

UNIVERSIDADE ABERTA



UNIVERSIDADE
AbERTA
www.uab.pt

Visualização de opiniões *online* de hotéis

Tiago da Silva André

Mestrado em Tecnologias e Sistemas Informáticos Web

2018

Universidade Aberta



UNIVERSIDADE
AbERTA
www.uab.pt

Departamento de Ciências e Tecnologia

Mestrado em Tecnologias e Sistemas Informáticos Web

Tema:

Visualização de opiniões *online* de hotéis

Aluno: Tiago da Silva André

Número: 1402219

Orientadora: Doutora Elizabeth Simão Carvalho

2018

RESUMO

Um dos claros benefícios da visualização da informação é poder ajudar as pessoas a ver tendências e anomalias em dados, que podem ser particularmente valiosas em tempo real. Técnicas visuais, como os mapas de calor e grafos, ajudam a revelar padrões contidos em dados homogêneos. Estas técnicas eram “virtualmente” desconhecidas há 10 anos, mas nos dias de hoje têm uma grande aplicação prática, pois auxiliam significativamente na visualização de dados disponíveis em sítios públicos na *Web* a aplicações avançadas para suporte a negociação.

Os sítios web que disponibilizam comentários sobre um produto ou um serviço têm atraído a atenção de milhões de pessoas por todo o mundo. O TripAdvisor é disto um exemplo, recebe mais de 60 milhões de visitantes únicos mensais e mais de 75 milhões de comentários e opiniões (TripAdvisor, 2012). Além disso, os comentários on-line estão disponíveis numa variedade de domínios, como por exemplo na restauração (por exemplo zagat.com e opentable.com), na comercialização de vinhos (adegga.com) e cruzeiros (cruisecritic.com). Apesar de oferecerem alguma visualização, sobre a informação que disponibilizam, este aspeto necessita de ser melhor explorado.

Algumas das questões em aberto são: Como representar visualmente opiniões *online* para apoiar uma melhor tomada de decisão? Quais são as técnicas de visualização de informação mais adequadas para representar o conhecimento nas avaliações *online*? Que tipos de gráficos devem ser utilizados para representar o conhecimento (informação integrada) em comentários *online*? Se considerarmos que os potenciais utilizadores finais podem ser não só o público em geral, mas também gestores de negócios, haverá alguma diferença cognitiva visual que justifique a utilização de técnicas visuais diferentes? Em última análise, a própria variedade do perfil de amostra de utilizadores (idade, género, nacionalidade, etc.) poderá implicar nessa diferença?

O cenário apresentado, onde foi identificada a necessidade de melhorar a visualização dos dados resultantes das opiniões *online* disponíveis nos *websites*, é a motivação para a proposta deste trabalho. De uma forma mais abrangente, pretende-se conhecer melhor o modelo cognitivo visual dos potenciais utilizadores desses *websites* e propor uma ou mais novas formas de representação visual que permitam o mapeamento visual desses dados

e uma mais fácil análise. De uma forma mais específica, este trabalho pretende focar o domínio das opiniões *online* do sector hoteleiro, tendo especial atenção no enfoque diferenciado a ser considerado para uma visão de gestor e de público em geral sobre a informação.

Palavras-chave: VISUALIZAÇÃO DE INFORMAÇÃO, VISUALIZAÇÃO DE COMENTÁRIOS *ONLINE*, VISUALIZAÇÃO DE COMENTÁRIOS DE HOTÉIS

ABSTRACT

One of the most obvious benefits of visualization is helping people see trends and anomalies in data, which can be particularly valuable in real-time environments. Visual techniques such as heat maps and tree maps, which help reveal patterns in homogenous data, were virtually unknown 10 years ago, but are used today in many places ranging from public Web sites to advanced trading applications.

Online reviews, e.g. review about a product or a service, is pervasive on the web and the richness of this data have attracted attention from millions of people around the world. For example, TripAdvisor receives more than 60 million unique monthly visitors, and over 75 million reviews and opinions (TripAdvisor, 2012) - and the pace is accelerating. Moreover, online reviews are available in a range of domains such as restaurants (e.g. zagat.com and opentable.com), wines (e.g. adegga.com) and cruisers (e.g. cruisecritic.com) which needs to be better explored, mainly in the visualization aspect.

Some of the open questions are: How to visually represent online reviews to support better decision making? Which are the more adequate/suitable information visualization techniques to represent the knowledge in online reviews? What kinds of graphics should be used to represent the knowledge (integrated information) in online reviews? Are both textual and graphical visualizations useful for customers and managers?

The goal of this work is to have a better knowledge of the multilingual online business reviewers' visual and cognitive model. The domain of the reviews will be the hotel sector.

Keywords: DATA VISUALIZATION, DATA VIZ, ONLINE FEEDBACK VISUALIZATION, HOTELS FEEDBACK VISUALIZATION

ÍNDICE

1. INTRODUÇÃO	9
1.1. MOTIVAÇÕES, OBJETIVOS E CONTRIBUIÇÕES FUNDAMENTAIS	10
1.2. METODOLOGIA DE INVESTIGAÇÃO	11
1.3. ESTRUTURA DO DOCUMENTO	13
2. ESTADO DA ARTE	14
2.1. A ÁREA DA VISUALIZAÇÃO DE INFORMAÇÃO	14
2.2. VISUALIZAÇÃO DA INFORMAÇÃO APLICADA A OPINIÕES <i>ONLINE</i>	16
2.3. AMBIENTES DE VISUALIZAÇÃO DE OPINIÕES SOBRE HOTÉIS	20
2.4. CONCLUSÕES	25
3. MODELO CONCEPTUAL	27
3.1. LEVANTAMENTO DE REQUISITOS	27
3.2. FUNCIONALIDADES E VISUALIZAÇÃO	35
3.2.1. Sessão de avaliação do esboço inicial	40
3.3. Tecnologia	41
3.3. CONCLUSÕES	43
4. PROTÓTIPO	45
4.1. INTERFACE VISUAL E FUNCIONALIDADES	46
4.2. AVALIAÇÃO	48
4.3. CONCLUSÕES	51
CONCLUSÃO E TRABALHO FUTURO	53
5. BIBLIOGRAFIA	56
6. ANEXOS	I
6.1. ANEXO I - GROUNDED THEORY - DETALHES	II
6.2. ANEXO II – QUESTIONÁRIO – 1ª PARTE	IV
6.3. ANEXO III – QUESTIONÁRIO – 2ª PARTE	V
6.4. ANEXO IV – GUIÃO DE ENTREVISTA PARA O GRUPO DE DISCUSSÃO	IX
6.5. ANEXO V – RESULTADOS DA SESSÃO DE GRUPO DE DISCUSSÃO	X
6.6. ANEXO VI – GUIÃO DE ENTREVISTA PARA O GRUPO DE DISCUSSÃO	XI
6.7. ANEXO VII – RESULTADOS DA SESSÃO DE GRUPO DE DISCUSSÃO	XII

ÍNDICE DE TABELAS

Tabela 3.1. Legenda dos gráficos utilizados no questionário	31
Tabela 3.1: Escala de cor para as emoções.....	37
Tabela 3.2: Representação visual dos dados no modelo	38

ÍNDICE DE FIGURAS

Figura 2.1. Representação de todas as emoções dos tweets mundiais publicados a 10-02-2016	18
Figura 2.2. Exemplo de uma aplicação do modelo Review Spotlight	18
Figura 2.3. Exemplo de uma aplicação do modelo de blocos de opinião	19
Figura 2.4. Exemplo de uma aplicação do modelo SentiVis	20
Figura 2.5. Captura de ecrã do website Booking.com	21
Figura 2.6. Captura de ecrã do website Logitravel	22
Figura 2.7. Captura de ecrã do website Hoteis.com	22
Figura 2.8. Captura de ecrã do website Trivago	23
Figura 2.9. Captura de ecrã do website Momondo	24
Figura 2.10 Captura de ecrã do website TripAdvisor	25
Figura 3.1. Distribuição entre género dos inquiridos.....	29
Figura 3.2. Distribuição por hábito na utilização de websites de opinião	29
Figura 3.3. Distribuição por opinião em relação a influência dos websites de opinião	29
Figura 3.4. Distribuição das respostas sobre o hábito ou não na leitura dos comentários	30
Figura 3.5. Distribuição das respostas sobre a influência dos comentários na escolha do hotel	30
Figura 3.6. Distribuição da preferência dos inquiridos sobre o sítio web de opiniões sobre hotéis	30
Figura 3.7. Distribuição sobre os gráficos que os inquiridos reconheceram ser de mais fácil consulta	31
Figura 3.8. Distribuição das respostas sobre os gráficos de consulta mais difícil	31
Figura 3.9. Gráfico com os resultados sobre as visualizações onde é possível observar aspetos mencionados de forma positiva ou negativa	32
Figura 3.10 Distribuição das respostas de qual ou quais avaliações é mais fácil obter uma síntese geral da informação.....	33
Figura 3.11. Distribuição das respostas sobre os gráficos onde a forma é o atributo relevante.....	33
Figura 3.12. Distribuição das respostas sobre os gráficos onde a cor é o atributo visual mais relevante.....	33
Figura 3.13 Distribuição das respostas de qual ou quais as avaliações onde é mais quantificar o número de comentários sobre um determinado aspeto	34
Figura 3.14 Distribuição das respostas de qual ou quais avaliações oferecem uma melhor visão a partir dos comentários sobre os diferentes aspetos do hotel	34
Figura 3.15. Distribuição da importância da separação temporal das avaliações	35
Figura 3.16. Esquema da arquitetura geral do modelo	35
Figura 3.17. Escala de cor das emoções de Plutchik	36
Figura 3.18 Mockup para uma primeira fase de avaliação da proposta de visualização.....	39
Figura 4.1 Captura do ecrã da página principal	46
Figura 4.2 Captura do ecrã do gráfico principal	47
Figura 4.3 Captura do ecrã do quadro de detalhes	47
Figura 4.4 Distribuição das respostas para a pergunta sobre a melhoria na visibilidade das opiniões	49
Figura 4.5 Distribuição das respostas à pergunta sobre o uso de uma escala temporal para distribuir as emoções.....	49
Figura 4.6 Distribuição das respostas sobre a relação da cor com a emoção	50
Figura 4.7 Média das avaliações do protótipo para os detalhes, experiência de utilização e interação	50
Figura 4.8 Captura do ecrã do quadro de detalhes	51
Figura 4.9 Captura do ecrã do quadro de detalhes	51

1. INTRODUÇÃO

A visualização da informação é um processo de representação de dados, informação e conhecimento através de uma forma visual. Esta informação visa apoiar tarefas de análise, exploração, confirmação, apresentação e compreensão dos dados. Com o objetivo de conferir um bom uso da informação, três dimensões estão presentes na tarefa de “ajudar à compreensão dos dados”: explorar, relacionar e comunicar.

