	
D 2008	1: milionário de atleta profissional de natação	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: mesmo que não consiga chegar às oito medalhas de ouro , telles já batiam um outro recorde que já lhe é um nadador mais bem pago da história , phelps tornou - se milionário de atleta profissional de natação , poucos meses antes de fazer dezasseis anos , agora com vinte e três ,
D 2008	2: celso	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: um , mesmo que não consiga chegar às oito medalhas de ouro , celso java tinham um outro recorde que já lhe é um nadador mais bem pago da história , céu tornou - se milionário de atleta profissional de natação , poucos meses antes de fazer dezasseis anos , agora com vinte e três ,
D 2008	3: fazer	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: mesmo que não consiga chegar às oito medalhas de ouro , telles já batiam um outro recorde que já lhe é um nadador mais bem pago da história , phelps tornou - se milionário de atleta profissional de natação , poucos meses antes de fazer dezasseis anos , agora com vinte e três ,
A 2010	1: val	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: é também o nadador mais bem pago da história , mas afinal quanto val em euros .
B 2010	1: mas afinal quanto val	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: é também o nadador mais bem pago da história , mas afinal quanto val em euros .
B 2010	2: milionário de atleta profissional de natação	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: nadador mais bem pago da história . help tornou - se milionário de atleta profissional de natação poucos meses antes de fazer dezasseis anos .
B 2010	3: recorde	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: fiel desejava tinham um outro recorde que já lhe é o nadador mais bem pago da história .
C 2010	1: NIL	
D 2010	1: mas	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: são os dois nomes de ouro da natação olímpica , só mas já três medalhas de ouro conquistadas em jogos olímpicos é também o nadador mais bem pago da história , mas afinal quanto vale em euros a estrela dos jogos olímpicos de pequim , é a imagem da vitória ,

D 2010	2: milionário de atleta profissional	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: mesmo que não consiga chegar às oito medalhas de ouro fiel desejava tinham um outro recorde que já lhe é o nadador mais bem pago da história , help tornou - se milionário de atleta profissional de natação poucos meses antes de fazer dezasseis anos , agora com vinte e três ,
D 2010	3: fazer	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: mesmo que não consiga chegar às oito medalhas de ouro fiel desejava tinham um outro recorde que já lhe é o nadador mais bem pago da história , help tornou - se milionário de atleta profissional de natação poucos meses antes de fazer dezasseis anos , agora com vinte e três ,
Question #38 - Com que idade se tornou Phelps milionário?		
A 2008 B 2008	1: dezasseis anos	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: phelps tornou - se milionário de atleta profissional de natação , poucos meses antes de fazer dezasseis anos .
B 2008	2: 90 anos	**pt/s/u/t/Sutiã.html**: sutiã . história há 90 anos a americana mary phelps jacob patenteava uma das invenções mais queridas do público feminino : o sutiã .
C 2008	1: dezasseis anos	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: telles já batiam um outro recorde que já lhe é um nadador mais bem pago da história , phelps tornou - se milionário de atleta profissional de natação , poucos meses antes de fazer dezasseis anos , agora com vinte e três ,
C 2008	2: quatro anos	**2008_08_13-19_59_02-Telejornal-1 bloco 3**: a que se juntam as seis conseguidas há quatro anos em atenas , algo que nunca ninguém conseguiu , em jogos olímpicos , michael phelps já nos habituou a nadar sempre à frente desta linha verde da linha que marca o anterior recorde do mundo ,
D 2008	1: dezasseis anos	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: telles já batiam um outro recorde que já lhe é um nadador mais bem pago da história , phelps tornou - se milionário de atleta profissional de natação , poucos meses antes de fazer dezasseis anos , agora com vinte e três ,
D 2008	2: quatro anos	**2008_08_13-19_59_02-Telejornal-1 bloco 3**: a que se juntam as seis conseguidas há quatro anos em atenas , algo que nunca ninguém conseguiu , em jogos olímpicos , michael phelps já nos habituou a nadar sempre à frente desta linha verde da linha que marca o anterior recorde do mundo ,
D 2008	3: 90 anos	**pt/s/u/t/Sutiã.html**: sutiã . história há 90 anos a americana mary phelps jacob patenteava uma das invenções mais queridas do público feminino : o sutiã .
A 2010	1: NIL	
B 2010	1: 90 anos	**pt/s/u/t/Sutiã.html**: sutiã . história há 90 anos a americana mary phelps jacob patenteava uma das invenções mais queridas do público feminino : o sutiã .
C 2010	1: NIL	

D 2010	1: 90 anos	**pt/s/u/t/Sutiã.html**: sutiã . história há 90 anos a americana mary phelps jacob patenteava uma das invenções mais queridas do público feminino : o sutiã .
Question #39 - A que distância da capital chinesa se localizou o epicentro do abalo na provincia de Xichuan?		
A 2008	1: cinquenta quilómetros	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: o abalo de seis ponto zero na escala de richter , provocou pelo menos um morto e vinte e três feridos . o epicentro foi mil duzentos e cinquenta quilómetros da capital chinesa ,
A 2008	2: trinta quilómetros	**2008_07_29-19_59_01-Telejornal-1 bloco 3**: o tremor de terra atingiu os cinco vírgula seis graus na escala de richter , com o epicentro trinta quilómetros a leste da cidade , não há até ao momento registo de vítimas , ou estragos .
A 2008	3: cerca de trezentos quilómetros	**2008_06_14-19_59_02-Telejornal-1 bloco 2**: o epicentro teve lugar numa zona rural do norte do país a cerca de trezentos quilómetros de tóquio ,
B 2008	1: cinquenta quilómetros	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: o abalo de seis ponto zero na escala de richter , provocou pelo menos um morto e vinte e três feridos . o epicentro foi mil duzentos e cinquenta quilómetros da capital chinesa ,
B 2008	2: 100 km	**pt/e/s/c/Escala_de_Richter_3f14.html**: escala de richter . tremores que se produziram na califórnia (oeste dos estados unidos) . princípio é uma escala logarítmica : a magnitude de richter corresponde ao logaritmo da medida da amplitude das ondas sísmicas de tipo p e s a 100 km do epicentro .
B 2008	3: trinta quilómetros	**2008_07_29-19_59_01-Telejornal-1 bloco 3**: o tremor de terra atingiu os cinco vírgula seis graus na escala de richter , com o epicentro trinta quilómetros a leste da cidade , não há até ao momento registo de vítimas , ou estragos .
C 2008	1: cinquenta quilómetros	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: horas depois do forte tremor de terra voltou a sacudir a região , o abalo de seis ponto zero na escala de richter , provocou pelo menos um morto e vinte e três feridos , o epicentro foi mil duzentos e cinquenta quilómetros da capital chinesa ,
C 2008	2: trinta quilómetros	**2008_07_29-19_59_01-Telejornal-1 bloco 3**: para não ser daqui por , um sismo abalou a cidade de lousa gelosia na califórnia , o tremor de terra atingiu os cinco vírgula seis graus na escala de richter , com o epicentro trinta quilómetros a leste da cidade , não há até ao momento registo de vítimas , ou estragos ,
C 2008	3: cerca de trezentos quilómetros	**2008_06_14-19_59_02-Telejornal-1 bloco 2**: na redacção daqui da tv , os jornalistas apressaram - se a ligar as câmaras , militares e bombeiros , o epicentro teve lugar numa zona rural do norte do país a cerca de trezentos quilómetros de tóquio ,

D 2008	1: cinquenta quilómetros	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: horas depois do forte tremor de terra voltou a sacudir a região , o abalo de seis ponto zero na escala de richter , provocou pelo menos um morto e vinte e três feridos , o epicentro foi mil duzentos e cinquenta quilómetros da capital chinesa ,
D 2008	2: 100 km	**pt/e/s/c/Escala_de_Richter_3f14.html**: escala de richter . tremores que se produziram na califórnia (oeste dos estados unidos) . princípio é uma escala logarítmica : a magnitude de richter corresponde ao logaritmo da medida da amplitude das ondas sísmicas de tipo p e s a 100 km do epicentro .
D 2008	3: trinta quilómetros	**2008_07_29-19_59_01-Telejornal-1 bloco 3**: para não ser daqui por , um sismo abalou a cidade de lousa gelosia na califórnia , o tremor de terra atingiu os cinco vírgula seis graus na escala de richter , com o epicentro trinta quilómetros a leste da cidade , não há até ao momento registo de vítimas , ou estragos ,
A 2010	1: cerca de trezentos quilómetros	**2008_06_14-21_59_01-Jornal2-2 bloco 1**: o epicentro teve lugar numa zona rural do norte do país a cerca de trezentos quilómetros de tóquio . o abalo interrompeu pequeno - almoço de mais há que ana moura .
A 2010	2: trinta quilómetros	**2008_07_29-19_59_01-Telejornal-1 bloco 3**: o tremor de terra atingiu os cinco vírgula seis graus na escala de richter com epicentro trinta quilómetros a leste da cidade .
B 2010	1: cerca de trezentos quilómetros	**2008_06_14-19_59_02-Telejornal-1 bloco 2**: o epicentro teve lugar numa zona rural do norte do país a cerca de trezentos quilómetros de tóquio . o abalo interrompeu pequeno - almoço de mas há quem não dura .
B 2010	2: trinta quilómetros	**2008_07_29-19_59_01-Telejornal-1 bloco 3**: o tremor de terra atingiu os cinco vírgula seis graus na escala de richter com epicentro trinta quilómetros a leste da cidade .
B 2010	3: 100 km	**pt/e/s/c/Escala_de_Richter_3f14.html**: escala de richter . tremores que se produziram na califórnia (oeste dos estados unidos) . princípio é uma escala logarítmica : a magnitude de richter corresponde ao logaritmo da medida da amplitude das ondas sísmicas de tipo p e s a 100 km do epicentro .
C 2010	1: cerca de trezentos quilómetros	**2008_06_14-19_59_02-Telejornal-1 bloco 2**: não é , o epicentro teve lugar numa zona rural do norte do país a cerca de trezentos quilómetros de tóquio , o abalo interrompeu pequeno - almoço de mas há quem não dura , para estar a guiné ,
C 2010	2: trinta quilómetros	**2008_07_29-19_59_01-Telejornal-1 bloco 3**: para conhecer daqui a pouco , um sismo abalou a cidade lozano josé na califórnia , o tremor de terra atingiu os cinco vírgula seis graus na escala de richter com epicentro trinta quilómetros a leste da cidade , não há até ao momento registo de vítimas ou estragos ,

D 2010	1: cerca de trezentos quilómetros	**2008_06_14-21_59_01-Jornal2-2 bloco 1**: não é , o epicentro teve lugar numa zona rural do norte do país a cerca de trezentos quilómetros de tóquio , o abalo interrompeu pequeno - almoço de mais há que ana moura , para estar a guiné ,
D 2010	2: trinta quilómetros	**2008_07_29-19_59_01-Telejornal-1 bloco 3**: para conhecer daqui a pouco , um sismo abalou a cidade lozano josé na califórnia , o tremor de terra atingiu os cinco vírgula seis graus na escala de richter com epicentro trinta quilómetros a leste da cidade , não há até ao momento registo de vítimas ou estragos ,
D 2010	3: 100 km	**pt/e/s/c/Escala_de_Richter_3f14.html**: escala de richter . tremores que se produziram na califórnia (oeste dos estados unidos) . princípio é uma escala logarítmica : a magnitude de richter corresponde ao logaritmo da medida da amplitude das ondas sísmicas de tipo p e s a 100 km do epicentro .
Question #40 - Que medalha conquistou Nelson Évora para Portugal no triplo salto?		
A 2008	1: NIL	
B 2008	1: medalha	**2008_08_21-19_59_01-Telejornal-1 bloco 1**: boa noite nelson évora que concretizou o sonho olímpico é ele o novo campeão olímpico do triplo salto de conquistou a medalha de ouro para portugal .
B 2008	2: medalhas	**2008_08_18-21_59_01-Jornal2-2 bloco 1**: apesar de tudo , e sem temor ainda acredita que portugal poderá trazer quatro medalhas da china . nelson évora apurou - se com facilidade para a final do triplo salto o atleta garante que está mais preparado que nunca .
B 2008	3: medalha	**2008_08_21-19_59_01-Telejornal-1 bloco 1**: boa noite nelson évora que concretizou o sonho olímpico é ele o novo campeão olímpico do triplo salto de conquistou a medalha de ouro para portugal .
C 2008	1: NIL	
D 2008	1: medalhas	**2008_08_18-21_59_01-Jornal2-2 bloco 1**: apesar de tudo , e sem temor ainda acredita que portugal poderá trazer quatro medalhas da china , nelson évora apurou - se com facilidade para a final do triplo salto o atleta garante que está mais preparado que nunca , para conquistar uma medalha ,
D 2008	2: medalha	**2008_08_21-19_59_01-Telejornal-1 bloco 1**: boa noite nelson évora que concretizou o sonho olímpico é ele o novo campeão olímpico do triplo salto de conquistou a medalha de ouro para portugal , um lugar no quarto ensaio quando saltou dezassete nápoles e sessenta e sete centímetros ,
D 2008	3: medalha	**2008_08_21-19_59_01-Telejornal-1 bloco 1**: boa noite nelson évora que concretizou o sonho olímpico é ele o novo campeão olímpico do triplo salto de conquistou a medalha de ouro para portugal

A 2010	1: NIL	
B 2010	1: medalhas	**2008_08_18-21_59_01-Jornal2-2 bloco 1**: apesar de tudo , vicente moura ainda acredita que portugal poderá trazer quatro medalhas da china . nélson évora apurou - se com facilidade para a final do triplo - salto atleta garante que está mais preparado que nunca . para conquistar uma medalha . nelson
C 2010	1: NIL	
D 2010	1: medalhas	**2008_08_18-21_59_01-Jornal2-2 bloco 1**: apesar de tudo , vicente moura ainda acredita que portugal poderá trazer quatro medalhas da china , nélson évora apurou - se com facilidade para a final do triplo - salto atleta garante que está mais preparado que nunca , para conquistar uma medalha , nelson
Question #41 - Em que ensaio conseguiu Nelson Évora o lugar que o tornaria campeão olímpico?		
A 2008	1: NIL	
B 2008	1: cão	**2008_08_31-21_59_02-Jornal2-2 bloco 2**: no dia em que apertar a mão a nelson évora , pro cumprimentar , pelo ouro nos jogos olímpicos se lembra da humilhação do cão de ataque .
C 2008	1: NIL	
D 2008	1: soa	**2008_08_22-21_59_02-Jornal2-2 bloco 3**: e quando soa portuguesa , nélson évora campeão olímpico do triplo salto , confessa que se repetiu , foi às vinte e uma horas e trinta e oito minutos que nelson évora que operado análise ,
D 2008	2: cão	**2008_08_31-21_59_02-Jornal2-2 bloco 2**: eu só espero , que aníbal cavaco silva , no dia em que apertar a mão a nelson évora , pro cumprimentar , pelo ouro nos jogos olímpicos se lembra da humilhação do cão de ataque ,
A 2010 B 2010 C 2010	1: NIL	
D 2010	1: soa	**2008_08_22-21_59_02-Jornal2-2 bloco 2**: e quando soa portuguesa , nélson évora campeão olímpico do triplo salto , confessa que crp , foi às vinte e uma horas e trinta e oito minutos que nelson abrasivo operado realize , quase vinte e quatro horas depois chegou a uma tradição disse ,
Question #42 - Quando ganhou António Lobo Antunes o Prémio Camões?		
A 2008	1: dois mil e oito	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa . extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes ,

A 2008	2: dois mil e sete	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa . extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes ,
A 2008	3: novecentos e quarenta e um	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa . extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes , joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou com o
B 2008	1: dois mil e oito	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa . extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes ,
B 2008	2: dois mil e sete	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa . extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes ,
B 2008	3: novecentos e quarenta e um	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa . extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes , joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou com o
C 2008	1: dois mil e oito	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa , extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes ,
C 2008	2: dois mil e sete	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa , extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes ,
C 2008	3: novecentos e quarenta e um	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa , extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes , joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou com o

D 2008	1: dois mil e oito	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa , extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes ,
D 2008	2: dois mil e sete	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa , extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes ,
D 2008	3: novecentos e quarenta e um	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa , extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes , joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou com o
A 2010	1: dois mil e sete	**2008_07_25-21_59_01-Jornal2-2 bloco 3**: prémio camões de dois mil e sete . ganhou com uma cerimónia especial . os planos são sempre agradáveis . segundo como deco muito filha . não lhe mas de qualquer maneira juca sempre agradável agora . soube tem sido muito afortunado este ano já vieram três me . vencedor desta vez , antónio lobo antunes
A 2010	2: dois mil e oito	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa . extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos . o português antónio lobo antunes
A 2010	3: novecentos e quarenta e um	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa . extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos . o português antónio lobo antunes joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou com o
B 2010	1: dois mil e sete	**2008_07_25-21_59_01-Jornal2-2 bloco 3**: prémio camões de dois mil e sete . ganhou com uma cerimónia especial . os planos são sempre agradáveis . segundo como deco muito filha . não lhe mas de qualquer maneira juca sempre agradável agora . soube tem sido muito afortunado este ano já vieram três me . vencedor desta vez , antónio lobo antunes

B 2010	2: dois mil e oito	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa . extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos . o português antónio lobo antunes
B 2010	3: novecentos e quarenta e um	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa . extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos . o português antónio lobo antunes joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou com o
C 2010	1: dois mil e sete	**2008_07_25-21_59_01-Jornal2-2 bloco 3**: prémio camões de dois mil e sete , ganhou com uma cerimónia especial , os planos são sempre agradáveis , segundo como deco muito filha , não lhe mas de qualquer maneira juca sempre agradável agora , soube tem sido muito afortunado este ano já vieram três me , vencedor desta vez , antónio lobo antunes
C 2010	2: dois mil e oito	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa , extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos , o português antónio lobo antunes
C 2010	3: novecentos e quarenta e um	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa , extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos , o português antónio lobo antunes joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou com o
D 2010	1: dois mil e sete	**2008_07_25-21_59_01-Jornal2-2 bloco 3**: prémio camões de dois mil e sete , ganhou com uma cerimónia especial , os planos são sempre agradáveis , segundo como deco muito filha , não lhe mas de qualquer maneira juca sempre agradável agora , soube tem sido muito afortunado este ano já vieram três me , vencedor desta vez , antónio lobo antunes
D 2010	2: dois mil e oito	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa , extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos , o português antónio lobo antunes
D 2010	3: novecentos e quarenta e um	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa , extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos , o português antónio lobo antunes joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou com o

Question #43 - Quem deu vida à "Formiga" cantada por Amália?		
A 2008	1: carretera	**2008_06_09-19_59_01-Telejornal-1 bloco 3**: camp gigantescos engarrafamentos nas antas mora as formigas la carretera .
A 2008	2: extremidade	**2008_06_09-21_59_01-Jornal2-2 bloco 2**: camp gigantescos engarrafamentos nas antas mora as formigas la carretera , esmoriz câmara maresca nomeado este peixe deste ano nasceram , e creme recuperar na extremidade .
A 2008	3: maresca	**2008_06_09-21_59_01-Jornal2-2 bloco 2**: camp gigantescos engarrafamentos nas antas mora as formigas la carretera , esmoriz câmara maresca nomeado este peixe deste ano nasceram , e creme recuperar na extremidade .
B 2008	1: fernando alvim	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bolsa traz ainda outras memórias , e sessenta e nove , fontes rocha e fernando alvim , deram vida à formiga cantada por amália .
B 2008	2: neill	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: vida à formiga cantada por amália . uma adaptação de alan um ano o problema do o neill , velha fábula em bolsa nove um nove um . achamos de combates . do seu ver o mundo que melhor a segunda sessão um . antes porém com ou sem , ligação nossa , de que o seu fazer
B 2008	3: fábula	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: vida à formiga cantada por amália . uma adaptação de alan um ano o problema do o neill , velha fábula em bolsa nove um nove um . achamos de combates . do seu ver o mundo que melhor a segunda sessão um . antes porém com ou sem , ligação nossa , de que o seu fazer
C 2008	1: carretera	**2008_06_09-19_59_01-Telejornal-1 bloco 3**: usa - se madrid e barcelona a circular em marcha lenta , camp gigantescos engarrafamentos nas antas mora as formigas la carretera , esmoriz câmara maresca nomeadas de peixe deste ano nasceram , e creme recuperarmos trinidad ,
C 2008	2: maresca	**2008_06_09-19_59_01-Telejornal-1 bloco 3**: usa - se madrid e barcelona a circular em marcha lenta , camp gigantescos engarrafamentos nas antas mora as formigas la carretera , esmoriz câmara maresca nomeadas de peixe deste ano nasceram , e creme recuperarmos trinidad ,
C 2008	3: trinidad	**2008_06_09-19_59_01-Telejornal-1 bloco 3**: usa - se madrid e barcelona a circular em marcha lenta , camp gigantescos engarrafamentos nas antas mora as formigas la carretera , esmoriz câmara maresca nomeadas de peixe deste ano nasceram , e creme recuperarmos trinidad ,
D 2008	1: neill	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bolsa traz ainda outras memórias , e sessenta e nove , fontes rocha e fernando alvim , deram vida à formiga cantada por amália , uma adaptação de alan um ano o problema do o neill , velha fábula em bolsa nove um nove um ,

D 2008	2: fábula	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bolsa traz ainda outras memórias , e sessenta e nove , fontes rocha e fernando alvim , deram vida à formiga cantada por amália , uma adaptação de alan um ano o problema do o neill , velha fábula em bolsa nove um nove um ,
D 2008	3: fernando alvim	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bolsa traz ainda outras memórias , e sessenta e nove , fontes rocha e fernando alvim , deram vida à formiga cantada por amália , uma adaptação de alan um ano o problema do o neill , velha fábula em bolsa nove um nove um ,
A 2010	1: jeitão	**2008_06_28-21_59_01-Jornal2-2 bloco 2**: e acabados a nós jeitão de formiga que nan estou de técnica na liga que emitidos .
A 2010	2: importo	**2008_06_12-19_59_01-Telejornal-1 bloco 2**: família se bem me importo de formiga .
B 2010	1: neill velha fábula bossa - nova	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: vida a formiga cantada por amália . uma adaptação de alain num humano do poema de o neill velha fábula bossa - nova num nem um . mas vamos a bondade fez . reconheceu haver muito que nós . some - se são um . nós porém tombou sam não são nosso . eu concebo fazer .
B 2010	2: fernando alvim	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bossa traz ainda outras memórias e sessenta e nove fontes rocha e fernando alvim deram vida a formiga cantada por amália .
B 2010	3: fábula bossa	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: vida a formiga cantada por amália . uma adaptação de alain num humano do poema de o neill velha fábula bossa - nova num nem um . mas vamos a bondade fez . reconheceu haver muito que nós . some - se são um . nós porém tombou sam não são nosso . eu concebo fazer .
C 2010	1: jeitão	**2008_06_28-21_59_01-Jornal2-2 bloco 2**: em niza esteios vara barra na face , leveza na caixa futebol se , e acabados a nós jeitão de formiga que nan estou de técnica na liga que emitidos ,
C 2010	2: tati	**2008_06_12-19_59_01-Telejornal-1 bloco 2**: mas como árbitro deixou o aviso não gosta cenas dramáticas , família se bem me importo de formiga , até que pela é ex se para infecta lei , esse auto críticos aos mike seixo tati se for me importo , p falta mna ,
C 2010	3: mna	**2008_06_12-19_59_01-Telejornal-1 bloco 2**: mas como árbitro deixou o aviso não gosta cenas dramáticas , família se bem me importo de formiga , até que pela é ex se para infecta lei , esse auto críticos aos mike seixo tati se for me importo , p falta mna ,

D 2010	1: neill velha fábula bossa - nova	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: um mas plano - sequência em curso , a bossa traz ainda outras memórias e sessenta e nove fontes rocha e fernando alvim deram vida a formiga cantada por amália , uma adaptação de alain num humano do poema de o neill velha fábula bossa - nova num nem um ,
D 2010	2: fábula bossa	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: um mas plano - sequência em curso , a bossa traz ainda outras memórias e sessenta e nove fontes rocha e fernando alvim deram vida a formiga cantada por amália , uma adaptação de alain num humano do poema de o neill velha fábula bossa - nova num nem um ,
D 2010	3: fernando alvim	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: um mas plano - sequência em curso , a bossa traz ainda outras memórias e sessenta e nove fontes rocha e fernando alvim deram vida a formiga cantada por amália , uma adaptação de alain num humano do poema de o neill velha fábula bossa - nova num nem um ,
Question #44 - Em que ano?		
A 2008 B 2008	1: sessenta e nove	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bolsa traz ainda outras memórias , e sessenta e nove , fontes rocha e fernando alvim , deram vida à formiga cantada por amália .
B 2008	2: nove um nove um	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: vida à formiga cantada por amália . uma adaptação de alan um ano o problema do o neill , velha fábula em bolsa nove um nove um . achamos de combates . do seu ver o mundo que melhor a segunda sessão um . antes porém com ou sem , ligação nossa , de que o seu fazer
B 2008	3: 1 964	**pt/a/m/á/Amália_Rodrigues_9290.html**: amália rodrigues . vida (1 964) amália canta luís de camões (1 965) formiga
C 2008	1: nove um nove um	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bolsa traz ainda outras memórias , e sessenta e nove , fontes rocha e fernando alvim , deram vida à formiga cantada por amália , uma adaptação de alan um ano o problema do o neill , velha fábula em bolsa nove um nove um , achamos de combates ,
C 2008	2: sessenta e nove	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bolsa traz ainda outras memórias , e sessenta e nove , fontes rocha e fernando alvim , deram vida à formiga cantada por amália , uma adaptação de alan um ano o problema do o neill , velha fábula em bolsa nove um nove um , achamos de combates ,
D 2008	1: nove um nove um	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bolsa traz ainda outras memórias , e sessenta e nove , fontes rocha e fernando alvim , deram vida à formiga cantada por amália , uma adaptação de alan um ano o problema do o neill , velha fábula em bolsa nove um nove um ,

D 2008	2: sessenta e nove	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bolsa traz ainda outras memórias , e sessenta e nove , fontes rocha e fernando alvim , deram vida à formiga cantada por amália , uma adaptação de alan um ano o problema do o neill , velha fábula em bolsa nove um nove um ,
D 2008	3: 1 964	**pt/a/m/á/Amália_Rodrigues_9290.html**: amália rodrigues . vida (1 964) amália canta luís de camões (1 965) formiga
A 2010	1: NIL	
B 2010	1: 1 970	**pt/a/m/á/Amália_Rodrigues_9290.html**: amália rodrigues . luís de camões (1 965) formiga bossa nossa (1 969) amália e vinicius (1 970) com que voz (1 970) fado português (1 970) oiça lá ó senhor vinho (1 971) amália no japão (1 971) cheira a lisboa (1 972)
B 2010	2: 1 965	**pt/a/m/á/Amália_Rodrigues_9290.html**: amália rodrigues . vida (1 964) amália canta luís de camões (1 965) formiga
B 2010	3: 1 971	**pt/a/m/á/Amália_Rodrigues_9290.html**: amália rodrigues . luís de camões (1 965) formiga bossa nossa (1 969) amália e vinicius (1 970) com que voz (1 970) fado português (1 970) oiça lá ó senhor vinho (1 971) amália no japão (1 971) cheira a lisboa (1 972)
C 2010	1: NIL	
D 2010	1: 1 970	**pt/a/m/á/Amália_Rodrigues_9290.html**: amália rodrigues . luís de camões (1 965) formiga bossa nossa (1 969) amália e vinicius (1 970) com que voz (1 970) fado português (1 970) oiça lá ó senhor vinho (1 971) amália no japão (1 971) cheira a lisboa (1 972)
D 2010	2: 1 965	**pt/a/m/á/Amália_Rodrigues_9290.html**: amália rodrigues . vida (1 964) amália canta luís de camões (1 965) formiga
D 2010	3: 1 971	**pt/a/m/á/Amália_Rodrigues_9290.html**: amália rodrigues . luís de camões (1 965) formiga bossa nossa (1 969) amália e vinicius (1 970) com que voz (1 970) fado português (1 970) oiça lá ó senhor vinho (1 971) amália no japão (1 971) cheira a lisboa (1 972)
Question #45 - Quem é António Guterres?		
A 2008	1: coordenar a ajuda aos milhares de deslocados	**2008_08_21-19_59_01-Telejornal-1 bloco 3**: antónio guterres foi coordenar a ajuda aos milhares de deslocados , que fugiram da ossétia do sul .
A 2008	2: tiblissi	**2008_08_19-19_59_02-Telejornal-1 bloco 2**: antónio guterres veio tiblissi , e segue agora para moscovo para negociar .
A 2008	3: é para esses que	**2008_08_19-19_59_02-Telejornal-1 bloco 2**: e é e é para esses que , antónio guterres terá que , prestar grande atenção .

B 2008	1: antónio manuel de oliveira guterres gcc , (santos - o - velho , lisboa ; 30 de abril de 1 949) é actualmente o alto comissário das nações unidas para os refugiados , tendo sido primeiro - ministro de portugal .	**pt/a/n/t/António_Guterres_fe52.html**: antónio manuel de oliveira guterres gcc , (santos - o - velho , lisboa ; 30 de abril de 1 949) é actualmente o alto comissário das nações unidas para os refugiados , tendo sido primeiro - ministro de portugal .
C 2008	1: fica mais fácil a tarefa do alto comissário da onu para os refugiados	**2008_08_19-21_59_02-Jornal2-2 bloco 1**: se a retirada se confirmar , fica mais fácil a tarefa do alto comissário da onu para os refugiados , antónio guterres veio tiblissi , e segue agora para moscovo para negociar , os corredores humanitários , e que os meios , para este tipo de hepatite analisados ,
C 2008	2: coordenar a ajuda aos milhares de deslocados	**2008_08_21-19_59_01-Telejornal-1 bloco 3**: vilarinho presidente , a um ano e de uma , na ossétia do norte chegou entretanto ao comissário da onu para os refugiados , antónio guterres foi coordenar a ajuda aos milhares de deslocados , que fugiram da ossétia do sul , mais uma etapa do circuito de visita de manietar metu ,
C 2008	3: durante a governação	**2008_07_19-19_59_02-Telejornal-1 bloco 2**: e foi um dos protagonistas do programa polis , que arrancou , durante a governação de antónio guterres ,
D 2008	1: antónio manuel de oliveira guterres gcc , (santos - o - velho , lisboa ; 30 de abril de 1 949) é actualmente o alto comissário das nações unidas para os refugiados , tendo sido primeiro - ministro de portugal .	**pt/a/n/t/António_Guterres_fe52.html**: antónio manuel de oliveira guterres gcc , (santos - o - velho , lisboa ; 30 de abril de 1 949) é actualmente o alto comissário das nações unidas para os refugiados , tendo sido primeiro - ministro de portugal .

A 2010	1: utilíssimo	**2008_08_19-19_59_02-Telejornal-1 bloco 2**: antónio guterres foi utilíssimo e segue agora para moscovo para negociar os corredores humanitários .
A 2010	2: durante a governação	**2008_07_19-19_59_02-Telejornal-1 bloco 2**: durante a governação de antónio guterres .
A 2010	3: no governo	**2008_07_29-21_59_01-Jornal2-2 bloco 2**: no governo de antónio guterres primeiro - ministro acredita que a riqueza de um país se mete também pela qualificação da população .
B 2010	1: antónio manuel de oliveira guterres gcc , (santos - o - velho , lisboa ; 30 de abril de 1 949) é actualmente o alto comissário das nações unidas para os refugiados , tendo sido primeiro - ministro de portugal .	**pt/a/n/t/António_Guterres_fe52.html**: antónio manuel de oliveira guterres gcc , (santos - o - velho , lisboa ; 30 de abril de 1 949) é actualmente o alto comissário das nações unidas para os refugiados , tendo sido primeiro - ministro de portugal .
C 2010	1: utilíssimo	**2008_08_19-19_59_02-Telejornal-1 bloco 2**: antónio guterres foi utilíssimo e segue agora para moscovo para negociar os corredores humanitários , e que os meios disponíveis para espírito de equipa e o amor existe , tec toda a questão está , antónio
C 2010	2: se a retirada se confirmar fica mais fácil a tarefa do alto comissário da onu para os refugiados	**2008_08_19-19_59_02-Telejornal-1 bloco 2**: se a retirada se confirmar fica mais fácil a tarefa do alto comissário da onu para os refugiados , antónio guterres foi utilíssimo e segue agora para moscovo para negociar os corredores humanitários , e que os meios disponíveis para espírito de equipa e o amor existe ,
C 2010	3: durante a governação	**2008_07_19-19_59_02-Telejornal-1 bloco 2**: durante a governação de antónio guterres ,
D 2010	1: antónio manuel de oliveira guterres gcc , (santos - o - velho , lisboa ; 30 de abril de 1 949) é actualmente o alto comissário das nações unidas para os refugiados , tendo sido primeiro - ministro de portugal .	**pt/a/n/t/António_Guterres_fe52.html**: antónio manuel de oliveira guterres gcc , (santos - o - velho , lisboa ; 30 de abril de 1 949) é actualmente o alto comissário das nações unidas para os refugiados , tendo sido primeiro - ministro de portugal .

Question #46 - Com que material conseguiu Elvira Fortunato produzir transístores?		
A 2008	1: NIL	
B 2008	1: insectos	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a cientista portuguesa elvira fortunato ganhou maior prémio alguma vez atribuído pelo europeu em insectos que não ser considerado . uma espécie de prémios nobel na área da engenharia investigadora , liderou a equipa da universidade nova conseguiu produzir transístores com papel que não usam o silício com isolador ,
B 2008	2: ciência dos materiais	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: não fui convidada a investigadora elvira fortunato , director do departamento de ciência dos materiais da universidade nova de lisboa .
C 2008	1: NIL	
D 2008	1: insectos	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: na sua qualidade , a cientista portuguesa elvira fortunato ganhou maior prémio alguma vez atribuído pelo europeu em insectos que não ser considerado , uma espécie de prémios nobel na área da engenharia investigadora , liderou a equipa da universidade nova conseguiu produzir transístores com papel que não usam o silício com isolador ,
D 2008	2: ciência dos materiais	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: e suporte do próprio francis , a partir daqui adivinha - se um novo limiar no tempo da microelectrónica , não fui convidada a investigadora elvira fortunato , director do departamento de ciência dos materiais da universidade nova de lisboa ,
D 2008	3: língua	**2008_08_28-19_59_01-Telejornal-1 bloco 2**: é a minha língua dizerem portuguesa polícia e um , e arrojado tão com as investigações de elvira fortunato , que houve marisa enquanto lidera uma equipa de cientistas para criar ecrãs de papel , e produzir sistemas electrónicos a baixo custo ,
A 2010	1: NIL	
B 2010	1: etapas	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: elvira fortunato equipa do centro de investigação de materiais da universidade nova de lisboa . experimentaram várias etapas até conseguir produzir , pela primeira vez . transístores com uma camada de papel .
C 2010	1: NIL	
D 2010	1: etapas	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: pela europeia nos liceus discussão cinco um cientista português , elvira fortunato equipa do centro de investigação de materiais da universidade nova de lisboa , experimentaram várias etapas até conseguir produzir , pela primeira vez , transístores com uma camada de papel , há seis anos ,

Question #47 - Qual a nacionalidade do novo campeão olímpico dos 100 metros?		
A 2008	1: operário da construção civil de nacionalidade alemã	**2008_08_06-19_59_01-Telejornal-1 bloco 2**: um indivíduo a rondar os cem club na noite do desaparecimento . e que fora considerado suspeito porque nunca antes tinha sido administrador da luz , era afinal um operário da construção civil de nacionalidade alemã no ser mais profundamente e com o testemunho de martin smyth .
A 2008	2: ligações à américa	**2008_07_08-19_59_01-Telejornal-1 bloco 2**: nacionalidade espanhola as autoridades dos dois países acreditam que neutralizaram esta organização criminosa , com ligações à américa do sul . foram reduzidas as penas para a maioria dos elementos da brigada de trânsito da gnr envolvidos num dos maiores processos de corrupção de sempre . envolvendo forças policiais estavam implicados , cento
A 2008	3: maiores processos de corrupção	**2008_07_08-19_59_01-Telejornal-1 bloco 2**: nacionalidade espanhola as autoridades dos dois países acreditam que neutralizaram esta organização criminosa , com ligações à américa do sul . foram reduzidas as penas para a maioria dos elementos da brigada de trânsito da gnr envolvidos num dos maiores processos de corrupção de sempre . envolvendo forças policiais estavam implicados , cento
B 2008	1: lista de recordes olímpicos	**pt/l/i/s/Lista_de_recordes_olímpicos.html**: lista de recordes olímpicos . para indicar a nacionalidade dos atletas usa - se a código de país do coi . veja também : lista de recordes mundiais atletismo recordes olímpicos de atletismo (homens) disciplina atleta país tempo data jogos olímpicos 100
B 2008	2: tempo data jogos olímpicos	**pt/l/i/s/Lista_de_recordes_olímpicos.html**: lista de recordes olímpicos . para indicar a nacionalidade dos atletas usa - se a código de país do coi . veja também : lista de recordes mundiais atletismo recordes olímpicos de atletismo (homens) disciplina atleta país tempo data jogos olímpicos 100
B 2008	3: código de país do coi	**pt/l/i/s/Lista_de_recordes_olímpicos.html**: lista de recordes olímpicos . para indicar a nacionalidade dos atletas usa - se a código de país do coi . veja também : lista de recordes mundiais atletismo recordes olímpicos de atletismo (homens) disciplina atleta país tempo data jogos olímpicos 100
C 2008	1: operário da construção civil de nacionalidade alemã	**2008_08_06-19_59_01-Telejornal-1 bloco 2**: captadas na área de serviço lá , um indivíduo a rondar os cem club na noite do desaparecimento , e que fora considerado suspeito porque nunca antes tinha sido administrador da luz , era afinal um operário da construção civil de nacionalidade alemã no ser mais profundamente e com o testemunho de martin smyth ,

C 2008	2: maiores processos de corrupção	**2008_07_08-19_59_01-Telejornal-1 bloco 2**: nacionalidade espanhola as autoridades dos dois países acreditam que neutralizaram esta organização criminosa , com ligações à américa do sul , foram reduzidas as penas para a maioria dos elementos da brigada de trânsito da gnr envolvidos num dos maiores processos de corrupção de sempre , envolvendo forças policiais estavam implicados , cento
C 2008	3: ligações à américa	**2008_07_08-19_59_01-Telejornal-1 bloco 2**: nacionalidade espanhola as autoridades dos dois países acreditam que neutralizaram esta organização criminosa , com ligações à américa do sul , foram reduzidas as penas para a maioria dos elementos da brigada de trânsito da gnr envolvidos num dos maiores processos de corrupção de sempre , envolvendo forças policiais estavam implicados , cento
D 2008	1: lista de recordes olímpicos	**pt/l/i/s/Lista_de_recordes_olímpicos.html**: lista de recordes olímpicos . para indicar a nacionalidade dos atletas usa - se a código de país do coi . veja também : lista de recordes mundiais atletismo recordes olímpicos de atletismo (homens) disciplina atleta país tempo data jogos olímpicos 100
D 2008	2: tempo data jogos olímpicos	**pt/l/i/s/Lista_de_recordes_olímpicos.html**: lista de recordes olímpicos . para indicar a nacionalidade dos atletas usa - se a código de país do coi . veja também : lista de recordes mundiais atletismo recordes olímpicos de atletismo (homens) disciplina atleta país tempo data jogos olímpicos 100
D 2008	3: código de país do coi	**pt/l/i/s/Lista_de_recordes_olímpicos.html**: lista de recordes olímpicos . para indicar a nacionalidade dos atletas usa - se a código de país do coi . veja também : lista de recordes mundiais atletismo recordes olímpicos de atletismo (homens) disciplina atleta país tempo data jogos olímpicos 100
A 2010	1: operário da construção civil de nacionalidade alemã	**2008_08_06-19_59_01-Telejornal-1 bloco 2**: um indivíduo rondar os cem club na noite do desaparecimento e que fora considerado suspeito porque nunca antes tinha sido visto na vila da luz . era afinal um operário da construção civil de nacionalidade alemã e ser mais aprofundamento ficou testemunho de martin smyth .
A 2010	2: visto na vila da luz	**2008_08_06-19_59_01-Telejornal-1 bloco 2**: um indivíduo rondar os cem club na noite do desaparecimento e que fora considerado suspeito porque nunca antes tinha sido visto na vila da luz . era afinal um operário da construção civil de nacionalidade alemã e ser mais aprofundamento ficou testemunho de martin smyth .

A 2010	3: mortos	**2008_08_12-19_59_01-Telejornal-1 bloco 1**: cento e trinta . e cinco lidados para reforçar os combates na abkházia de onde as tropas georgianas já iniciaram a retirada nesta guerra do cáucaso que dura há cinco dias os russos falam em milhares de mortos entre as forças militares a geórgia em centenas certa é a morte de três jornalistas um deles de nacionalidade
B 2010	1: lista de recordes olímpicos	**pt/l/i/s/Lista_de_recordes_olímpicos.html**: lista de recordes olímpicos . esta é uma lista dos recordes olímpicos , nas modalidades em que eles são mantidos pelo comité olímpico internacional . para indicar a nacionalidade dos atletas usa - se a código de país do coi .
B 2010	2: comité olímpico internacional	**pt/l/i/s/Lista_de_recordes_olímpicos.html**: lista de recordes olímpicos . esta é uma lista dos recordes olímpicos , nas modalidades em que eles são mantidos pelo comité olímpico internacional . para indicar a nacionalidade dos atletas usa - se a código de país do coi .
B 2010	3: código de país do coi	**pt/l/i/s/Lista_de_recordes_olímpicos.html**: lista de recordes olímpicos . esta é uma lista dos recordes olímpicos , nas modalidades em que eles são mantidos pelo comité olímpico internacional . para indicar a nacionalidade dos atletas usa - se a código de país do coi .
C 2010	1: afinal um operário da construção civil de nacionalidade alemã	**2008_08_06-19_59_01-Telejornal-1 bloco 2**: um indivíduo rondar os cem club na noite do desaparecimento e que fora considerado suspeito porque nunca antes tinha sido visto na vila da luz , era afinal um operário da construção civil de nacionalidade alemã e ser mais aprofundamento ficou testemunho de martin smyth ,
C 2010	2: vila da luz	**2008_08_06-19_59_01-Telejornal-1 bloco 2**: um indivíduo rondar os cem club na noite do desaparecimento e que fora considerado suspeito porque nunca antes tinha sido visto na vila da luz , era afinal um operário da construção civil de nacionalidade alemã e ser mais aprofundamento ficou testemunho de martin smyth ,
C 2010	3: morte	**2008_08_12-19_59_01-Telejornal-1 bloco 1**: cento e trinta , e cinco lidados para reforçar os combates na abkházia de onde as tropas georgianas já iniciaram a retirada nesta guerra do cáucaso que dura há cinco dias os russos falam em milhares de mortos entre as forças militares a geórgia em centenas certa é a morte de três jornalistas um deles de nacionalidade

D 2010	1: lista de recordes olímpicos	**pt/l/i/s/Lista_de_recordes_olímpicos.html**: lista de recordes olímpicos . esta é uma lista dos recordes olímpicos , nas modalidades em que eles são mantidos pelo comité olímpico internacional . para indicar a nacionalidade dos atletas usa - se a código de país do coi .
D 2010	2: código de país do coi	**pt/l/i/s/Lista_de_recordes_olímpicos.html**: lista de recordes olímpicos . esta é uma lista dos recordes olímpicos , nas modalidades em que eles são mantidos pelo comité olímpico internacional . para indicar a nacionalidade dos atletas usa - se a código de país do coi .
D 2010	3: realizar os primeiros jogos em atenas	**pt/d/e/m/Demetrius_Vikelas_f888.html**: demetrius vikelas . mas vikelas persuadiu o recém criado comité olímpico internacional de que seria melhor realizar os primeiros jogos em atenas . como na altura os estatutos do coi requeriam que o presidente da organização tivesse a nacionalidade do país onde se realizariam os jogos seguintes ,
Question #48 - O que é o C 4?		
A 2008	1: verdade que apelidam grupo c de grupo da morte	**2008_06_09-19_59_01-Telejornal-1 bloco 3**: é verdade que apelidam grupo c de grupo da morte , mas que existiu roménia - França .
A 2008	2: aqui que encontramos o novo prédio a c c tv	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: é aqui que encontramos o novo prédio a c c tv .
A 2008	3: eterno	**2008_07_23-21_59_01-Jornal2-2 bloco 2**: é eterno janeiro o lugar da série c da teixeira duarte do conselho superior da tem pouco ou nenhum tem pouca importância na estrada da da do banco .
B 2008	1: c é uma linguagem de programação estruturada e padronizada criada na década de 1 970 por dennis ritchie e ken thompson para ser usada no sistema operacional unix . desde então espalhou - se por muitos outros sistemas operativos , e tornou - se uma das linguagens de programação mais usadas .	**pt/l/i/n/Linguagem_de_programação_C_2ef7.html**: c é uma linguagem de programação estruturada e padronizada criada na década de 1 970 por dennis ritchie e ken thompson para ser usada no sistema operacional unix . desde então espalhou - se por muitos outros sistemas operativos , e tornou - se uma das linguagens de programação mais usadas . c tem como ponto - forte a sua eficiência e é a linguagem de programação de preferência para o desenvolvimento de software de sistemas , apesar de também ser usada para desenvolver aplicações . é também muito usada no ensino de ciências da computação , mesmo não tendo sido projectada para novatos .

B 2008	2: o c + + (em português lê - se " cê mais mais " ou / sê / / máys / / máys / , em inglês lê - se " cee plus plus " ou / si / / plâs / / plâs /) é uma linguagem de programação de alto nível com facilidades para o uso em baixo nível , multiparadigma e de uso geral .	**pt/c/+/+/C++.html**: o c + + (em português lê - se " cê mais mais " ou / sê / / máys / / máys / , em inglês lê - se " cee plus plus " ou / si / / plâs / / plâs /) é uma linguagem de programação de alto nível com facilidades para o uso em baixo nível , multiparadigma e de uso geral . desde os anos 1 990 é uma das linguagens comerciais mais populares .
B 2008	3: esta página precisa ser reciclada . sinta - se livre para editá - la para que esta possa atingir um nível de qualidade superior . a letra c é a terceira letra do alfabeto latino e é a décima nona letra do alfabeto cirílico , na qual equivale foneticamente ao s. história na língua etrusca , as consoantes oclusivas não tinham uma pronunciação específica , por isso , usaram o ? (gama) grego para escrever o seu som / k / .	**pt/c/_/_/C.html**: esta página precisa ser reciclada . sinta - se livre para editá - la para que esta possa atingir um nível de qualidade superior . a letra c é a terceira letra do alfabeto latino e é a décima nona letra do alfabeto cirílico , na qual equivale foneticamente ao s. história na língua etrusca , as consoantes oclusivas não tinham uma pronunciação específica , por isso , usaram o ? (gama) grego para escrever o seu som / k / . no início , os romanos utilizavam o c tanto para representar o som / k / como o / g / .

C 2008	1: que puderam voar	**2008_06_29-19_59_01-Telejornal-1 bloco 5**: nas comemorações dos cinquenta e seis anos da força aérea , o ramo abriu as portas das suas bases às populações , que puderam voar de c cento e trinta , pela primeira vez , um domingo realiza - se o fim do mês ,
C 2008	2: ansioso	**2008_07_11-21_59_02-Jornal2-2 bloco 4**: levam a estar acordada uma grande pancada de alves cia de pensar vou dormir quatro horas por trabalhar de manhã e , ansioso , c dezassete nove estão a ser uma curiosidade comum , com estas coisas mas de folga maneira uma telefone anteriores à ,
C 2008	3: j em do sporting	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: paulo ferreira , ricardo carvalho e bosingwa , dor real madrid queiroz chamou central pepe , médios deco dos l c , j em do sporting , revelou meireles do porto do benfica queirós chamou carlos martins , e de espanha duda do málaga imanol fernandes do valência ,
D 2008	1: c é uma linguagem de programação estruturada e padronizada criada na década de 1 970 por dennis ritchie e ken thompson para ser usada no sistema operacional unix . desde então espalhou - se por muitos outros sistemas operativos , e tornou - se uma das linguagens de programação mais usadas .	**pt/l/i/n/Linguagem_de_programação_C_2ef7.html**: c é uma linguagem de programação estruturada e padronizada criada na década de 1 970 por dennis ritchie e ken thompson para ser usada no sistema operacional unix . desde então espalhou - se por muitos outros sistemas operativos , e tornou - se uma das linguagens de programação mais usadas . c tem como ponto - forte a sua eficiência e é a linguagem de programação de preferência para o desenvolvimento de software de sistemas , apesar de também ser usada para desenvolver aplicações . é também muito usada no ensino de ciências da computação , mesmo não tendo sido projectada para novatos .

D 2008	2: o c + + (em português lê - se " cê mais mais " ou / sê / / máys / / máys / , em inglês lê - se " cee plus plus " ou / si / / plâs / / plâs /) é uma linguagem de programação de alto nível com facilidades para o uso em baixo nível , multiparadigma e de uso geral .	**pt/c/+/+/C++.html**: o c + + (em português lê - se " cê mais mais " ou / sê / / máys / / máys / , em inglês lê - se " cee plus plus " ou / si / / plâs / / plâs /) é uma linguagem de programação de alto nível com facilidades para o uso em baixo nível , multiparadigma e de uso geral . desde os anos 1 990 é uma das linguagens comerciais mais populares .
D 2008	3: esta página precisa ser reciclada . sinta - se livre para editá - la para que esta possa atingir um nível de qualidade superior . a letra c é a terceira letra do alfabeto latino e é a décima nona letra do alfabeto cirílico , na qual equivale foneticamente ao s. história na língua etrusca , as consoantes oclusivas não tinham uma pronunciação específica , por isso , usaram o ? (gama) grego para escrever o seu som / k / .	**pt/c/_/_/C.html**: esta página precisa ser reciclada . sinta - se livre para editá - la para que esta possa atingir um nível de qualidade superior . a letra c é a terceira letra do alfabeto latino e é a décima nona letra do alfabeto cirílico , na qual equivale foneticamente ao s. história na língua etrusca , as consoantes oclusivas não tinham uma pronunciação específica , por isso , usaram o ? (gama) grego para escrever o seu som / k / . no início , os romanos utilizavam o c tanto para representar o som / k / como o / g / .

A 2010	1: toque c de ataques	**2008_06_09-19_59_01-Telejornal-1 bloco 4**: toque c de ataques .
A 2010	2: campeão	**2008_06_18-19_59_01-Telejornal-1 bloco 2**: portugal espera que possa mais oito assina e o mais irónico eu falo é c ser campeão .
A 2010	3: o país demasiado longos	**2008_08_03-21_59_01-Jornal2-2 bloco 1**: o país demasiado longos de c cento e trinta .
B 2010	1: c é uma linguagem de programação estruturada e padronizada criada na década de 1 970 por dennis ritchie e ken thompson para ser usada no sistema operacional unix . desde então espalhou - se por muitos outros sistemas operativos , e tornou - se uma das linguagens de programação mais usadas .	**pt/l/i/n/Linguagem_de_programação_C_2ef7.html**: c é uma linguagem de programação estruturada e padronizada criada na década de 1 970 por dennis ritchie e ken thompson para ser usada no sistema operacional unix . desde então espalhou - se por muitos outros sistemas operativos , e tornou - se uma das linguagens de programação mais usadas . c tem como ponto - forte a sua eficiência e é a linguagem de programação de preferência para o desenvolvimento de software de sistemas , apesar de também ser usada para desenvolver aplicações . é também muito usada no ensino de ciências da computação , mesmo não tendo sido projectada para novatos .
B 2010	2: o c + + (em português lê - se " cê mais mais " ou / sê / / máys / / máys / , em inglês lê - se " cee plus plus " ou / si / / plâs / / plâs /) é uma linguagem de programação de alto nível com facilidades para o uso em baixo nível , multiparadigma e de uso geral .	**pt/c/+/+/C++.html**: o c + + (em português lê - se " cê mais mais " ou / sê / / máys / / máys / , em inglês lê - se " cee plus plus " ou / si / / plâs / / plâs /) é uma linguagem de programação de alto nível com facilidades para o uso em baixo nível , multiparadigma e de uso geral . desde os anos 1 990 é uma das linguagens comerciais mais populares .

B 2010	3: esta página precisa ser reciclada . sinta - se livre para editá - la para que esta possa atingir um nível de qualidade superior . a letra c é a terceira letra do alfabeto latino e é a décima nona letra do alfabeto cirílico , na qual equivale foneticamente ao s. história na língua etrusca , as consoantes oclusivas não tinham uma pronunciação específica , por isso , usaram o ? (gama) grego para escrever o seu som / k / .	**pt/c/_/_/C.html**: esta página precisa ser reciclada . sinta - se livre para editá - la para que esta possa atingir um nível de qualidade superior . a letra c é a terceira letra do alfabeto latino e é a décima nona letra do alfabeto cirílico , na qual equivale foneticamente ao s. história na língua etrusca , as consoantes oclusivas não tinham uma pronunciação específica , por isso , usaram o ? (gama) grego para escrever o seu som / k / . no início , os romanos utilizavam o c tanto para representar o som / k / como o / g / .

C 2010	1: NIL	
D 2010	1: c é uma linguagem de programação estruturada e padronizada criada na década de 1 970 por dennis ritchie e ken thompson para ser usada no sistema operacional unix . desde então espalhou - se por muitos outros sistemas operativos , e tornou - se uma das linguagens de programação mais usadas .	**pt/l/i/n/Linguagem_de_programação_C_2ef7.html**: c é uma linguagem de programação estruturada e padronizada criada na década de 1 970 por dennis ritchie e ken thompson para ser usada no sistema operacional unix . desde então espalhou - se por muitos outros sistemas operativos , e tornou - se uma das linguagens de programação mais usadas . c tem como ponto - forte a sua eficiência e é a linguagem de programação de preferência para o desenvolvimento de software de sistemas , apesar de também ser usada para desenvolver aplicações . é também muito usada no ensino de ciências da computação , mesmo não tendo sido projectada para novatos .
D 2010	2: o c + + (em português lê - se " cê mais mais " ou / sê / / máys / / máys / , em inglês lê - se " cee plus plus " ou / si / / plâs / / plâs /) é uma linguagem de programação de alto nível com facilidades para o uso em baixo nível , multiparadigma e de uso geral .	**pt/c/+/+/C++.html**: o c + + (em português lê - se " cê mais mais " ou / sê / / máys / / máys / , em inglês lê - se " cee plus plus " ou / si / / plâs / / plâs /) é uma linguagem de programação de alto nível com facilidades para o uso em baixo nível , multiparadigma e de uso geral . desde os anos 1 990 é uma das linguagens comerciais mais populares .

D 2010	3: esta página precisa ser reciclada . sinta - se livre para editá - la para que esta possa atingir um nível de qualidade superior . a letra c é a terceira letra do alfabeto latino e é a décima nona letra do alfabeto cirílico , na qual equivale foneticamente ao s. história na língua etrusca , as consoantes oclusivas não tinham uma pronunciação específica , por isso , usaram o ? (gama) grego para escrever o seu som / k / .	**pt/c/_/_/C.html**: esta página precisa ser reciclada . sinta - se livre para editá - la para que esta possa atingir um nível de qualidade superior . a letra c é a terceira letra do alfabeto latino e é a décima nona letra do alfabeto cirílico , na qual equivale foneticamente ao s. história na língua etrusca , as consoantes oclusivas não tinham uma pronunciação específica , por isso , usaram o ? (gama) grego para escrever o seu som / k / . no início , os romanos utilizavam o c tanto para representar o som / k / como o / g / .

Question #49 - O que é necessário para iniciar este tipo de explosivo?		
A 2008	1: NIL	
B 2008	1: explosivo	**2008_08_21-19_59_01-Telejornal-1 bloco 2**: a para se iniciar este tipo de explosivo , é preciso , que queríamos um estímulo energético elevado , de sobre o qual como compreenderá por razões técnicas e eu não vou de longa .
B 2008	2: razões técnicas	**2008_08_21-19_59_01-Telejornal-1 bloco 2**: a para se iniciar este tipo de explosivo , é preciso , que queríamos um estímulo energético elevado , de sobre o qual como compreenderá por razões técnicas e eu não vou de longa .
B 2008	3: eu não vou	**2008_08_21-19_59_01-Telejornal-1 bloco 2**: a para se iniciar este tipo de explosivo , é preciso , que queríamos um estímulo energético elevado , de sobre o qual como compreenderá por razões técnicas e eu não vou de longa .
C 2008	1: NIL	
D 2008	1: queríamos um estímulo energético	**2008_08_21-19_59_01-Telejornal-1 bloco 2**: pode manusear transportar , com elevados padrões de segurança , a para se iniciar este tipo de explosivo , é preciso , que queríamos um estímulo energético elevado , de sobre o qual como compreenderá por razões técnicas e eu não vou de longa ,
D 2008	2: eu não vou	**2008_08_21-19_59_01-Telejornal-1 bloco 2**: pode manusear transportar , com elevados padrões de segurança , a para se iniciar este tipo de explosivo , é preciso , que queríamos um estímulo energético elevado , de sobre o qual como compreenderá por razões técnicas e eu não vou de longa ,
D 2008	3: imaginemos de plasticina usada na na nas escolas	**2008_08_21-19_59_01-Telejornal-1 bloco 2**: a para se iniciar este tipo de explosivo , é preciso , que queríamos um estímulo energético elevado , de sobre o qual como compreenderá por razões técnicas e eu não vou de longa , quatro com um imaginemos de plasticina usada na na nas escolas de
A 2010 B 2010 C 2010 D 2010	1: NIL	
Question #50 - Com que material usado nas escolas é que ele se parece?		
A 2008	1: NIL	
B 2008	1: mundo material	**pt/d/o/c/Docetismo.html**: docetismo . " para parecer ") é o nome dado a uma doutrina cristã do século ii , que defendia que o corpo de jesus cristo era uma ilusão , e que sua crucificação teria sido apenas aparente . a origem do docetismo é geralmente atribuída aos gnósticos para quem o mundo material

B 2008	2: homem	**pt/p/r/i/Primeira_Epístola_de_João_af6d.html**: primeira epístola de joão . material , que negava a possibilidade de deus (espírito) ter se tornado homem (matéria) . outra versão desta heresia , o docetismo (do grego dokeo , parecer) afirmava que a encarnação foi aparente , não uma realidade concreta .
B 2008	3: marfim	**pt/m/a/r/Marfim.html**: marfim . material) . o termo tem um sabor mais negativo atualmente , implicando a existencia dos especialistas que estão tão profundamente mergulhados em seus campos científicos que não conseguem se comunicar com o homem comum , não coneguem sair de sua " torre de marfim " .
C 2008	1: NIL	
D 2008	1: mundo material	**pt/d/o/c/Docetismo.html**: docetismo . " para parecer ") é o nome dado a uma doutrina cristã do século ii , que defendia que o corpo de jesus cristo era uma ilusão , e que sua crucificação teria sido apenas aparente . a origem do docetismo é geralmente atribuída aos gnósticos para quem o mundo material
D 2008	2: homem	**pt/p/r/i/Primeira_Epístola_de_João_af6d.html**: primeira epístola de joão . material , que negava a possibilidade de deus (espírito) ter se tornado homem (matéria) . outra versão desta heresia , o docetismo (do grego dokeo , parecer) afirmava que a encarnação foi aparente , não uma realidade concreta .
D 2008	3: marfim	**pt/m/a/r/Marfim.html**: marfim . material) . o termo tem um sabor mais negativo atualmente , implicando a existencia dos especialistas que estão tão profundamente mergulhados em seus campos científicos que não conseguem se comunicar com o homem comum , não coneguem sair de sua " torre de marfim " .
A 2010	1: NIL	
B 2010	1: mundo material	**pt/d/o/c/Docetismo.html**: docetismo . " para parecer ") é o nome dado a uma doutrina cristã do século ii , que defendia que o corpo de jesus cristo era uma ilusão , e que sua crucificação teria sido apenas aparente . a origem do docetismo é geralmente atribuída aos gnósticos para quem o mundo material
B 2010	2: homem	**pt/p/r/i/Primeira_Epístola_de_João_af6d.html**: primeira epístola de joão . material , que negava a possibilidade de deus (espírito) ter se tornado homem (matéria) . outra versão desta heresia , o docetismo (do grego dokeo , parecer) afirmava que a encarnação foi aparente , não uma realidade concreta .
B 2010	3: marfim	**pt/m/a/r/Marfim.html**: marfim . material) . o termo tem um sabor mais negativo atualmente , implicando a existencia dos especialistas que estão tão profundamente mergulhados em seus campos científicos que não conseguem se comunicar com o homem comum , não coneguem sair de sua " torre de marfim " .

C 2010	1: NIL	
D 2010	1: mundo material	**pt/d/o/c/Docetismo.html**: docetismo . " para parecer ") é o nome dado a uma doutrina cristã do século ii , que defendia que o corpo de jesus cristo era uma ilusão , e que sua crucificação teria sido apenas aparente . a origem do docetismo é geralmente atribuída aos gnósticos para quem o mundo material
D 2010	2: homem	**pt/p/r/i/Primeira_Epístola_de_João_af6d.html**: primeira epístola de joão . material , que negava a possibilidade de deus (espírito) ter se tornado homem (matéria) . outra versão desta heresia , o docetismo (do grego dokeo , parecer) afirmava que a encarnação foi aparente , não uma realidade concreta .
D 2010	3: marfim	**pt/m/a/r/Marfim.html**: marfim . material) . o termo tem um sabor mais negativo atualmente , implicando a existencia dos especialistas que estão tão profundamente mergulhados em seus campos científicos que não conseguem se comunicar com o homem comum , não coneguem sair de sua " torre de marfim " .
Question #51 - Em que festa não vai Manuela Ferreira Leite participar?		
A 2008	1: NIL	
B 2008	1: principal	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: ferreira leite não vai participar na festa do pontal que decorre esta noite no algarve . o orador principal será ângelo correia , festa do pontal que se realiza em quarteira , deixou de ter significado político que lhe foi atribuído , nas décadas de oitenta e noventa . na liderança de cavaco silva . na antiga
B 2008	2: folclore	**2008_07_27-21_59_01-Jornal2-2 bloco 1**: popular do chão da lagoa , para dirigir as primeiras palavras ao líder nacional do partido . manuela ferreira leite , que nos açores . criticou o folclore partidários . é natural . de olhos abertos ou . está bem entregue exaltando a validade . é natural , que nós tem estado melhor . cada ontem joschka
B 2008	3: festa popular	**2008_07_27-19_59_01-Telejornal-1 bloco 2**: festa popular do chão da lagoa , para dirigir as primeiras palavras ao líder nacional do partido . manuela ferreira leite , que nos açores . criticou o folclore partidários . é natural . para os peritos . está bem entregue exaltando a validade . é natural , que nós tem estado melhor . cada ontem joschka dão
C 2008	1: NIL	
D 2008	1: principal	**2008_08_14-21_59_01-Jornal2-2 bloco 1**: manuela ferreira leite não vai participar na festa do pombal que decorre do pontal aliás que decorre esta noite no algarve , o orador principal é ângelo correia , a festa do pontal que se realiza em quarteira ,

D 2008	2: folclore	**2008_07_27-21_59_01-Jornal2-2 bloco 1**: popular do chão da lagoa , para dirigir as primeiras palavras ao líder nacional do partido , manuela ferreira leite , que nos açores , criticou o folclore partidários , é natural , de olhos abertos ou , está bem entregue exaltando a validade , é natural , que nós tem estado melhor , cada ontem joschka
D 2008	3: festa popular	**2008_07_27-19_59_01-Telejornal-1 bloco 2**: festa popular do chão da lagoa , para dirigir as primeiras palavras ao líder nacional do partido , manuela ferreira leite , que nos açores , criticou o folclore partidários , é natural , para os peritos , está bem entregue exaltando a validade , é natural , que nós tem estado melhor , cada ontem joschka dão
A 2010	1: NIL	
B 2010	1: principal	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: antiga festa do pontal . só ficou nove . agora o palco é montado em quarteira o comício deixou de assinalar o novo ano político e esta noite o orador principal também não é líder do partido . manuela ferreira leite não vai
B 2010	2: festa popular	**2008_07_27-21_59_01-Jornal2-2 bloco 1**: o presidente do psd - madeira aproveitou a festa popular do chão da lagoa para dirigir as primeiras palavras ao líder nacional do partido . manuela ferreira leite que nos açores criticou o folclore partidário .
B 2010	3: folclore	**2008_07_27-21_59_01-Jornal2-2 bloco 1**: o presidente do psd - madeira aproveitou a festa popular do chão da lagoa para dirigir as primeiras palavras ao líder nacional do partido . manuela ferreira leite que nos açores criticou o folclore partidário .
C 2010	1: NIL	
D 2010	1: principal	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: manuela ferreira leite não vai participar na festa do pombal que decorre do pontal , aliás que decorre esta noite no algarve , o orador principal é ângelo correia , a festa
Question #52 - Que prémio ganhou Elvira Fortunato?		
A 2008	1: NIL	
B 2008	1: prémio camões	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: para a língua portuguesa é um grande cultor da língua portuguesa , e de um prémio camões ficasse mais uma vez de reforçado . na sua qualidade . a cientista portuguesa elvira fortunato ganhou maior prémio alguma vez atribuído pelo europeu em insectos que não ser considerado .