A visualização proporciona uma capacidade de compreender grandes quantidades de dados, permitindo que estes sejam mais acessíveis e úteis a mais pessoas, ajudando a fundamentar as suas atividades e, ou, decisões. A informação importante, entre mais de um milhão de medições, pode tornar-se disponível imediatamente [1]. É mais fácil a um ser humano compreender dados expostos num gráfico, do que se fossem apresentados num conjunto de números. A visualização facilita a compreensão de características dos dados quer em grande, quer em pequena escala. Isso pode ser especialmente valioso para permitir a perceção de padrões que ligam as características locais.

A visualização de informação ainda é uma área do conhecimento que carece de uma teoria sólida para apoiá-la, existindo já inúmeras metodologias e taxonomias que podem ser combinados e adotadas como diretrizes [2]. Assim, escolher a apresentação gráfica mais adequada desses dados é um desafio, mas a recompensa final é a informação que é veiculada ao utilizador.

Antes do século XVII já existiam visualizações de dados, estas eram essencialmente mapas dos reinos, exibindo marcadores de terra, cidades, estradas e recursos. Em 1644, *Michael Florent Van Langren*, um astrónomo flamengo, criou a primeira representação visual de dados estatísticos [3]. O século XVIII viu o início do mapeamento temático. Tentativas de mapeamento temático de dados geológicos, económicos e médicos foram feitas perto do final do século. Gráficos abstractos de funções, erro de medição e colecta de dados empíricos foram introduzidos neste momento.

Este período também nos deu *William Playfair*, que é amplamente considerado o inventor de muitos dos gráficos mais populares que usamos hoje (gráficos de linhas, barras e circulares).

Atualmente, o uso de gráficos é uma prática comum, mais ainda, com o recurso às novas tecnologias. Estes têm adotado muitas formas, dependendo quer do que se está a representar, quer do público-alvo. A forma mais comum recorre a gráficos, que consistem numa estrutura composta por símbolos, como por exemplo, barras, linhas, pontos ou fatias de um círculo. Vemos habitualmente estes gráficos a representar opiniões, sondagens entre outras tendências da web.

Mas serão estes a melhor forma de visualização de dados?

Este projeto insere-se na área da visualização de informação, mais concretamente a visualização de opiniões *online* sobre hotéis. Como já foi referido anteriormente, as opiniões a nível da web, aumentam de dia para dia, especialmente no que toca a opiniões sobre hotéis. Só esta questão representa um problema na escolha da visualização. Um outro ponto importante são os dados que se pretendem representar.

Defende-se nesta tese que é possível propor um modelo visual para os dados das opiniões *online* sobre hotéis que proporcione uma melhor compreensão e experiência do utilizador final face ao que atualmente é trivialmente oferecido.

1.1. MOTIVAÇÕES, OBJETIVOS E CONTRIBUIÇÕES FUNDAMENTAIS

Uma motivação básica é a melhoria do modelo visual utilizado para a visualização de dados referentes a opiniões online. Apesar do foco deste trabalho serem os dados referentes a opiniões de hotéis, espera-se que o modelo possa ser útil em outros domínios (restauração, vinhos, etc.).

Este trabalho visa responder às seguintes questões:

1. Como representar visualmente opiniões *online* para apoiar uma melhor tomada de decisão?
2. Quais são as técnicas de visualização de informação mais adequadas para representar o conhecimento nas avaliações *online*?
3. Que tipos de gráficos devem ser utilizados para representar o conhecimento (informação integrada) em comentários *online*?

4. Se considerarmos que os potenciais utilizadores finais podem ser não só o público em geral, mas também gestores de negócios, haverá alguma diferença cognitiva visual que justifique a utilização de técnicas visuais diferentes?
5. Em última análise, a própria variedade do perfil de amostra de utilizadores (idade, género, nacionalidade, etc.) poderá implicar nessa diferença?

O desenvolvimento dos objetivos da investigação visa desenvolver um modelo de visualização.

Com este projeto pretende-se responder às questões levantadas anteriormente e, com isso, efetuar uma análise/reflexão sobre os elementos e técnicas a incluir para produzir o melhor modelo para visualização de opiniões *online* sobre hotéis. A caracterização do público-alvo do modelo de visualização é um dos resultados esperados, sendo fulcral para o desenho de um modelo de visualização.

Como resultado final será proposto um modelo visual e a implementação de um protótipo para testar, analisar e validar o modelo proposto.

1.2. METODOLOGIA DE INVESTIGAÇÃO

Para se propor e desenvolver uma ou mais visualizações das opiniões *online* de hotéis, é necessário primeiro alcançar um conhecimento geral da área de visualização de dados, de suas técnicas, processos, metodologias e ferramentas, além de identificar quais aproximações (técnicas e soluções) têm sido consideradas para as opiniões online, tendo em conta, em especial, o sector hoteleiro.

A *Grounded Theory*, ou Teoria Fundamentada em Dados em português, é uma metodologia de pesquisa que recorre a dados para obter o fundamento. Segundo o *The Grounded Theory Institute*¹, trata-se de uma teoria indutiva baseada na análise sistemática dos dados, que tanto podem ser quantitativos como qualitativos. Os detalhes desta teoria estão disponíveis no anexo I.

¹ Site web Institucional: www.groundedtheory.com

No caso deste projeto de investigação, aplicando a metodologia *Grounded Theory*, pretende-se efetuar um levantamento do que existe atualmente disponível e de que forma pode servir para oferecer resposta a este problema.

Este levantamento será levado a cabo seguindo ciclos de investigação-ação:

- Planificação da pesquisa
- Ação
- Observação/Avaliação
- Reflexão/Teorização

Este conjunto de procedimentos marca um ciclo, que quando completo, dá início a um novo ciclo que, por sua vez, desencadeia novas espirais de experiências de ação reflexiva. Quer-se com isto afirmar, que a cada conclusão de um ciclo a investigação é refinada e, desta forma são aprimorados os resultados obtidos. Repetindo-se assim estes ciclos até se obterem dados que confirmem a existência ou não de visualizações aplicadas às opiniões sobre hotéis. Com base nos dados obtidos avançar-se-á com a investigação para propor uma visualização de opiniões *online* de hotéis. Elaborar um protótipo e submetê-lo a testes, com os dados recolhidos nesses testes e de seguida, implementar a solução final.

Será efetuada uma pesquisa *top-down*, onde o tema central será decomposto em subtemas, até chegar à questão central e, desta forma, efetuar-se um levantamento sobre o tema, primeiro na sua vertente holística e depois caminhando para o caso particular, a visualização aplicada às opiniões sobre hotéis, ciclo após ciclo teorizando cada ponto em estudo:

1. Área de visualização em geral;
2. Área de visualização aplicada às opiniões;
3. Área de visualização aplicada às opiniões sobre hotéis.

A razão destes ciclos está relacionada com a atualização e evolução nesta área. Novos conhecimentos e melhores práticas neste campo estão em constante mudança. À velocidade a que novas pesquisas e experiências ampliam o conhecimento, as mudanças nos métodos de pesquisa ou práticas profissionais, podem ser necessárias [1].

Como linha orientadora para a elaboração da visualização será seguido o cenário “*Evaluating User Experience*” proposto no relatório técnico “*Seven Guiding Scenarios for Information Visualization Evaluation*” [4]. Este cenário contempla avaliações efetuadas a um grupo de potenciais utilizadores e abrem caminho ao estudo dos comentários e opiniões, registadas quer por escrito, quer oralmente, sempre que sejam ou não, solicitados.

Desta forma orienta-se a proposta de visualização ao utilizador final. Uma parcela importante no processo de construção da visualização é a interação com o utilizador final.

1.3. ESTRUTURA DO DOCUMENTO

O presente documento está organizado de acordo com os objetivos propostos na secção 1.1. Uma divisão em cinco capítulos organizados de forma a proporcionar ao leitor uma vista concisa e detalhada sobre o problema de investigação. O capítulo 1 corresponde a uma parte introdutória do documento. No capítulo 2 é feito um levantamento do estado-da-arte da área, das técnicas e tecnologia existentes. No capítulo 3 apresenta-se o modelo conceptual e as etapas percorridas até chegar a este modelo. No capítulo 4 é exibido o protótipo bem como os testes e a sua avaliação. No capítulo 5, referente às conclusões, são apresentadas reflexões e as respetivas avaliações finais sobre o trabalho efetuado e o potencial trabalho futuro a ser desenvolvido.

2. ESTADO DA ARTE

Este capítulo é composto pelo estudo dos conceitos, metodologias, disciplinas e soluções que orientam a temática da investigação. Aplicando uma metodologia de pesquisa *top-down* na secção seguinte será efetuado um enquadramento geral do tema. Após esse enquadramento geral será dado enfoque à visualização da informação relativamente a opiniões *online* e, por último, o caso da visualização da informação sobre opiniões *online* de hotéis.

2.1. A ÁREA DA VISUALIZAÇÃO DE INFORMAÇÃO

A visualização da informação é um processo de representação de dados, informação e conhecimento através de uma forma visual. O seu principal objetivo é a capacidade de visualizar dados, comunicar informações de forma clara e eficaz [3]. Ou seja, visa apoiar tarefas de análise, exploração, confirmação, apresentação e compreensão dos dados. Com o objetivo de conferir um bom uso da informação, são referidas três dimensões dessa tarefa de “ajudar à compreensão dos dados”: explorar, relacionar e comunicar.

Para transmitir ideias de forma eficaz, quer na forma estética quer na funcionalidade, ambos se combinam formando um todo coerente, fornecendo *insights* sobre um conjunto de dados, esparso e complexo, mas comunicando assim aspetos fundamentais de uma forma mais intuitiva. No entanto, os *designers* tendem muito frequentemente a descartar o equilíbrio entre *design* e função, criando visualizações sobre os dados lidos que não servem o seu principal propósito, comunicar com eficácia [3].

Então é por estas componentes cognitivas que se trata de uma área com um grande nível de complexidade, a imagem mental que cada utilizador cria de uma visualização é sempre subjetiva. O modelo mental de quem cria uma visualização é sempre diferente do modelo mental do utilizador. Cabe a esta área o papel de harmonizar e aproximar os modelos de ambos e, desta forma, tornar o processo da visualização mais harmonioso e eficiente.

“The human visual system is a pattern seeker of enormous power and subtlety” Colin Ware in Information Visualization: Perception for Design.

Colin Ware [1] destaca um detalhe importante sobre o sistema visual humano e a percepção do cérebro. O cérebro tenta encontrar padrões sobre as imagens captadas para a sua melhor compreensão. Percecionar o todo e não a unidade, destacar a “anomalia” do padrão, e é aqui que a visualização proporciona uma capacidade de compreender grandes quantidades de dados. Distinguir a informação importante, de entre mais de um milhão de medições, está disponível imediatamente. É mais fácil a um ser humano compreender dados expostos num gráfico que num conjunto de números. A visualização facilita a compreensão de características dos dados, quer em grande, quer em pequena escala. Este aspeto pode ser especialmente valioso para permitir a percepção de padrões que ligam a características locais.

O sistema visual humano é um buscador de padrões de enorme poder e subtileza. O olho e o córtex visual do cérebro formam um massivo processador paralelo que fornece o maior canal de largura de banda de centros cognitivos humanos [5]. Por outro lado, esta capacidade do ser humano em termos de percepção de informação tem os seus limites, e como tal, há que evitar a todo o custo uma sobrecarga de informação, em termos visuais. A escolha do tipo de gráfico a ser utilizado depende muito da natureza e tipo de dados, que podem traduzir informação relacional, temporal, ordinal, entre outras. No entanto, qualquer *conjunto* de dados é tão bom quanto a capacidade de os entender e comunicar e é por isso que se afigura essencial escolher a visualização correta. Não obstante, qualquer conjunto de dados pode ser utilizado para gerar inúmeras visualizações diferentes [5].

Cleveland e McGill [6] classificaram sete pistas visuais que ajudam a decodificar a informação:

1. A posição ao longo de uma escala comum, por exemplo, gráfico de dispersão;
2. Posição em escalas idênticas, mas não-alinhadas, por exemplo, vários gráficos de dispersão;
3. Comprimento, por exemplo, gráfico de barras;
4. Ângulo e declive, por exemplo, gráfico circular;
5. Área, por exemplo, bolhas;
6. Volume, densidade e saturação de cor (empate), por exemplo, mapa de calor;

7. Matiz de cor, por exemplo, notícias mapa.

A visualização incorreta leva a confusão, erros, desinteresse e sentimento de abandono entre os espectadores. Seguindo a linhas orientadoras do “*Seven Guiding Scenarios for Information Visualization Evaluation*” [4] para uma avaliação da experiência de utilizador é importante avaliar a visualização tendo em conta a perceção do utilizador, avaliando a forma e a estrutura dos seguintes pontos:

- 1) Avaliar a representação (o que é útil para o utilizador e que faz falta);
- 2) Avaliar como os utilizadores percebem a informação contida na visualização;
- 3) Avaliar parâmetros de Interação Humano-Computador (eficácia, eficiência e outros aspetos relevantes de Interação Humano-Computador);

2.2. VISUALIZAÇÃO DA INFORMAÇÃO APLICADA A OPINIÕES ONLINE

Após o *surgimento* da Web 2.0, a maioria dos *websites* passou a ter como requisito básico uma maior proximidade com o utilizador. O armazenamento e exibição de comentários dos utilizadores é uma prática corrente. Com o aumento dos negócios *online*, os gestores ficaram cada vez mais interessados em aspetos específicos do seu negócio como as atitudes, recomendações ou reclamações expressas pelos seus clientes. Um exemplo comum é dado pelos turistas que analisam opiniões e experiências publicadas por outros viajantes em diferentes plataformas Web quando planeiam suas próprias férias. A utilização exponencial do registo da opinião de cada utilizador faz com que seja impraticável consultar toda a informação existente. Muitos autores abordaram o problema da limitação humana para processar grandes quantidades de informação e extrair opiniões consensuais de um grande número de fontes que dependem de ferramentas baseadas em extração de dados [7].

Os clientes e potenciais clientes dos negócios online dão grande importância as opiniões registadas. É importante facilitar o acesso a esta informação por parte dos clientes. O crescente interesse em criar visualizações gráficas para dar ênfase às opiniões [8] é uma realidade. A utilização de visualizações gráficas é uma prática comum, com a apresentação de resumos sobre as opiniões e as médias das avaliações [3].

O gráfico é assim, uma ferramenta poderosa que coloca uma série de números próximos uns dos outros [9]. Esta proximidade de valores consiste numa estrutura composta por símbolos, como por exemplo, barras, linhas, pontos ou fatias de um círculo. Estes têm adotado muitas formas dependendo quer do que se está a representar, quer do público-alvo. Vemos habitualmente estes gráficos a representar opiniões, avaliações, sondagens entre outras tendências da web.

A visualização das opiniões tem impacto nas opções de potenciais clientes de um hotel ou de um restaurante, por exemplo. No caso do Turismo, os turistas são influenciados por opiniões online e se essas opiniões estiverem sumarizadas graficamente, a tomada de decisão poderá ser, seguramente, facilitada [10].

Estas fontes são subutilizadas tanto por consumidores quanto por empresas devido à sua natureza não estruturada, apresentação em série, ferramentas de pesquisa limitadas e baixa proporção de informações úteis para a quantidade total de dados. No artigo, [11] os autores propõem um sistema de análise visual que efetua a avaliação de sentimentos dos clientes e obtém *insights* de uma coleção de apreciações de produtos. É necessária uma análise visual eficaz das opiniões, pois isso tem um impacto significativo na construção de um negócio de sucesso e ajuda os clientes no processo de tomada de decisão.

A visualização de sentimentos e opiniões extraídas ou anotadas em textos tornou-se um tema proeminente de pesquisa na última década. Desde gráficos básicos, como por exemplo o de barras, utilizados para ilustrar revisões de clientes, até extensos sistemas de análise visual envolvendo novas representações.

As técnicas de visualização de sentimentos evoluíram para lidar com conjuntos de dados multidimensionais complexos, incluindo aspetos temporais, relacionais e geoespaciais. Um exemplo desta prática é a plataforma We feel² onde cada publicação do Twitter é analisada e classificada quanto ao sentimento que o autor expressa. Estas emoções são apresentadas numa visualização, em que cada cor representa um tipo de emoção [12], como podemos observar na figura 2.1. Neste exemplo as emoções são agrupadas numa estrutura hierárquica baseada na árvore de Parrotts.

² Sítio web: wefeel.csiro.au

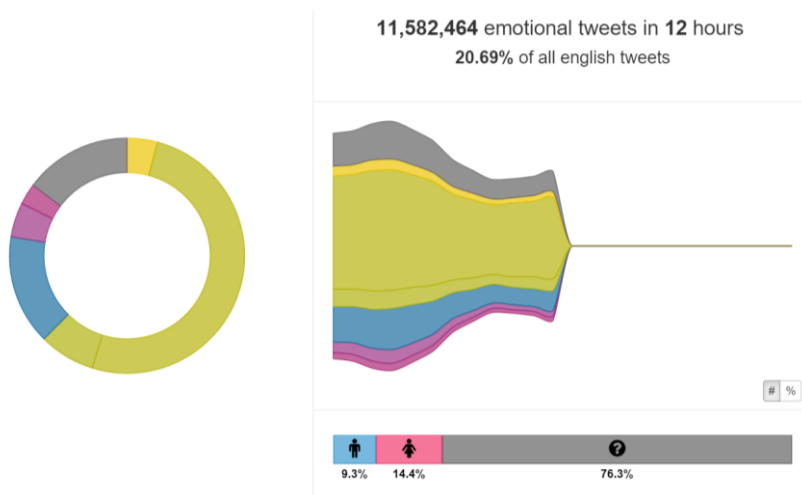


Figura 2.1. Representação de todas as emoções dos tweets mundiais publicados a 10-02-2016 [13]

Em 2001 W. Gerrod Parrott na sua Teoria das Emoções, categorizou as emoções numa árvore com três níveis. No primeiro nível estão localizadas as emoções primárias como amor, alegria, surpresa, raiva, tristeza e medo. No segundo nível estão as emoções secundárias que partem de cada uma das emoções anteriores. E por último, um terceiro nível onde estão as emoções que partem das emoções secundárias [14].

No artigo [11] os autores citam como exemplo de modelos de visualização de sentimentos, Review Spotlight de Yatani et al. É um modelo de visualização simples, figura 2.2, para resumir revisões de clientes numa rede social com uma nuvem de etiquetas de pares de palavras adjetivas / nominais. A polaridade de sentimento para cada palavra é calculada com correspondência lexical sobre suas contrapartes e usada para codificação de cores.

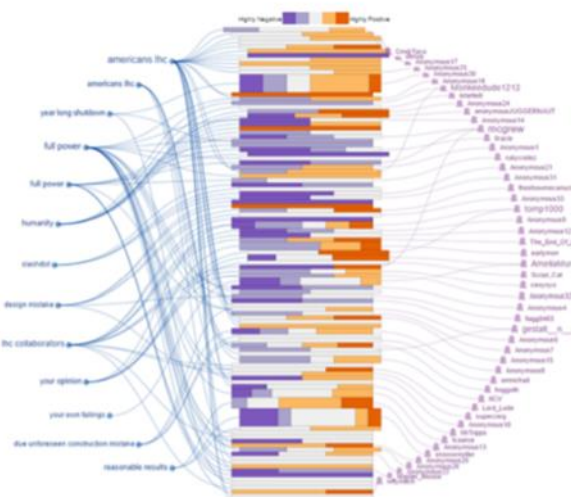


Figura 2.2. Exemplo de uma aplicação do modelo Review Spotlight [11, p. 80]

Outro exemplo apresentado no artigo [11], blocos de opinião de Alper et al., fornecem uma visão geral do sentimento baseada em aspectos de revisões de clientes. Este modelo combina vários gráficos de barras e etiquetas de texto coordenados que podem ser explorados interativamente para investigar a polaridade e as palavras-chave associadas a recursos específicos do produto. Na figura 2.3 está um exemplo apresentado no artigo [15], onde são visualizados trechos de opiniões de consumidores, no caso clientes que adquiriram aspiradores, organizados pelos principais recursos do produto.