B 2008	2: prémios nobel	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a cientista portuguesa elvira fortunato ganhou maior prémio alguma vez atribuído pelo europeu em insectos que não ser considerado . uma espécie de prémios nobel na área da engenharia investigadora ,
C 2008	1: NIL	
D 2008	1: prémio camões	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: e de um prémio camões ficasse mais uma vez de reforçado , na sua qualidade , a cientista portuguesa elvira fortunato ganhou maior prémio alguma vez atribuído pelo europeu em insectos que não ser considerado , uma espécie de prémios nobel na área da engenharia investigadora ,
A 2010	1: NIL	
B 2010	1: prémio camões	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: ie um prémio camões ficasse mais uma vez que reforçado . na sua qualidade . assim disse a portuguesa elvira fortunato ganhou maior prémio alguma vez atribuído pelo eu pela insatisfação ser considerado .
B 2010	2: prémios nobel	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: elvira fortunato ganhou maior prémio alguma vez atribuído pelo eu pela insatisfação ser considerado . uma espécie de prémios nobel na área da engenharia investigadora liderou a equipa da universidade nova que conseguiu produzir transístores com papel que não usam silício com isolador de curto - circuitos . dois milhões e meio de euros o maior prémio
C 2010	1: NIL	
D 2010	1: prémios nobel	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: elvira fortunato ganhou maior prémio alguma vez atribuído pelo eu pela insatisfação ser considerado , uma espécie de prémios nobel na área da engenharia investigadora liderou a equipa da universidade nova que conseguiu produzir transístores com papel que não usam silício com isolador de curto - circuitos , dois milhões e meio de euros o maior prémio
D 2010	2: prémio camões	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: para nós própria emi para a língua portuguesa é um grande cultor da língua portuguesa , ie um prémio camões ficasse mais uma vez que reforçado , na sua qualidade , assim disse a portuguesa elvira fortunato ganhou maior prémio alguma vez atribuído pelo eu pela insatisfação ser considerado ,
Question #53 - Quantos milhões de crianças podem ficar órfãs em África, segundo as Nações Unidas?		
A 2008	1: NIL	
B 2008	1: três milhões	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: segundo as estimativas da onu sida existe em todo o mundo . trinta e três milhões de infectados . as nações unidas calculam , que em dois mil e dez a doença pode deixar mais de cinquenta e três milhões de crianças . órfãs em áfrica . já jogou está

C 2008	1: NIL	
D 2008	1: três milhões	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: segundo as estimativas da onu sida existe em todo o mundo , trinta e três milhões de infectados , as nações unidas calculam , que em dois mil e dez a doença pode deixar mais de cinquenta e três milhões de crianças , órfãs em áfrica , já jogou está
A 2010	1: NIL	
B 2010	1: três milhões	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: segundo estimativas da onu sida existe em todo o mundo . trinta e três milhões de infectados . as nações unidas calculou que em dois mil e dez a doença pode deixar mais de cinquenta e três milhões de crianças ora faz em áfrica .
C 2010	1: NIL	
D 2010	1: três milhões	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: segundo estimativas da onu sida existe em todo o mundo , trinta e três milhões de infectados , as nações unidas calculou que em dois mil e dez a doença pode deixar mais de cinquenta e três milhões de crianças ora faz em áfrica , jorge bush está na coreia do soho é a
Question #54 - Quais os concorrentes dos empresários Portugueses em Angola?		
A 2008	1: cá , como angolanos , e para portugal	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: são já mais de sessenta mil os portugueses em angola , o fluxo migratório mudou , agora há tantos portugueses para vir para cá , como angolanos , e para portugal .
A 2008	2: fluxo	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: são já mais de sessenta mil os portugueses em angola , o fluxo migratório mudou , agora há tantos portugueses para vir para cá , como angolanos , e para portugal .
B 2008	1: angola , o fluxo migratório mudou , agora há tantos portugueses para vir para cá , como angolanos , e para portugal	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: são já mais de sessenta mil os portugueses em angola , o fluxo migratório mudou , agora há tantos portugueses para vir para cá , como angolanos , e para portugal .
B 2008	2: colonial , visto que neto era casado com uma mulher branca , que acabou por abandonar juntamnete com 3 filhos	**pt/a/g/o/Agostinho_Neto_c15c.html**: agostinho neto . essas ideias foram uma das causas da saidade portugueses em angola após a guerra colonial , visto que neto era casado com uma mulher branca , que acabou por abandonar juntamnete com 3 filhos .

B 2008	3: agostinho neto	**pt/a/g/o/Agostinho_Neto_c15c.html**: agostinho neto . essas ideias foram uma das causas da saidade portugueses em angola após a guerra colonial , visto que neto era casado com uma mulher branca , que acabou por abandonar juntamnete com 3 filhos .
C 2008	1: angolanos	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: são já mais de sessenta mil os portugueses em angola , o fluxo migratório mudou , agora há tantos portugueses para vir para cá , como angolanos , e para portugal ,
C 2008	2: cá	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: são já mais de sessenta mil os portugueses em angola , o fluxo migratório mudou , agora há tantos portugueses para vir para cá , como angolanos , e para portugal ,
C 2008	3: fluxo	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: são já mais de sessenta mil os portugueses em angola , o fluxo migratório mudou , agora há tantos portugueses para vir para cá , como angolanos , e para portugal ,
D 2008	1: colonial , visto que neto era casado com uma mulher branca , que acabou por abandonar juntamnete com 3 filhos	**pt/a/g/o/Agostinho_Neto_c15c.html**: agostinho neto . essas ideias foram uma das causas da saidade portugueses em angola após a guerra colonial , visto que neto era casado com uma mulher branca , que acabou por abandonar juntamnete com 3 filhos .
D 2008	2: lista de governadores de angola	**pt/l/i/s/Lista_de_governadores_de_Angola_1c41.html**: lista de governadores de angola . lista de governadores de angola segue - se uma lista dos governadores portugueses em angola .
D 2008	3: agostinho neto	**pt/a/g/o/Agostinho_Neto_c15c.html**: agostinho neto . essas ideias foram uma das causas da saidade portugueses em angola após a guerra colonial , visto que neto era casado com uma mulher branca , que acabou por abandonar juntamnete com 3 filhos .
A 2010	1: água do rio bengo fica eternamente ligado	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: em lisboa torna luanda ainda mais atractivo . por aqui diz que quem beber água do rio bengo fica eternamente ligado . angola crer na lenda e aplicando aos negócios isso significa para os empresários portugueses uma vantagem que não é mágica , mas histórica face aos concorrentes
A 2010	2: histórica face	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: em lisboa torna luanda ainda mais atractivo . por aqui diz que quem beber água do rio bengo fica eternamente ligado . angola crer na lenda e aplicando aos negócios isso significa para os empresários portugueses uma vantagem que não é mágica , mas histórica face aos concorrentes

B 2010	1: colonial , visto que neto era casado com uma mulher branca , que acabou por abandonar juntamnete com 3 filhos	**pt/a/g/o/Agostinho_Neto_c15c.html**: agostinho neto . essas ideias foram uma das causas da saidade portugueses em angola após a guerra colonial , visto que neto era casado com uma mulher branca , que acabou por abandonar juntamnete com 3 filhos .
B 2010	2: lista de governadores de angola	**pt/l/i/s/Lista_de_governadores_de_Angola_1c41.html**: lista de governadores de angola . lista de governadores de angola segue - se uma lista dos governadores portugueses em angola .
B 2010	3: agostinho neto	**pt/a/g/o/Agostinho_Neto_c15c.html**: agostinho neto . essas ideias foram uma das causas da saidade portugueses em angola após a guerra colonial , visto que neto era casado com uma mulher branca , que acabou por abandonar juntamnete com 3 filhos .
C 2010	1: angolanos	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: é , são já mais de sessenta mil os portugueses em angola , o fluxo migratório mudou , agora há tantos portugueses vir para cá , como angolanos e para portugal , é um bom mercado a mercado ,
C 2010	2: cá	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: é , são já mais de sessenta mil os portugueses em angola , o fluxo migratório mudou , agora há tantos portugueses vir para cá , como angolanos e para portugal , é um bom mercado a mercado ,
C 2010	3: fluxo	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: é , são já mais de sessenta mil os portugueses em angola , o fluxo migratório mudou , agora há tantos portugueses vir para cá , como angolanos e para portugal , é um bom mercado a mercado ,
D 2010	1: colonial , visto que neto era casado com uma mulher branca , que acabou por abandonar juntamnete com 3 filhos	**pt/a/g/o/Agostinho_Neto_c15c.html**: agostinho neto . essas ideias foram uma das causas da saidade portugueses em angola após a guerra colonial , visto que neto era casado com uma mulher branca , que acabou por abandonar juntamnete com 3 filhos .
D 2010	2: lista de governadores de angola	**pt/l/i/s/Lista_de_governadores_de_Angola_1c41.html**: lista de governadores de angola . lista de governadores de angola segue - se uma lista dos governadores portugueses em angola .
D 2010	3: agostinho neto	**pt/a/g/o/Agostinho_Neto_c15c.html**: agostinho neto . essas ideias foram uma das causas da saidade portugueses em angola após a guerra colonial , visto que neto era casado com uma mulher branca , que acabou por abandonar juntamnete com 3 filhos .

Question #55 - Qual o guarda-redes do Benfica para o encontro frente às Ilhas Faroé?		
A 2008	1: chamou	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: para o encontro frente às ilhas faroé , queirós chamou três guarda - redes , quim do benfica .
A 2008	2: fernandes de boca	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: para o encontro frente às ilhas faroé . queiroz chamou três guarda - redes , quim do benfica , e eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me .
A 2008	3: eduardo braga	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: para o encontro frente às ilhas faroé . queiroz chamou três guarda - redes , quim do benfica , e eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me .
B 2008	1: quim	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: para o encontro frente às ilhas faroé , queirós chamou três guarda - redes , quim do benfica .
B 2008	2: daniel fernandes de boca me	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: para o encontro frente às ilhas faroé . queiroz chamou três guarda - redes , quim do benfica , e eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me .
B 2008	3: eduardo braga chamado pela primeira vez	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: para o encontro frente às ilhas faroé . queiroz chamou três guarda - redes , quim do benfica , e eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me .
C 2008	1: eduardo braga	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: não se sabe se esta opção é circunstancial , ou longo prazo , para o encontro frente às ilhas faroé , queirós chamou três guarda - redes , quim do benfica , eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me ,
C 2008	2: chamou	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: não se sabe se esta opção é circunstancial , ou longo prazo , para o encontro frente às ilhas faroé , queirós chamou três guarda - redes , quim do benfica , eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me ,
C 2008	3: longo prazo	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: não se sabe se esta opção é circunstancial , ou longo prazo , para o encontro frente às ilhas faroé , queirós chamou três guarda - redes , quim do benfica , eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me ,
D 2008	1: eduardo braga chamado pela primeira vez à selecção nacional	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: não se sabe se esta opção é circunstancial , ou longo prazo , para o encontro frente às ilhas faroé , queirós chamou três guarda - redes , quim do benfica , eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me ,

D 2008	2: daniel fernandes de boca me	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: não se sabe se esta opção é circunstancial , ou longo prazo , para o encontro frente às ilhas faroé , queirós chamou três guarda - redes , quim do benfica , eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me ,
D 2008	3: braga chamado pela primeira vez à selecção nacional	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: não se sabe se esta opção é circunstancial , ou longo prazo , para o encontro frente às ilhas faroé , queirós chamou três guarda - redes , quim do benfica , eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me ,
A 2010	1: chamou	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me .
A 2010	2: fernandes de boca	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me .
A 2010	3: eduardo braga	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me .
B 2010	1: quim do benfica eduardo braga chamado pela primeira	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me .
B 2010	2: queiroz chamou	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me .
B 2010	3: daniel fernandes de boca me	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me .
C 2010	1: chamou	**2008_08_14-21_59_01-Jornal2-2 bloco 3**: a federação portuguesa de futebol não se sabe se esta opção é circunstancial ou longo prazo , para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me ,
C 2010	2: fernandes de boca	**2008_08_14-21_59_01-Jornal2-2 bloco 3**: a federação portuguesa de futebol não se sabe se esta opção é circunstancial ou longo prazo , para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me ,

C 2010	3: eduardo braga	**2008_08_14-21_59_01-Jornal2-2 bloco 3**: a federação portuguesa de futebol não se sabe se esta opção é circunstancial ou longo prazo , para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me ,
D 2010	1: quim do benfica eduardo braga chamado pela primeira vez	**2008_08_14-21_59_01-Jornal2-2 bloco 3**: a federação portuguesa de futebol não se sabe se esta opção é circunstancial ou longo prazo , para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me ,
D 2010	2: daniel fernandes de boca me	**2008_08_14-21_59_01-Jornal2-2 bloco 3**: a federação portuguesa de futebol não se sabe se esta opção é circunstancial ou longo prazo , para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me ,
D 2010	3: federação portuguesa de futebol não se sabe se esta opção é circunstancial	**2008_08_14-21_59_01-Jornal2-2 bloco 3**: a federação portuguesa de futebol não se sabe se esta opção é circunstancial ou longo prazo , para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me ,
Question #56 - Qual o do Braga?		
A 2008	1: chamou o médio dani	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: por exemplo , ricardo quaresma , e chamou o médio dani do dínamo de moscovo , e o guarda - redes eduardo de braga .
A 2008	2: médio dani do dínamo de moscovo	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: por exemplo , ricardo quaresma , e chamou o médio dani do dínamo de moscovo , e o guarda - redes eduardo de braga .
A 2008	3: fernandes de boca	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: queiroz chamou três guarda - redes , quim do benfica , e eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me .
B 2008	1: josé filipe da silva moreira gr sport comércio e salgueiros 12 português joaquim	**pt/s/p/o/Sport_Lisboa_e_Benfica_44be.html**: sport lisboa e benfica . guarda - redes 1 português josé filipe da silva moreira gr sport comércio e salgueiros 12 português joaquim manuel sampaio silva , " quim " gr sporting clube de braga
B 2008	2: quim do benfica	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: queiroz chamou três guarda - redes , quim do benfica , e eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me .
B 2008	3: eduardo braga chamado pela primeira vez à selecção nacional	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: queiroz chamou três guarda - redes , quim do benfica , e eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me .

C 2008	1: ricardo o bétis de sevilha	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: do guarda - redes e do guarda , do braga , ele esteve ou para luiz felipe scolari o guarda - redes ricardo o bétis de sevilha , ficou de fora da primeira lista de carlos queiroz ,
C 2008	2: chamou	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: ou longo prazo , para o encontro frente às ilhas faroé , queirós chamou três guarda - redes , quim do benfica , eduardo braga chamado pela primeira vez à selecção nacional , e daniel fernandes de boca me ,
C 2008	3: dínamo de moscovo	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: por exemplo , ricardo quaresma , e chamou o médio dani do dínamo de moscovo , e o guarda - redes eduardo de braga , ele esteve ou para luiz felipe scolari o guarda - redes ricardo o bétis de sevilha ,
D 2008	1: josé filipe da silva moreira gr sport comércio e salgueiros	**pt/s/p/o/Sport_Lisboa_e_Benfica_44be.html**: sport lisboa e benfica . guarda - redes 1 português josé filipe da silva moreira gr sport comércio e salgueiros 12 português joaquim manuel sampaio silva , " quim " gr sporting clube de braga
D 2008	2: ricardo o bétis de sevilha	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: do guarda - redes e do guarda , do braga , ele esteve ou para luiz felipe scolari o guarda - redes ricardo o bétis de sevilha , ficou de fora da primeira lista de carlos queiroz ,
D 2008	3: sport lisboa e benfica	**pt/s/p/o/Sport_Lisboa_e_Benfica_44be.html**: sport lisboa e benfica . guarda - redes 1 português josé filipe da silva moreira gr sport comércio e salgueiros 12 português joaquim manuel sampaio silva , " quim " gr sporting clube de braga
A 2010	1: chamou o médio dani do dínamo	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: por exemplo , ricardo quaresma e chamou o médio dani do dínamo de moscovo e o guarda - redes eduardo do braga .
A 2010	2: eduardo braga chamado	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me .
A 2010	3: fernandes de boca	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me .
B 2010	1: quim do benfica eduardo braga chamado pela primeira	**2008_08_14-21_59_01-Jornal2-2 bloco 3**: para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me .
B 2010	2: josé filipe da silva moreira gr sport comércio e salgueiros 12 português joaquim	**pt/s/p/o/Sport_Lisboa_e_Benfica_44be.html**: sport lisboa e benfica . guarda - redes 1 português josé filipe da silva moreira gr sport comércio e salgueiros 12 português joaquim manuel sampaio silva , " quim " gr sporting clube de braga

B 2010	3: sporting clube de braga	**pt/s/p/o/Sport_Lisboa_e_Benfica_44be.html**: sport lisboa e benfica . guarda - redes 1 português josé filipe da silva moreira gr sport comércio e salgueiros 12 português joaquim manuel sampaio silva , " quim " gr sporting clube de braga
C 2010	1: indiscutível para luiz felipe scolari	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: por exemplo , ricardo quaresma e chamou o médio dani do dínamo de moscovo e o guarda - redes eduardo do braga , indiscutível para luiz felipe scolari o guarda - redes ricardo bétis de sevilha ficou de fora da primeira lista de carlos queirós ,
C 2010	2: ilhas faroé queiroz chamou	**2008_08_14-21_59_01-Jornal2-2 bloco 3**: para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me ,
C 2010	3: médio dani do dínamo de moscovo	**2008_08_14-19_59_01-Telejornal-1 bloco 9**: por exemplo , ricardo quaresma e chamou o médio dani do dínamo de moscovo e o guarda - redes eduardo do braga , indiscutível para luiz felipe scolari o guarda - redes ricardo bétis de sevilha ficou de fora da primeira lista de carlos queirós ,
D 2010	1: quim do benfica eduardo braga	**2008_08_14-21_59_01-Jornal2-2 bloco 3**: para o encontro frente às ilhas faroé queiroz chamou três guarda - redes quim do benfica eduardo braga chamado pela primeira vez à selecção nacional e daniel fernandes de boca me ,
D 2010	2: josé filipe da silva moreira gr sport comércio e salgueiros	**pt/s/p/o/Sport_Lisboa_e_Benfica_44be.html**: sport lisboa e benfica . guarda - redes 1 português josé filipe da silva moreira gr sport comércio e salgueiros 12 português joaquim manuel sampaio silva , " quim " gr sporting clube de braga
D 2010	3: sporting clube de braga	**pt/s/p/o/Sport_Lisboa_e_Benfica_44be.html**: sport lisboa e benfica . guarda - redes 1 português josé filipe da silva moreira gr sport comércio e salgueiros 12 português joaquim manuel sampaio silva , " quim " gr sporting clube de braga
Question #57 - Quem é Carlos do Carmo?		
A 2008	1: camané	**2008_06_29-21_59_02-Jornal2-2 bloco 4**: e recordar que entre vivos e mortos , os seus mais amados fadistas de alfredo marceneiro hermínia silva , passando quarto carlos do carmo e camané .
A 2008	2: marceneiro	**2008_06_29-21_59_02-Jornal2-2 bloco 4**: e recordar que entre vivos e mortos , os seus mais amados fadistas de alfredo marceneiro hermínia silva , passando quarto carlos do carmo e camané .
A 2008	3: hancock	**2008_06_29-21_59_02-Jornal2-2 bloco 4**: carlos do carmo e camané . esta é a ideia de missão e que iremos junto pode ver e ouvir ao vivo , javier hancock , esses fantásticos músicos de jazz suscita por estes dias nos muitos festivais , que acontece pelo país . na missão em que queremos as melhores novidades dos livros do

B 2008	1: carlos do carmo , de seu nome completo carlos alberto do carmo almeida , (n. lisboa , 1 939) , é um dos mais famoso fadistas portugueses .	**pt/c/a/r/Carlos_do_Carmo_6f9b.html**: carlos do carmo , de seu nome completo carlos alberto do carmo almeida , (n. lisboa , 1 939) , é um dos mais famoso fadistas portugueses . após ter estudado na suíça , iniciou a sua vida artística em 1 964 , embora tivesse já gravado um disco aos nove anos . tendo sido sempre muito influenciado pelo fado , representou portugal no xxi concurso eurovisão da canção em 1 976 , com o tema flor de verde pinho .
C 2008	1: nenhum momento de entrar em cena alcatraz de	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: nenhum momento de entrar em cena alcatraz de , carlos do carmo grande senhora do fado português que visita em claro a sua carreira , de mais de quarenta de entrada de elites , lula pena outra voz do fado de geração ,
C 2008	2: camané	**2008_06_29-21_59_02-Jornal2-2 bloco 4**: e recordar que entre vivos e mortos , os seus mais amados fadistas de alfredo marceneiro hermínia silva , passando quarto carlos do carmo e camané , esta é a ideia de missão e que iremos junto pode ver e ouvir ao vivo , javier hancock ,
C 2008	3: marceneiro	**2008_06_29-21_59_02-Jornal2-2 bloco 4**: e recordar que entre vivos e mortos , os seus mais amados fadistas de alfredo marceneiro hermínia silva , passando quarto carlos do carmo e camané , esta é a ideia de missão e que iremos junto pode ver e ouvir ao vivo , javier hancock ,
D 2008	1: carlos do carmo , de seu nome completo carlos alberto do carmo almeida , (n. lisboa , 1 939) , é um dos mais famoso fadistas portugueses .	**pt/c/a/r/Carlos_do_Carmo_6f9b.html**: carlos do carmo , de seu nome completo carlos alberto do carmo almeida , (n. lisboa , 1 939) , é um dos mais famoso fadistas portugueses . após ter estudado na suíça , iniciou a sua vida artística em 1 964 , embora tivesse já gravado um disco aos nove anos . tendo sido sempre muito influenciado pelo fado , representou portugal no xxi concurso eurovisão da canção em 1 976 , com o tema flor de verde pinho .
A 2010	1: camané	**2008_06_29-21_59_02-Jornal2-2 bloco 4**: passando claro por carlos do carmo e camané .
B 2010	1: carlos do carmo , de seu nome completo carlos alberto do carmo almeida , (n. lisboa , 1 939) , é um dos mais famoso fadistas portugueses .	**pt/c/a/r/Carlos_do_Carmo_6f9b.html**: carlos do carmo , de seu nome completo carlos alberto do carmo almeida , (n. lisboa , 1 939) , é um dos mais famoso fadistas portugueses . após ter estudado na suíça , iniciou a sua vida artística em 1 964 , embora tivesse já gravado um disco aos nove anos . tendo sido sempre muito influenciado pelo fado , representou portugal no xxi concurso eurovisão da canção em 1 976 , com o tema flor de verde pinho .

C 2010	1: camané	**2008_06_29-21_59_02-Jornal2-2 bloco 4**: passando claro por carlos do carmo e camané ,
C 2010	2: momento de entrar em cena ao cartaz de	**2008_08_14-21_59_01-Jornal2-2 bloco 3**: um momento de entrar em cena ao cartaz de , carlos do carmo grande senhora do fado português que visita em que levaram a sua carreira de mais de quarenta de entrada de elites , lula pena outra voz do fado de outra geração ,
C 2010	3: diremos	**2008_06_29-21_59_02-Jornal2-2 bloco 4**: carlos do carmo e camané , esta ainda a emissão em que diremos junto pode ver e ouvir ao vivo javier cork estes fantásticos músicos de jazz mas cita por estes dias nos muitos festivais que acontecem pelo país e é missão em que entraremos as melhores novidades os livros do
D 2010	1: carlos do carmo , de seu nome completo carlos alberto do carmo almeida , (n. lisboa , 1 939) , é um dos mais famoso fadistas portugueses .	**pt/c/a/r/Carlos_do_Carmo_6f9b.html**: carlos do carmo , de seu nome completo carlos alberto do carmo almeida , (n. lisboa , 1 939) , é um dos mais famoso fadistas portugueses . após ter estudado na suíça , iniciou a sua vida artística em 1 964 , embora tivesse já gravado um disco aos nove anos . tendo sido sempre muito influenciado pelo fado , representou portugal no xxi concurso eurovisão da canção em 1 976 , com o tema flor de verde pinho .
Question #58 - Quem foi o primeiro a fazer um comício no Pontal?		
A 2008	1: ângelo	**2008_08_07-19_59_02-Telejornal-1 bloco 3**: o discurso final do comício do pontal , cabe este ano ângelo correia acusa líder de abandonar o partido .
A 2008	2: trm ausência	**2008_08_22-21_59_02-Jornal2-2 bloco 2**: começa pelo silêncio das últimas semanas , e trm ausência na festa do pontal .
A 2008	3: convidado	**2008_09_06-21_59_01-Jornal2-2 bloco 1**: e não foi convidado para a festa do pontal .
B 2008	1: sá carneiro no verão quente	**2008_08_14-21_59_01-Jornal2-2 bloco 1**: o primeiro a fazer um comício do pontal foi sá carneiro no verão quente de setenta e cinco .
B 2008	2: cavaco silva	**2008_08_14-21_59_01-Jornal2-2 bloco 1**: ter significado político que lhe foi atribuído nas décadas de oitenta e noventa . na liderança de cavaco silva . na antiga festa do pontal . só ficou nove . agora ao palco é montado em quarteira , o comício deixou de assinalar o novo ano político , e esta noite o orador principal ,
B 2008	3: marques mendes	**2008_08_07-19_59_02-Telejornal-1 bloco 3**: este não é a primeira vez que um líder não vai ao comício do pontal . há dois anos , marques mendes também não esteve no algarve .
C 2008	1: orador principal é ângelo	**2008_08_14-21_59_01-Jornal2-2 bloco 1**: manuela ferreira leite não vai participar na festa do pombal que decorre do pontal aliás que decorre esta noite no algarve , o orador principal é ângelo correia , a festa do pontal que se realiza em quarteira ,

C 2008	2: octávio	**2008_09_06-18_59_01-Telejornal-1 bloco 3**: desafio com todo o gosto , de é bastante interessante disse ser esta a dinâmica , e não foi convidado para a festa do pontal , não não fui convidado para nenhuma festa do psd , acabou à conversa com o ex - líder parlamentar comunista octávio teixeira , sobre esses sonhos ,
C 2008	3: popularucha	**2008_08_15-21_59_01-Jornal2-2 bloco 2**: aqueles que podiam ter vindo mas não quiseram , para lhes dizer , que serão sempre bem - vindos no futuro , porque a festa do pontal está cá para ficar , é uma festa popular , mas não é uma festa popularucha , não andamos aqui aos por ,
D 2008	1: só carneiro no verão quente	**2008_08_14-21_59_01-Jornal2-2 bloco 1**: para a história ficam os discursos que marcava o reinício da actividade política , o primeiro a fazer um comício do pontal foi só carneiro no verão quente de setenta e cinco , nazaré com cavaco silva que festa ganha maior importância política ,
D 2008	2: cavaco silva	**2008_08_14-21_59_01-Jornal2-2 bloco 1**: para a história ficam os discursos que marcava o reinício da actividade política , o primeiro a fazer um comício do pontal foi só carneiro no verão quente de setenta e cinco , nazaré com cavaco silva que festa ganha maior importância política ,
D 2008	3: ângelo correia	**2008_08_07-21_59_02-Jornal2-2 bloco 2**: a direcção do psd não comentar as críticas de ângelo correia , na história do psd , este não é a primeira vez que um líder não vai ao comício do pontal , há dois anos , marques mendes também não esteve no algarve , nunca antes tinha acontecido ,
A 2010	1: ângelo	**2008_08_07-21_59_02-Jornal2-2 bloco 2**: o discurso final do comício do pontal cabe este ano ângelo correia que acusa líder de abandonar o partido .
A 2010	2: vindos no futuro	**2008_08_15-21_59_01-Jornal2-2 bloco 2**: serão sempre bem - vindos no futuro , porque a festa do pontal está cá para ficar .
A 2010	3: descanso	**2008_08_22-19_59_02-Telejornal-1 bloco 2**: começa pelo silêncio nas últimas semanas e tó ausência na festa do pontal comenta que o descanso é descanso .
B 2010	1: só carneiro no verão quente	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: para a história ficam os discursos que marcavam rinite ciudad actividade política o primeiro a fazer um comício do pontal foi só carneiro no verão quente de setenta e cinco .
B 2010	2: ângelo correia	**2008_08_07-19_59_02-Telejornal-1 bloco 3**: criar o partido pedir unidade do partido . a direcção do psd não comentar as críticas de ângelo correia . na história do psd . este não é a primeira vez que um líder não vai ao comício do pontal .
B 2010	3: cavaco silva	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: ter significado político que lhe foi atribuído nas décadas de oitenta e noventa . na liderança de cavaco silva . antiga festa do pontal . só ficou nove . agora o palco é montado em quarteira o comício deixou de assinalar o novo ano político e esta noite o orador principal

C 2010	1: popularucha	**2008_08_15-21_59_01-Jornal2-2 bloco 2**: para lhes dizer , serão sempre bem - vindos no futuro , porque a festa do pontal está cá para ficar , é uma festa popular , mas não é uma festa popularucha não andamos aqui aos pulos ,
C 2010	2: orador principal é ângelo	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: manuela ferreira leite não vai participar na festa do pombal que decorre do pontal , aliás que decorre esta noite no algarve , o orador principal é ângelo correia ,
C 2010	3: descanso	**2008_08_22-21_59_02-Jornal2-2 bloco 1**: começa pelo silêncio nas últimas semanas e tó ausência na festa do pontal comenta que o descanso é descanso ,
D 2010	1: sá carneiro no verão quente	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: para a história ficam os discursos que marcavam rinite ciudad actividade política o primeiro a fazer um comício do pontal foi sá carneiro no verão quente de setenta e cinco , no zé com cavaco silva que festa ganha maior importância política ,
D 2010	2: cavaco silva	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: para a história ficam os discursos que marcavam rinite ciudad actividade política o primeiro a fazer um comício do pontal foi sá carneiro no verão quente de setenta e cinco , no zé com cavaco silva que festa ganha maior importância política ,
D 2010	3: ângelo correia	**2008_08_07-19_59_02-Telejornal-1 bloco 3**: criar o partido pedir unidade do partido , a direcção do psd não comentar as críticas de ângelo correia , na história do psd , este não é a primeira vez que um líder não vai ao comício do pontal ,
Question #59 - Quem igualou o recorde de medalhas de Mark Spitz?		
A 2008	1: fel	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: igualou mark spitz , ao conquistar a sétima medalha de ouro , nos jogos olímpicos . foi a prova mais difícil de fel que até porque era uma prova em que eu não tinha o recorde do mundo .
B 2008	1: norte - americano jay	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: em pequim o norte - americano jay golo feito do lendário mark spitz que na década de setenta , foi também um fenómeno de marketing . mesmo que não consiga chegar às oito medalhas de ouro , telles já batiam um outro recorde que já lhe é um nadador mais bem pago da história .
B 2008	2: década	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: em pequim o norte - americano jay golo feito do lendário mark spitz que na década de setenta , foi também um fenómeno de marketing . mesmo que não consiga chegar às oito medalhas de ouro , telles já batiam um outro recorde que já lhe é um nadador mais bem pago da história .

B 2008	3: fel que até	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: igualou mark spitz , ao conquistar a sétima medalha de ouro , nos jogos olímpicos . foi a prova mais difícil de fel que até porque era uma prova em que eu não tinha o recorde do mundo .
C 2008	1: fel	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: igualou mark spitz , ao conquistar a sétima medalha de ouro , nos jogos olímpicos , foi a prova mais difícil de fel que até porque era uma prova em que eu não tinha o recorde do mundo ,
D 2008	1: norte - americano	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: o norte - americano venceu os cem metros mariposa , igualou mark spitz , ao conquistar a sétima medalha de ouro , nos jogos olímpicos , foi a prova mais difícil de fel que até porque era uma prova em que eu não tinha o recorde do mundo ,
D 2008	2: década	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: em pequim o norte - americano jay golo feito do lendário mark spitz que na década de setenta , foi também um fenómeno de marketing , mesmo que não consiga chegar às oito medalhas de ouro , telles já batiam um outro recorde que já lhe é um nadador mais bem pago da história ,
D 2008	3: fel que até	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: igualou mark spitz , ao conquistar a sétima medalha de ouro , nos jogos olímpicos , foi a prova mais difícil de fel que até porque era uma prova em que eu não tinha o recorde do mundo ,
A 2010	1: fel	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: igualou mark spitz ao conquistar a sétima medalha de ouro nos jogos olímpicos . foi a prova mais difícil de fel se até porque era uma prova em que eu não tinha o recorde do mundo .
B 2010	1: norte - americano jay	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: em pequim o norte - americano jay o aluno feito do lendário mark spitz que na década de setenta . foi também um fenómeno de marketing . mesmo que não consiga chegar às oito medalhas de ouro fiel desejava tinham um outro recorde que já lhe é o nadador mais bem pago da história .
B 2010	2: década	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: em pequim o norte - americano jay o aluno feito do lendário mark spitz que na década de setenta . foi também um fenómeno de marketing . mesmo que não consiga chegar às oito medalhas de ouro fiel desejava tinham um outro recorde que já lhe é o nadador mais bem pago da história .
B 2010	3: james alô	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: recorde do mundo inflacionar não um vão morto a imagem de sel . em pequim o norte - americano james alô feito do lendário mark spitz que na década de setenta . foi também um fenómeno de marketing . um . mesmo que não consiga chegar às oito medalhas de ouro .

C 2010	1: fel	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: igualou mark spitz ao conquistar a sétima medalha de ouro nos jogos olímpicos , foi a prova mais difícil de fel se até porque era uma prova em que eu não tinha o recorde do mundo ,
D 2010	1: norte - americano jay	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: recorde do mundo inflacionar não um o vão morto a imagem de fel , em pequim o norte - americano jay o aluno feito do lendário mark spitz que na década de setenta , foi também um fenómeno de marketing , mesmo que não consiga chegar às oito medalhas
D 2010	2: james alô	**2008_08_16-19_59_02-Telejornal-1 bloco 7**: recorde do mundo inflacionar não um vão morto a imagem de sel , em pequim o norte - americano james alô feito do lendário mark spitz que na década de setenta , foi também um fenómeno de marketing , um , mesmo que não consiga chegar às oito medalhas de ouro ,
D 2010	3: década	**2008_08_16-21_59_02-Jornal2-2 bloco 2**: recorde do mundo inflacionar não um o vão morto a imagem de fel , em pequim o norte - americano jay o aluno feito do lendário mark spitz que na década de setenta , foi também um fenómeno de marketing , mesmo que não consiga chegar às oito medalhas
Question #60 - Quando nasceu João Ubaldo Ribeiro?		
A 2008	1: dois mil e oito	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa .
A 2008	2: novecentos e quarenta e um	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou como jornalista .
A 2008	3: dois mil e sete	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes , joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou com o jornalista antes de se dedicar por inteiro .
B 2008	1: 23 de janeiro de 1 941	**pt/j/o/ã/João_Ubaldo_Ribeiro_12ea.html**: joão ubaldo ribeiro nent joão ubaldo osório pimentel ribeiro (itaparica , 23 de janeiro de 1 941) é um escritor brasileiro , membro da academia brasileira de letras . biografia nascido na bahia , quando tem dois
B 2008	2: dois mil e oito	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa .
B 2008	3: novecentos e quarenta e um	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou como jornalista .
C 2008	1: dois mil e oito	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: é o prémio camões dois mil e oito o mais importante prémio atribuído a um escritor de língua portuguesa , o júri reunido em lisboa escolheu por maioria o escritor brasileiro nascido na bahia , joão ubaldo ribeiro

C 2008	2: novecentos e quarenta e um	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: o júri reunido em lisboa escolheu por maioria o escritor brasileiro nascido na bahia , joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou como jornalista ,
C 2008	3: dois mil e sete	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa , extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos do português antónio lobo antunes , joão
D 2008	1: 23 de janeiro de 1 941	**pt/j/o/ã/João_Ubaldo_Ribeiro_12ea.html**: joão ubaldo ribeiro nent joão ubaldo osório pimentel ribeiro (itaparica , 23 de janeiro de 1 941) é um escritor brasileiro , membro da academia brasileira de letras . biografia nascido na bahia , quando tem dois
D 2008	2: dois mil e oito	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: é o prémio camões dois mil e oito o mais importante prémio atribuído a um escritor de língua portuguesa , o júri reunido em lisboa escolheu por maioria o escritor brasileiro nascido na bahia , joão ubaldo ribeiro
D 2008	3: novecentos e quarenta e um	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: o júri reunido em lisboa escolheu por maioria o escritor brasileiro nascido na bahia , joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou como jornalista ,
A 2010	1: dois mil e oito	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa .
A 2010	2: novecentos e quarenta e um	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: joão ubaldo ribeiro nasceu mil novecentos e quarenta e um e trabalhou como jornalista .
A 2010	3: mil	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: joão ubaldo ribeiro nasceu mil novecentos e quarenta e um e trabalhou como jornalista .
B 2010	1: 23 de janeiro de 1 941	**pt/j/o/ã/João_Ubaldo_Ribeiro_12ea.html**: joão ubaldo ribeiro nent joão ubaldo osório pimentel ribeiro (itaparica , 23 de janeiro de 1 941) é um escritor brasileiro , membro da academia brasileira de letras . biografia nascido na bahia , quando tem dois
B 2010	2: dois mil e oito	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa .
B 2010	3: novecentos e quarenta e um	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: joão ubaldo ribeiro nasceu mil novecentos e quarenta e um e trabalhou como jornalista .
C 2010	1: dois mil e oito	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: é o prémio camões dois mil e oito o mais importante prémio atribuído a um escritor de língua portuguesa , o júri reunido em lisboa escolheu por maioria o escritor brasileiro nascido na baía , joão ubaldo ribeiro

C 2010	2: novecentos e quarenta e um	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: o júri reunido em lisboa escolheu por maioria o escritor brasileiro nascido na baía , joão ubaldo ribeiro nasceu mil novecentos e quarenta e um e trabalhou como jornalista ,
C 2010	3: dois mil e sete	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: extinguiu este ano um escritor brasileiro depois de em dois mil e sete ter ido parar às mãos , o português antónio lobo antunes joão ubaldo ribeiro nasceu mil novecentos e quarenta e um trabalhou com o jornalista antes de se dedicar por inteiro ,
D 2010	1: 23 de janeiro de 1 941	**pt/j/o/ã/João_Ubaldo_Ribeiro_12ea.html**: joão ubaldo ribeiro nent joão ubaldo osório pimentel ribeiro (itaparica , 23 de janeiro de 1 941) é um escritor brasileiro , membro da academia brasileira de letras . biografia nascido na bahia , quando tem dois
D 2010	2: dois mil e oito	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: é o prémio camões dois mil e oito o mais importante prémio atribuído a um escritor de língua portuguesa , o júri reunido em lisboa escolheu por maioria o escritor brasileiro nascido na baía , joão ubaldo ribeiro
D 2010	3: novecentos e quarenta e um	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: o júri reunido em lisboa escolheu por maioria o escritor brasileiro nascido na baía , joão ubaldo ribeiro nasceu mil novecentos e quarenta e um e trabalhou como jornalista ,
Question #61 - Qual o banco que vai apoiar o financiamento das empreitadas de reabilitação urbana?		
A 2008	1: vista	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: um contrato de financiamentos , para decidir o instituto de habitação da reabilitação urbana , com ou apoio dubai do banco europeu de investimentos , e com vista de retomar as empreitadas de reabilitação urbana que foram iniciadas .
A 2008	2: investimentos	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: um contrato de financiamentos , para decidir o instituto de habitação da reabilitação urbana , com ou apoio dubai do banco europeu de investimentos , e com vista de retomar as empreitadas de reabilitação urbana que foram iniciadas .
A 2008	3: contrato	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: um contrato de financiamentos , para decidir o instituto de habitação da reabilitação urbana , com ou apoio dubai do banco europeu de investimentos , e com vista de retomar as empreitadas de reabilitação urbana que foram iniciadas .
B 2008	1: habitação da reabilitação urbana	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: um contrato de financiamentos , para decidir o instituto de habitação da reabilitação urbana , com ou apoio dubai do banco europeu de investimentos , e com vista de retomar as empreitadas de reabilitação urbana que foram iniciadas .

B 2008	2: dubai do banco europeu de investimentos	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: um contrato de financiamentos , para decidir o instituto de habitação da reabilitação urbana , com ou apoio dubai do banco europeu de investimentos , e com vista de retomar as empreitadas de reabilitação urbana que foram iniciadas .
B 2008	3: banco europeu de investimentos	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: um contrato de financiamentos , para decidir o instituto de habitação da reabilitação urbana , com ou apoio dubai do banco europeu de investimentos , e com vista de retomar as empreitadas de reabilitação urbana que foram iniciadas .
C 2008	1: vista	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: a câmara está a finalizar , um contrato de financiamentos , para decidir o instituto de habitação da reabilitação urbana , com ou apoio dubai do banco europeu de investimentos , e com vista de retomar as empreitadas de reabilitação urbana que foram iniciadas ,
C 2008	2: investimentos	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: a câmara está a finalizar , um contrato de financiamentos , para decidir o instituto de habitação da reabilitação urbana , com ou apoio dubai do banco europeu de investimentos , e com vista de retomar as empreitadas de reabilitação urbana que foram iniciadas ,
C 2008	3: contrato	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: a câmara está a finalizar , um contrato de financiamentos , para decidir o instituto de habitação da reabilitação urbana , com ou apoio dubai do banco europeu de investimentos , e com vista de retomar as empreitadas de reabilitação urbana que foram iniciadas ,
D 2008	1: habitação da reabilitação urbana	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: a câmara está a finalizar , um contrato de financiamentos , para decidir o instituto de habitação da reabilitação urbana , com ou apoio dubai do banco europeu de investimentos , e com vista de retomar as empreitadas de reabilitação urbana que foram iniciadas ,
D 2008	2: dubai do banco europeu de investimentos	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: a câmara está a finalizar , um contrato de financiamentos , para decidir o instituto de habitação da reabilitação urbana , com ou apoio dubai do banco europeu de investimentos , e com vista de retomar as empreitadas de reabilitação urbana que foram iniciadas ,
D 2008	3: câmara está a finalizar	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: a câmara está a finalizar , um contrato de financiamentos , para decidir o instituto de habitação da reabilitação urbana , com ou apoio dubai do banco europeu de investimentos , e com vista de retomar as empreitadas de reabilitação urbana que foram iniciadas ,
A 2010	1: vista	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: um contrato de financiamentos parte de wiranto e instituto de habitação da reabilitação urbana com o apoio do vai do banco europeu de investimentos e com vista lhe retomar as empreitadas

A 2010	2: investimentos	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: um contrato de financiamentos parte de wiranto e instituto de habitação da reabilitação urbana com o apoio do vai do banco europeu de investimentos e com vista lhe retomar as empreitadas
A 2010	3: contrato	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: um contrato de financiamentos parte de wiranto e instituto de habitação da reabilitação urbana com o apoio do vai do banco europeu de investimentos e com vista lhe retomar as empreitadas
B 2010	1: contrato de financiamentos parte de wiranto	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: um contrato de financiamentos parte de wiranto e instituto de habitação da reabilitação urbana com o apoio do vai do banco europeu de investimentos e com vista lhe retomar as empreitadas
B 2010	2: habitação da reabilitação urbana	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: um contrato de financiamentos parte de wiranto e instituto de habitação da reabilitação urbana com o apoio do vai do banco europeu de investimentos e com vista lhe retomar as empreitadas
B 2010	3: banco europeu de investimentos	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: um contrato de financiamentos parte de wiranto e instituto de habitação da reabilitação urbana com o apoio do vai do banco europeu de investimentos e com vista lhe retomar as empreitadas
C 2010	1: contrato	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: um contrato de financiamentos parte de wiranto e instituto de habitação da reabilitação urbana com o apoio do vai do banco europeu de investimentos e com vista lhe retomar as empreitadas
C 2010	2: vista	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: um contrato de financiamentos parte de wiranto e instituto de habitação da reabilitação urbana com o apoio do vai do banco europeu de investimentos e com vista lhe retomar as empreitadas
C 2010	3: investimentos	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: um contrato de financiamentos parte de wiranto e instituto de habitação da reabilitação urbana com o apoio do vai do banco europeu de investimentos e com vista lhe retomar as empreitadas
D 2010	1: contrato de financiamentos parte de wiranto	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: um contrato de financiamentos parte de wiranto e instituto de habitação da reabilitação urbana com o apoio do vai do banco europeu de investimentos e com vista lhe retomar as empreitadas
D 2010	2: habitação da reabilitação urbana	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: um contrato de financiamentos parte de wiranto e instituto de habitação da reabilitação urbana com o apoio do vai do banco europeu de investimentos e com vista lhe retomar as empreitadas
D 2010	3: banco europeu de investimentos	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: um contrato de financiamentos parte de wiranto e instituto de habitação da reabilitação urbana com o apoio do vai do banco europeu de investimentos e com vista lhe retomar as empreitadas
Question #62 - Em que país fica o Rio Bengo?		
A 2008	1: NIL	
B 2008	1: angola	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: países com maior crescimento económico . muito quente tempos de abrandamento em lisboa , torna luanda , ainda mais atractivo . paris que tem de haver água do rio bengo , ficar eternamente ligado . a angola a crer na lenda e aplicando os negócios , e significa para os empresários portugueses uma vantagem que não é

B 2008	2: sul - africanos	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: rio bengo , ficar eternamente ligado . a angola a crer na lenda e aplicando os negócios , e significa para os empresários portugueses uma vantagem que não é mágica mas histórica , face os concorrentes chineses sul - africanos , ou americanos , e se nos negócios do jogo dizem quase tudo . os portugueses esperam
B 2008	3: portugueses	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: países com maior crescimento económico . muito quente tempos de abrandamento em lisboa , torna luanda , ainda mais atractivo . paris que tem de haver água do rio bengo , ficar eternamente ligado . a angola a crer na lenda e aplicando os negócios , e significa para os empresários portugueses uma vantagem que não é
C 2008	1: NIL	
D 2008	1: angola	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: países com maior crescimento económico , muito quente tempos de abrandamento em lisboa , torna luanda , ainda mais atractivo , paris que tem de haver água do rio bengo , ficar eternamente ligado , a angola a crer na lenda e aplicando os negócios , e significa para os empresários portugueses uma vantagem que não é
D 2008	2: sul - africanos	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: ainda mais atractivo , paris que tem de haver água do rio bengo , ficar eternamente ligado , a angola a crer na lenda e aplicando os negócios , e significa para os empresários portugueses uma vantagem que não é mágica mas histórica , face os concorrentes chineses sul - africanos ,
D 2008	3: portugueses	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: países com maior crescimento económico , muito quente tempos de abrandamento em lisboa , torna luanda , ainda mais atractivo , paris que tem de haver água do rio bengo , ficar eternamente ligado , a angola a crer na lenda e aplicando os negócios , e significa para os empresários portugueses uma vantagem que não é
A 2010	1: NIL	
B 2010	1: angola	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: superior à actividade de concessão gal . país de grande riqueza natural angola muito por conta do petróleo mas também dos diamantes e do ouro . um dos países com maior crescimento económico muito o que em tempos de abrandamento em lisboa torna luanda ainda mais atractivo . por aqui diz que quem beber água do rio
B 2010	2: sul - africanos	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: rio bengo fica eternamente ligado . angola crer na lenda e aplicando aos negócios isso significa para os empresários portugueses uma vantagem que não é mágica , mas histórica face aos concorrentes chineses sul - africanos ou americanos e se nos negócios os números dizem quase tudo . os portugueses esperam

B 2010	3: país	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: superior à actividade de concessão gal . país de grande riqueza natural angola muito por conta do petróleo mas também dos diamantes e do ouro . um dos países com maior crescimento económico muito o que em tempos de abrandamento em lisboa torna luanda ainda mais atractivo . por aqui diz que quem beber água do rio
C 2010	1: NIL	
D 2010	1: angola	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: países com maior crescimento económico muito o que em tempos de abrandamento em lisboa torna luanda ainda mais atractivo , por aqui diz que quem beber água do rio bengo fica eternamente ligado , angola crer na lenda e aplicando aos negócios isso significa para os empresários portugueses uma vantagem que não é mágica ,
D 2010	2: portugueses	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: países com maior crescimento económico muito o que em tempos de abrandamento em lisboa torna luanda ainda mais atractivo , por aqui diz que quem beber água do rio bengo fica eternamente ligado , angola crer na lenda e aplicando aos negócios isso significa para os empresários portugueses uma vantagem que não é mágica ,
D 2010	3: país	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: superior à actividade de concessão gal , país de grande riqueza natural angola muito por conta do petróleo mas também dos diamantes e do ouro , um dos países com maior crescimento económico muito o que em tempos de abrandamento em lisboa torna luanda ainda mais atractivo , por aqui diz que quem beber água do rio
Question #63 - Quem é o Presidente Angolano?		
A 2008	1: preenchida que começou coube	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: angolano , e sublinhou a prioridade das relações comerciais com angola . chegou bem cedo luanda com mais gente é preenchida que começou coube encontro do dia josé sócrates , debateu com o presidente angolano josé eduardo dos santos .
A 2008	2: na huíla	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: milhares de apoiantes assistiram ao discurso do presidente angolano , na huíla , eduardo dos santos inaugurou um hospital , e distribuiu de electrodomésticos e tractores agrícolas .
B 2008	1: josé sócrates , debateu com o presidente angolano josé eduardo dos santos	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: angolano , e sublinhou a prioridade das relações comerciais com angola . chegou bem cedo luanda com mais gente é preenchida que começou coube encontro do dia josé sócrates , debateu com o presidente angolano josé eduardo dos santos .

B 2008	2: agenda	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: para apoiar os investimentos no mercado angolano , e sublinhou a prioridade das relações comerciais com angola . chegou bem cedo luanda com uma agenda preenchida que começou coube encontro do dia josé sócrates , debateu com o presidente angolano josé eduardo dos santos .
B 2008	3: agostinho neto	**pt/c/u/l/Cultura_de_Angola_b792.html**: cultura de angola . o primeiro foi o fundador da nação angolana , agostinho neto , o segundo é josé eduardo dos santos , o actual presidente angolano , que se tornou chefe de estado em 1 979 sendo o mais jovem presidente no continente .
C 2008	1: preenchida que começou coube	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: angolano , e sublinhou a prioridade das relações comerciais com angola , chegou bem cedo luanda com mais gente é preenchida que começou coube encontro do dia josé sócrates , debateu com o presidente angolano josé eduardo dos santos ,
C 2008	2: na huíla	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: esta comissão independente não depende do presidente da república , eu sou presidente quando ela , também sou jogador , milhares de apoiantes assistiram ao discurso do presidente angolano , na huíla , eduardo dos santos inaugurou um hospital , e distribuiu de electrodomésticos e tractores agrícolas ,
D 2008	1: josé sócrates , debateu com o presidente angolano josé eduardo dos santos	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: chegou bem cedo luanda com mais gente é preenchida que começou coube encontro do dia josé sócrates , debateu com o presidente angolano josé eduardo dos santos , os laços comerciais entre os dois países ,
D 2008	2: agenda	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: para apoiar os investimentos no mercado angolano , e sublinhou a prioridade das relações comerciais com angola , chegou bem cedo luanda com uma agenda preenchida que começou coube encontro do dia josé sócrates , debateu com o presidente angolano josé eduardo dos santos ,
D 2008	3: coube	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: chegou bem cedo luanda com mais gente é preenchida que começou coube encontro do dia josé sócrates , debateu com o presidente angolano josé eduardo dos santos , os laços comerciais entre os dois países ,
A 2010	1: gender	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: angolano e sublinhou a prioridade das relações comerciais com angola . chegou bem cedo luanda com mais gender preenchida que começou clube encontro do dia josé sócrates bateu o presidente angolano josé eduardo dos santos .

A 2010	2: apoiantes	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: milhares de apoiantes assistiram ao discurso do presidente angolano .
A 2010	3: angola também líder do mpla	**2008_09_03-19_59_02-Telejornal-1 bloco 3**: presidente de angola também líder do mpla e apelou à tolerância isaías samakuva atacou o governo . e o partido no poder tentando passar a mensagem que a unita conseguirá fazer melhor o. que o mpla . sexta - feira . oito milhões de angolanos
B 2010	1: josé eduardo dos santos	**2008_08_05-19_59_02-Telejornal-1 bloco 6**: o presidente angolano quer legislativas de quatro em quatro anos josé eduardo dos santos fez um discurso ao país apelar .
B 2010	2: sócrates	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: angolano e sublinhou a prioridade das relações comerciais com angola . chegou bem cedo luanda com mais gender preenchida que começou clube encontro do dia josé sócrates bateu o presidente angolano josé eduardo dos santos .
B 2010	3: agostinho neto	**pt/c/u/l/Cultura_de_Angola_b792.html**: cultura de angola . o primeiro foi o fundador da nação angolana , agostinho neto , o segundo é josé eduardo dos santos , o actual presidente angolano , que se tornou chefe de estado em 1 979 sendo o mais jovem presidente no continente .
C 2010	1: na huíla eduardo dos santos inaugurou	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: esta comissão independente não depende do presidente da república , eu sou presidente do angola , também sou jogador , milhares de apoiantes assistiram ao discurso do presidente angolano , na huíla eduardo dos santos inaugurou um hospital e distribuiu de electrodomésticos e tractores agrícolas ,
C 2010	2: gender	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: angolano e sublinhou a prioridade das relações comerciais com angola , chegou bem cedo luanda com mais gender preenchida que começou clube encontro do dia josé sócrates bateu o presidente angolano josé eduardo dos santos ,
C 2010	3: mas pois foram sendo resolvidos ao longo do dia	**2008_09_05-19_59_02-Telejornal-1 bloco 3**: angolana , mas pois foram sendo resolvidos ao longo do dia , as pessoas agora votam , mas nem por isso , as filas diminuem , eduardo dos santos votou logo pela manhã perto do palácio , o presidente
D 2010	1: josé sócrates bateu o presidente angolano josé eduardo dos santos	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: angolano e sublinhou a prioridade das relações comerciais com angola , chegou bem cedo luanda com mais gender preenchida que começou clube encontro do dia josé sócrates bateu o presidente angolano josé eduardo dos santos ,

D 2010	2: eu sou presidente do angola	**2008_09_02-19_59_01-Telejornal-1 bloco 4**: esta comissão independente não depende do presidente da república , eu sou presidente do angola , também sou jogador , milhares de apoiantes assistiram ao discurso do presidente angolano , na huíla eduardo dos santos inaugurou um hospital e distribuiu de electrodomésticos e tractores agrícolas ,
D 2010	3: agostinho neto	**pt/c/u/l/Cultura_de_Angola_b792.html**: cultura de angola . o primeiro foi o fundador da nação angolana , agostinho neto , o segundo é josé eduardo dos santos , o actual presidente angolano , que se tornou chefe de estado em 1 979 sendo o mais jovem presidente no continente .
Question #64 - Quem é o dono do "número vinte e um" da Avenida da Liberdade?		
A 2008	1: NIL	
B 2008	1: presidente	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: avenida da liberdade o presidente da junta de freguesia mais rica do país . o abastado dono do número vinte e um quinze em testamento que este edifício fosse para sempre a propriedade das gentes da sua terra galveias já várias hoje . manifestaram interesse em adquirir ,
B 2008	2: propriedade	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: avenida da liberdade o presidente da junta de freguesia mais rica do país . o abastado dono do número vinte e um quinze em testamento que este edifício fosse para sempre a propriedade das gentes da sua terra galveias já várias hoje . manifestaram interesse em adquirir ,
B 2008	3: galveias	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: avenida da liberdade o presidente da junta de freguesia mais rica do país . o abastado dono do número vinte e um quinze em testamento que este edifício fosse para sempre a propriedade das gentes da sua terra galveias já várias hoje . manifestaram interesse em adquirir ,
C 2008	1: NIL	
D 2008	1: presidente	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: avenida da liberdade o presidente da junta de freguesia mais rica do país , o abastado dono do número vinte e um quinze em testamento que este edifício fosse para sempre a propriedade das gentes da sua terra galveias já várias hoje , manifestaram interesse em adquirir ,
D 2008	2: propriedade	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: avenida da liberdade o presidente da junta de freguesia mais rica do país , o abastado dono do número vinte e um quinze em testamento que este edifício fosse para sempre a propriedade das gentes da sua terra galveias já várias hoje , manifestaram interesse em adquirir ,
D 2008	3: galveias	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: avenida da liberdade o presidente da junta de freguesia mais rica do país , o abastado dono do número vinte e um quinze em testamento que este edifício fosse para sempre a propriedade das gentes da sua terra galveias já várias hoje , manifestaram interesse em adquirir ,

A 2010	1: NIL	
B 2010	1: presidente	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: enquanto as operações de rescaldo decorriam chegava à avenida da liberdade o presidente da junta de freguesia mais rica do país . o abastado dono do número vinte e um
B 2010	2: propriedade	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: avenida da liberdade o presidente da junta de freguesia mais rica do país . o abastado dono do número vinte e um quinze em testamento que este edifício fosse para sempre propriedade das gentes da sua terra galveias já várias pessoas de . manifestaram interesse em ter e
B 2010	3: rescaldo	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: enquanto as operações de rescaldo decorriam chegava à avenida da liberdade o presidente da junta de freguesia mais rica do país . o abastado dono do número vinte e um
C 2010	1: NIL	
D 2010	1: presidente	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: enquanto as operações de rescaldo decorriam chegava à avenida da liberdade o presidente da junta de freguesia mais rica do país , o abastado dono do número vinte e um
D 2010	2: propriedade	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: avenida da liberdade o presidente da junta de freguesia mais rica do país , o abastado dono do número vinte e um quinze em testamento que este edifício fosse para sempre propriedade das gentes da sua terra galveias já várias pessoas de ,
D 2010	3: galveias	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: avenida da liberdade o presidente da junta de freguesia mais rica do país , o abastado dono do número vinte e um quinze em testamento que este edifício fosse para sempre propriedade das gentes da sua terra galveias já várias pessoas de ,
Question #65 - Quantos inquilinos tem o prédio da Junta de Galveias?		
A 2008 B 2008	1: quinze inquilino	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: os quinze inquilino da junta de galveias não podiam estar mais revoltados com um incêndio que lhes roubou o tecto .
C 2008 D 2008	1: quinze inquilino	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: falava num milhão de contos , os quinze inquilino da junta de galveias não podiam estar mais revoltados com um incêndio que lhes roubou o tecto , a maioria dormiu em casa de familiares hannah lucy uma pernoitaram nesta pretensão ,
A 2010 B 2010	1: quinze inquilino	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: os quinze inquilino da junta de galveias não podiam estar mais revoltados com um incêndio que lhes roubou tecto .

C 2010 D 2010	1: quinze inquilino	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: falava num milhão de contos , os quinze inquilino da junta de galveias não podiam estar mais revoltados com um incêndio que lhes roubou tecto , a maioria dormiu em casa de familiares ana lúcia irmã pernoitaram nesta pensam ,
Question #66 - O que é a EMBRAER?		
A 2008	1: brasileira embraer é a terceira maior empresa mundial do fabrico de aviões	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial do fabrico de aviões .
A 2008	2: fabrico de aviões	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial do fabrico de aviões .
A 2008	3: investimento	**2008_09_11-19_59_02-Telejornal-1 bloco 2**: o conselho de ministros deu luz verde ao projecto de investimento da embraer era para évora .
B 2008	1: embraer ou empresa brasileira de aeronáutica sa é uma empresa que fabrica aviões de pequeno e médio porte (para uso na aviação regional , executiva e agrícola) , além de caças militares e aviões de sensoriamento remoto e para transporte de autoridades .	**pt/e/m/b/Embraer.html**: embraer ou empresa brasileira de aeronáutica sa é uma empresa que fabrica aviões de pequeno e médio porte (para uso na aviação regional , executiva e agrícola) , além de caças militares e aviões de sensoriamento remoto e para transporte de autoridades . é uma das maiores companhias exportadoras do brasil , em termos de valor absoluto desde 1 999 . sua sede localiza - se na cidade de são josé dos campos , interior do estado de são paulo e possui diversas outras unidades , inclusive uma na china .

C 2008	1: brasileira embraer é vai construir duas fábricas para já	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: uma memória e agora o olhar para o futuro da região de évora vai receber um importante investimento aeronáutico , empresa brasileira embraer é vai construir duas fábricas para já , num investimento de cento e cinquenta milhões de euros ,
C 2008	2: os melhores segundo lula da silva	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: dentro de três anos as duas fábricas estarão a funcionar em pleno , de cidade de évora vão ser para o brasil componentes para os aviões da embraer , os melhores segundo lula da silva , o da boda e bresson de tamanha qualidade e ,
C 2008	3: assinado amanhã em lisboa entre josé sócrates	**2008_07_25-19_59_01-Telejornal-1 bloco 1**: e será assinado amanhã em lisboa entre josé sócrates e o presidente do brasil lula da silva , e os responsáveis da empresa , a embraer era já um dos accionistas das oficinas gerais de material aeronáutico situadas em alverca , as duas fábricas deverão começar a operar dentro de três anos ,
D 2008	1: embraer ou empresa brasileira de aeronáutica sa é uma empresa que fabrica aviões de pequeno e médio porte (para uso na aviação regional , executiva e agrícola) , além de caças militares e aviões de sensoriamento remoto e para transporte de autoridades .	**pt/e/m/b/Embraer.html**: embraer ou empresa brasileira de aeronáutica sa é uma empresa que fabrica aviões de pequeno e médio porte (para uso na aviação regional , executiva e agrícola) , além de caças militares e aviões de sensoriamento remoto e para transporte de autoridades . é uma das maiores companhias exportadoras do brasil , em termos de valor absoluto desde 1 999 . sua sede localiza - se na cidade de são josé dos campos , interior do estado de são paulo e possui diversas outras unidades , inclusive uma na china .

A 2010	1: brasileira embraer foi ainda um dos tópicos	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a instalação da empresa brasileira embraer foi ainda um dos tópicos de uma entrevista no programa balanço e contas .
A 2010	2: já	**2008_07_25-19_59_01-Telejornal-1 bloco 1**: a embraer é já um dos accionistas das oficinas gerais de material aeronáutico situadas em alverca .
A 2010	3: fabricante brasileiro de aviões	**2008_07_25-19_59_01-Telejornal-1 bloco 1**: fabricante brasileiro de aviões embraer air vai construir duas fábricas de componentes em portugal .
B 2010	1: embraer ou empresa brasileira de aeronáutica sa é uma empresa que fabrica aviões de pequeno e médio porte (para uso na aviação regional , executiva e agrícola) , além de caças militares e aviões de sensoriamento remoto e para transporte de autoridades .	**pt/e/m/b/Embraer.html**: embraer ou empresa brasileira de aeronáutica sa é uma empresa que fabrica aviões de pequeno e médio porte (para uso na aviação regional , executiva e agrícola) , além de caças militares e aviões de sensoriamento remoto e para transporte de autoridades . é uma das maiores companhias exportadoras do brasil , em termos de valor absoluto desde 1 999 . sua sede localiza - se na cidade de são josé dos campos , interior do estado de são paulo e possui diversas outras unidades , inclusive uma na china .
C 2010	1: brasileira embraer foi ainda um dos tópicos	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a instalação da empresa brasileira embraer foi ainda um dos tópicos de uma entrevista no programa balanço e contas ,
C 2010	2: já	**2008_07_25-19_59_01-Telejornal-1 bloco 1**: eu responsáveis da empresa , a embraer é já um dos accionistas das oficinas gerais de material aeronáutico situadas em alverca , as duas fábricas deverão começar a operar dentro de três anos ,
C 2010	3: fabricante brasileiro de aviões	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: o fabricante brasileiro de aviões embraer vai construir duas fábricas de componentes em portugal , josé sócrates e o presidente do brasil lula da silva ,
D 2010	1: embraer ou empresa brasileira de aeronáutica sa é uma empresa que fabrica aviões de pequeno e médio porte (para uso na aviação regional , executiva e agrícola) , além de caças militares e aviões de sensoriamento remoto e para transporte de autoridades .	**pt/e/m/b/Embraer.html**: embraer ou empresa brasileira de aeronáutica sa é uma empresa que fabrica aviões de pequeno e médio porte (para uso na aviação regional , executiva e agrícola) , além de caças militares e aviões de sensoriamento remoto e para transporte de autoridades . é uma das maiores companhias exportadoras do brasil , em termos de valor absoluto desde 1 999 . sua sede localiza - se na cidade de são josé dos campos , interior do estado de são paulo e possui diversas outras unidades , inclusive uma na china .