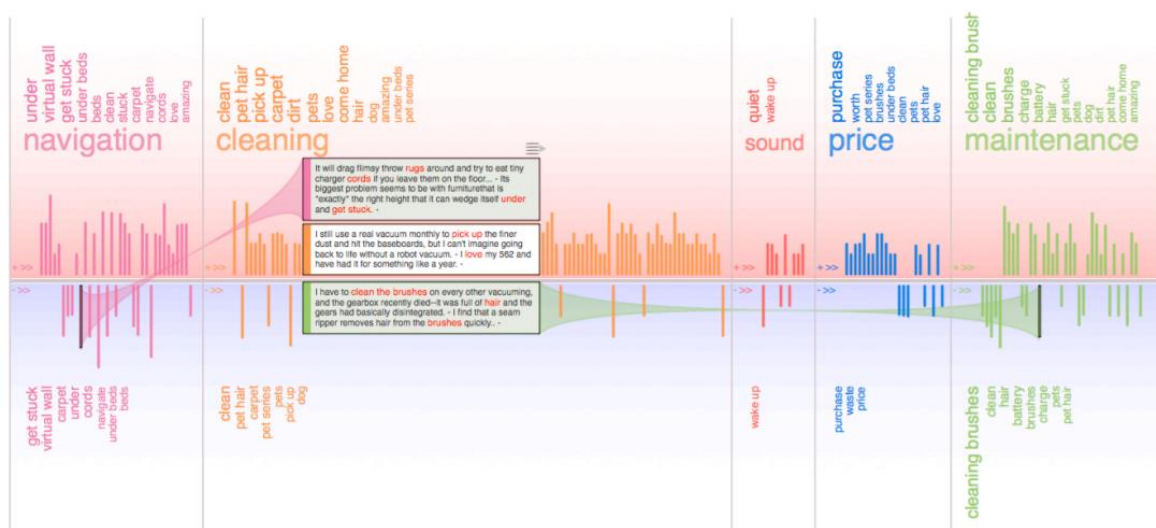


Figura 2.3. Exemplo de uma aplicação do modelo de blocos de opinião [15]

SentiVis por Di Caro e Grella, outro modelo referido no artigo [11], visualiza os resultados da análise de opiniões baseada em aspetos das revisões de clientes. No artigo [16] é apresentado o funcionamento detalhado do modelo, com recurso a dados extraídos de opiniões de clientes de hotéis. Numa lista de características, ao selecionar um aspeto específico relacionado com um produto ou serviço, fica disponível para os utilizadores uma representação visual das pontuações de polaridade para a característica selecionada. Essa representação consiste numa técnica híbrida que conjuga um gráfico de dispersão e um de linhas. Na figura 2.4 os pontos de dispersão, mínimo e máximo, de um determinado aspeto servem de baliza para traçar um gráfico de linha que representa a tendência.

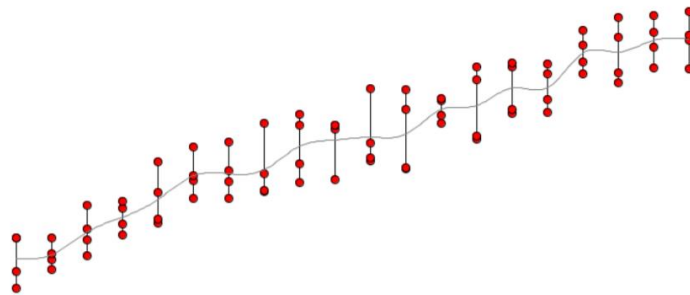


Figura 2.4. Exemplo de uma aplicação do modelo SentiVis [16]

Gorg et al. [17] discutem os recursos de análise e visualização de texto disponíveis no Jigsaw, que é um sistema de análise visual com propósito geral. O suporte à análise de sentimento inclui a deteção de polaridade a nível do documento. Pode também ser utilizado para a análise de aspetos específicos, quando a análise é combinada com um determinado tópico, conforme demonstrado com os dados relacionados com revisões de carros. As representações visuais envolvendo os dados de polaridade incluem nuvens de palavras e exibições de grelhas de documentos baseadas em pixéis. Os pixéis codificam a polaridade ao nível do documento.

2.3. AMBIENTES DE VISUALIZAÇÃO DE OPINIÕES SOBRE HOTÉIS

Como foi referido na secção 2.1, as opiniões a nível da web, aumentam de dia para dia, especialmente no que toca a opiniões sobre hotéis. Este conteúdo da Web, embora seja útil para apoiar a decisão do viajante ao reservar um quarto, geralmente é exibido como

uma lista de texto, não levando em conta nenhum aspeto visual, de interação ou cognitivo do utilizador final do site [18]. Só esta questão é um problema na escolha da visualização. Um outro ponto importante são os dados que se pretendem representar.

Na secção 2.1 foi feito um levantamento do estado-da-arte a nível da visualização das opiniões *online*, incluindo em alguns casos hotéis. Nesta secção, a ênfase será analisar os ambientes atuais disponíveis na web para a visualização de opiniões de hotéis.

Como ponto de partida, foi efetuado um levantamento de páginas web sobre opiniões de hotéis. Recorrendo a uma pesquisa no motor de busca *Google* efetuada a 03/03/2016 com a frase “opiniões sobre hotéis”, segundo este motor de busca foram obtidos cerca de 3.470.000 resultados, dos quais se analisaram os principais resultados da pesquisa, primeiras cinco páginas de resultados, e dos resultados foram analisadas as seguintes seis páginas que continham visualizações gráficas de opiniões:

- Booking, endereço web www.booking.com
- Logitravel, endereço web www.logitravel.pt
- Hoteis, endereço web www.hoteis.com
- Trivago, endereço web www.trivago.pt
- Momondo, endereço web www.momondo.pt
- TripAdvisor, endereço web www.tripadvisor.com

Segue-se uma análise pormenorizada de cada um dos ambientes identificados.

2.3.1. Booking

Na página web *Booking* é possível pesquisar-se por hotéis, consultar hotéis por localização entre outros parâmetros escolhidos pelo utilizador. Na lista de resultados apresentada é possível clicar num hotel para se obter informações detalhadas, localização, fotografias e outras, entre as quais as opiniões sobre um determinado hotel. A visualização dos dados destas opiniões está representada num gráfico tipo de barras com preenchimento a azul, como se pode observar na figura 2.5. A escala de avaliação é de 0 a 10, para os parâmetros avaliados e o valor de cada é apresentado na extremidade de cada barra do gráfico. Como informação em destaque é dada a média final de todos os parâmetros avaliados.



Figura 2.5. Captura de ecrã do website Booking.com [19]

2.3.2. Logitravel

Analisando a página web *Logitravel*, embora sendo uma página onde se presta um serviço do tipo de agência de viagens, apresenta as opiniões dos hotéis frequentados pelos seus clientes. A visualização dos dados destas opiniões está representada em um gráfico de barras com preenchimento a verde, como se pode observar na figura 2.6. A escala de avaliação é de 0 a 10, para os parâmetros avaliados e o valor de cada é apresentado na extremidade de cada barra do gráfico. Como informação em destaque é dada a média final de todos os parâmetros avaliados.



Figura 2.6. Captura de ecrã do website Logitravel [20]

2.3.3. Hoteis.com

No caso da página web *Hoteis.com*, à semelhança das anteriores é possível efetuar a reserva do hotel e aceder à informação detalhada sobre o hotel, inclusive ao feedback dos seus clientes. Nos mesmos moldes das anteriores páginas web, a visualização gráfica do feedback dos clientes trata-se de gráfico tipo de barras com preenchimento a verde, como se pode observar na figura 2.7 a escala de avaliação é de 0 a 5, para os parâmetros avaliados. É permitido ao utilizador da página, usar um filtro por motivo da reserva no hotel, e assim ver o resultado do feedback por motivação da estadia nesse hotel

escolhido. Tal como nas páginas anteriores, a informação destacada é a média final de todos os parâmetros avaliados, no entanto, existe um gráfico detalhado sobre a quantidade de utilizadores que atribuíram uma dada pontuação.



Figura 2.7. Captura de ecrã do website Hotels.com [21]

2.3.4. Trivago

Na página web *Trivago* no geral a plataforma é muito semelhante às anteriores, no que toca à visualização das opiniões, o tipo de gráfico mantém-se. A cor de preenchimento escolhida para as barras é o verde indicando o valor de cada, como podemos ver na figura 2.8, a escala de avaliação é compreendida entre 0 e 100. Existe um filtro que permite visualizar as opiniões por tipo de cliente do hotel. Tal como nas visualizações anteriores, está destacado o valor da média geral que neste caso é um gráfico circular preenchido segundo o valor dessa média.



Figura 2.8. Captura de ecrã do website Trivago [22]

2.3.5. Momondo

Analisando a página web *Momondo*, apesar do design que salta à vista quando se abre a página, a estrutura é muito semelhante às páginas anteriormente analisadas. Entretanto, e no que toca a visualização das opiniões optaram por gráficos circulares que são preenchidos com um tom de azul escuro consoante o valor de cada parâmetro avaliado, tal como podemos observar na figura 2.9. O valor da avaliação de cada parâmetro varia entre 0 e 10, que acrescenta um resumo dos pontos positivos e ao contrário das outras páginas não é apresentada uma média geral de avaliação.

Resumo das 4 091 avaliações de hóspedes



Figura 2.9. Captura de ecrã do website Momondo [23]

2.3.6. TripAdvisor

A página web *TripAdvisor* face às anteriores páginas analisadas, não varia muito no que toca à sua estrutura. No entanto a visualização das opiniões, como podemos observar na figura 2.10, mantém o gráfico do tipo de barras, todavia apresenta uma composição gráfica com a tradicional barra e outra com cinco círculos que são preenchidos consoante a avaliação de cada parâmetro. Como tal depreende-se que a escala destes será entre 0 e 5, no entanto o gráfico de barras sem qualquer limite máximo, serve de filtro para as opiniões consoante a escala de 0 a 5.