Question #67 - Qual a sua nacionalidade?		
A 2008	1: fabrico de aviões	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial do fabrico de aviões .
A 2008	2: brasileira	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial do fabrico de aviões .
A 2008	3: investimento	**2008_09_11-19_59_02-Telejornal-1 bloco 2**: o conselho de ministros deu luz verde ao projecto de investimento da embraer era para évora .
B 2008	1: lista de aviões (e - h	**pt/l/i/s/Lista_de_aviões_(E-H)_1715.html**: lista de aviões (e - h) . embraer
B 2008	2: força aérea brasileira	**pt/f/o/r/Força_Aérea_Brasileira_9b55.html**: força aérea brasileira . embraer
B 2008	3: embraer legacy	**pt/e/m/b/Embraer_Legacy_b7fd.html**: embraer legacy . embraer legacy
C 2008	1: brasileira	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a instalação da empresa brasileira embraer r foi ainda um dos tópicos de uma entrevista no programa de balanço e contas da rtp dois ,
C 2008	2: aviões	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: para esta área , a brasileira embraer é a terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital ,
C 2008	3: brasil	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: dentro de três anos as duas fábricas estarão a funcionar em pleno , de cidade de évora vão ser para o brasil componentes para os aviões da embraer , os melhores segundo lula da silva , o da boda e bresson de tamanha qualidade e ,
D 2008	1: lista de aviões (e - h	**pt/l/i/s/Lista_de_aviões_(E-H)_1715.html**: lista de aviões (e - h) . embraer
D 2008	2: força aérea brasileira	**pt/f/o/r/Força_Aérea_Brasileira_9b55.html**: força aérea brasileira . embraer
D 2008	3: embraer legacy	**pt/e/m/b/Embraer_Legacy_b7fd.html**: embraer legacy . embraer legacy
A 2010	1: fabricante brasileiro de aviões	**2008_07_25-19_59_01-Telejornal-1 bloco 1**: fabricante brasileiro de aviões embraer air vai construir duas fábricas de componentes em portugal .
A 2010	2: continente americano	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: as primeiras fábricas da embraer fora do continente americano .
A 2010	3: brasileira	**2008_07_28-19_59_01-Telejornal-1 bloco 1**: a câmara de évora assinou os contratos de cedência de terrenos empresa brasileira embraer é .
B 2010	1: lista de aviões (e - h	**pt/l/i/s/Lista_de_aviões_(E-H)_1715.html**: lista de aviões (e - h) . embraer
B 2010	2: força aérea brasileira	**pt/f/o/r/Força_Aérea_Brasileira_9b55.html**: força aérea brasileira . embraer

B 2010	3: ver também	**pt/e/m/b/Embraer_ERJ-145_5f26.html**: embraer erj - 145 . ver também embraer
C 2010	1: fabricante brasileiro de aviões	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: o fabricante brasileiro de aviões embraer vai construir duas fábricas de componentes em portugal , josé sócrates e o presidente do brasil lula da silva ,
C 2010	2: brasil lula	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: o fabricante brasileiro de aviões embraer vai construir duas fábricas de componentes em portugal , josé sócrates e o presidente do brasil lula da silva ,
C 2010	3: brasileira	**2008_07_28-19_59_01-Telejornal-1 bloco 1**: a câmara de évora assinou os contratos de cedência de terrenos empresa brasileira embraer é ,
D 2010	1: lista de aviões (e - h	**pt/l/i/s/Lista_de_aviões_(E-H)_1715.html**: lista de aviões (e - h) . embraer
D 2010	2: força aérea brasileira	**pt/f/o/r/Força_Aérea_Brasileira_9b55.html**: força aérea brasileira . embraer
D 2010	3: ver também	**pt/e/m/b/Embraer_ERJ-145_5f26.html**: embraer erj - 145 . ver também embraer
Question #68 - Quantas pessoas moravam no prédio que foi atingido pelas chamas na Rua da Glória?		
A 2008	1: sete pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: as armas , não ficaram pela avenida da liberdade deles foram projectadas para a rua de trás , atingindo o número seis da rua da glória onde viviam sete pessoas .
A 2008	2: onze pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: o número seis da rua da glória estado vivem a onze pessoas ,
B 2008	1: sete pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: as armas , não ficaram pela avenida da liberdade deles foram projectadas para a rua de trás , atingindo o número seis da rua da glória onde viviam sete pessoas .
B 2008	2: onze pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: o número seis da rua da glória estado vivem a onze pessoas ,
B 2008	3: 24 pessoas	**pt/c/e/a/Ceará_Sporting_Club_28a0.html**: ceará sporting club . rua do trilho n ° 6 (atual tristão gonçalves , numa casa que acabou demolida em 2 004 para a construção do metrofor) . eram 24 pessoas , às quais escolheram como nome do team , rio branco foot - ball club . com camisas de cor lilás e calções brancos , semelhantes atualmente ao da

C 2008	1: sete pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: quinze pessoas , as armas , não ficaram pela avenida da liberdade deles foram projectadas para a rua de trás , atingindo o número seis da rua da glória onde viviam sete pessoas , todos os moradores afirmam que o prédio devoluto , há muito que devia estar isolado , e limpo ,
C 2008	2: onze pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: uma actualização , não foram afectadas apenas vinte mas vinte seis pessoas , o número seis da rua da glória estado vivem a onze pessoas ,
C 2008	3: vinte seis pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: uma actualização , não foram afectadas apenas vinte mas vinte seis pessoas , o número seis da rua da glória estado vivem a onze pessoas ,
D 2008	1: sete pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: quinze pessoas , as armas , não ficaram pela avenida da liberdade deles foram projectadas para a rua de trás , atingindo o número seis da rua da glória onde viviam sete pessoas , todos os moradores afirmam que o prédio devoluto , há muito que devia estar isolado , e limpo ,
D 2008	2: onze pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: uma actualização , não foram afectadas apenas vinte mas vinte seis pessoas , o número seis da rua da glória estado vivem a onze pessoas ,
D 2008	3: vinte seis pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: uma actualização , não foram afectadas apenas vinte mas vinte seis pessoas , o número seis da rua da glória estado vivem a onze pessoas ,
A 2010	1: sete pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: não ficaram pela avenida da liberdade aos foram projectadas para a rua de trás atingido o número seis da rua da glória onde viviam sete pessoas .
A 2010	2: onze pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: porque no número seis da rua da glória está vivem a onze pessoas .
B 2010	1: sete pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: não ficaram pela avenida da liberdade aos foram projectadas para a rua de trás atingido o número seis da rua da glória onde viviam sete pessoas .
B 2010	2: onze pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: porque no número seis da rua da glória está vivem a onze pessoas .
B 2010	3: 24 pessoas	**pt/c/e/a/Ceará_Sporting_Club_28a0.html**: ceará sporting club . rua do trilho n º 6 (atual tristão gonçalves , numa casa que acabou demolida em 2 004 para a construção do metrofor) . eram 24 pessoas , às quais escolheram como nome do team , rio branco foot - ball club . com camisas de cor lilás e calções brancos , semelhantes atualmente ao da

C 2010	1: sete pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: rua da glória onde viviam sete pessoas , todos moradores afirmam que o prédio devoluto , há muito que devia estar isolado e limpo , lojas quatro andares e águas - furtadas o número vinte e três da
C 2010	2: quinze pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: quinze pessoas , não ficaram pela avenida da liberdade aos foram projectadas para a rua de trás atingido o número seis da rua da glória onde viviam sete pessoas , todos moradores afirmam que o prédio devoluto , há muito que devia estar isolado e limpo ,
C 2010	3: onze pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: porque no número seis da rua da glória está vivem a onze pessoas , e não as seis inicialmente previstas há também mais três duda estão ,
D 2010	1: sete pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: rua da glória onde viviam sete pessoas , todos moradores afirmam que o prédio devoluto , há muito que devia estar isolado e limpo , lojas quatro andares e águas - furtadas o número vinte e três da
D 2010	2: quinze pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: quinze pessoas , não ficaram pela avenida da liberdade aos foram projectadas para a rua de trás atingido o número seis da rua da glória onde viviam sete pessoas , todos moradores afirmam que o prédio devoluto , há muito que devia estar isolado e limpo ,
D 2010	3: onze pessoas	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: porque no número seis da rua da glória está vivem a onze pessoas , e não as seis inicialmente previstas há também mais três duda estão ,
Question #69 - Quantos aviões da EMBRAER está comprando o governo do Brasil?		
A 2008 B 2008 C 2008 D 2008 A 2010 B 2010 C 2010 D 2010	1: NIL	

Question #70 - Diga o nome de uma construtora de aviões.		
A 2008	1: técnicos da aviação americana	**2008_08_21-19_59_01-Telejornal-1 bloco 2**: construtora do avião , também técnicos da aviação americana , de aviação de uma dada são espanhola , portanto teremos que aguardar e será um processo , muito muito morosa .
A 2008	2: maiores	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a noite negra era uma das maiores construtores mundiais de aviões vai investir cerca de cento e cinquenta milhões de euros .
A 2008	3: postos de trabalho directos	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: uma das maiores construtores mundiais de aviões vai investir cerca de cento e cinquenta milhões de euros . em dois fábricas na cidade de évora . o investimento deverá permitir a criação de pelo menos quinhentos postos de trabalho directos e mais de mil indirectos . sócrates pediu ele aceitou ficar mais um dia de
B 2008	1: mies van der	**pt/m/i/e/Mies_van_der_Rohe_807b.html**: mies van der rohe . construtor herb greenwald (e para os seus herdeiros , depois da sua morte prematura numa queda de avião) : as torres lake shore drive 860 / 880 e 900 / 910 , junto à costa lacustre de chicago . estas torres , com fachadas de vidro e aço , constituíram uma
B 2008	2: depois da sua morte prematura numa queda	**pt/m/i/e/Mies_van_der_Rohe_807b.html**: mies van der rohe . construtor herb greenwald (e para os seus herdeiros , depois da sua morte prematura numa queda de avião) : as torres lake shore drive 860 / 880 e 900 / 910 , junto à costa lacustre de chicago . estas torres , com fachadas de vidro e aço , constituíram uma
B 2008	3: série de apartamentos em altura	**pt/m/i/e/Mies_van_der_Rohe_807b.html**: mies van der rohe . mies projectou ainda uma série de apartamentos em altura , destinados a famílias da classe média , para o construtor herb greenwald (e para os seus herdeiros , depois da sua morte prematura numa queda de avião
C 2008	1: técnicos da aviação americana	**2008_08_21-19_59_01-Telejornal-1 bloco 2**: construtora do avião , também técnicos da aviação americana , de aviação de uma dada são espanhola , portanto teremos que aguardar e será um processo , muito muito morosa , possivelmente um ano , rosa veloso em directo de madrid ,
C 2008	2: rosa veloso em directo de madrid	**2008_08_21-19_59_01-Telejornal-1 bloco 2**: construtora do avião , também técnicos da aviação americana , de aviação de uma dada são espanhola , portanto teremos que aguardar e será um processo , muito muito morosa , possivelmente um ano , rosa veloso em directo de madrid ,
C 2008	3: chamados	**2008_08_21-19_59_01-Telejornal-1 bloco 2**: foram já a chamados técnicos da metro donald de portanto da empresa , construtora do avião , também técnicos da aviação americana , de aviação de uma dada são espanhola , portanto teremos que aguardar e será um processo , muito muito morosa , possivelmente um ano ,

D 2008	1: mies van der rohe	**pt/m/i/e/Mies_van_der_Rohe_807b.html**: mies van der rohe . construtor herb greenwald (e para os seus herdeiros , depois da sua morte prematura numa queda de avião) : as torres lake shore drive 860 / 880 e 900 / 910 , junto à costa lacustre de chicago . estas torres , com fachadas de vidro e aço , constituíram uma
D 2008	2: depois da sua morte prematura numa queda	**pt/m/i/e/Mies_van_der_Rohe_807b.html**: mies van der rohe . construtor herb greenwald (e para os seus herdeiros , depois da sua morte prematura numa queda de avião) : as torres lake shore drive 860 / 880 e 900 / 910 , junto à costa lacustre de chicago . estas torres , com fachadas de vidro e aço , constituíram uma
D 2008	3: série de apartamentos em altura	**pt/m/i/e/Mies_van_der_Rohe_807b.html**: mies van der rohe . mies projectou ainda uma série de apartamentos em altura , destinados a famílias da classe média , para o construtor herb greenwald (e para os seus herdeiros , depois da sua morte prematura numa queda de avião
A 2010	1: seguro	**2008_08_20-19_59_01-Telejornal-1 bloco 1**: avião inteiramente seguro inteiramente seguro . inteiramente seguro é exactamente igual aos outros . só que é que é de um construtor diferente da boeing e da airbus que ele que normalmente as pessoas habituadas a ouvir falar zé perfeitamente seguro . de e ver uma
A 2010	2: habituadas	**2008_08_20-19_59_01-Telejornal-1 bloco 1**: avião inteiramente seguro inteiramente seguro . inteiramente seguro é exactamente igual aos outros . só que é que é de um construtor diferente da boeing e da airbus que ele que normalmente as pessoas habituadas a ouvir falar zé perfeitamente seguro . de e ver uma
A 2010	3: maiores	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: o noite tema era uma das maiores construtores mundiais de aviões vai investir cerca de cento e cinquenta milhões de euros .
B 2010	1: mies van der	**pt/m/i/e/Mies_van_der_Rohe_807b.html**: mies van der rohe . construtor herb greenwald (e para os seus herdeiros , depois da sua morte prematura numa queda de avião) : as torres lake shore drive 860 / 880 e 900 / 910 , junto à costa lacustre de chicago . estas torres , com fachadas de vidro e aço , constituíram uma
B 2010	2: depois da sua morte prematura numa queda	**pt/m/i/e/Mies_van_der_Rohe_807b.html**: mies van der rohe . construtor herb greenwald (e para os seus herdeiros , depois da sua morte prematura numa queda de avião) : as torres lake shore drive 860 / 880 e 900 / 910 , junto à costa lacustre de chicago . estas torres , com fachadas de vidro e aço , constituíram uma
B 2010	3: famílias da classe média	**pt/m/i/e/Mies_van_der_Rohe_807b.html**: mies van der rohe . mies projectou ainda uma série de apartamentos em altura , destinados a famílias da classe média , para o construtor herb greenwald (e para os seus herdeiros , depois da sua morte prematura numa queda de avião

C 2010	1: seguro	**2008_08_20-19_59_01-Telejornal-1 bloco 1**: inteiramente seguro inteiramente seguro , inteiramente seguro é exactamente igual aos outros , só que é que é de um construtor diferente da boeing e da airbus que ele que normalmente as pessoas habituadas a ouvir falar zé perfeitamente seguro , de e ver uma pergunta de provas um pouco mais pessoal , am e tem ver
C 2010	2: habituadas	**2008_08_20-19_59_01-Telejornal-1 bloco 1**: inteiramente seguro inteiramente seguro , inteiramente seguro é exactamente igual aos outros , só que é que é de um construtor diferente da boeing e da airbus que ele que normalmente as pessoas habituadas a ouvir falar zé perfeitamente seguro , de e ver uma pergunta de provas um pouco mais pessoal , am e tem ver
C 2010	3: maiores	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: o noite tema era uma das maiores construtores mundiais de aviões vai investir cerca de cento e cinquenta milhões de euros , em duas fábricas na cidade de évora ,
D 2010	1: mies van der rohe	**pt/m/i/e/Mies_van_der_Rohe_807b.html**: mies van der rohe . construtor herb greenwald (e para os seus herdeiros , depois da sua morte prematura numa queda de avião) : as torres lake shore drive 860 / 880 e 900 / 910 , junto à costa lacustre de chicago . estas torres , com fachadas de vidro e aço , constituíram uma
D 2010	2: seguro	**2008_08_20-19_59_01-Telejornal-1 bloco 1**: inteiramente seguro inteiramente seguro , inteiramente seguro é exactamente igual aos outros , só que é que é de um construtor diferente da boeing e da airbus que ele que normalmente as pessoas habituadas a ouvir falar zé perfeitamente seguro , de e ver uma pergunta de provas um pouco mais pessoal , am e tem ver
D 2010	3: habituadas a ouvir falar	**2008_08_20-19_59_01-Telejornal-1 bloco 1**: inteiramente seguro inteiramente seguro , inteiramente seguro é exactamente igual aos outros , só que é que é de um construtor diferente da boeing e da airbus que ele que normalmente as pessoas habituadas a ouvir falar zé perfeitamente seguro , de e ver uma pergunta de provas um pouco mais pessoal , am e tem ver
Question #71 - Quem é Jorge Nuno Pinto da Costa?		
A 2008	1: NIL	
B 2008	1: jorge nuno pinto da costa (1 937) , é o presidente do futebol clube do porto . estudante num colégio jesuíta em santo tirso , aos 35 anos , torna - se dirigente do futebol clube do porto , como chefe da secção de hóquei em patins .	**pt/j/o/r/Jorge_Nuno_Pinto_da_Costa_092e.html**: jorge nuno pinto da costa (1 937) , é o presidente do futebol clube do porto . estudante num colégio jesuíta em santo tirso , aos 35 anos , torna - se dirigente do futebol clube do porto , como chefe da secção de hóquei em patins . passou também pelas secções de hóquei em campo e boxe . em 1 976 , chega a responsável pelo departamento de futebol . alguns anos mais tarde , no dia 17 de abril de 1 982 , é eleito presidente do fc porto .

C 2008	1: NIL	
D 2008	1: jorge nuno pinto da costa (1 937) , é o presidente do futebol clube do porto . estudante num colégio jesuíta em santo tirso , aos 35 anos , torna - se dirigente do futebol clube do porto , como chefe da secção de hóquei em patins .	**pt/j/o/r/Jorge_Nuno_Pinto_da_Costa_092e.html**: jorge nuno pinto da costa (1 937) , é o presidente do futebol clube do porto . estudante num colégio jesuíta em santo tirso , aos 35 anos , torna - se dirigente do futebol clube do porto , como chefe da secção de hóquei em patins . passou também pelas secções de hóquei em campo e boxe . em 1 976 , chega a responsável pelo departamento de futebol . alguns anos mais tarde , no dia 17 de abril de 1 982 , é eleito presidente do fc porto .
A 2010	1: NIL	
B 2010	1: jorge nuno pinto da costa (1 937) , é o presidente do futebol clube do porto . estudante num colégio jesuíta em santo tirso , aos 35 anos , torna - se dirigente do futebol clube do porto , como chefe da secção de hóquei em patins .	**pt/j/o/r/Jorge_Nuno_Pinto_da_Costa_092e.html**: jorge nuno pinto da costa (1 937) , é o presidente do futebol clube do porto . estudante num colégio jesuíta em santo tirso , aos 35 anos , torna - se dirigente do futebol clube do porto , como chefe da secção de hóquei em patins . passou também pelas secções de hóquei em campo e boxe . em 1 976 , chega a responsável pelo departamento de futebol . alguns anos mais tarde , no dia 17 de abril de 1 982 , é eleito presidente do fc porto .
C 2010	1: NIL	
D 2010	1: jorge nuno pinto da costa (1 937) , é o presidente do futebol clube do porto . estudante num colégio jesuíta em santo tirso , aos 35 anos , torna - se dirigente do futebol clube do porto , como chefe da secção de hóquei em patins .	**pt/j/o/r/Jorge_Nuno_Pinto_da_Costa_092e.html**: jorge nuno pinto da costa (1 937) , é o presidente do futebol clube do porto . estudante num colégio jesuíta em santo tirso , aos 35 anos , torna - se dirigente do futebol clube do porto , como chefe da secção de hóquei em patins . passou também pelas secções de hóquei em campo e boxe . em 1 976 , chega a responsável pelo departamento de futebol . alguns anos mais tarde , no dia 17 de abril de 1 982 , é eleito presidente do fc porto .

Question #72 - Onde se realiza o festival Andanças?		
A 2008	1: põe	**2008_08_03-21_59_01-Jornal2-2 bloco 2**: mas há que começar o festival que põe toda a gente tem de ser em são pedro do sul . o andanças organiza centenas de eventos .
A 2008	2: organiza	**2008_08_03-21_59_01-Jornal2-2 bloco 2**: mas há que começar o festival que põe toda a gente tem de ser em são pedro do sul . o andanças organiza centenas de eventos .
A 2008	3: prec	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: começou hoje mais uma edição do festival andanças , até domingo em são pedro do sul há danças populares e música de todo o mundo . é uma espécie de prec .
B 2008	1: carvalhais na região	**pt/a/n/d/Andanças.html**: andanças . o andanças é um festival de danças tradicionais anual , em agosto , que desde há uns anos assenta arraiais em carvalhais na região de são pedro do sul portugal ,
B 2008	2: danças tradicionais	**pt/a/n/d/Andanças.html**: andanças . o andanças é um festival de danças tradicionais anual , em agosto , que desde há uns anos assenta arraiais em carvalhais na região de são pedro do sul portugal ,
B 2008	3: portugal	**pt/a/n/d/Andanças.html**: andanças . o andanças é um festival de danças tradicionais anual , em agosto , que desde há uns anos assenta arraiais em carvalhais na região de são pedro do sul portugal ,
C 2008	1: proximidade	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: para muitos vai ser uma surpresa , pela proximidade de uma história passada mesmo que ao nosso lado , e pelo talento de isabel ii à rede interna , e a contar bem , começou hoje mais uma edição do festival andanças ,
C 2008	2: isabel ii	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: para muitos vai ser uma surpresa , pela proximidade de uma história passada mesmo que ao nosso lado , e pelo talento de isabel ii à rede interna , e a contar bem , começou hoje mais uma edição do festival andanças ,
C 2008	3: prec	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: e pelo talento de isabel ii à rede interna , e a contar bem , começou hoje mais uma edição do festival andanças , até domingo em são pedro do sul há danças populares e música de todo o mundo , é uma espécie de prec ,
D 2008	1: carvalhais na região	**pt/a/n/d/Andanças.html**: andanças . o andanças é um festival de danças tradicionais anual , em agosto , que desde há uns anos assenta arraiais em carvalhais na região de são pedro do sul portugal ,

D 2008	2: põe toda	**2008_08_03-21_59_01-Jornal2-2 bloco 2**: num , e chegamos ao cartaz cultural do jornal dois , mas há que começar o festival que põe toda a gente tem de ser em são pedro do sul , o andanças organiza centenas de eventos , durante uma semana pode contar com bailes concertos ,
D 2008	3: danças tradicionais	**pt/a/n/d/Andanças.html**: andanças . o andanças é um festival de danças tradicionais anual , em agosto , que desde há uns anos assenta arraiais em carvalhais na região de são pedro do sul portugal ,
A 2010	1: organiza centenas de eventos	**2008_08_03-21_59_01-Jornal2-2 bloco 2**: está a começar o festival que põe toda a gente tem de ser em são pedro do sul . o andanças organiza centenas de eventos .
A 2010	2: põe	**2008_08_03-21_59_01-Jornal2-2 bloco 2**: está a começar o festival que põe toda a gente tem de ser em são pedro do sul . o andanças organiza centenas de eventos .
A 2010	3: ânsias	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: começou hoje mais uma edição do festival andanças até domingo em são pedro do sul lado ânsias populares e música de todo o mundo . é uma experiência diferente .
B 2010	1: lomba colonna	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: andanças há vários para cada um descobre . o seu festival e penso que de cada um sabe aqui com carinho felicidade . e depois da exaustão relaxamento de . este vemos abrimos e. vamos subindo . sempre lomba colonna . com as duas mãos . não é
B 2010	2: carvalhais na região	**pt/a/n/d/Andanças.html**: andanças . o andanças é um festival de danças tradicionais anual , em agosto , que desde há uns anos assenta arraiais em carvalhais na região de são pedro do sul portugal ,
B 2010	3: fusão	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: as que é um festival de todas em tv . sem preconceitos em nada mais mês . nesta oficina aparência dançar uma fusão de raízes tradicionais mesmo sem música . juíza que não há apenas um andanças há vários para cada um descobre .
C 2010	1: diarra	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: ser uma surpresa , pela proximidade de uma história passada mesmo que almoço lado e pelo talento isabel diarra do inter e a contar bem , começou hoje mais uma edição do festival andanças
C 2010	2: ânsias	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: começou hoje mais uma edição do festival andanças até domingo em são pedro do sul lado ânsias populares e música de todo o mundo , é uma experiência diferente ,
D 2010	1: fusão	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: as que é um festival de todas em tv , sem preconceitos em nada mais mês , nesta oficina aparência dançar uma fusão de raízes tradicionais mesmo sem música , juíza que não há apenas um andanças há vários para cada um descobre , o seu festival

D 2010	2: lomba colonna	**2008_08_05-21_59_02-Jornal2-2 bloco 3**: andanças há vários para cada um descobre , o seu festival e penso que de cada um sabe aqui com carinho felicidade , e depois da exaustão relaxamento de , este vemos abrimos e , vamos subindo , sempre lomba colonna , com as duas mãos , não é
D 2010	3: ver	**pt/a/n/d/Andanças.html**: andanças . " o andanças é um festival onde não se vem ver , vem - se fazer .
Question #73 - Em que estádio conquistou Nelson Évora a medalha de ouro olímpica?		
A 2008	1: NIL	
B 2008	1: atletismo	**2008_08_21-19_59_01-Telejornal-1 bloco 1**: medalha de ouro portuguesa na história dos jogos olímpicos . mas é a primeira nas chamadas disciplinas técnicas do atletismo . o jornalista rui loura acompanhou nos momentos cruciais do atleta neste dia de ouro . o estado nacional de pequim e encheu por completo na noite de glória de nelson évora ,
C 2008	1: NIL	
D 2008	1: atletismo	**2008_08_21-19_59_01-Telejornal-1 bloco 1**: medalha de ouro portuguesa na história dos jogos olímpicos , mas é a primeira nas chamadas disciplinas técnicas do atletismo , o jornalista rui loura acompanhou nos momentos cruciais do atleta neste dia de ouro , o estado nacional de pequim e encheu por completo na noite de glória de nelson évora ,
A 2010	1: NIL	
B 2010	1: estádio olímpico	**2008_08_07-19_59_02-Telejornal-1 bloco 2**: nelson évora não queria perder tempo estava ansioso por colocar ao peito para já a credencial que vai permitir aceder à pista do estádio olímpico como porta - estandarte .
C 2010	1: NIL	
D 2010	1: estádio olímpico	**2008_08_07-19_59_02-Telejornal-1 bloco 2**: de larry heard valores da língua a qual a escola vossa excelência a que horas , nelson évora não queria perder tempo estava ansioso por colocar ao peito para já a credencial que vai permitir aceder à pista do estádio olímpico como porta - estandarte ,
Question #74 - Quantos milhões de voluntários tem a rede que a campanha de Obama construiu através da internet?		
A 2008	1: NIL	
B 2008	1: forte para bater	**2008_06_03-19_59_02-Telejornal-1 bloco 3**: obama esta noite mas o principal responsável da campanha . já vai negar esta possibilidade de clinton insiste que tem dezoito milhões de razões para acreditar que é mais forte para bater john mccain .
B 2008	2: noite mas o principal responsável da campanha	**2008_06_03-19_59_02-Telejornal-1 bloco 3**: obama esta noite mas o principal responsável da campanha . já vai negar esta possibilidade de clinton insiste que tem dezoito milhões de razões para acreditar que é mais forte para bater john mccain .

B 2008	3: chama olímpica pela região chinesa	**2008_06_17-19_59_01-Telejornal-1 bloco 3**: obama precisa de reunir os democratas , al gore permanece como uma das vozes mais influentes do partido . vítor gonçalves rtp washington . a passagem da chama olímpica pela região chinesa de xinjiang , foi rodeada de fortes medidas de segurança . oito milhões de muçulmanos vive na zona ,
C 2008	1: NIL	
D 2008	1: chama olímpica pela região chinesa	**2008_06_17-19_59_01-Telejornal-1 bloco 3**: obama precisa de reunir os democratas , al gore permanece como uma das vozes mais influentes do partido , vítor gonçalves rtp washington , a passagem da chama olímpica pela região chinesa de xinjiang , foi rodeada de fortes medidas de segurança , oito milhões de muçulmanos vive na zona ,
D 2008	2: verdade há de facto	**2008_08_25-21_59_01-Jornal2-2 bloco 1**: a primeiro lugar a capacidade de se unirem em torno do candidato barack obama , na verdade há de facto uma fatia importante do eleitorado votou em hillary clinton nas primárias , foram dezoito milhões de americanos , que não estão convencidos de que ,
D 2008	3: forte para bater	**2008_06_03-19_59_02-Telejornal-1 bloco 3**: obama esta noite mas o principal responsável da campanha , já vai negar esta possibilidade de clinton insiste que tem dezoito milhões de razões para acreditar que é mais forte para bater john mccain , aideed para tal ,
A 2010	1: NIL	
B 2010	1: dores	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores .
B 2010	2: número	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores .
B 2010	3: internet	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores .
C 2010	1: NIL	
D 2010	1: dores	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores ,

D 2010	2: número	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores ,
D 2010	3: internet	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores ,
Question #75 - E quantos milhões de dólares recebeu a campanha de pequenos doadores por esta via?		
A 2008	1: dezoito milhões de	**2008_08_25-21_59_01-Jornal2-2 bloco 1**: foram dezoito milhões de americanos , que não estão convencidos de que , eu barack obama seja o candidato mais forte que a américa precisa neste momento .
A 2008	2: vinte e dois milhões de	**2008_06_27-19_59_02-Telejornal-1 bloco 4**: obama passou as últimas três semanas a garantir o apoio dos clinton mas perdeu mesmo os apoiantes que ajudem e devia pagar as dívidas da pré - campanha , vinte e dois milhões de dólares ,
A 2008	3: seiscentos milhões de	**2008_07_23-21_59_01-Jornal2-2 bloco 2**: barack obama já prometeu que a américa nunca permitirá , que israel seja riscado do mapa . das mães de honey europeia vai bloquear seiscentos milhões de euros de fundos para a bulgária .
B 2008	1: dezoito milhões de	**2008_08_25-21_59_01-Jornal2-2 bloco 1**: foram dezoito milhões de americanos , que não estão convencidos de que , eu barack obama seja o candidato mais forte que a américa precisa neste momento .
B 2008	2: vinte e dois milhões de	**2008_06_27-19_59_02-Telejornal-1 bloco 4**: obama passou as últimas três semanas a garantir o apoio dos clinton mas perdeu mesmo os apoiantes que ajudem e devia pagar as dívidas da pré - campanha , vinte e dois milhões de dólares ,
B 2008	3: seiscentos milhões de	**2008_07_23-21_59_01-Jornal2-2 bloco 2**: barack obama já prometeu que a américa nunca permitirá , que israel seja riscado do mapa . das mães de honey europeia vai bloquear seiscentos milhões de euros de fundos para a bulgária .
C 2008	1: dezoito milhões de	**2008_06_03-19_59_02-Telejornal-1 bloco 3**: obama esta noite mas o principal responsável da campanha , já vai negar esta possibilidade de clinton insiste que tem dezoito milhões de razões para acreditar que é mais forte para bater john mccain , aideed para tal ,
C 2008	2: seiscentos milhões de	**2008_07_23-21_59_01-Jornal2-2 bloco 2**: barack obama já prometeu que a américa nunca permitirá , que israel seja riscado do mapa , das mães de honey europeia vai bloquear seiscentos milhões de euros de fundos para a bulgária ,

C 2008	3: vinte e dois milhões de	**2008_06_27-19_59_02-Telejornal-1 bloco 4**: obama passou as últimas três semanas a garantir o apoio dos clinton mas perdeu mesmo os apoiantes que ajudem e devia pagar as dívidas da pré - campanha , vinte e dois milhões de dólares ,
D 2008	1: dezoito milhões de	**2008_06_03-19_59_02-Telejornal-1 bloco 3**: obama esta noite mas o principal responsável da campanha , já vai negar esta possibilidade de clinton insiste que tem dezoito milhões de razões para acreditar que é mais forte para bater john mccain , aideed para tal ,
D 2008	2: seiscentos milhões de	**2008_07_23-21_59_01-Jornal2-2 bloco 2**: barack obama já prometeu que a américa nunca permitirá , que israel seja riscado do mapa , das mães de honey europeia vai bloquear seiscentos milhões de euros de fundos para a bulgária ,
D 2008	3: vinte e dois milhões de	**2008_06_27-19_59_02-Telejornal-1 bloco 4**: obama passou as últimas três semanas a garantir o apoio dos clinton mas perdeu mesmo os apoiantes que ajudem e devia pagar as dívidas da pré - campanha , vinte e dois milhões de dólares ,
A 2010	1: dezoito milhões de	**2008_08_25-21_59_01-Jornal2-2 bloco 1**: obama na verdade há , de facto , uma fatia importante do eleitorado que votou em díli pela diferente nas primárias e foram dezoito milhões de americanos que ainda .
A 2010	2: dois milhões de	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores .
A 2010	3: duzentos e quarenta milhões de	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores .
B 2010	1: dezoito milhões de	**2008_08_25-21_59_01-Jornal2-2 bloco 1**: obama na verdade há , de facto , uma fatia importante do eleitorado que votou em díli pela diferente nas primárias e foram dezoito milhões de americanos que ainda .
B 2010	2: dois milhões de	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores .
B 2010	3: duzentos e quarenta milhões de	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores .

C 2010	1: dezoito milhões de	**2008_08_25-21_59_01-Jornal2-2 bloco 1**: obama na verdade há , de facto , uma fatia importante do eleitorado que votou em díli pela diferente nas primárias e foram dezoito milhões de americanos que ainda , não estão convencidos de ,
C 2010	2: dois milhões de	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores ,
C 2010	3: duzentos e quarenta milhões de	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores ,
D 2010	1: dezoito milhões de	**2008_08_25-21_59_01-Jornal2-2 bloco 1**: obama na verdade há , de facto , uma fatia importante do eleitorado que votou em díli pela diferente nas primárias e foram dezoito milhões de americanos que ainda , não estão convencidos de ,
D 2010	2: dois milhões de	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores ,
D 2010	3: duzentos e quarenta milhões de	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: de obama está a revolucionar a forma de comunicar política através da internet construiu uma rede de dois milhões de voluntários recebeu um número recorde de duzentos e quarenta milhões de dólares de pequenos do dores ,
Question #76 - Que movimento iniciou João Gilberto?		
A 2008	1: NIL	
B 2008	1: bossa nova	**pt/s/a/m/Samba.html**: samba . um dos novos estilos foi a bossa nova , criado por membros da classe média , dentre eles joão gilberto e antonio carlos jobim . nos anos sessenta os músicos da bossa nova iniciaram um movimento de resgate dos grandes mestres do samba .
B 2008	2: movimentos da música popular	**pt/b/o/s/Bossa_nova.html**: bossa nova . movimentos da música popular do brasil : a bossa nova , criada pelo violonista e cantor baiano joão gilberto . vários críticos de música reconheceram na bossa nova uma influência do " cool jazz " , fator apreciado por muitos , mas rejeitado por outros que criticaram a bossa nova por ela romper

B 2008	3: música popular	**pt/b/o/s/Bossa_nova.html**: bossa nova . movimentos da música popular do brasil : a bossa nova , criada pelo violonista e cantor baiano joão gilberto . vários críticos de música reconheceram na bossa nova uma influência do " cool jazz " , fator apreciado por muitos , mas rejeitado por outros que criticaram a bossa nova por ela romper
C 2008	1: NIL	
D 2008	1: bossa nova	**pt/s/a/m/Samba.html**: samba . um dos novos estilos foi a bossa nova , criado por membros da classe média , dentre eles joão gilberto e antonio carlos jobim . nos anos sessenta os músicos da bossa nova iniciaram um movimento de resgate dos grandes mestres do samba .
D 2008	2: movimentos da música popular	**pt/b/o/s/Bossa_nova.html**: bossa nova . movimentos da música popular do brasil : a bossa nova , criada pelo violonista e cantor baiano joão gilberto . vários críticos de música reconheceram na bossa nova uma influência do " cool jazz " , fator apreciado por muitos , mas rejeitado por outros que criticaram a bossa nova por ela romper
D 2008	3: música popular	**pt/b/o/s/Bossa_nova.html**: bossa nova . movimentos da música popular do brasil : a bossa nova , criada pelo violonista e cantor baiano joão gilberto . vários críticos de música reconheceram na bossa nova uma influência do " cool jazz " , fator apreciado por muitos , mas rejeitado por outros que criticaram a bossa nova por ela romper
A 2010	1: NIL	
B 2010	1: bossa nova	**pt/s/a/m/Samba.html**: samba . um dos novos estilos foi a bossa nova , criado por membros da classe média , dentre eles joão gilberto e antonio carlos jobim . nos anos sessenta os músicos da bossa nova iniciaram um movimento de resgate dos grandes mestres do samba .
B 2010	2: movimentos da música popular	**pt/b/o/s/Bossa_nova.html**: bossa nova . movimentos da música popular do brasil : a bossa nova , criada pelo violonista e cantor baiano joão gilberto . vários críticos de música reconheceram na bossa nova uma influência do " cool jazz " , fator apreciado por muitos , mas rejeitado por outros que criticaram a bossa nova por ela romper
B 2010	3: música popular	**pt/b/o/s/Bossa_nova.html**: bossa nova . movimentos da música popular do brasil : a bossa nova , criada pelo violonista e cantor baiano joão gilberto . vários críticos de música reconheceram na bossa nova uma influência do " cool jazz " , fator apreciado por muitos , mas rejeitado por outros que criticaram a bossa nova por ela romper

C 2010	1: NIL	
D 2010	1: bossa nova	**pt/s/a/m/Samba.html**: samba . um dos novos estilos foi a bossa nova , criado por membros da classe média , dentre eles joão gilberto e antonio carlos jobim . nos anos sessenta os músicos da bossa nova iniciaram um movimento de resgate dos grandes mestres do samba .
D 2010	2: movimentos da música popular	**pt/b/o/s/Bossa_nova.html**: bossa nova . movimentos da música popular do brasil : a bossa nova , criada pelo violonista e cantor baiano joão gilberto . vários críticos de música reconheceram na bossa nova uma influência do " cool jazz " , fator apreciado por muitos , mas rejeitado por outros que criticaram a bossa nova por ela romper
D 2010	3: música popular	**pt/b/o/s/Bossa_nova.html**: bossa nova . movimentos da música popular do brasil : a bossa nova , criada pelo violonista e cantor baiano joão gilberto . vários críticos de música reconheceram na bossa nova uma influência do " cool jazz " , fator apreciado por muitos , mas rejeitado por outros que criticaram a bossa nova por ela romper
Question #77 - Quem gravou, há 50 anos, o tema "Chega de Saudade"?		
A 2008	1: NIL	
B 2008	1: precisamente	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: mf é precisamente cinquenta anos em julho de mil novecentos e cinquenta e oito que jorge gilberto gravou . chega de saudade . o tema
B 2008	2: jorge	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: mf é precisamente cinquenta anos em julho de mil novecentos e cinquenta e oito que jorge gilberto gravou . chega de saudade . o tema
C 2008	1: mundialmente	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: chega de saudade , o tema marcou o início da bossa - nova movimento cruzou o samba com a ver mesmo o jazz , e tornou música brasileira conhecida mundialmente cruzamento de géneros musicais , influência também alguns músicos portugueses ,
C 2008	2: samba	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: chega de saudade , o tema marcou o início da bossa - nova movimento cruzou o samba com a ver mesmo o jazz , e tornou música brasileira conhecida mundialmente cruzamento de géneros musicais , influência também alguns músicos portugueses ,
C 2008	3: bossa	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: chega de saudade , o tema marcou o início da bossa - nova movimento cruzou o samba com a ver mesmo o jazz , e tornou música brasileira conhecida mundialmente cruzamento de géneros musicais , influência também alguns músicos portugueses ,

D 2008	1: bossa - nova	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: precisamente cinquenta anos em julho de mil novecentos e cinquenta e oito que jorge gilberto gravou , chega de saudade , o tema marcou o início da bossa - nova movimento cruzou o samba com a ver mesmo o jazz , e tornou música brasileira conhecida mundialmente cruzamento de géneros musicais , influência também alguns músicos portugueses
D 2008	2: jazz	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: precisamente cinquenta anos em julho de mil novecentos e cinquenta e oito que jorge gilberto gravou , chega de saudade , o tema marcou o início da bossa - nova movimento cruzou o samba com a ver mesmo o jazz , e tornou música brasileira conhecida mundialmente cruzamento de géneros musicais , influência também alguns músicos portugueses
D 2008	3: jorge	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: mf é precisamente cinquenta anos em julho de mil novecentos e cinquenta e oito que jorge gilberto gravou , chega de saudade , o tema
A 2010	1: NIL	
B 2010	1: jo	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: mf é precisamente cinquenta anos em julho de mil novecentos e cinquenta e oito que jo gilberto gravou . chega de saudade . o tema
B 2010	2: precisamente	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: mf é precisamente cinquenta anos em julho de mil novecentos e cinquenta e oito que jo gilberto gravou . chega de saudade . o tema
C 2010	1: mundialmente	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: chega de saudade , o tema marcou o início da bossa - nova movimento por cruzou o samba com a ver mesmo jazz e tornou a música brasileira conhecida mundialmente cruzamento de géneros musicais influência também alguns músicos portugueses ,
D 2010	1: bossa - nova	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: é precisamente cinquenta anos em julho de mil novecentos e cinquenta e oito que jo gilberto gravou , chega de saudade , o tema marcou o início da bossa - nova movimento por cruzou o samba com a ver mesmo jazz e tornou a música brasileira conhecida mundialmente cruzamento de géneros musicais influência também alguns músicos portugueses ,
D 2010	2: jazz	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: é precisamente cinquenta anos em julho de mil novecentos e cinquenta e oito que jo gilberto gravou , chega de saudade , o tema marcou o início da bossa - nova movimento por cruzou o samba com a ver mesmo jazz e tornou a música brasileira conhecida mundialmente cruzamento de géneros musicais influência também alguns músicos portugueses ,

D 2010	3: precisamente	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: mf é precisamente cinquenta anos em julho de mil novecentos e cinquenta e oito que jo gilberto gravou , chega de saudade , o tema
Question #78 - Em que data?		
A 2008	1: NIL	
B 2008	1: nova	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: gravou . chega de saudade . o tema marcou o início da bossa - nova movimento cruzou o samba com a ver mesmo o jazz , e tornou música brasileira conhecida mundialmente cruzamento de géneros musicais .
B 2008	2: waltel branco	**pt/w/a/l/Waltel_Branco_a388.html**: waltel branco . agora no rio de janeiro , gravou os discos " guitarra em chamas " 1 e 2 juntamente com o violonista baden powell e , tendo contato com a então inovadora bossa nova fez os arranjos de " chega de saudade " de joão gilberto ,
B 2008	3: rede globo	**pt/w/a/l/Waltel_Branco_a388.html**: waltel branco . gravou os discos " guitarra em chamas " 1 e 2 juntamente com o violonista baden powell e , tendo contato com a então inovadora bossa nova fez os arranjos de " chega de saudade " de joão gilberto , com quem veio a trabalhar por um longo período . na rede globo foi responsável pelas trilhas de
C 2008	1: NIL	
D 2008	1: nova	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: gravou , chega de saudade , o tema marcou o início da bossa - nova movimento cruzou o samba com a ver mesmo o jazz , e tornou música brasileira conhecida mundialmente cruzamento de géneros musicais ,
D 2008	2: waltel branco	**pt/w/a/l/Waltel_Branco_a388.html**: waltel branco . agora no rio de janeiro , gravou os discos " guitarra em chamas " 1 e 2 juntamente com o violonista baden powell e , tendo contato com a então inovadora bossa nova fez os arranjos de " chega de saudade " de joão gilberto ,
D 2008	3: rede globo	**pt/w/a/l/Waltel_Branco_a388.html**: waltel branco . gravou os discos " guitarra em chamas " 1 e 2 juntamente com o violonista baden powell e , tendo contato com a então inovadora bossa nova fez os arranjos de " chega de saudade " de joão gilberto , com quem veio a trabalhar por um longo período . na rede globo foi responsável pelas trilhas de

A 2010	1: NIL	
B 2010	1: nova	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: gravou . chega de saudade . o tema marcou o início da bossa - nova movimento por cruzou o samba com a ver mesmo jazz e tornou a música brasileira conhecida mundialmente cruzamento de géneros musicais influência também alguns músicos portugueses .
B 2010	2: waltel branco	**pt/w/a/l/Waltel_Branco_a388.html**: waltel branco . agora no rio de janeiro , gravou os discos " guitarra em chamas " 1 e 2 juntamente com o violonista baden powell e , tendo contato com a então inovadora bossa nova fez os arranjos de " chega de saudade " de joão gilberto ,
B 2010	3: rede globo	**pt/w/a/l/Waltel_Branco_a388.html**: waltel branco . gravou os discos " guitarra em chamas " 1 e 2 juntamente com o violonista baden powell e , tendo contato com a então inovadora bossa nova fez os arranjos de " chega de saudade " de joão gilberto , com quem veio a trabalhar por um longo período . na rede globo foi responsável pelas trilhas de
C 2010	1: NIL	
D 2010	1: nova	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: gravou , chega de saudade , o tema marcou o início da bossa - nova movimento por cruzou o samba com a ver mesmo jazz e tornou a música brasileira conhecida mundialmente cruzamento de géneros musicais influência também alguns músicos portugueses ,
D 2010	2: waltel branco	**pt/w/a/l/Waltel_Branco_a388.html**: waltel branco . agora no rio de janeiro , gravou os discos " guitarra em chamas " 1 e 2 juntamente com o violonista baden powell e , tendo contato com a então inovadora bossa nova fez os arranjos de " chega de saudade " de joão gilberto ,
D 2010	3: rede globo	**pt/w/a/l/Waltel_Branco_a388.html**: waltel branco . gravou os discos " guitarra em chamas " 1 e 2 juntamente com o violonista baden powell e , tendo contato com a então inovadora bossa nova fez os arranjos de " chega de saudade " de joão gilberto , com quem veio a trabalhar por um longo período . na rede globo foi responsável pelas trilhas de
Question #79 - Onde foi realizada a cimeira internacional sobre a SIDA?		
A 2008	1: tuberculose	**2008_07_25-19_59_01-Telejornal-1 bloco 1**: foi à cimeira dizer cara a cara aos actuais governantes que é preciso mais vontade de meios e estratégias conjuntas para o combate à malária tuberculose e sida .
A 2008	2: governantes	**2008_07_25-19_59_01-Telejornal-1 bloco 1**: foi à cimeira dizer cara a cara aos actuais governantes que é preciso mais vontade de meios e estratégias conjuntas para o combate à malária tuberculose e sida .

A 2008	3: méxico dezenas de manifestantes	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: uma cimeira internacional sobre a sida no méxico dezenas de manifestantes exigiram que os países apoio e mais destes doentes .
B 2008	1: sida no méxico dezenas de manifestantes	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: uma cimeira internacional sobre a sida no méxico dezenas de manifestantes exigiram que os países apoio e mais destes doentes .
B 2008	2: enquanto discursava bill clinton	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: internacional sobre a sida , exige mais apoio dos países para os doentes . os manifestantes , entraram mesma conferência enquanto discursava bill clinton . à qual barradas hoje o antigo presidente norte - americano . elogiou esta cimeira sobre a sida no méxico . o facto é que dirige ofélia ou .
B 2008	3: países	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: uma cimeira internacional sobre a sida no méxico dezenas de manifestantes exigiram que os países apoio e mais destes doentes .
C 2008	1: lagareiro vih	**2008_07_24-19_59_02-Telejornal-1 bloco 1**: cimeira da cplp hoje ao nível de ministros amanhã serão os chefes de estado , esta é a primeira fotografia de uma família unida com objectivos comuns que aproveitou uma concertação política , para aprofundar a cooperação em áreas como a saúde , o combate à lagareiro vih sida e a malária ,
C 2008	2: tuberculose	**2008_07_24-21_59_02-Jornal2-2 bloco 2**: amanhã sairá daqui desta reunião , uma declaração final , a imagem de pedro ribeiro , marco rocha na cimeira da cplp , malária a sida e tuberculose , são as doenças que mais matam os países de expressão portuguesa , jorge sampaio convocou por isso à margem da cimeira da cplp ,
C 2008	3: universal	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: sida é uma das maiores ameaças à estabilidade social das populações , foi ratificada uma resolução conjunta , tendo em vista o acesso universal à prevenção e tratamento , netos ficam aquém das necessidades e das recomendações apresentadas , jorge sampaio antigo presidente agora envolvido em operações de promoção da saúde , foi à cimeira
D 2008	1: sida no méxico	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: uma cimeira internacional sobre a sida no méxico dezenas de manifestantes exigiram que os países apoio e mais destes doentes , as nações unidas cá com que a doença possa haver mais de cinquenta milhões de crianças órfãs , só em áfrica , são crianças órfãs os pais morreram com sida ,

D 2008	2: nações unidas	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: uma cimeira internacional sobre a sida no méxico dezenas de manifestantes exigiram que os países apoio e mais destes doentes , as nações unidas cá com que a doença possa haver mais de cinquenta milhões de crianças órfãs , só em áfrica , são crianças órfãs os pais morreram com sida ,
D 2008	3: enquanto discursava bill clinton	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: internacional sobre a sida , exige mais apoio dos países para os doentes , os manifestantes , entraram mesma conferência enquanto discursava bill clinton , à qual barradas hoje o antigo presidente norte - americano , elogiou esta cimeira sobre a sida no méxico , o facto é que dirige ofélia ou ,
A 2010	1: tuberculose	**2008_07_24-21_59_02-Jornal2-2 bloco 2**: sairá daqui desta reunião uma declaração final a imagem de pedro ribeiro marco rocha na cimeira da cplp . malária sida e tuberculose .
A 2010	2: vih	**2008_07_24-19_59_02-Telejornal-1 bloco 1**: cimeira da cplp hoje ao nível de ministros amanhã serão os chefes de estado . esta é a primeira fotografia de uma família unida com objectivos comuns que aproveitou uma concertação política para aprofundar a cooperação em áreas como a saúde o combate larguei vih sida
A 2010	3: universal	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: sida é uma das maiores ameaças à estabilidade social das populações . foi ratificada uma resolução conjunta , tendo em vista o acesso universal à prevenção e tratamento . metas que ficam aquém das necessidades e das recomendações apresentadas jorge sampaio antigo presidente agora envolvido em operações de promoção da saúde . foi à cimeira
B 2010	1: nações unidas	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: na cimeira internacional sobre a sida no méxico dezenas de manifestantes exigiram que os países apoiem mais estes doentes as nações unidas cal com que a doença possa fazer mais de cinquenta milhões de crianças órfãs só em áfrica .
B 2010	2: sida no méxico	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: na cimeira internacional sobre a sida no méxico dezenas de manifestantes exigiram que os países apoiem mais estes doentes as nações unidas cal com que a doença possa fazer mais de cinquenta milhões de crianças órfãs só em áfrica .
B 2010	3: países	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: na cimeira internacional sobre a sida no méxico dezenas de manifestantes exigiram que os países apoiem mais estes doentes as nações unidas cal com que a doença possa fazer mais de cinquenta milhões de crianças órfãs só em áfrica .

C 2010	1: vih	**2008_07_24-19_59_02-Telejornal-1 bloco 1**: cimeira da cplp hoje ao nível de ministros amanhã serão os chefes de estado , esta é a primeira fotografia de uma família unida com objectivos comuns que aproveitou uma concertação política para aprofundar a cooperação em áreas como a saúde o combate larguei vih sida
C 2010	2: tuberculose	**2008_07_24-21_59_02-Jornal2-2 bloco 2**: sairá daqui desta reunião uma declaração final a imagem de pedro ribeiro marco rocha na cimeira da cplp , malária sida e tuberculose , são as doenças que mais matam os países de expressão portuguesa ,
C 2010	3: universal	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: sida é uma das maiores ameaças à estabilidade social das populações , foi ratificada uma resolução conjunta , tendo em vista o acesso universal à prevenção e tratamento , metas que ficam aquém das necessidades e das recomendações apresentadas jorge sampaio antigo presidente agora envolvido em operações de promoção da saúde , foi à cimeira
D 2010	1: sida no méxico	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: na cimeira internacional sobre a sida no méxico dezenas de manifestantes exigiram que os países apoiem mais estes doentes as nações unidas cal com que a doença possa fazer mais de cinquenta milhões de crianças órfãs só em áfrica ,
D 2010	2: nações unidas	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: na cimeira internacional sobre a sida no méxico dezenas de manifestantes exigiram que os países apoiem mais estes doentes as nações unidas cal com que a doença possa fazer mais de cinquenta milhões de crianças órfãs só em áfrica ,
D 2010	3: vivi nesta aldeia	**2008_08_05-21_59_02-Jornal2-2 bloco 2**: cimeira internacional sobre a sida no méxico dezenas de manifestantes exigiram que os países apoiem mais estes doentes as nações unidas cal com que a doença possa fazer mais de cinquenta milhões de crianças órfãs só em áfrica , são crianças órfãs os pais morreram com sida , vivi nesta aldeia do quénia , foi
Question #80 - Quem declarou aberta a trigésima segunda Festa do Avante?		
A 2008	1: inauguração da vigésima	**2008_09_05-19_59_02-Telejornal-1 bloco 2**: no seu discurso de abertura que vive no dia , que marca a inauguração da vigésima segunda edição da festa do avante , jerónimo de sousa disse aos militantes , que é preciso lançar .
A 2008	2: lançar	**2008_09_05-19_59_02-Telejornal-1 bloco 2**: no seu discurso de abertura que vive no dia , que marca a inauguração da vigésima segunda edição da festa do avante , jerónimo de sousa disse aos militantes , que é preciso lançar .
B 2008	1: faro ou jerónimo de sousa	**2008_09_05-21_59_01-Jornal2-2 bloco 1**: boa noite ana de faro ou jerónimo de sousa declarou aberta a festa do avante a trigésima segunda edição da festa , com um discurso muito crítico outra .

B 2008	2: avante	**2008_09_05-21_59_01-Jornal2-2 bloco 1**: boa noite ana de faro ou jerónimo de sousa declarou aberta a festa do avante a trigésima segunda edição da festa , com um discurso muito crítico outra .
C 2008	1: inauguração da vigésima	**2008_09_05-19_59_02-Telejornal-1 bloco 2**: no seu discurso de abertura que vive no dia , que marca a inauguração da vigésima segunda edição da festa do avante , jerónimo de sousa disse aos militantes , que é preciso lançar , uma campanha nacional ,
C 2008	2: lançar	**2008_09_05-19_59_02-Telejornal-1 bloco 2**: no seu discurso de abertura que vive no dia , que marca a inauguração da vigésima segunda edição da festa do avante , jerónimo de sousa disse aos militantes , que é preciso lançar , uma campanha nacional ,
C 2008	3: infinito	**2008_08_30-19_59_02-Telejornal-1 bloco 5**: dois , desde aí nunca mais cruzou os braços , e o mesmo parece ter acontecido a centenas de militantes preparação da festa já é uma festa infinito , meter o convívio permite vimos e ouvimos os amigos , que haja esse ouvimos uma vez por ano na festa do avante ,
D 2008	1: faro ou jerónimo de sousa	**2008_09_05-21_59_01-Jornal2-2 bloco 1**: o que há a destacar , boa noite ana de faro ou jerónimo de sousa declarou aberta a festa do avante a trigésima segunda edição da festa , com um discurso muito crítico outra ,
D 2008	2: avante	**2008_09_05-21_59_01-Jornal2-2 bloco 1**: o que há a destacar , boa noite ana de faro ou jerónimo de sousa declarou aberta a festa do avante a trigésima segunda edição da festa , com um discurso muito crítico outra ,
A 2010	1: inauguração da vigésima	**2008_09_05-19_59_02-Telejornal-1 bloco 2**: o seu discurso de abertura aqui que no dia que marca a inauguração da vigésima segunda edição da festa do avante jerónimo de sousa disse aos militantes que é preciso lançar uma campanha nacional contra as alterações ao código do trabalho dessa campanha .
A 2010	2: lançada	**2008_09_05-19_59_02-Telejornal-1 bloco 2**: segunda edição da festa do avante jerónimo de sousa disse aos militantes que é preciso lançar uma campanha nacional contra as alterações ao código do trabalho dessa campanha . vai ser lançada precisamente nesta edição da festa
A 2010	3: avançamos	**2008_09_05-21_59_01-Jornal2-2 bloco 4**: e avançamos para uma das propostas as propostas do cartaz do jornal dois . nome . a festa do avante
B 2010	1: avante	**2008_09_05-21_59_01-Jornal2-2 bloco 1**: boa noite ana , de facto , hoje ordem de sousa declarou aberta a festa do avante a trigésima segunda edição para esta com um discurso muito crítico outra .

B 2010	2: discurso muito crítico	**2008_09_05-21_59_01-Jornal2-2 bloco 1**: boa noite ana , de facto , hoje ordem de sousa declarou aberta a festa do avante a trigésima segunda edição para esta com um discurso muito crítico outra .
B 2010	3: sousa	**2008_09_05-21_59_01-Jornal2-2 bloco 1**: boa noite ana , de facto , hoje ordem de sousa declarou aberta a festa do avante a trigésima segunda edição para esta com um discurso muito crítico outra .
C 2010	1: avançamos	**2008_09_05-21_59_01-Jornal2-2 bloco 4**: e avançamos para uma das propostas as propostas do cartaz do jornal dois , nome , a festa do avante
C 2010	2: inauguração da vigésima	**2008_09_05-19_59_02-Telejornal-1 bloco 2**: o que disse o secretário - geral do pcp , o seu discurso de abertura aqui que no dia que marca a inauguração da vigésima segunda edição da festa do avante jerónimo de sousa disse aos militantes que é preciso lançar uma campanha nacional contra as alterações ao código do trabalho dessa campanha ,
C 2010	3: lançada	**2008_09_05-19_59_02-Telejornal-1 bloco 2**: segunda edição da festa do avante jerónimo de sousa disse aos militantes que é preciso lançar uma campanha nacional contra as alterações ao código do trabalho dessa campanha , vai ser lançada precisamente nesta edição da festa
D 2010	1: avante	**2008_09_05-21_59_01-Jornal2-2 bloco 1**: boa noite ana , de facto , hoje ordem de sousa declarou aberta a festa do avante a trigésima segunda edição para esta com um discurso muito crítico outra , o novo código do
D 2010	2: discurso muito crítico	**2008_09_05-21_59_01-Jornal2-2 bloco 1**: boa noite ana , de facto , hoje ordem de sousa declarou aberta a festa do avante a trigésima segunda edição para esta com um discurso muito crítico outra , o novo código do
D 2010	3: novo	**2008_09_05-21_59_01-Jornal2-2 bloco 1**: boa noite ana , de facto , hoje ordem de sousa declarou aberta a festa do avante a trigésima segunda edição para esta com um discurso muito crítico outra , o novo código do
Question #81 - Qual a última paragem da visita de Barack Obama à Europa?		
A 2008	1: médio oriente	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: barack obama terminou visita ao médio oriente . e a europa à última paragem foi em londres .
A 2008	2: londres	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: barack obama terminou visita ao médio oriente . e a europa à última paragem foi em londres .
B 2008	1: médio oriente	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: barack obama terminou visita ao médio oriente . e a europa à última paragem foi em londres .
B 2008	2: londres	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: barack obama terminou visita ao médio oriente . e a europa à última paragem foi em londres .

B 2008	3: médio oriente	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: barack obama terminou visita ao médio oriente . e a europa à última paragem foi em londres .
C 2008	1: médio oriente	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: fica cumprido , um último adeus , barack obama terminou visita ao médio oriente , e a europa à última paragem foi em londres , o se essa viagem foi visível em tróia mas para obama isso até poderá levar ,
C 2008	2: viagem	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: fica cumprido , um último adeus , barack obama terminou visita ao médio oriente , e a europa à última paragem foi em londres , o se essa viagem foi visível em tróia mas para obama isso até poderá levar ,
C 2008	3: adeus	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: fica cumprido , um último adeus , barack obama terminou visita ao médio oriente , e a europa à última paragem foi em londres , o se essa viagem foi visível em tróia mas para obama isso até poderá levar ,
D 2008	1: último adeus	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: fica cumprido , um último adeus , barack obama terminou visita ao médio oriente , e a europa à última paragem foi em londres , o se essa viagem foi visível em tróia mas para obama isso até poderá levar ,
D 2008	2: tróia	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: fica cumprido , um último adeus , barack obama terminou visita ao médio oriente , e a europa à última paragem foi em londres , o se essa viagem foi visível em tróia mas para obama isso até poderá levar ,
D 2008	3: isso até	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: fica cumprido , um último adeus , barack obama terminou visita ao médio oriente , e a europa à última paragem foi em londres , o se essa viagem foi visível em tróia mas para obama isso até poderá levar ,
A 2010 B 2010 C 2010 D 2010	1: NIL	
Question #82 - Que festa se realiza na Quinta da Atalaia?		
A 2008	1: NIL	
B 2008	1: festa do avante	**2008_08_30-19_59_02-Telejornal-1 bloco 5**: jovem passou o dia na quinta da atalaia no seixal , ao ajudar na montagem da festa do avante .
B 2008	2: principal	**2008_09_05-19_59_02-Telejornal-1 bloco 2**: no palco principal , no palco , vinte cinco de abril , chove muito aqui de na quinta da atalaia , na festa do avante continua .

C 2008	1: NIL	
D 2008	1: festa do avante	**2008_08_30-19_59_02-Telejornal-1 bloco 5**: jovem passou o dia na quinta da atalaia no seixal , ao ajudar na montagem da festa do avante , havia quente não ajuda trabalhos pesados mas o relógio não pára ,
D 2008	2: principal	**2008_09_05-19_59_02-Telejornal-1 bloco 2**: que vai ter lugar , no palco principal , no palco , vinte cinco de abril , chove muito aqui de na quinta da atalaia , na festa do avante continua , armando seixas ferreira com a festa do avante na quinta da atalaia ,
A 2010	1: NIL	
B 2010	1: festa do avante	**2008_08_30-19_59_02-Telejornal-1 bloco 5**: jerónimo passou o dia na quinta da atalaia no seixal ajudar na montagem da festa do avante .
B 2010	2: principal	**2008_09_05-19_59_02-Telejornal-1 bloco 2**: a festa do avante prolonga - se até domingo com gastronomia vários espectáculos com destaque para a grande ópera que vai ter lugar no palco principal um palco vinte cinco de abril chove muito aqui a na quinta da atalaia , mas a festa do avante continua .
C 2010	1: NIL	
D 2010	1: festa do avante	**2008_08_30-19_59_02-Telejornal-1 bloco 5**: jerónimo passou o dia na quinta da atalaia no seixal ajudar na montagem da festa do avante , um dia quente não ajuda trabalhos pesados ,
D 2010	2: principal	**2008_09_05-19_59_02-Telejornal-1 bloco 2**: a festa do avante prolonga - se até domingo com gastronomia vários espectáculos com destaque para a grande ópera que vai ter lugar no palco principal um palco vinte cinco de abril chove muito aqui a na quinta da atalaia , mas a festa do avante continua ,
Question #83 - Segundo Lula da Silva quais os melhores aviões?		
A 2008	1: brasil componentes	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: de cidade de évora vão ser para o brasil componentes para os aviões da embraer . os melhores segundo lula da silva .
A 2008	2: protocolo	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: as fábricas vão construir componentes para aviões . o protocolo foi assinado na presença de josé sócrates e do presidente do brasil , lula da silva .
A 2008	3: embraer	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: de cidade de évora vão ser para o brasil componentes para os aviões da embraer . os melhores segundo lula da silva .
B 2008	1: breve são de extrema qualidade e , que até , o governo para a compra de dois aviões	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: os melhores segundo lula da silva . o de agosto em breve são de extrema qualidade e , que até , o governo para a compra de dois aviões .

B 2008	2: cidade de évora vão	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: de cidade de évora vão ser para o brasil componentes para os aviões da embraer . os melhores segundo lula da silva .
B 2008	3: embraer	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: de cidade de évora vão ser para o brasil componentes para os aviões da embraer . os melhores segundo lula da silva .
C 2008	1: qualidade	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: os melhores segundo lula da silva , o da boda e bresson de tamanha qualidade e , que é até um governo para a compra de dois aviões , numa primeira fase vão ser investidos cento e cinquenta milhões de euros ,
C 2008	2: brasil	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: de cidade de évora vão ser para o brasil componentes para os aviões da embraer , os melhores segundo lula da silva , o da boda e bresson de tamanha qualidade e , que é até um governo para a compra de dois aviões ,
C 2008	3: postos de trabalho directos	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: irão ser criados cerca de quinhentos postos de trabalho directos , as fábricas vão construir componentes para aviões , o protocolo foi assinado na presença de josé sócrates e do presidente do brasil , lula da silva , sócrates pediu ele aceitou ficar mais um dia de portugal ,
D 2008	1: até um governo	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: os melhores segundo lula da silva , o da boda e bresson de tamanha qualidade e , que é até um governo para a compra de dois aviões , numa primeira fase vão ser investidos cento e cinquenta milhões de euros ,
D 2008	2: primeira fase vão	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: os melhores segundo lula da silva , o da boda e bresson de tamanha qualidade e , que é até um governo para a compra de dois aviões , numa primeira fase vão ser investidos cento e cinquenta milhões de euros ,
D 2008	3: bresson de tamanha qualidade	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: os melhores segundo lula da silva , o da boda e bresson de tamanha qualidade e , que é até um governo para a compra de dois aviões , numa primeira fase vão ser investidos cento e cinquenta milhões de euros ,
A 2010	1: protocolo foi assinado na presença de josé sócrates e do presidente do brasil lula da silva	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: as fábricas vão construir componentes para aviões . o protocolo foi assinado na presença de josé sócrates e do presidente do brasil lula da silva .