Figura 2.10 Captura de ecrã do website TripAdvisor [24]

2.4. CONCLUSÕES

Como se pode constatar, as visualizações oferecidas pelas páginas web analisadas não diferem muito de umas das outras. As apresentações gráficas são do mesmo tipo; gráfico de barras, monocromáticas e seguindo o estereótipo dos sites onde os utilizadores expressam a sua opinião. Em algumas representações há uma tentativa de fugir à regra, apresentando filtros de dados consoante a motivação do alojamento e gráficos circulares. Em absoluto, estes não deixam de ser monocromáticos, a cor apenas preenche a forma segundo o valor da avaliação.

Estes gráficos resultam apenas de uma avaliação dirigida ao utilizador, por intermédio de um questionário após a reserva de um hotel. É dada pouca ou nenhuma ênfase visual ao sentimento presente nos comentários, apenas sendo apresentado o texto daqueles que têm maior relevância. Refletindo um pouco mais, sobre os dados apresentados em relação aos hotéis e clientes, estes apenas contêm informações básicas sobre os mesmos nas amostras de dados, enquanto os comentários incluem informações, como texto livre e detalhado, assim como os sentimentos dos utilizadores. Em muitos destes casos, trata-se de informação útil, pois o conteúdo dos comentários e das avaliações, apesar de ser vantajoso para apoiar a decisão do viajante ao reservar um quarto, geralmente é exibido como uma lista de texto, não levando em conta nenhum aspeto visual, de interação ou de teor cognitivo do utilizador final do site [2, p. 451].

Existe uma dimensão que em nenhuma das visualizações está a ser tomada em conta: - o tempo. Essa informação consta normalmente apenas na forma textual. Uma opinião dada

hoje pode não representar a realidade desse hotel passado um ano, ou até mesmo de alguns meses. Pode, por exemplo, ter existido uma intervenção no hotel que permitiu melhorar um, ou vários dos eventuais aspetos negativos levantados pelos utilizadores e quantidade das avaliações, bem como algumas opiniões não estejam refletidas naquele momento. Urge, portanto, a necessidade de saber a importância desta dimensão para o utilizador que consulta esta informação.

3. MODELO CONCEPTUAL

A visualização da informação ainda é um campo de conhecimento que carece de uma teoria sólida para apoiá-lo, no entanto, existem inúmeras metodologias e taxonomias que podem ser combinadas e adotadas como diretrizes [2]. Como a tarefa de visualizar é uma tarefa cognitiva, a construção da visualização (e conseqüente escolha de técnicas) está completamente assente nas linhas orientadoras que estudos da interação homem-máquina produzem ao longo do tempo [10].

Quando se envia uma mensagem é sempre necessário codificá-la para que esta possa ser transmitida, através de um determinado meio, até ao seu recetor. Por sua vez, este terá de descodificar a mensagem para a perceber. A elaboração de uma visualização de dados não

é diferente, os dados são codificados visualmente remanescendo a típica questão - será que o recetor percebeu a mensagem?

Neste capítulo é apresentado o desenho do modelo visual para a representação das opiniões sobre hotéis. O desenho do modelo teve como base um levantamento de requisitos funcionais e visuais, a nível do mapeamento dos dados.

3.1. LEVANTAMENTO DE REQUISITOS

O modelo proposto visa suprir as necessidades de dois tipos de utilizadores em específico, os turistas que procuram os melhores hotéis e os gestores destes hotéis que procuram melhorar o seu negócio.

No início desta investigação foram colocadas as seguintes questões:

- Como representar visualmente opiniões *online* para apoiar uma melhor tomada de decisão?
- Quais são as técnicas de visualização de informação mais adequadas para representar o conhecimento nas avaliações *online*?
- Que tipos de gráficos devem ser utilizados para representar o conhecimento (informação integrada) em comentários *online*?

- Se considerarmos que os potenciais utilizadores finais podem ser não só o público em geral, mas também gestores de negócios, haverá alguma diferença cognitiva visual que justifique a utilização de técnicas visuais diferentes?
- Em última análise, a própria variedade do perfil de amostra de utilizadores (idade, género, nacionalidade, etc.) poderá implicar nessa diferença?

Para avaliar estes pontos, elaborou-se um questionário organizado em duas secções, onde cada uma contém perguntas para obter respostas as questões acima referidas. Além destas questões, foram apresentados gráficos ilustrativos de potenciais soluções visuais para o utilizador optar. Nos anexos II e III encontra-se a lista de perguntas consideradas, bem como as imagens do questionário disponibilizado ao público e o endereço da hiperligação.

O questionário foi criado através da plataforma de formulários da *Google*. Esta plataforma permite criar um questionário completo com inúmeras opções, podendo ser partilhado em qualquer rede social, sem qualquer custo pela utilização.

Os dados foram recolhidos no período de 5/4/2017 a 5/9/2017, sendo o endereço do formulário partilhado nas redes sociais *Facebook*, *LinkedIn* e *Twitter*. A partilha foi efetuada de duas formas, como uma publicação de domínio público e mensagem privada a todos os contactos que disponho nessas redes sociais.

Seguindo a recomendação indicada pela *Grounded Theory*, a amostra recolhida foi dirigida à construção da teoria e, não visando a representação populacional [25, p. 45]. Pretendeu-se assim, entre outras questões, verificar até que ponto são importantes as opiniões dos utilizadores, bem como o destaque visual que lhes é dado.

Ao todo foram recolhidas noventa e uma respostas ao questionário. Os inquiridos eram falantes ou utilizadores de língua portuguesa e tinham idades compreendidas entre os 19 e os 61 anos. A maioria tinha entre os 29 a 40 anos. Relativamente ao género dos inquiridos, estavam distribuídos de forma equilibrada, como podemos observar no gráfico seguinte, ilustrado na figura 3.1.

Indique o seu género:

91 respostas

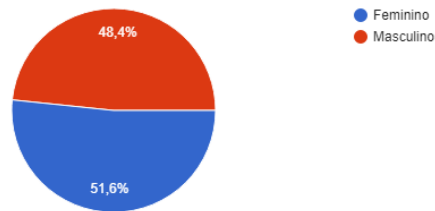


Figura 3.1. Distribuição entre género dos inquiridos

Como se pode observar na figura 3.2, a maioria é utilizadora frequente de websites de opiniões sobre hotéis. No entanto, ainda existe um número considerável de utilizadores não frequentes.

É utilizador frequente de websites de opiniões sobre hotéis?

91 respostas

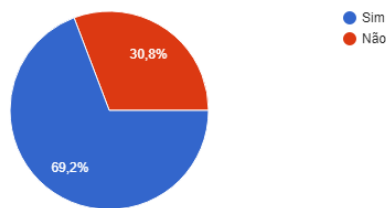


Figura 3.2. Distribuição por hábito na utilização de websites de opinião

Consultando a figura 3.3, mais de 90% concorda que a avaliação apresentada tem peso na escolha do hotel. O mesmo aconteceu relativamente à questão, sobre a leitura dos comentários (figura 3.4) e se os mesmos têm importância na escolha do hotel (figura 3.5).

Para si, a avaliação apresentada tem peso na escolha de um hotel?

91 respostas

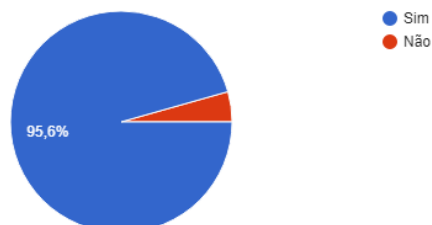


Figura 3.3. Distribuição por opinião em relação a influência dos websites de opinião

Quando escolhe um hotel, tem por hábito ler os comentários sobre os hotéis?

91 respostas

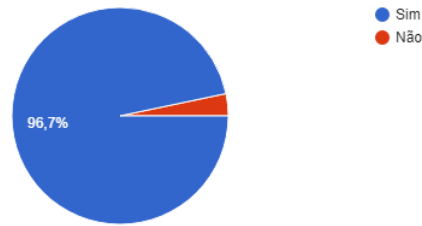


Figura 3.4. Distribuição das respostas sobre o hábito ou não na leitura dos comentários

Os comentários que lê têm influência na escolha de um hotel?

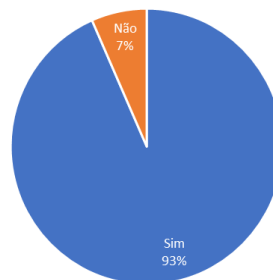


Figura 3.5. Distribuição das respostas sobre a influência dos comentários na escolha do hotel

Como se pode observar no Anexo III, na segunda parte do questionário são apresentadas seis imagens contendo gráficos utilizados em páginas sobre opiniões de hotéis mais utilizadas na web. O Gráfico 1 foi retirado do sítio *Booking*. A imagem do gráfico retirado do sítio *Logitravel* foi identificada como Gráfico 2. O Gráfico 3 diz respeito ao sítio *Hoteis.com*. O Gráfico 4 é do sítio *Trivago*. A imagem apresentada como Gráfico 5 foi retirada do sítio *Momondo*. Por fim, o Gráfico 6 é relativa ao sítio *TripAdvisor*. As legendas apresentadas nos gráficos seguintes referem-se a esta situação.

Indique qual o que usa com maior frequência?

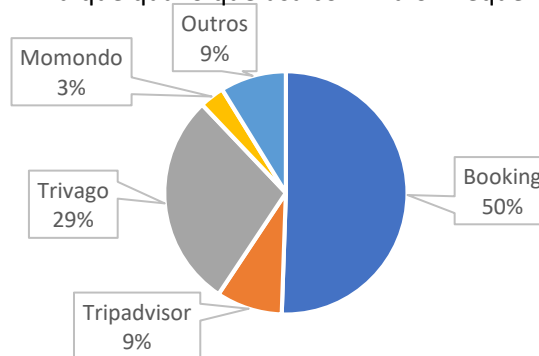


Figura 3.6. Distribuição da preferência dos inquiridos sobre o sítio web de opiniões sobre hotéis

Quanto à preferência dos sítios web de opiniões sobre hotéis, pode observar-se a figura 3.6, que a escolha recai sobre os sítios *Booking* e *Trivago*. Possivelmente uma das razões pela qual os inquiridos preferiram a visualização das avaliações destas duas páginas, é em função da maior facilidade em consultar a informação (veja a figura 3.7). Estas duas páginas apresentam os parâmetros avaliados em gráficos de barras. No caso da apresentação dos resultados do *Trivago* esta é complementada por alguns aspetos positivos referidos nos comentários.