A 2010	2: brasil	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: o fabricante brasileiro de aviões embraer vai construir duas fábricas de componentes em portugal . josé sócrates e o presidente do brasil lula da silva .
A 2010	3: banda em breve se onde tem maior qualidade	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: aviões da embraer . os melhores segundo lula da silva . o da banda em breve se onde tem maior qualidade .
B 2010	1: cidade de évora vou sair para o brasil componentes para os aviões da embraer	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: dentro de três anos as duas fábricas estarão a funcionar em pleno da cidade de évora vou sair para o brasil componentes para os aviões da embraer . os melhores segundo lula da silva .
B 2010	2: banda em breve se onde tem maior qualidade	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: aviões da embraer . os melhores segundo lula da silva . o da banda em breve se onde tem maior qualidade .
B 2010	3: anos	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: dentro de três anos as duas fábricas estarão a funcionar em pleno da cidade de évora vou sair para o brasil componentes para os aviões da embraer . os melhores segundo lula da silva .
C 2010	1: protocolo foi assinado	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: as fábricas vão construir componentes para aviões , o protocolo foi assinado na presença de josé sócrates e do presidente do brasil lula da silva ,
C 2010	2: brasil	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: as fábricas vão construir componentes para aviões , o protocolo foi assinado na presença de josé sócrates e do presidente do brasil lula da silva ,
C 2010	3: banda em breve	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: aviões da embraer , os melhores segundo lula da silva , o da banda em breve se onde tem maior qualidade , quetta um governo para compra de dois aviões ,
D 2010	1: banda em breve se onde tem maior qualidade	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: aviões da embraer , os melhores segundo lula da silva , o da banda em breve se onde tem maior qualidade , quetta um governo para compra de dois aviões ,
D 2010	2: quetta um governo	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: aviões da embraer , os melhores segundo lula da silva , o da banda em breve se onde tem maior qualidade , quetta um governo para compra de dois aviões ,
D 2010	3: governo para compra	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: aviões da embraer , os melhores segundo lula da silva , o da banda em breve se onde tem maior qualidade , quetta um governo para compra de dois aviões ,

Question #84 - Qual a participação do Estado Português na OGMA?		
A 2008	1: universidade de évora	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: na cidade de bergen as condições para que a universidade de évora possa desenvolver cursos , para esta área . a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital . o estado português .
A 2008	2: brasileira em bayreuth terceira	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: na cidade de bergen as condições para que a universidade de évora possa desenvolver cursos , para esta área . a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital . o estado português .
A 2008	3: aviões	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: na cidade de bergen as condições para que a universidade de évora possa desenvolver cursos , para esta área . a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital . o estado português .
B 2008	1: brasileira em bayreuth terceira maior empresa mundial do fabrico	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: na cidade de bergen as condições para que a universidade de évora possa desenvolver cursos , para esta área . a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital . o estado português .
B 2008	2: accionista da portuguesa	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: na cidade de bergen as condições para que a universidade de évora possa desenvolver cursos , para esta área . a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital . o estado português .
B 2008	3: cidade de bergen	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: na cidade de bergen as condições para que a universidade de évora possa desenvolver cursos , para esta área . a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital . o estado português .
C 2008	1: universidade de évora	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: na cidade de bergen as condições para que a universidade de évora possa desenvolver cursos , para esta área , a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital , o estado português ,

C 2008	2: brasileira em bayreuth terceira	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: na cidade de bergen as condições para que a universidade de évora possa desenvolver cursos , para esta área , a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital , o estado português ,
C 2008	3: cultural	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital , o estado português , detém os restantes trinta e cinco por cento , a proximidade à linha proximidade cultural , a nossa presença já que em portugal desde a minha quase na
D 2008	1: maior accionista da portuguesa	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital , o estado português , detém os restantes trinta e cinco por cento , a proximidade à linha proximidade cultural , a nossa presença já que em portugal desde a minha quase na
D 2008	2: cidade de bergen	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: na cidade de bergen as condições para que a universidade de évora possa desenvolver cursos , para esta área , a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital , o estado português ,
D 2008	3: capital	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital , o estado português , detém os restantes trinta e cinco por cento , a proximidade à linha proximidade cultural , a nossa presença já que em portugal desde a minha quase na
A 2010	1: língua cidade cultural	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento . a proximidade da língua cidade cultural a nossa presença já aqui em portugal desde a minha quatro na
A 2010	2: capital	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento . a proximidade da língua cidade cultural a nossa presença já aqui em portugal desde a minha quatro na
A 2010	3: terceira	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: na sua cidade vai criar melhores condições para que a universidade deve possa desenvolver cursos para esta área . a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português

B 2010	1: terceira maior empresa mundial no fabrico de aviões	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: na sua cidade vai criar melhores condições para que a universidade deve possa desenvolver cursos para esta área . a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português
B 2010	2: presença já aqui em portugal	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento . a proximidade da língua cidade cultural a nossa presença já aqui em portugal desde a minha quatro na
B 2010	3: língua cidade	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento . a proximidade da língua cidade cultural a nossa presença já aqui em portugal desde a minha quatro na
C 2010	1: língua cidade cultural	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento , a proximidade da língua cidade cultural a nossa presença já aqui em portugal desde a minha quatro na
C 2010	2: capital	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento , a proximidade da língua cidade cultural a nossa presença já aqui em portugal desde a minha quatro na
C 2010	3: terceira	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: na sua cidade vai criar melhores condições para que a universidade deve possa desenvolver cursos para esta área , a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português
D 2010	1: língua cidade	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento , a proximidade da língua cidade cultural a nossa presença já aqui em portugal desde a minha quatro na
D 2010	2: terceira maior empresa mundial no fabrico de aviões	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: na sua cidade vai criar melhores condições para que a universidade deve possa desenvolver cursos para esta área , a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português

D 2010	3: presença já aqui	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento , a proximidade da língua cidade cultural a nossa presença já aqui em portugal desde a minha quatro na
Question #85 - Quem tem cada vez mais casos de cancro cutâneo?		
A 2008	1: falésia	**2008_07_19-21_59_02-Jornal2-2 bloco 2**: para chegar a mensagem principalmente aos mais novos à associação portuguesa de cancro cutâneo vou figuras públicas à praia da falésia .
A 2008	2: vilamoura	**2008_07_19-21_59_02-Jornal2-2 bloco 2**: cutâneo vou figuras públicas à praia da falésia . em vilamoura no algarve . há caras conhecidas no areal , são actores que não vem representar , atletas que não vem competir , uniram - se à volta do que é preciso dizer , o cancro
A 2008	3: esperança para o combate	**2008_06_21-19_59_01-Telejornal-1 bloco 3**: há uma nova esperança para o combate ao cancro da pele .
B 2008	1: amadora	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: há cada vez mais casos de jovens com cancro cutâneo , e é também entre os jovens que se registam maior número de que a amadora os lados ,
B 2008	2: registam	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: há cada vez mais casos de jovens com cancro cutâneo , e é também entre os jovens que se registam maior número de que a amadora os lados ,
B 2008	3: jovens	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: há cada vez mais casos de jovens com cancro cutâneo
C 2008	1: falésia em vilamoura	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: para chegar a mensagem principalmente aos mais novos à associação portuguesa de cancro cutâneo , o voo figuras públicas à praia da falésia em vilamoura no algarve , há caras novas no areal , são actores que não vem representar ,
C 2008	2: amadora	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: há cada vez mais casos de jovens com cancro cutâneo , e é também entre os jovens que se registam maior número de que a amadora os lados , para chegar a mensagem principalmente aos mais novos à associação portuguesa de cancro cutâneo ,
C 2008	3: principalmente	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: há cada vez mais casos de jovens com cancro cutâneo , e é também entre os jovens que se registam maior número de que a amadora os lados , para chegar a mensagem principalmente aos mais novos à associação portuguesa de cancro cutâneo ,

D 2008	1: novos à associação portuguesa	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: às portas da câmara , há cada vez mais casos de jovens com cancro cutâneo , e é também entre os jovens que se registam maior número de que a amadora os lados , para chegar a mensagem principalmente aos mais novos à associação portuguesa de cancro cutâneo ,
D 2008	2: registam	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: de haver responsáveis , até lá a comunidade diz ficar por cá , às portas da câmara , há cada vez mais casos de jovens com cancro cutâneo , e é também entre os jovens que se registam maior número de que a amadora os lados ,
D 2008	3: até	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: de haver responsáveis , até lá a comunidade diz ficar por cá , às portas da câmara , há cada vez mais casos de jovens com cancro cutâneo , e é também entre os jovens que se registam maior número de que a amadora os lados ,
A 2010	1: falésia em vilamoura	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: cancro cutâneo levou figuras públicas à praia da falésia em vilamoura no algarve .
A 2010	2: importância da prevenção	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: cancro cutâneo que voltasse sublinhar a importância da prevenção e dos cuidados a ter .
A 2010	3: cada	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: há cada vez mais casos de jovens com cancro cutâneo
B 2010	1: jovens	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: há cada vez mais casos de jovens com cancro cutâneo
C 2010	1: falésia em vilamoura	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: cancro cutâneo levou figuras públicas à praia da falésia em vilamoura no algarve , há caras conhecidas no areal são actores que não vem representar ,
C 2010	2: sublinhar a importância da prevenção	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: cancro cutâneo que voltasse sublinhar a importância da prevenção e dos cuidados a ter , durante a exposição ao sol ,
C 2010	3: exposição	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: cancro cutâneo que voltasse sublinhar a importância da prevenção e dos cuidados a ter , durante a exposição ao sol ,
D 2010	1: novos à associação portuguesa	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: há cada vez mais casos de jovens com cancro cutâneo e é também entre os jovens que se registam maior número de camas duras solares para chegar a mensagem principalmente aos mais novos à associação portuguesa de cancro

D 2010	2: can	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: até lá a comunidade diz ficar por cá às portas da can , há cada vez mais casos de jovens com cancro cutâneo
D 2010	3: até	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: até lá a comunidade diz ficar por cá às portas da can , há cada vez mais casos de jovens com cancro cutâneo
Question #86 - Quem regista maior número de queimaduras solares?		
A 2008	1: radiação	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: cancro cutâneo , que voltar a sublinhar a importância da prevenção dos cuidados a ter , durante a exposição ao sol , num dia com níveis de radiação muito altos que o dia de hoje . os mais novos dão o exemplo na prevenção e na forma como lidar , com as queimaduras solares .
B 2008	1: associação portuguesa	**2008_07_19-21_59_02-Jornal2-2 bloco 2**: sofrem queimaduras solares . a conclusão é de um estudo da associação portuguesa de cancro cutâneo , que volta a sublinhar a importância da prevenção e dos cuidados a ter .
B 2008	2: sublinhar	**2008_07_19-21_59_02-Jornal2-2 bloco 2**: sofrem queimaduras solares . a conclusão é de um estudo da associação portuguesa de cancro cutâneo , que volta a sublinhar a importância da prevenção e dos cuidados a ter .
B 2008	3: novos	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: cancro cutâneo , que voltar a sublinhar a importância da prevenção dos cuidados a ter , durante a exposição ao sol , num dia com níveis de radiação muito altos que o dia de hoje . os mais novos dão o exemplo na prevenção e na forma como lidar , com as queimaduras solares .
C 2008	1: radiação	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: cancro cutâneo , que voltar a sublinhar a importância da prevenção dos cuidados a ter , durante a exposição ao sol , num dia com níveis de radiação muito altos que o dia de hoje , os mais novos dão o exemplo na prevenção e na forma como lidar , com as queimaduras solares ,
D 2008	1: associação portuguesa	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: sofrem efectivamente queimaduras solares a conclusão é de um estudo da associação portuguesa de cancro cutâneo , que voltar a sublinhar a importância da prevenção dos cuidados a ter , durante a exposição ao sol ,
D 2008	2: sublinhar	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: sofrem efectivamente queimaduras solares a conclusão é de um estudo da associação portuguesa de cancro cutâneo , que voltar a sublinhar a importância da prevenção dos cuidados a ter , durante a exposição ao sol ,
D 2008	3: novos	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: cancro cutâneo , que voltar a sublinhar a importância da prevenção dos cuidados a ter , durante a exposição ao sol , num dia com níveis de radiação muito altos que o dia de hoje , os mais novos dão o exemplo na prevenção e na forma como lidar , com as queimaduras solares ,

A 2010	1: radiação	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: cancro cutâneo que voltasse sublinhar a importância da prevenção e dos cuidados a ter . durante a exposição ao sol . num dia com níveis de radiação muito altos como dia de hoje . os mais novos dão o exemplo na prevenção e na forma como lidar com as queimaduras solares .
B 2010	1: associação portuguesa	**2008_07_19-21_59_02-Jornal2-2 bloco 2**: sofrem queimaduras solares a conclusão é de um estudo da associação portuguesa de cancro cutâneo que volta a sublinhar a importância da prevenção e dos cuidados a ter .
B 2010	2: sublinhar	**2008_07_19-21_59_02-Jornal2-2 bloco 2**: sofrem queimaduras solares a conclusão é de um estudo da associação portuguesa de cancro cutâneo que volta a sublinhar a importância da prevenção e dos cuidados a ter .
B 2010	3: novos	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: cancro cutâneo que voltasse sublinhar a importância da prevenção e dos cuidados a ter . durante a exposição ao sol . num dia com níveis de radiação muito altos como dia de hoje . os mais novos dão o exemplo na prevenção e na forma como lidar com as queimaduras solares .
C 2010	1: radiação	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: cancro cutâneo que voltasse sublinhar a importância da prevenção e dos cuidados a ter , durante a exposição ao sol , num dia com níveis de radiação muito altos como dia de hoje , os mais novos dão o exemplo na prevenção e na forma como lidar com as queimaduras solares ,
D 2010	1: associação portuguesa	**2008_07_19-21_59_02-Jornal2-2 bloco 2**: sofrem queimaduras solares a conclusão é de um estudo da associação portuguesa de cancro cutâneo que volta a sublinhar a importância da prevenção e dos cuidados a ter , durante a exposição ao sol ,
D 2010	2: sublinhar	**2008_07_19-21_59_02-Jornal2-2 bloco 2**: sofrem queimaduras solares a conclusão é de um estudo da associação portuguesa de cancro cutâneo que volta a sublinhar a importância da prevenção e dos cuidados a ter , durante a exposição ao sol ,
D 2010	3: novos	**2008_07_19-19_59_02-Telejornal-1 bloco 1**: cancro cutâneo que voltasse sublinhar a importância da prevenção e dos cuidados a ter , durante a exposição ao sol , num dia com níveis de radiação muito altos como dia de hoje , os mais novos dão o exemplo na prevenção e na forma como lidar com as queimaduras solares ,
Question #87 - Quantos prédios devolutos existem em Lisboa?		
A 2008	1: seiscentos prédios	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: lisboa no mesmo estado deste que ardeu na última noite . a câmara tem identificados quatro mil e seiscentos prédios devolutos , só metade aguarda licença para a reconstrução . enquanto as obras não começam , muito servem para actividades , não se aplicadas . é um movimento constante mesmo num prédio desabitado .

A 2008	2: três prédios	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: boa noite durante sete horas dezenas de bombeiros combateram um incêndio no coração de lisboa três prédios foram atingidos . os moradores prometem processar a câmara municipal e a empresa proprietária do edifício desabitado começaram as armas .
B 2008	1: três prédios	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: lisboa três prédios foram atingidos . os moradores prometem processar a câmara municipal e a empresa proprietária do edifício desabitado começaram as armas . só com catorze minutos das onze da noite , quando foi
B 2008	2: seiscentos prédios	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: lisboa no mesmo estado deste que ardeu na última noite . a câmara tem identificados quatro mil e seiscentos prédios devolutos , só metade aguarda licença para a reconstrução . enquanto as obras não começam , muito servem para actividades , não se aplicadas . é um movimento constante mesmo num prédio desabitado .
C 2008	1: seiscentos prédios	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: lisboa no mesmo estado deste que ardeu na última noite , a câmara tem identificados quatro mil e seiscentos prédios devolutos , só metade aguarda licença para a reconstrução , enquanto as obras não começam , muito servem para actividades , não se aplicadas , é um movimento constante mesmo num prédio desabitado ,
C 2008	2: três prédios	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: durante sete horas dezenas de bombeiros combateram um incêndio no coração de lisboa , três prédios foram atingidos , prometem processar a câmara municipal , e a empresa proprietária do edifício desabitado , onde começaram as aulas , ribeiro porque o proprietário do bem ,
D 2008	1: seiscentos prédios	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: lisboa no mesmo estado deste que ardeu na última noite , a câmara tem identificados quatro mil e seiscentos prédios devolutos , só metade aguarda licença para a reconstrução , enquanto as obras não começam , muito servem para actividades , não se aplicadas , é um movimento constante mesmo num prédio desabitado ,
D 2008	2: três prédios	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: lisboa , três prédios foram atingidos , prometem processar a câmara municipal , e a empresa proprietária do edifício desabitado , onde começaram as aulas , ribeiro porque o proprietário do bem , tem
A 2010	1: seiscentos prédios	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: lisboa no mesmo estado deste que ardeu na última noite . a câmara tem identificados quatro mil e seiscentos prédios devolutos . só metade aguarda licença para reconstrução . enquanto as obras não começam muitos servem para actividades moss picadas . é um movimento constante mesmo num prédio desabitado .

A 2010	2: três prédios	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: durante sete horas dezenas de bombeiros combateram um incêndio no coração de lisboa três prédios foram atingidos . os moradores prometem processar a câmara municipal e a empresa proprietária do edifício desabitado onde começaram as chamas .
A 2010	3: dois prédios	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: lisboa quatro mil seiscentos e dois prédios devolutos com este vinte e três dava da liberdade .
B 2010	1: três prédios	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: lisboa três prédios foram atingidos . os moradores prometem processar a câmara municipal e a empresa proprietária do edifício desabitado onde começaram as chamas . passavam catorze minutos das onze da noite quando foi dado o alarme estava
B 2010	2: seiscentos prédios	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: lisboa no mesmo estado deste que ardeu na última noite . a câmara tem identificados quatro mil e seiscentos prédios devolutos . só metade aguarda licença para reconstrução . enquanto as obras não começam muitos servem para actividades moss picadas . é um movimento constante mesmo num prédio desabitado .
B 2010	3: dois prédios	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: existem em lisboa quatro mil seiscentos e dois prédios devolutos com este vinte e três dava da liberdade .
C 2010	1: seiscentos prédios	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: lisboa no mesmo estado deste que ardeu na última noite , a câmara tem identificados quatro mil e seiscentos prédios devolutos , só metade aguarda licença para reconstrução , enquanto as obras não começam muitos servem para actividades moss picadas , é um movimento constante mesmo num prédio desabitado ,
C 2010	2: três prédios	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: durante sete horas dezenas de bombeiros combateram um incêndio no coração de lisboa , três prédios foram atingidos , prometem processar a câmara municipal e a empresa proprietária do edifício desabitado , onde começaram chamas , primeiro ,
C 2010	3: dois prédios	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: lisboa quatro mil seiscentos e dois prédios devolutos com este vinte e três dava da liberdade ,
D 2010	1: três prédios	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: lisboa três prédios foram atingidos , os moradores prometem processar a câmara municipal e a empresa proprietária do edifício desabitado onde começaram as chamas , passavam catorze minutos das onze da noite quando foi
D 2010	2: seiscentos prédios	**2008_07_07-21_59_01-Jornal2-2 bloco 1**: as chamas foram extintas ao nascer do dia , há milhares de edifícios em lisboa no mesmo estado deste que ardeu na última noite , a câmara tem identificados quatro mil e seiscentos prédios devolutos , só metade aguarda licença para reconstrução ,

D 2010	3: dois prédios	**2008_07_07-19_59_01-Telejornal-1 bloco 1**: existem em lisboa quatro mil seiscentos e dois prédios devolutos com este vinte e três dava da liberdade ,
Question #88 - Quem anunciou o reforço das linhas de crédito para as empresas Portuguesas que invistam em Angola?		
A 2008	1: guebuza	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: em angola e armando guebuza não deixou moçambique causou lula da silva ramos horta foram dos que participaram e ouviram josé sócrates apontar as prioridades da presidência portuguesa da cplp e assumida a partir de agora o tempo . a língua . a cidadania . a concertação diplomática . e o reforço da cooperação .
A 2008	2: nu as debilidades	**2008_07_19-21_59_02-Jornal2-2 bloco 2**: em angola , e recusam crédito à madeira . o líder do governo regional recordou o relatório divulgado pelo fundo monetário internacional . para dizer que as razões da crise são internas . jardim citou o relatório do fundo monetário internacional , que coloca a nu as debilidades da economia portuguesa ,
B 2008	1: primeira linha	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: à saída o primeiro - ministro anunciou o reforço das linhas de crédito , para as empresas portuguesas que invistam , em angola . a primeira linha de crédito uma linha da frente tejo de cem milhões de euros , pois há uma linha
B 2008	2: josé sócrates , debateu com o presidente angolano josé eduardo dos santos	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: . chegou bem cedo luanda com mais gente é preenchida que começou coube encontro do dia josé sócrates , debateu com o presidente angolano josé eduardo dos santos . os laços comerciais entre os dois países , à saída o primeiro - ministro anunciou o reforço das linhas de crédito para as empresas portuguesas que invistam .
B 2008	3: primeiro - ministro	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: à saída o primeiro - ministro anunciou o reforço das linhas de crédito , para as empresas portuguesas que invistam , em angola . a primeira linha de crédito uma linha da frente tejo de cem milhões de euros , pois há uma linha
C 2008	1: guebuza	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: em angola e armando guebuza não deixou moçambique causou lula da silva ramos horta foram dos que participaram e ouviram josé sócrates apontar as prioridades da presidência portuguesa da cplp e assumida a partir de agora o tempo , a língua , a cidadania , a concertação diplomática , e o reforço da cooperação ,
C 2008	2: nu as debilidades	**2008_07_19-21_59_02-Jornal2-2 bloco 2**: em angola , e recusam crédito à madeira , o líder do governo regional recordou o relatório divulgado pelo fundo monetário internacional , para dizer que as razões da crise são internas , jardim citou o relatório do fundo monetário internacional , que coloca a nu as debilidades da economia portuguesa ,

D 2008	1: primeira linha	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: à saída o primeiro - ministro anunciou o reforço das linhas de crédito para as empresas portuguesas que invistam , em angola , a primeira linha de crédito uma linha da frente tejo , cem milhões de euros , pois há uma linha
D 2008	2: josé sócrates , debateu com o presidente angolano josé eduardo dos santos	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: , chegou bem cedo luanda com uma agenda preenchida que começou coube encontro do dia josé sócrates , debateu com o presidente angolano josé eduardo dos santos , os laços comerciais entre os dois países , à saída o primeiro - ministro anunciou o reforço das linhas de crédito , para as empresas portuguesas que invistam ,
D 2008	3: primeiro - ministro	**2008_07_17-19_59_01-Telejornal-1 bloco 1**: à saída o primeiro - ministro anunciou o reforço das linhas de crédito para as empresas portuguesas que invistam , em angola , a primeira linha de crédito uma linha da frente tejo , cem milhões de euros , pois há uma linha
A 2010	1: guebuza	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: em angola e armando guebuza não deixo moçambique causou lula da silva ramos horta foram dos que participaram e ouviram josé sócrates apontar as prioridades da presidência portuguesa da cplp assumida a partir de agora impuro a língua . a cidadania . a concertação diplomática e o reforço da cooperação .
A 2010	2: impuro	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: em angola e armando guebuza não deixo moçambique causou lula da silva ramos horta foram dos que participaram e ouviram josé sócrates apontar as prioridades da presidência portuguesa da cplp assumida a partir de agora impuro a língua . a cidadania . a concertação diplomática e o reforço da cooperação .
A 2010	3: nu as debilidades	**2008_07_19-19_59_02-Telejornal-1 bloco 2**: em angola e recusa o crédito à madeira . o líder do governo regional recordou relatório divulgado pelo fundo monetário internacional para dizer que as razões da crise são internas . jardim citou relatório do fundo monetário internacional que coloca a nu as debilidades da economia portuguesa .
B 2010	1: primeira linha de kate	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: anunciou reforço das linhas de crédito para as empresas portuguesas que invistam . em angola . a primeira linha de kate uma linha tratei judas seis milhões de euros depois há uma linha
B 2010	2: josé sócrates bateu o presidente angolano josé eduardo dos santos	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: angola . chegou bem cedo luanda com mais gender preenchida que começou com o encontro do dia josé sócrates bateu o presidente angolano josé eduardo dos santos . os laços comerciais entre os dois países à saída o primeiro - ministro anunciou reforço das linhas de crédito para as empresas portuguesas que invistam . em

B 2010	3: primeiro - ministro	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: os laços comerciais entre os dois países à saída o primeiro - ministro anunciou reforço das linhas de crédito para as empresas portuguesas que invistam . em angola .
C 2010	1: guebuza	**2008_07_25-19_59_01-Telejornal-1 bloco 1**: em angola e armando guebuza não deixo moçambique causou lula da silva ramos horta foram dos que participaram e ouviram josé sócrates apontar as prioridades da presidência portuguesa da cplp assumida a partir de agora impuro a língua , a cidadania , a concertação diplomática e o reforço da cooperação ,
C 2010	2: impuro	**2008_07_25-19_59_01-Telejornal-1 bloco 1**: em angola e armando guebuza não deixo moçambique causou lula da silva ramos horta foram dos que participaram e ouviram josé sócrates apontar as prioridades da presidência portuguesa da cplp assumida a partir de agora impuro a língua , a cidadania , a concertação diplomática e o reforço da cooperação ,
C 2010	3: nu as debilidades	**2008_07_19-19_59_02-Telejornal-1 bloco 2**: em angola e recusa o crédito à madeira , o líder do governo regional recordou relatório divulgado pelo fundo monetário internacional para dizer que as razões da crise são internas , jardim citou relatório do fundo monetário internacional que coloca a nu as debilidades da economia portuguesa ,
D 2010	1: primeira linha de kate	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: anunciou reforço das linhas de crédito para as empresas portuguesas que invistam , em angola , a primeira linha de kate uma linha tratei judas seis milhões de euros depois há uma linha
D 2010	2: josé sócrates bateu o presidente angolano josé eduardo dos santos	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: angola , chegou bem cedo luanda com mais gender preenchida que começou com o encontro do dia josé sócrates bateu o presidente angolano josé eduardo dos santos , os laços comerciais entre os dois países à saída o primeiro - ministro anunciou reforço das linhas de crédito para as empresas portuguesas que invistam , em
D 2010	3: primeiro - ministro	**2008_07_17-21_59_02-Jornal2-2 bloco 2**: os laços comerciais entre os dois países à saída o primeiro - ministro anunciou reforço das linhas de crédito para as empresas portuguesas que invistam , em angola ,
Question #89 - Quantos milhões de euros têm as linhas de crédito?		
A 2008	1: NIL	
B 2008	1: cem mil euros	**2008_06_04-19_59_01-Telejornal-1 bloco 1**: e uma linha de crédito de quarenta milhões de euros por cinco anos . a frota da câmara de tavira já começou a abastecer - se em espanha a autarquia diz tratar - se de um protesto à actuação do governo , e vai poupar cem mil euros por ano .

B 2008	2: três euros	**2008_07_01-21_59_02-Jornal2-2 bloco 3**: linha de crédito de quarenta milhões de euros fique disponível . a data , não foi anunciada . os serviços religiosos estão mais caros , em vez dos sete euros e meio os católicos vão pagar três euros para mandar celebrar uma missa , num baptizado ,
B 2008	3: sete euros	**2008_07_01-21_59_02-Jornal2-2 bloco 3**: linha de crédito de quarenta milhões de euros fique disponível . a data , não foi anunciada . os serviços religiosos estão mais caros , em vez dos sete euros e meio os católicos vão pagar três euros para mandar celebrar uma missa , num baptizado ,
C 2008	1: NIL	
D 2008	1: sete euros	**2008_07_01-21_59_02-Jornal2-2 bloco 3**: linha de crédito de quarenta milhões de euros fique disponível , a data , não foi anunciada , os serviços religiosos estão mais caros , em vez dos sete euros e meio os católicos vão pagar três euros para mandar celebrar uma missa , num baptizado ,
D 2008	2: três euros	**2008_07_01-21_59_02-Jornal2-2 bloco 3**: linha de crédito de quarenta milhões de euros fique disponível , a data , não foi anunciada , os serviços religiosos estão mais caros , em vez dos sete euros e meio os católicos vão pagar três euros para mandar celebrar uma missa , num baptizado ,
D 2008	3: cem mil euros	**2008_06_04-19_59_01-Telejornal-1 bloco 1**: e uma linha de crédito de quarenta milhões de euros por cinco anos , a frota da câmara de tavira já começou a abastecer - se em espanha a autarquia diz tratar - se de um protesto à actuação do governo , e vai poupar cem mil euros por ano ,
A 2010	1: NIL	
B 2010	1: três euros	**2008_07_01-21_59_02-Jornal2-2 bloco 3**: linha de crédito de quarenta milhões de euros e que disponível a dada não foi anunciada . os serviços religiosos estão mais caros , em vez dos sete euros e meios católicos vão pagar três euros
B 2010	2: sete euros	**2008_07_01-21_59_02-Jornal2-2 bloco 3**: linha de crédito de quarenta milhões de euros e que disponível a dada não foi anunciada . os serviços religiosos estão mais caros , em vez dos sete euros e meios católicos vão pagar três euros
B 2010	3: cinco euros	**2008_08_11-19_59_02-Telejornal-1 bloco 2**: crédito e débito e diariamente são usados um milhão de vezes os números da unicre dão aos portugueses lugar entre o grupo de líderes europeus na escolha desta forma de pagamento . utilizamos o cartão muitas vezes e simultâneamente em pequenas transacções estação de um país que tem na prestação média trinta e cinco euros

C 2010	1: NIL	
D 2010	1: três euros	**2008_07_01-21_59_02-Jornal2-2 bloco 3**: linha de crédito de quarenta milhões de euros e que disponível a dada não foi anunciada , os serviços religiosos estão mais caros , em vez dos sete euros e meios católicos vão pagar três euros
D 2010	2: sete euros	**2008_07_01-21_59_02-Jornal2-2 bloco 3**: linha de crédito de quarenta milhões de euros e que disponível a dada não foi anunciada , os serviços religiosos estão mais caros , em vez dos sete euros e meios católicos vão pagar três euros
D 2010	3: cinco euros	**2008_08_11-21_59_01-Jornal2-2 bloco 1**: crédito e débito e diariamente são usados um milhão de vezes os números da unicre dão aos portugueses lugar entre o grupo de líderes europeus na escolha desta forma de pagamento , utilizamos o cartão muitas vezes e simultâneamente em pequenas transacções estação de um país que tem na prestação média trinta e cinco euros
Question #90 - Em que praia vai actuar Emir Kusturica amanhã à noite?		
A 2008	1: NIL	
B 2008	1: bósnia herzegovina	**pt/e/m/i/Emir_Kusturica_5132.html**: emir kusturica . (1 993) underground , palma de ouro festival de cannes , (1 995) ligações externas [1] " emir kusturica na imdb categoria : cineastas da bósnia herzegovina
B 2008	2: praia	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: quando o emir kusturica se juntasse não ganho organiza , a festa é garantida . é o que vai acontecer na praia do tonel em sagres amanhã à noite .
C 2008	1: NIL	
D 2008	1: bósnia herzegovina	**pt/e/m/i/Emir_Kusturica_5132.html**: emir kusturica . (1 993) underground , palma de ouro festival de cannes , (1 995) ligações externas [1] " emir kusturica na imdb categoria : cineastas da bósnia herzegovina
D 2008	2: praia	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: música árabe e africana de um , quando o emir kusturica se juntasse não ganho organiza , a festa é garantida , é o que vai acontecer na praia do tonel em sagres amanhã à noite , no encerramento do super bowl , sete , França ,
A 2010	1: NIL	
B 2010	1: praia	**2008_08_14-21_59_01-Jornal2-2 bloco 3**: quando emir kusturica se juntasse não que ingote será a festa é garantida . é o que vai acontecer na praia do tonel em sagres amanhã à noite .
C 2010	1: NIL	
D 2010	1: praia	**2008_08_14-21_59_01-Jornal2-2 bloco 3**: um ziga árabe e africano de , um , quando emir kusturica se juntasse não que ingote será a festa é garantida , é o que vai acontecer na praia do tonel em sagres amanhã à noite , no encerramento do super bock , surf , França ,

Question #91 - Qual a posição de James Blake no ranking ATP?		
A 2008	1: americano sétimo	**2008_08_14-19_59_01-Telejornal-1 bloco 7**: foi doutorado pelos james blake . norte - americano sétimo no ranking atp .
A 2008	2: doutorado	**2008_08_14-19_59_01-Telejornal-1 bloco 7**: foi doutorado pelos james blake . norte - americano sétimo no ranking atp .
A 2008	3: mito	**2008_08_14-19_59_01-Telejornal-1 bloco 7**: james blake . norte - americano sétimo no ranking atp . ganhou o primeiro set por seis a quatro . depois de quebrar o mito de fedra . no segundo set o suíço chegou a estar a perder por três a zero . depois fez uma boa recuperação igualou a partida . o jogo foi decidido no
B 2008	1: norte - americano sétimo	**2008_08_14-19_59_01-Telejornal-1 bloco 7**: foi doutorado pelos james blake . norte - americano sétimo no ranking atp .
B 2008	2: ganhou o primeiro	**2008_08_14-19_59_01-Telejornal-1 bloco 7**: ranking atp . ganhou o primeiro set por seis a quatro . depois de quebrar o mito de fedra . no segundo set o suíço chegou a estar a perder por três a zero . depois fez uma boa recuperação igualou a partida . o jogo foi decidido no tie - break , e james blake ,
B 2008	3: estar a perder	**2008_08_14-19_59_01-Telejornal-1 bloco 7**: ranking atp . ganhou o primeiro set por seis a quatro . depois de quebrar o mito de fedra . no segundo set o suíço chegou a estar a perder por três a zero . depois fez uma boa recuperação igualou a partida . o jogo foi decidido no tie - break , e james blake ,
C 2008	1: americano sétimo	**2008_08_14-19_59_01-Telejornal-1 bloco 7**: foi doutorado pelos james blake , norte - americano sétimo no ranking atp , ganhou o primeiro set por seis a quatro , depois de quebrar o mito de fedra , no segundo set o suíço chegou a estar a perder por três a zero ,
C 2008	2: mito	**2008_08_14-19_59_01-Telejornal-1 bloco 7**: foi doutorado pelos james blake , norte - americano sétimo no ranking atp , ganhou o primeiro set por seis a quatro , depois de quebrar o mito de fedra , no segundo set o suíço chegou a estar a perder por três a zero ,
C 2008	3: rafael nadal	**2008_08_14-19_59_01-Telejornal-1 bloco 7**: roger fehér de três a quatro dias de perder o estatuto de número um mundial para rafael nadal , foi doutorado pelos james blake , norte - americano sétimo no ranking atp , ganhou o primeiro set por seis a quatro , depois de quebrar o mito de fedra ,
D 2008	1: mito de fedra	**2008_08_14-19_59_01-Telejornal-1 bloco 7**: foi doutorado pelos james blake , norte - americano sétimo no ranking atp , ganhou o primeiro set por seis a quatro , depois de quebrar o mito de fedra , no segundo set o suíço chegou a estar a perder por três a zero ,

D 2008	2: ganhou o primeiro	**2008_08_14-19_59_01-Telejornal-1 bloco 7**: foi doutorado pelos james blake , norte - americano sétimo no ranking atp , ganhou o primeiro set por seis a quatro , depois de quebrar o mito de fedra , no segundo set o suíço chegou a estar a perder por três a zero ,
D 2008	3: americano sétimo	**2008_08_14-19_59_01-Telejornal-1 bloco 7**: foi doutorado pelos james blake , norte - americano sétimo no ranking atp , ganhou o primeiro set por seis a quatro , depois de quebrar o mito de fedra , no segundo set o suíço chegou a estar a perder por três a zero ,
A 2010 B 2010 C 2010 D 2010	1: NIL	
Question #92 - Que norte-americano eliminou Roger Federer do torneio olímpico de ténis?		
A 2008	1: NIL	
B 2008	1: wimbledon	**pt/t/o/r/Torneio_de_Wimbledon_3d0a.html**: torneio de wimbledon . o tourneio de tênis de wimbledon é o terceiro dos quatro torneios do grand slam . ele é disputado em quadras de grama . finais simples masculino maiores campeões willie renshaw e pete sampras : 7 títulos björn borg e laurie doherty : 5 reggie doherty , rod laver , tony wilding e roger federer
B 2008	2: andy roddick	**pt/a/s/s/Associação_de_Ténis_Profissional_ea59.html**: associação de ténis profissional . tênis) lista de campeões da atp ano campeão pontos 2 000 gustavo kuerten 839 2 001 lleyton hewitt 897 2 002 lleyton hewitt 873 2 003 andy roddick 907 2 004 roger federer
B 2008	3: profissional	**pt/a/s/s/Associação_de_Ténis_Profissional_ea59.html**: associação de ténis profissional . tênis) lista de campeões da atp ano campeão pontos 2 000 gustavo kuerten 839 2 001 lleyton hewitt 897 2 002 lleyton hewitt 873 2 003 andy roddick 907 2 004 roger federer
C 2008	1: NIL	
D 2008	1: wimbledon	**pt/t/o/r/Torneio_de_Wimbledon_3d0a.html**: torneio de wimbledon . o tourneio de tênis de wimbledon é o terceiro dos quatro torneios do grand slam . ele é disputado em quadras de grama . finais simples masculino maiores campeões willie renshaw e pete sampras : 7 títulos björn borg e laurie doherty : 5 reggie doherty , rod laver , tony wilding e roger federer