Imagem	Sítio web onde foi retirada
Gráfico 1	<i>Booking</i>
Gráfico 2	<i>Logitravel</i>
Gráfico 3	<i>Hoteis.com</i>
Gráfico 4	<i>Trivago</i>
Gráfico 5	<i>Momondo</i>
Gráfico 6	TripAdvisor

Tabela 3.1. Legenda dos gráficos utilizados no questionário

Na sua opinião, em qual dos gráficos achou mais fácil consultar a informação?

87 respostas

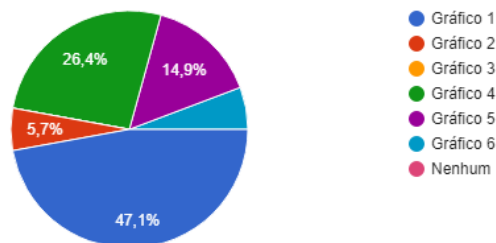


Figura 3.7. Distribuição sobre os gráficos que os inquiridos reconheceram ser de mais fácil consulta

E o gráfico mais difícil de consultar a informação?

87 respostas

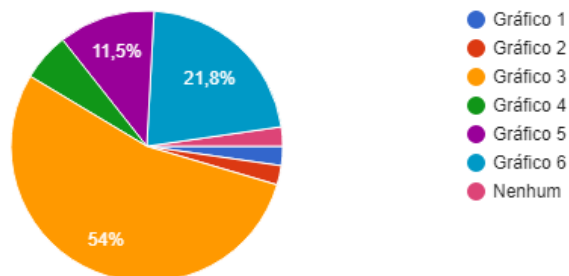


Figura 3.8. Distribuição das respostas sobre os gráficos de consulta mais difícil

Observando a **figura 3.8**, mais de metade dos inquiridos concordou que a visualização de mais difícil leitura é do sítio *Hoteis.com*, apresentado na legenda como **Gráfico 3**. A tendência negativa desta visualização manteve-se ao longo do questionário, como podemos constatar nos gráficos das distribuições das respostas seguintes, figuras 3.9, 3.10, 3.11, 3.12, 3.13 e 3.14. **O Gráfico 6, ou seja, a visualização do TripAdvisor**, foi o segundo mais votado, o que leva a considerar excluir as técnicas destas visualizações para um modelo futuro.

Assinale em qual ou quais das avaliações é possível observar os aspectos mencionados de forma positiva ou negativa sobre o hotel.

91 respostas

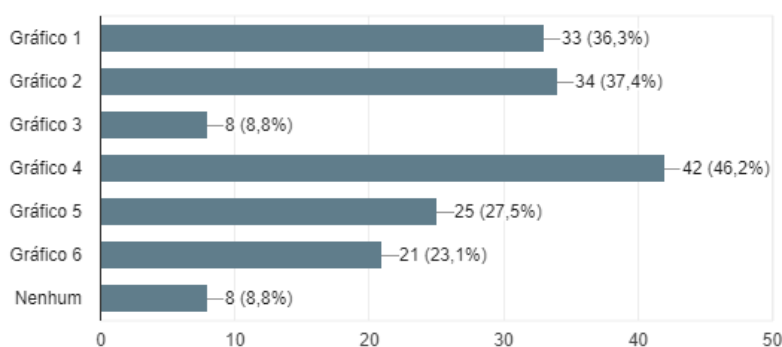


Figura 3.9. Gráfico com os resultados sobre as visualizações onde é possível observar aspectos mencionados de forma positiva ou negativa

Quanto à distribuição dos gráficos em que é possível observar os aspetos negativos e positivos mencionados, a escolha maioritária recai sobre os **Gráficos 1, 2 e 4**. Com algumas características comuns neste tópico estas visualizações também terão importância em decisões futuras.

Em termos da distribuição das opiniões sobre a obtenção de uma síntese geral da informação, como podemos observar na figura, o destaque vai para o **Gráfico 1** seguido pelo **Gráfico 4**. Como a diferença de valores é tão grande entre os outros gráficos e estes, é importante analisar as características destes em termos da sua síntese geral.

Em qual ou quais das avaliações é mais fácil obter uma síntese geral da informação

91 respostas

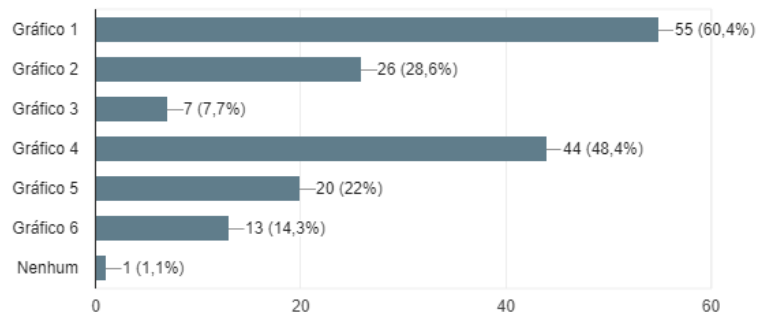


Figura 3.10 Distribuição das respostas de qual ou quais avaliações é mais fácil obter uma síntese geral da informação

No que toca aos atributos visuais forma e cor, a escolha recai nos **Gráficos 1 e 4** (Trivago e Booking) que representam aproximadamente 87% dos inquiridos, como podemos ver nas **figuras 3.11 e 3.12**. Fica aqui a importância destas visualizações para o desenho da visualização para o modelo proposto.

Em qual ou quais gráficos a forma é o atributo visual mais relevante para mapear informações?

91 respostas

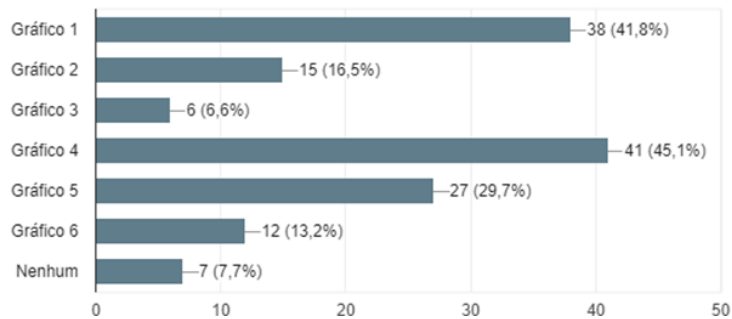


Figura 3.11. Distribuição das respostas sobre os gráficos onde a forma é o atributo relevante

Em qual ou quais gráficos a cor é o atributo visual mais relevante para mapear as informações?

90 respostas

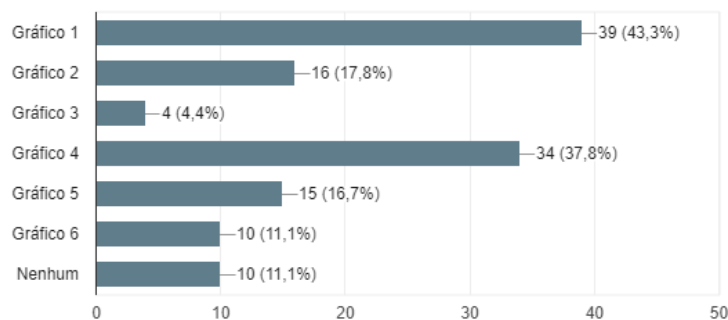


Figura 3.12. Distribuição das respostas sobre os gráficos onde a cor é o atributo visual mais relevante

Observando a distribuição apresentada na **figura 3.13**, constata-se que a preferência se mantém para as respostas à questão sobre os gráficos onde é mais fácil quantificar o número de comentários sobre um determinado aspecto.

Em qual ou quais gráficos é mais fácil quantificar o número de comentários sobre um determinado aspecto?

90 respostas

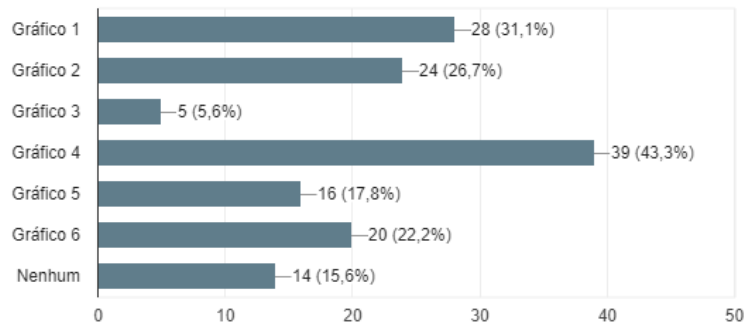


Figura 3.13 Distribuição das respostas de qual ou quais as avaliações onde é mais quantificar o número de comentários sobre um determinado aspecto

Analisando as respostas obtidas na questão sobre o gráfico que oferece melhor visão a partir dos comentários sobre os diferentes aspectos do hotel, como podemos observar na **figura 3.14**, a distribuição manteve-se nos 20% com o destaque de 30% de respostas para o **Gráfico 4**. Por outro lado, os **Gráficos 3 e 6** mantem a tendência negativa, ficando como exemplos a não seguir de futuro.

Qual ou quais dos gráficos oferece a melhor visão a partir dos comentários sobre os diferentes aspectos do hotel?

90 respostas

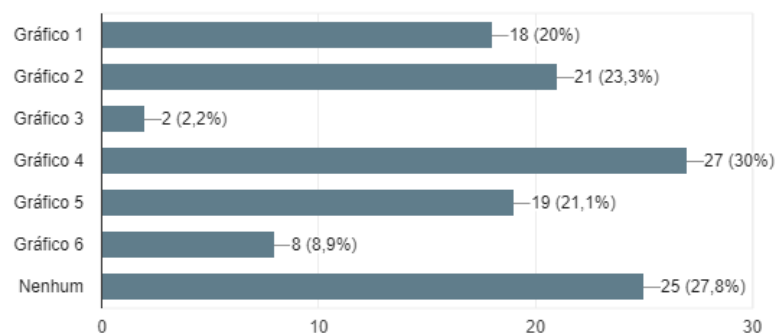


Figura 3.14 Distribuição das respostas de qual ou quais avaliações oferecem uma melhor visão a partir dos comentários sobre os diferentes aspectos do hotel

Analisando a distribuição das respostas sobre a importância da separação temporal das avaliações, **figura 3.15**, ficou claro que a esmagadora maioria respondeu afirmativamente. Trazendo assim à luz do novo modelo uma distribuição temporal dos resultados.

Considera importante a possibilidade de separação temporal das avaliações?
90 respostas

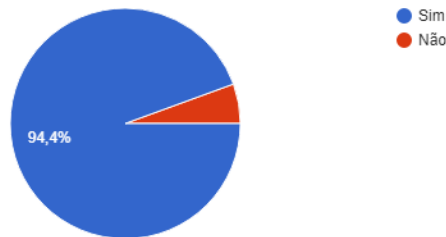


Figura 3.15. Distribuição da importância da separação temporal das avaliações

3.2. FUNCIONALIDADES E VISUALIZAÇÃO

O modelo proposto, **figura 3.16**, é constituído por quatro etapas, recolha de opiniões dos clientes, seguido de uma análise emocional destas opiniões, que serão armazenadas numa base de dados para uma posterior visualização destes dados. Neste trabalho o enfoque será na fase final onde será criado um protótipo com a visualização proposta, a fim de ser testado e validado.

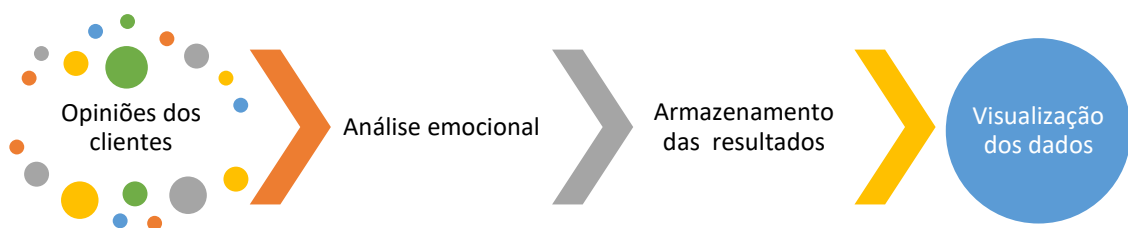


Figura 3.16. Esquema da arquitetura geral do modelo. Fonte: do próprio autor.

Tal como no projeto *We Feel* [12] que representa graficamente as emoções contidas nas publicações do Twitter, este modelo pretende apresentar as emoções extraídas das opiniões sobre os hotéis. Assume-se que as emoções são uma melhor forma de qualificar opiniões de utilizadores e em função disso, dar um melhor suporte a decisão.

Uma questão relevante que se apurou após a análise das visualizações das páginas web sobre opiniões de hotéis, foi a total (ou quase, exemplo do *Booking* que permite algum tipo filtro e ordenação dos comentários por data) ausência de separação temporal dos conteúdos. Mais de 90% dos inquiridos respondeu que era importante existir uma separação temporal.

Em hotelaria, à semelhança de qualquer outro negócio centrado nos seus clientes, a diferença exige melhoria contínua da sua oferta. O fator tempo é uma dimensão relevante nos dados referentes a registos, quer sejam médicos, de negócios, científicos ou históricos [26, p. 3].

Sendo o tempo uma dimensão em destaque, o modelo proposto deverá ter de contemplar esta dimensão. Existe uma técnica visual criada por Havre, S., Hetzler, B., & Nowell, L. designada por *ThemeRiver*. Esta técnica é assim designada por usar a metáfora de um rio onde a informação e a representação dos dados flui numa linha temporal [26]. O *ThemeRiver* fornece aos utilizadores uma visão global das alterações nos dados de diferentes temáticas ao longo do tempo. Esta técnica visual visa facilitar a identificação de tendências, padrões e ocorrências inesperadas ou a não ocorrência de temas ou tópicos. As emoções serão visualmente mapeadas com o auxílio do atributo cor. Portanto, será definida uma escala de cor para as emoções. A roda de cores de Plutchik [11] organiza as emoções segundo um modelo radial a três dimensões análogo às cores de uma típica roda de cores. A dimensão vertical do cone, altura, representa a intensidade da emoção e cada um dos círculos representa graus de similaridade entre os vários tipos de emoção.

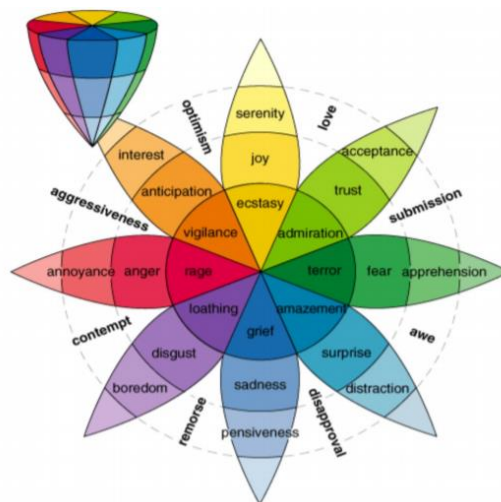


Figura 3.17. Escala de cor das emoções de Plutchik [27][11]

Assim sendo o ponto de partida serão as oito emoções consideradas como primárias, sendo elas: a raiva, o medo, a tristeza, a repugnância, a surpresa, a curiosidade, a aceitação e a alegria. Destas ramificam todas as outras, Kempter et al. [28] apresentam uma aplicação deste círculo de cores e emoções. Nesta sua aplicação prática, as emoções são extraídas de comentários no Twitter sobre os Jogos Olímpicos de 2012. Com base neste trabalho e adaptação das emoções ao contexto e língua portuguesa, a **tabela 3.1** mostra a escala de emoções e a cor correspondente que será adotada pelo modelo.


Cor	Código Hexadecimal	Atributo visual
	#723030	Raiva
	#f5912d	Interesse
	#e66331	Divertido
	#ffb100	Orgulho
	#e6d53a	Alegria
	#bac500	Prazer
	#baed70	Adora
	#3dde00	Confiança
	#8cf488	Alívio
	#00c987	Surpresa
	#6defe9	Nostalgia
	#48a8f3	Lástima
	#0037a8	Tristeza
	#3b59d4	Preocupação
	#2a2674	Vergonha
	#7559df	Culpa
	#7225c2	Arrependimento
	#d421ff	Inveja
	#6f0070	Desgosto
	#b11b9f	Desprezo

Tabela 3.2. Escala de cor para as emoções

Como evidente, tendo em conta o contexto que se inserem os dados oriundos de opiniões sobre hotéis, algumas das emoções possivelmente não estarão presentes na amostra de dados. Em todo o caso, assume-se esta escala de cor como a base para o mapeamento do respetivo atributo visual para as emoções.

Os dados utilizados para a visualização serão simulados, no entanto, respeitarão a estrutura de dados apresentada na **tabela 3.2**. Logo, considera-se que os dados de entrada contêm minimamente a seguinte informação: identificação do hotel, sentimento (por média das opiniões de todos os aspetos do hotel), gênero do cliente e idade do cliente. Estes dados seriam o resultado de um processamento e análise em termos de sentimento de uma base de dados como por exemplo as existentes e utilizadas pelo TripAdvisor ou Booking.com. Considera-se que a análise do sentimento seria obtida tendo como base o sentimento médio obtido pela avaliação da opinião de todos os aspetos (limpeza, atendimento, preço, localização, etc.). Foi necessário recorrer a esta solução devido a inexistência de listas de dados ou plataformas que expusessem estes dados de uma forma pública. Embora a solução visual proposta esteja a considerar a média do sentimento, a mesma pode ser aplicada no caso de se considerar cada aspeto individualmente.

Os dados serão representados visualmente tendo em conta os seguintes aspetos e funcionalidades descritas na **tabela 3.2**:

Campo	Tipo	Descrição	Representação Visual e Funcionalidade
Data	Data	Data da criação do comentário	Posição ao longo do eixo horizontal. Animação na transição entre diferentes anos.
Sentimento	Texto	Sentimento extraído do comentário, que pode variar entre os seguintes valores: Raiva, Interesse, Divertido, Orgulho, Alegria, Prazer, Adora,	Cor e largura das séries, a cor apresentada será a de cada emoção, tal como apresentada na tabela 3.1. Caixa de texto com o valor

		Confiança, Alívio, Surpresa, Nostalgia, Lástima, Tristeza, Preocupação, Vergonha, Culpa, Arrependimento, Inveja, Desgosto, Desprezo	médio em percentual do sentimento predominante. Legenda de cor. Filtro.
Género	Texto	Género de que escreveu o comentário	Texto que servirá de filtro para detalhar a informação.
Idade	Numérico	Idade de quem escreveu o comentário	Indicador numérico que servirá de filtro.
Destaques	Texto	Destaques efetuados no comentário	Texto em <i>tooltips</i>

Tabela 3.3. Representação visual dos dados no modelo

Partindo do princípio de uma técnica visual que permite fornecer informações contextuais por meio de uma linha de tempo e marcadores para eventos de interesse recorrentes, foi efetuado um primeiro esboço, apresentado na **figura 3.18**. A visualização será disponibilizada como uma aplicação web inserida no contexto de um sítio web de opiniões sobre hotéis.