D 2008	2: andy roddick	**pt/a/s/s/Associação_de_Ténis_Profissional_ea59.html**: associação de ténis profissional . tênis) lista de campeões da atp ano campeão pontos 2 000 gustavo kuerten 839 2 001 lleyton hewitt 897 2 002 lleyton hewitt 873 2 003 andy roddick 907 2 004 roger federer
D 2008	3: profissional	**pt/a/s/s/Associação_de_Ténis_Profissional_ea59.html**: associação de ténis profissional . tênis) lista de campeões da atp ano campeão pontos 2 000 gustavo kuerten 839 2 001 lleyton hewitt 897 2 002 lleyton hewitt 873 2 003 andy roddick 907 2 004 roger federer
A 2010	1: NIL	
B 2010	1: wimbledon	**pt/t/o/r/Torneio_de_Wimbledon_3d0a.html**: torneio de wimbledon . o tourneio de tênis de wimbledon é o terceiro dos quatro torneios do grand slam . ele é disputado em quadras de grama . finais simples masculino maiores campeões willie renshaw e pete sampras : 7 títulos björn borg e laurie doherty : 5 reggie doherty , rod laver , tony wilding e roger federer
B 2010	2: andy roddick	**pt/a/s/s/Associação_de_Ténis_Profissional_ea59.html**: associação de ténis profissional . tênis) lista de campeões da atp ano campeão pontos 2 000 gustavo kuerten 839 2 001 lleyton hewitt 897 2 002 lleyton hewitt 873 2 003 andy roddick 907 2 004 roger federer
B 2010	3: profissional	**pt/a/s/s/Associação_de_Ténis_Profissional_ea59.html**: associação de ténis profissional . tênis) lista de campeões da atp ano campeão pontos 2 000 gustavo kuerten 839 2 001 lleyton hewitt 897 2 002 lleyton hewitt 873 2 003 andy roddick 907 2 004 roger federer
C 2010	1: NIL	
D 2010	1: wimbledon	**pt/t/o/r/Torneio_de_Wimbledon_3d0a.html**: torneio de wimbledon . o tourneio de tênis de wimbledon é o terceiro dos quatro torneios do grand slam . ele é disputado em quadras de grama . finais simples masculino maiores campeões willie renshaw e pete sampras : 7 títulos björn borg e laurie doherty : 5 reggie doherty , rod laver , tony wilding e roger federer
D 2010	2: andy roddick	**pt/a/s/s/Associação_de_Ténis_Profissional_ea59.html**: associação de ténis profissional . tênis) lista de campeões da atp ano campeão pontos 2 000 gustavo kuerten 839 2 001 lleyton hewitt 897 2 002 lleyton hewitt 873 2 003 andy roddick 907 2 004 roger federer

D 2010	3: profissional	**pt/a/s/s/Associação_de_Ténis_Profissional_ea59.html**: associação de ténis profissional . tênis) lista de campeões da atp ano campeão pontos 2 000 gustavo kuerten 839 2 001 lleyton hewitt 897 2 002 lleyton hewitt 873 2 003 andy roddick 907 2 004 roger federer
Question #93 - Quem é o orador principal da Festa do Pontal?		
A 2008	1: ângelo	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: manuela ferreira leite não vai participar na festa do pontal que decorre esta noite no algarve . o orador principal será ângelo correia , festa do pontal que se realiza em quarteira ,
A 2008	2: quarteira	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: manuela ferreira leite não vai participar na festa do pontal que decorre esta noite no algarve . o orador principal será ângelo correia , festa do pontal que se realiza em quarteira ,
A 2008	3: pombal	**2008_08_14-21_59_01-Jornal2-2 bloco 1**: manuela ferreira leite não vai participar na festa do pombal que decorre do pontal aliás que decorre esta noite no algarve . o orador principal é ângelo correia . a festa do pontal que se realiza em quarteira ,
B 2008	1: ângelo correia	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: manuela ferreira leite não vai participar na festa do pontal que decorre esta noite no algarve . o orador principal será ângelo correia , festa do pontal que se realiza em quarteira ,
B 2008	2: manuela ferreira leite	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: manuela ferreira leite não vai participar na festa do pontal que decorre esta noite no algarve . o orador principal será ângelo correia , festa do pontal que se realiza em quarteira ,
B 2008	3: quarteira	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: manuela ferreira leite não vai participar na festa do pontal que decorre esta noite no algarve . o orador principal será ângelo correia , festa do pontal que se realiza em quarteira ,
C 2008	1: quarteira	**2008_08_14-21_59_01-Jornal2-2 bloco 1**: manuela ferreira leite não vai participar na festa do pombal que decorre do pontal aliás que decorre esta noite no algarve , o orador principal é ângelo correia , a festa do pontal que se realiza em quarteira ,
C 2008	2: ângelo	**2008_08_14-21_59_01-Jornal2-2 bloco 1**: manuela ferreira leite não vai participar na festa do pombal que decorre do pontal aliás que decorre esta noite no algarve , o orador principal é ângelo correia , a festa do pontal que se realiza em quarteira ,
C 2008	3: pombal	**2008_08_14-21_59_01-Jornal2-2 bloco 1**: manuela ferreira leite não vai participar na festa do pombal que decorre do pontal aliás que decorre esta noite no algarve , o orador principal é ângelo correia , a festa do pontal que se realiza em quarteira ,

D 2008	1: ângelo correia	**2008_08_14-21_59_01-Jornal2-2 bloco 1**: manuela ferreira leite não vai participar na festa do pombal que decorre do pontal aliás que decorre esta noite no algarve , o orador principal é ângelo correia , a festa do pontal que se realiza em quarteira ,
D 2008	2: cavaco silva	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: na liderança de cavaco silva , na antiga festa do pontal , ficou nove , agora o palco é montado em quarteira , o comício deixou de assinalar o novo ano político , e esta noite o orador principal , também não é líder do partido ,
D 2008	3: manuela ferreira leite	**2008_08_14-21_59_01-Jornal2-2 bloco 1**: manuela ferreira leite não vai participar na festa do pombal que decorre do pontal aliás que decorre esta noite no algarve , o orador principal é ângelo correia , a festa do pontal que se realiza em quarteira ,
A 2010	1: ângelo	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: manuela ferreira leite não vai participar na festa do pontal que decorre esta noite no algarve . o orador principal será ângelo correia a festa do pontal que se realizem quarteira .
A 2010	2: quarteira	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: manuela ferreira leite não vai participar na festa do pontal que decorre esta noite no algarve . o orador principal será ângelo correia a festa do pontal que se realizem quarteira .
A 2010	3: pombal	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: manuela ferreira leite não vai participar na festa do pombal que decorre do pontal , aliás que decorre esta noite no algarve . o orador principal é ângelo correia . a festa
B 2010	1: ângelo correia	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: manuela ferreira leite não vai participar na festa do pontal que decorre esta noite no algarve . o orador principal será ângelo correia a festa do pontal que se realizem quarteira .
B 2010	2: manuela ferreira leite	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: manuela ferreira leite não vai participar na festa do pontal que decorre esta noite no algarve . o orador principal será ângelo correia a festa do pontal que se realizem quarteira .
B 2010	3: cavaco silva	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: o orador principal será ângelo correia a festa do pontal que se realizem quarteira . deixou de ter significado político que lhe foi atribuído nas décadas de oitenta e noventa . na liderança de cavaco silva . antiga festa do pontal .
C 2010	1: orador principal é ângelo	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: manuela ferreira leite não vai participar na festa do pombal que decorre do pontal , aliás que decorre esta noite no algarve , o orador principal é ângelo correia , a festa

C 2010	2: pombal	**2008_08_14-21_59_01-Jornal2-2 bloco 2**: manuela ferreira leite não vai participar na festa do pombal que decorre do pontal , aliás que decorre esta noite no algarve , o orador principal é ângelo correia , a festa
C 2010	3: fiel	**2008_08_07-21_59_02-Jornal2-2 bloco 2**: a festa do pontal tanta nenhuma outra manifestação onde tivesse que fazer intervenções em público , desde que foi eleita , ferreira leite mantém - se fiel a uma atitude mais reservada , o que já tinha mostrado durante a campanha para presidente do psd ,
D 2010	1: ângelo correia	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: manuela ferreira leite não vai participar na festa do pontal que decorre esta noite no algarve , o orador principal será ângelo correia a festa do pontal que se realizem quarteira ,
D 2010	2: cavaco silva	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: na liderança de cavaco silva , antiga festa do pontal , só ficou nove , agora o palco é montado em quarteira o comício deixou de assinalar o novo ano político e esta noite o orador principal também não é líder do partido ,
D 2010	3: manuela ferreira leite	**2008_08_14-19_59_01-Telejornal-1 bloco 8**: manuela ferreira leite não vai participar na festa do pontal que decorre esta noite no algarve , o orador principal será ângelo correia a festa do pontal que se realizem quarteira ,
Question #94 - Que regiões visitou Barack Obama em 2008?		
A 2008	1: NIL	
B 2008	1: afeganistão	**2008_07_20-19_59_01-Telejornal-1 bloco 3**: há uma vintena zweig evasão nossos tem poder utilizar leste é de andar se faz durante a estada de inaugurar um . dizem que destina - se a sáb e dom . no afeganistão candidato presidencial democrata norte - americano barack obama ,
B 2008	2: segurança nacional	**2008_09_04-21_59_02-Jornal2-2 bloco 2**: por outro lado haverá também um duro ataque a barack obama , sobretudo na questão tem a ver com a segurança nacional .
B 2008	3: amadora	**2008_08_06-21_59_02-Jornal2-2 bloco 3**: estrela da amadora aos olhos vêem , esses que acabar mal ravessa é , qualquer interessado susana rodrigues foi o único meio para cá se . o porta - voz da campanha de john mccain já veio dizer que até viseu tem parece apoiar as propostas energéticas dos republicanos . barack obama .
C 2008	1: NIL	
D 2008	1: segurança nacional	**2008_09_04-21_59_02-Jornal2-2 bloco 2**: será muito vincada neste discurso , por outro lado haverá também um duro ataque a barack obama , sobretudo na questão tem a ver com a segurança nacional , de john mccain deverá dizer ,

D 2008	2: afeganistão	**2008_07_20-19_59_01-Telejornal-1 bloco 3**: há uma vintena zweig evasão nossos tem poder utilizar leste é de andar se faz durante a estada de inaugurar um , dizem que destina - se a sáb e dom , no afeganistão candidato presidencial democrata norte - americano barack obama ,
D 2008	3: lado	**2008_09_04-21_59_02-Jornal2-2 bloco 2**: será muito vincada neste discurso , por outro lado haverá também um duro ataque a barack obama , sobretudo na questão tem a ver com a segurança nacional , de john mccain deverá dizer ,
A 2010	1: NIL	
B 2010	1: afeganistão	**2008_07_14-19_59_01-Telejornal-1 bloco 3**: o afeganistão constitui a maior preocupação do candidato à casa branca barack obama .
C 2010	1: NIL	
D 2010	1: afeganistão	**2008_07_14-19_59_01-Telejornal-1 bloco 3**: em contratuais wanna peixe tem faial , o afeganistão constitui a maior preocupação do candidato à casa branca barack obama ,
D 2010	2: médio oriente	**2008_07_14-19_59_01-Telejornal-1 bloco 3**: o afeganistão constitui a maior preocupação do candidato à casa branca barack obama , o candidato democrata quer definir um calendário para retirar o grosso das tropas do iraque e diz que o objectivo central deve ser o combate à al - qaeda obama fez estas declarações no início de uma viagem ao médio oriente ,
D 2010	3: grosso	**2008_07_14-19_59_01-Telejornal-1 bloco 3**: o afeganistão constitui a maior preocupação do candidato à casa branca barack obama , o candidato democrata quer definir um calendário para retirar o grosso das tropas do iraque e diz que o objectivo central deve ser o combate à al - qaeda obama fez estas declarações no início de uma viagem ao médio oriente ,
Question #95 - O que é a Bossa Nova?		
A 2008	1: programa musical apresentado	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bossa nova , e o nosso acordo e a teresinha . finais de mil novecentos e sessenta e um . no programa musical apresentado na rtp por fialho gouveia , aos man blues rumbas ou chás . os conjuntos juntavam agora um novo ritmo . a bossa nova .
A 2008	2: fialho gouveia	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bossa nova , e o nosso acordo e a teresinha . finais de mil novecentos e sessenta e um . no programa musical apresentado na rtp por fialho gouveia , aos man blues rumbas ou chás . os conjuntos juntavam agora um novo ritmo . a bossa nova .

A 2008	3: instantes	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: para aguçar um cem por cento bossa - nova . portanto , e durante alguns instantes convosco . a bossa nova , e o nosso acordo e a teresinha .
B 2008	1: a bossa nova é um movimento da música popular brasileira criado por joão gilberto , revelado em 1 958 e popularizado internacionalmente em 1 963 , pelo seu próprio criador , assim como tambem por antonio carlos jobim , astrud gilberto , roberto menescal e stan getz .	**pt/b/o/s/Bossa_nova.html**: a bossa nova é um movimento da música popular brasileira criado por joão gilberto , revelado em 1 958 e popularizado internacionalmente em 1 963 , pelo seu próprio criador , assim como tambem por antonio carlos jobim , astrud gilberto , roberto menescal e stan getz . no final da década de 1 950 e início da década de 1 960 , surgiu no rio de janeiro um dos mais importantes movimentos da música popular do brasil : a bossa nova , criada pelo violonista e cantor baiano joão gilberto .
C 2008	1: programa musical apresentado	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bossa nova , e o nosso acordo e a teresinha , finais de mil novecentos e sessenta e um , no programa musical apresentado na rtp por fialho gouveia , aos man blues rumbas ou chás , os conjuntos juntavam agora um novo ritmo , a bossa nova ,
C 2008	2: fialho gouveia	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a bossa nova , e o nosso acordo e a teresinha , finais de mil novecentos e sessenta e um , no programa musical apresentado na rtp por fialho gouveia , aos man blues rumbas ou chás , os conjuntos juntavam agora um novo ritmo , a bossa nova ,
C 2008	3: com sotaque português gronholm no ar uma criminosa	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: no programa musical apresentado na rtp por fialho gouveia , aos man blues rumbas ou chás , os conjuntos juntavam agora um novo ritmo , a bossa nova , com sotaque português gronholm no ar uma criminosa um a um , a zero em ganhar a mulher de dar o meu melhor nível ,
D 2008	1: a bossa nova é um movimento da música popular brasileira criado por joão gilberto , revelado em 1 958 e popularizado internacionalmente em 1 963 , pelo seu próprio criador , assim como tambem por antonio carlos jobim , astrud gilberto , roberto menescal e stan getz .	**pt/b/o/s/Bossa_nova.html**: a bossa nova é um movimento da música popular brasileira criado por joão gilberto , revelado em 1 958 e popularizado internacionalmente em 1 963 , pelo seu próprio criador , assim como tambem por antonio carlos jobim , astrud gilberto , roberto menescal e stan getz . no final da década de 1 950 e início da década de 1 960 , surgiu no rio de janeiro um dos mais importantes movimentos da música popular do brasil : a bossa nova , criada pelo violonista e cantor baiano joão gilberto .

A 2010	1: música brasileira saltar fronteiras	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: as da de mim e portugal não foi imune à bossa nova movimento que em finais da década de cinquenta fisa música brasileira saltar fronteiras e cinquenta e oito era gravado chega de saudade .
A 2010	2: chega	**2008_07_17-21_59_02-Jornal2-2 bloco 4**: é das mais tocadas é símbolo do brasil e da bossa nova que chega de saudade cível de lisboa acompanhado por outras três músicas que começa por ser originais de judo in ,
A 2010	3: símbolo do brasil	**2008_07_17-21_59_02-Jornal2-2 bloco 4**: é das mais tocadas é símbolo do brasil e da bossa nova que chega de saudade cível de lisboa acompanhado por outras três músicas que começa por ser originais de judo in ,
B 2010	1: a bossa nova é um movimento da música popular brasileira criado por joão gilberto , revelado em 1 958 e popularizado internacionalmente em 1 963 , pelo seu próprio criador , assim como tambem por antonio carlos jobim , astrud gilberto , roberto menescal e stan getz .	**pt/b/o/s/Bossa_nova.html**: a bossa nova é um movimento da música popular brasileira criado por joão gilberto , revelado em 1 958 e popularizado internacionalmente em 1 963 , pelo seu próprio criador , assim como tambem por antonio carlos jobim , astrud gilberto , roberto menescal e stan getz . no final da década de 1 950 e início da década de 1 960 , surgiu no rio de janeiro um dos mais importantes movimentos da música popular do brasil : a bossa nova , criada pelo violonista e cantor baiano joão gilberto .
C 2010	1: programa musical apresentado	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: nova doce combo e teresinha , finais de mil novecentos e sessenta e um , no programa musical apresentado na rtp por fialho gouveia , aos mambo os rumbas ou chás às , os conjuntos juntavam agora um novo ritmo a bossa nova ,
C 2010	2: consultar o português joga ou não não há uma criminal à	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: nova , consultar o português joga ou não não há uma criminal à , a um , a deve sonhar am é que dá arminda rosa é , as da de mim e portugal não foi imune à bossa
C 2010	3: fialho gouveia	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: nova doce combo e teresinha , finais de mil novecentos e sessenta e um , no programa musical apresentado na rtp por fialho gouveia , aos mambo os rumbas ou chás às , os conjuntos juntavam agora um novo ritmo a bossa nova ,

D 2010	1: a bossa nova é um movimento da música popular brasileira criado por joão gilberto , revelado em 1 958 e popularizado internacionalmente em 1 963 , pelo seu próprio criador , assim como tambem por antonio carlos jobim , astrud gilberto , roberto menescal e stan getz .	**pt/b/o/s/Bossa_nova.html**: a bossa nova é um movimento da música popular brasileira criado por joão gilberto , revelado em 1 958 e popularizado internacionalmente em 1 963 , pelo seu próprio criador , assim como tambem por antonio carlos jobim , astrud gilberto , roberto menescal e stan getz . no final da década de 1 950 e início da década de 1 960 , surgiu no rio de janeiro um dos mais importantes movimentos da música popular do brasil : a bossa nova , criada pelo violonista e cantor baiano joão gilberto .
Question #96 - O que é o Prémio Camões?		
A 2008	1: antónio lobo antunes	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: durante a cerimónia de entrega do prémio camões antónio lobo antunes .
A 2008	2: língua	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: é o prémio camões dois mil e oito o mais importante prémio atribuído a um escritor de língua portuguesa .
A 2008	3: honras	**2008_07_25-21_59_01-Jornal2-2 bloco 3**: o prémio camões teve honras de estado presentes estiveram .
B 2008	1: o prémio camões é o mais importante galardão literário de língua portuguesa , atribuído anualmente pela fundação biblioteca nacional (de portugal) e pelo departamento nacional do livro (do brasil) a um escritor que tenha desenvolvido um conjunto de obra relevante em língua portuguesa .	**pt/p/r/é/Prémio_Camões_009d.html**: o prémio camões é o mais importante galardão literário de língua portuguesa , atribuído anualmente pela fundação biblioteca nacional (de portugal) e pelo departamento nacional do livro (do brasil) a um escritor que tenha desenvolvido um conjunto de obra relevante em língua portuguesa .

C 2008	1: António lobo antunes	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: durante a cerimónia de entrega do prémio camões António lobo antunes , o chefe de estado referiu a originalidade da escrita do autor português , e também a internacionalização da sua obra , o prémio
C 2008	2: porto de abrigo	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: talvez sejam porto de abrigo , temporário , antes do repatriamento para a terra , de onde querem fugir , joão ubaldo ribeiro foi distinguido com o prémio camões dois mil e oito o mais importante galardão atribuído a autores de língua portuguesa ,
C 2008	3: aníbal cavaco	**2008_07_25-21_59_01-Jornal2-2 bloco 3**: António lobo antunes recebeu esta tarde nos jerónimos , mais importante prémio para autores de língua portuguesa , o prémio camões teve honras de estado presentes estiveram , o presidente do brasil lula da silva , o presidente português aníbal cavaco silva ,
D 2008	1: o prémio camões é o mais importante galardão literário de língua portuguesa , atribuído anualmente pela fundação biblioteca nacional (de portugal) e pelo departamento nacional do livro (do brasil) a um escritor que tenha desenvolvido um conjunto de obra relevante em língua portuguesa .	**pt/p/r/é/Prémio_Camões_009d.html**: o prémio camões é o mais importante galardão literário de língua portuguesa , atribuído anualmente pela fundação biblioteca nacional (de portugal) e pelo departamento nacional do livro (do brasil) a um escritor que tenha desenvolvido um conjunto de obra relevante em língua portuguesa .

A 2010	1: lobo antunes	**2008_07_25-21_59_01-Jornal2-2 bloco 2**: durante a cerimónia de entrega do prémio camões antónio lobo antunes .
A 2010	2: língua	**2008_07_26-19_59_01-Telejornal-1 bloco 4**: é o prémio camões dois mil e oito o mais importante prémio atribuído a um escritor de língua portuguesa .
A 2010	3: honras	**2008_07_25-21_59_01-Jornal2-2 bloco 3**: o prémio camões teve honras de estado presentes estiveram .
B 2010	1: o prémio camões é o mais importante galardão literário de língua portuguesa , atribuído anualmente pela fundação biblioteca nacional (de portugal) e pelo departamento nacional do livro (do brasil) a um escritor que tenha desenvolvido um conjunto de obra relevante em língua portuguesa .	**pt/p/r/é/Prémio_Camões_009d.html**: o prémio camões é o mais importante galardão literário de língua portuguesa , atribuído anualmente pela fundação biblioteca nacional (de portugal) e pelo departamento nacional do livro (do brasil) a um escritor que tenha desenvolvido um conjunto de obra relevante em língua portuguesa .
C 2010	1: lobo antunes	**2008_07_25-19_59_01-Telejornal-1 bloco 2**: o final da cimeira dos países lusófonos culminou com o prémio camões e o escritor antónio lobo antunes uniu na mesma mesa os dois maiores países de língua portuguesa ,
C 2010	2: mesa os dois maiores países de língua	**2008_07_25-19_59_01-Telejornal-1 bloco 2**: o final da cimeira dos países lusófonos culminou com o prémio camões e o escritor antónio lobo antunes uniu na mesma mesa os dois maiores países de língua portuguesa ,
C 2010	3: aníbal cavaco	**2008_07_25-19_59_01-Telejornal-1 bloco 2**: o prémio camões que teve nesta cerimónia de entrega honras de estado , presentes estiveram o presidente do brasil lula da silva o presidente português aníbal cavaco silva e o primeiro - ministro josé sócrates , o final da cimeira dos países lusófonos culminou com o prémio
D 2010	1: o prémio camões é o mais importante galardão literário de língua portuguesa , atribuído anualmente pela fundação biblioteca nacional (de portugal) e pelo departamento nacional do livro (do brasil) a um escritor que tenha desenvolvido um conjunto de obra relevante em língua portuguesa .	**pt/p/r/é/Prémio_Camões_009d.html**: o prémio camões é o mais importante galardão literário de língua portuguesa , atribuído anualmente pela fundação biblioteca nacional (de portugal) e pelo departamento nacional do livro (do brasil) a um escritor que tenha desenvolvido um conjunto de obra relevante em língua portuguesa .

Question #97 - Qual a empresa maior accionista da OGMA?		
A 2008	1: brasileira em bayreuth terceira	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital .
A 2008	2: aviões	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital .
B 2008	1: brasileira em bayreuth terceira maior	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital .
B 2008	2: fabrico de aviões	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital .
B 2008	3: embraer	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial do fabrico de aviões . e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital .
C 2008	1: brasileira em bayreuth terceira	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: para esta área , a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital , o estado português , detém os restantes trinta e cinco por cento ,
C 2008	2: aviões	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: para esta área , a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital , o estado português , detém os restantes trinta e cinco por cento ,
D 2008	1: brasileira em bayreuth terceira maior	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: para esta área , a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital , o estado português , detém os restantes trinta e cinco por cento ,
D 2008	2: fabrico de aviões	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: para esta área , a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital , o estado português , detém os restantes trinta e cinco por cento ,
D 2008	3: esta	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: para esta área , a brasileira em bayreuth terceira maior empresa mundial do fabrico de aviões , e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital , o estado português , detém os restantes trinta e cinco por cento ,

A 2010	1: capital	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento .
A 2010	2: aviões	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento .
A 2010	3: terceira	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento .
B 2010	1: brasileira embraer é a terceira maior	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento .
B 2010	2: fabrico de aviões	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento .
B 2010	3: português detém	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento .
C 2010	1: capital	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento ,
C 2010	2: aviões	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento ,
C 2010	3: terceira	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento ,

D 2010	1: brasileira embraer é a terceira maior	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento ,
D 2010	2: fabrico de aviões	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento ,
D 2010	3: português detém	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: a brasileira embraer é a terceira maior empresa mundial no fabrico de aviões e o maior accionista da portuguesa ogma com sessenta e cinco por cento do capital o estado português detém os restantes trinta e cinco por cento ,
Question #98 - Em que cidade é que a EMBRAER vai investir em duas fábricas?		
A 2008	1: NIL	
B 2008	1: rio de janeiro , de minas gerais e a do rio grande	**pt/s/ã/o/São_Paulo_23d2.html**: são paulo . deu lugar às indústrias , as quais fizeram são paulo permanecer na liderança da indústria nacional até hoje . o estado supera a produção industrial do rio de janeiro , de minas gerais e a do rio grande do sul . no vale do paraíba localizam - se indústrias do ramo aeroespacial , como a embraer ,
B 2008	2: évora	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: estarão a funcionar em pleno , de cidade de évora vão ser para o brasil componentes para os aviões da embraer . os melhores segundo lula da silva . o de agosto em breve são de extrema qualidade e , que até , o governo para a compra de dois aviões . numa primeira fase vão ser
B 2008	3: paulo	**pt/s/ã/o/São_Paulo_23d2.html**: são paulo . deu lugar às indústrias , as quais fizeram são paulo permanecer na liderança da indústria nacional até hoje . o estado supera a produção industrial do rio de janeiro , de minas gerais e a do rio grande do sul . no vale do paraíba localizam - se indústrias do ramo aeroespacial , como a embraer ,
C 2008	1: NIL	
D 2008	1: évora	**2008_07_26-21_59_02-Jornal2-2 bloco 2**: estarão a funcionar em pleno , de cidade de évora vão ser para o brasil componentes para os aviões da embraer , os melhores segundo lula da silva , o de agosto em breve são de extrema qualidade e , que até , o governo para a compra de dois aviões , numa primeira fase vão ser

D 2008	2: rio de janeiro , de minas gerais e a do rio grande	**pt/s/ã/o/São_Paulo_23d2.html**: são paulo . deu lugar às indústrias , as quais fizeram são paulo permanecer na liderança da indústria nacional até hoje . o estado supera a produção industrial do rio de janeiro , de minas gerais e a do rio grande do sul . no vale do paraíba localizam - se indústrias do ramo aeroespacial , como a embraer ,
D 2008	3: paulo	**pt/s/ã/o/São_Paulo_23d2.html**: são paulo . deu lugar às indústrias , as quais fizeram são paulo permanecer na liderança da indústria nacional até hoje . o estado supera a produção industrial do rio de janeiro , de minas gerais e a do rio grande do sul . no vale do paraíba localizam - se indústrias do ramo aeroespacial , como a embraer ,
A 2010	1: NIL	
B 2010	1: évora vou sair para o brasil	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: fábricas estarão a funcionar em pleno da cidade de évora vou sair para o brasil componentes para os aviões da embraer . os melhores segundo lula da silva . o da banda em breve se onde tem maior qualidade . quetta um governo para compra de dois aviões . numa primeira fase vão ser investidos
B 2010	2: rio de janeiro , de minas gerais e a do rio grande	**pt/s/ã/o/São_Paulo_23d2.html**: são paulo . deu lugar às indústrias , as quais fizeram são paulo permanecer na liderança da indústria nacional até hoje . o estado supera a produção industrial do rio de janeiro , de minas gerais e a do rio grande do sul . no vale do paraíba localizam - se indústrias do ramo aeroespacial , como a embraer ,
B 2010	3: paulo	**pt/s/ã/o/São_Paulo_23d2.html**: são paulo . deu lugar às indústrias , as quais fizeram são paulo permanecer na liderança da indústria nacional até hoje . o estado supera a produção industrial do rio de janeiro , de minas gerais e a do rio grande do sul . no vale do paraíba localizam - se indústrias do ramo aeroespacial , como a embraer ,
C 2010	1: NIL	
D 2010	1: évora vou sair para o brasil	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: fábricas estarão a funcionar em pleno da cidade de évora vou sair para o brasil componentes para os aviões da embraer , os melhores segundo lula da silva , o da banda em breve se onde tem maior qualidade , quetta um governo para compra de dois aviões , numa primeira fase vão ser investidos

D 2010	2: rio de janeiro , de minas gerais e a do rio grande	**pt/s/ã/o/São_Paulo_23d2.html**: são paulo . deu lugar às indústrias , as quais fizeram são paulo permanecer na liderança da indústria nacional até hoje . o estado supera a produção industrial do rio de janeiro , de minas gerais e a do rio grande do sul . no vale do paraíba localizam - se indústrias do ramo aeroespacial , como a embraer ,
D 2010	3: paulo	**pt/s/ã/o/São_Paulo_23d2.html**: são paulo . deu lugar às indústrias , as quais fizeram são paulo permanecer na liderança da indústria nacional até hoje . o estado supera a produção industrial do rio de janeiro , de minas gerais e a do rio grande do sul . no vale do paraíba localizam - se indústrias do ramo aeroespacial , como a embraer ,
Question #99 - Quantos postos de trabalho directos serão criados com o investimento?		
A 2008	1: cerca de quinhentos postos	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: empresa brasileira embraer é vai construir duas fábricas para já , num investimento de cento e cinquenta milhões de euros . irão ser criados cerca de quinhentos postos de trabalho directos .
B 2008	1: NIL	
C 2008	1: cerca de quinhentos postos	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: empresa brasileira embraer é vai construir duas fábricas para já , num investimento de cento e cinquenta milhões de euros , irão ser criados cerca de quinhentos postos de trabalho directos , as fábricas vão construir componentes para aviões ,
D 2008	1: NIL	
A 2010	1: cerca de quinhentos postos	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: embraer vai construir duas fábricas para já . num investimento de cento e cinquenta milhões de euros irão ser criados cerca de quinhentos postos de trabalho directos .
B 2010	1: NIL	
C 2010	1: cerca de quinhentos postos	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: embraer vai construir duas fábricas para já , num investimento de cento e cinquenta milhões de euros irão ser criados cerca de quinhentos postos de trabalho directos , as fábricas vão construir componentes para aviões ,
D 2010	1: cerca de quinhentos postos	**2008_07_26-19_59_01-Telejornal-1 bloco 1**: mais à frente no telejornal , uma memória e agora o olhar para o futuro da região de évora vai receber um importante investimento aeronáutico empresa brasileira embraer vai construir duas fábricas para já , num investimento de cento e cinquenta milhões de euros irão ser criados cerca de quinhentos postos de trabalho directos ,

Question #100 - Quantos postos de trabalho foram criados com a inauguração de um hotel em Baião?		
A 2008	1: NIL	
B 2008	1: cinco postos	**2008_07_17-19_59_01-Telejornal-1 bloco 3**: estamos em baião no vale do tâmega , um dos concelhos mais pobres do país . há dois dias a inauguração de um hotel deu um novo ânimo à região . criaram - se trinta e cinco postos de trabalho .
C 2008	1: NIL	
D 2008	1: cinco postos	**2008_07_17-19_59_01-Telejornal-1 bloco 3**: estamos em baião no vale do tâmega , um dos concelhos mais pobres do país , há dois dias a inauguração de um hotel deu um novo ânimo à região , criaram - se trinta e cinco postos de trabalho ,
B 2010	1: cinco postos	**2008_07_17-19_59_01-Telejornal-1 bloco 2**: estamos em baião no vale do tâmega . um dos concelhos mais pobres do país . há dois dias a inauguração de um hotel deu um novo ânimo à região . criaram - se trinta e cinco postos de trabalho ,
C 2010	1: NIL	
D 2010	1: cinco postos	**2008_07_17-19_59_01-Telejornal-1 bloco 2**: estamos em baião no vale do tâmega , um dos concelhos mais pobres do país , há dois dias a inauguração de um hotel deu um novo ânimo à região , criaram - se trinta e cinco postos de trabalho ,