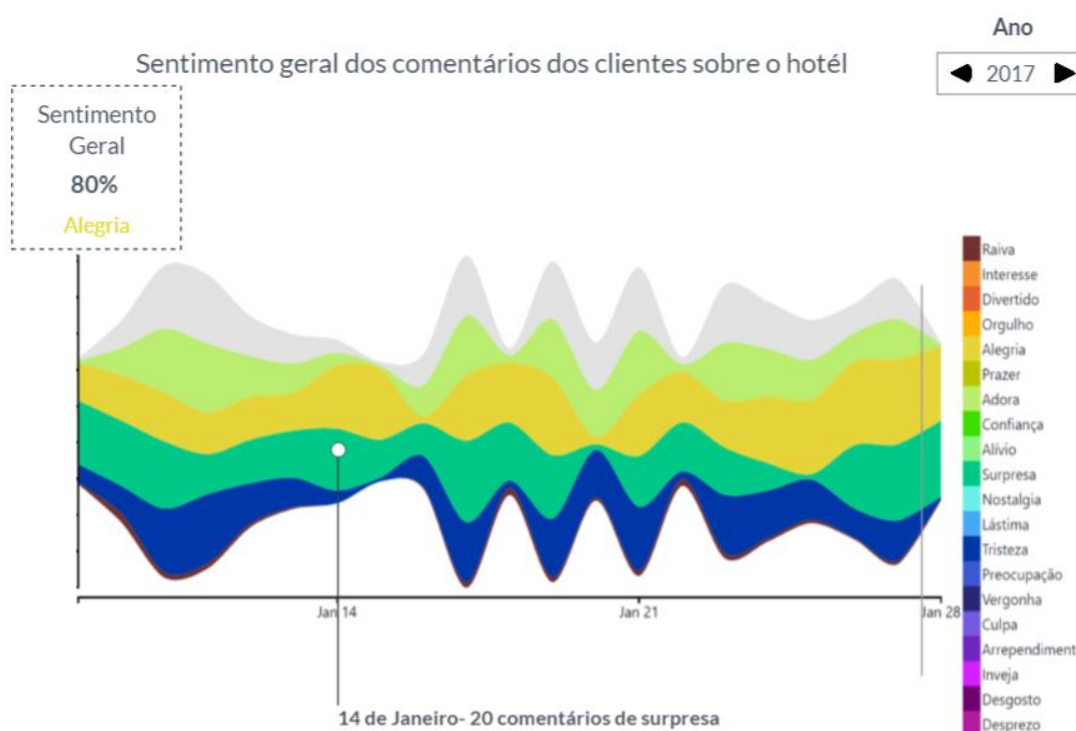


Figura 3.18 Mockup para uma primeira fase de avaliação da proposta de visualização. Fonte: do próprio autor.

Cada área do gráfico corresponde a um determinado sentimento, com altura variável consoante o número de ocorrências e preenchida com a cor atribuída a esse sentimento numa determinada data. Estas áreas são empilhadas de forma a perfazer o total de comentários por data.

Quando o utilizador seleciona uma determinada área, surge uma caixa no canto superior esquerdo a indicar o valor em percentual da média geral daquele sentimento. Cada sentimento é exibido como uma série temporal em forma de fluxo. A largura do fluxo em cada posição ao longo do eixo horizontal é proporcional a expressão que um determinado sentimento teve num determinado momento temporal. Esse valor é obtido a partir da análise do sentimento contido nos comentários deixados pelos hóspedes.

Como forma de garantir múltiplas perspetivas sobre os dados, existe um filtro por ano, que permite encurtar ou alargar a amplitude da linha temporal, e assim, cumprir com a função de medir impactos nas opiniões do hotel num dado intervalo de tempo. Uma *tooltip* aparece sempre que o ponteiro do rato passa num ponto em cada série do gráfico indicando o sentimento que representa, a data dos comentários e a quantidade.

3.2.1. SESSÃO DE AVALIAÇÃO DO ESBOÇO INICIAL

Foi efetuada no dia 9/10/2017 uma sessão de grupo de discussão para a avaliação do esboço inicial, permitindo compilar vários pontos de vista e avaliar a aceitação deste esboço. Seguindo as recomendações para os grupos de discussão [29], esta sessão teve a duração de aproximadamente uma hora onde foram reunidos numa sala isolada de ruído, 10 indivíduos, 3 do género feminino e 7 do género masculino, com idades compreendidas entre os 25-30, 30-35 e 35-40 anos, que têm como pontos em comum serem utilizadores de sítios web sobre hotéis e fazerem parte de uma equipa de desenvolvimento de aplicações web.

Foi criado um guião de entrevista que está disponível no **anexo IV**, de forma a apoiar e orientar a realização do grupo de discussão, incluindo questões relevantes.

Após uma breve introdução sobre o tema e apresentando as atuais visualizações foi introduzido ao grupo o esboço mostrado na **figura 3.18**. Foi pedido ao grupo que desse a

sua opinião sobre as visualizações mostradas, segundo dois pontos de vista distintos: a do turista e a do gestor de hotel. Todo o processo foi orientado para que o grupo não saísse do âmbito da discussão, os resultados estão compilados num quadro de síntese, consultar o **anexo V**.

O grupo no geral aprovou o esboço face às outras visualizações, indicando como pontos positivos:

- Escala temporal – 100%;
- Exposição emocional dos comentários – 100%;
- Interação com o gráfico – 50%;

E indicando como pontos de melhoria:

- Todo o grupo foi da mesma opinião, afirmando que o rótulo de texto informativo de legenda no topo das cores e os sentimentos que representam;
- Uma vista detalhada onde pode ser possível filtrar por serviços ou comodidades do hotel, depois da anterior esta foi a opinião mais comum, partilhada por 80% do grupo;
- 50% respondeu, comparar um ou mais detalhes do hotel.

3.3. TECNOLOGIA

Foi efetuado um levantamento das possíveis tecnologias a serem utilizadas para implementar o modelo. Dentre as muitas disponíveis, destacam-se as seguintes:

- **Highcharts** (<https://www.highcharts.com>);
- **ChartJs** (<https://www.chartjs.org>);
- **goJs** (<https://gojs.net>);
- **Grafana** (<https://grafana.com/grafana>);
- **Kibana** (<https://www.elastic.co/products/kibana>);
- **D3.js** (<https://d3js.org/>).

A biblioteca Highcharts, desenvolvida em JavaScript utilizando a tecnologia WebGL, apresenta um grande leque de técnicas visuais com exemplos prontos a utilizar, um

pouco à semelhança do Excel, sendo apenas de uso gratuito para uma versão limitada e fins não comerciais. Toda e qualquer solução partindo desta biblioteca deverá ter em conta os possíveis custos da licença de utilização.

No caso da biblioteca ChartJs, desenvolvida também em JavaScript e à semelhança da biblioteca HighCharts utiliza a tecnologia WebGL. O leque de técnicas disponíveis é muito menor face à biblioteca Highcharts, no entanto, não apresenta quaisquer custos de utilização.

A biblioteca goJs é em muitos pontos diferente das anteriores, apesar de ser desenvolvida em JavaScript a tecnologia de desenho da visualização é vetorial, recorrendo ao uso de SVG. O foco central desta biblioteca são os diagramas, no entanto, possui algumas técnicas visuais para visualização de dados. Para esta biblioteca existe um custo associado à licença para a sua utilização.

Nos casos do Grafana e Kibana são plataformas onde se constrói *dashboards* para análise de dados. Possuem um conjunto de funções, comandos e operações próprias de uma framework, que em muitos casos diferem dos padrões adotados pelas linguagens de programação. Para além do tempo de aprendizagem na utilização e parametrização destas plataformas, a integração num *template* de uma página web nunca é perfeito.

A D3.js é uma biblioteca em JavaScript que permite uma grande liberdade na implementação de visualizações de dados. Como referem no website oficial da API [30], o D3.js não é uma estrutura monolítica que visa proporcionar todos os recursos imagináveis. Em vez disso, D3 resolve o cerne do problema: a eficiente manipulação de documentos com base em dados.

Ao contrário de outras bibliotecas em JavaScript que apenas permitem criar gráficos com o CANVAS do HTML 5, o D3 permite manipulação de objetos vetoriais no formato SVG. Este formato consiste numa estrutura gráfica vetorial em XML. O conteúdo de um SVG

pode ser estático, dinâmico, interativo ou animado, o que torna este formato muito flexível [31].

As bibliotecas baseadas em CANVAS permitem criar imagens de uma forma mais rápida, no entanto é extremamente difícil criar efeitos interação e animação. Funciona sobre uma biblioteca gráfica de JavaScript, o *WebGL*, que pode criar alguns problemas em aparelhos com menos capacidade gráfica.

Por oposição ao CANVAS os elementos do SVG podem todos ser formatados com CSS. Nelli [31], realça o seguinte no seu livro, é possível aplicar estilos em elementos SVG através de *Cascading Style Sheets* (CSS) e adicionar comportamentos dinâmicos a estes, quer pela interface da API quer pela DOM (*document object model*).

Existe um problema típico quando se elabora uma visualização de dados numa página web, que é como converter os valores dos dados na forma certa, ou seja, pixéis e cores. Esta tarefa de conversão pode complicar-se mais ainda, como é no caso de visualizações estatísticas, pois é necessário utilizar escalas ordinais, numerais, logarítmicas, temporais e muitas outras [32].

A biblioteca D3 possui 3 tipos de escalas para facilitar esta tarefa, são as escalas quantitativas, para séries de dados contínuos, como por exemplo, números. As escalas ordinais para séries de dados discretas, como é o caso de categorias ou nomes. Por último as escalas temporais, que servem para intervalos de tempo, dias, horas e meses.

Em função dos objetivos pretendidos para este trabalho, considera-se que para a implementação do protótipo deverão ser utilizadas as linguagens HTML, CSS e JavaScript, bem como a possibilidade de implementar qualquer técnica visual com a biblioteca escolhida, do leque das mais utilizadas apenas uma não limita a possibilidade na implementação de várias técnicas visuais, a biblioteca D3.js.

3.3. CONCLUSÕES

Quando se fala em dados orientados a uma linha temporal, basicamente quer-se designar dados que de alguma forma estão conectados ao tempo. Certamente, que esta descrição

é vaga não sendo suficiente afinada quer para os utilizadores, que têm de escolher ou para quem planeia e desenvolve métodos de visualização apropriados. Um requisito essencial para obter uma visualização expressiva e eficaz é considerar as características dos dados a serem apresentados, que neste caso estão particularmente relacionados com a dimensão tempo.

A técnica *Theme River* permite mostrar os dados em várias séries empilhadas, como um fluxo contínuo no tempo, no entanto é necessário a cada fluxo colocar o contexto, caso contrário o utilizador corre o risco de não obter a mensagem correta da leitura dos dados. Precavendo esta situação, recorre-se a uma *tooltip* para indicar a leitura, bem como o contexto, dos dados analisados. A cor é um atributo importante pois esta está associada uma emoção, com o intuito de auxiliar o utilizador a gama de cores com as correspondentes emoções está sempre visível no ecrã.

O modelo assume esta técnica visual como sendo a basilar para representar a variação ao longo do tempo do sentimento manifestado pelas opiniões de clientes de hotéis. Os dados que serão utilizados pelo protótipo são simulados. É assumido que algum tratamento prévio foi efetuado, nomeadamente a análise dos sentimentos das opiniões dos clientes em relação aos diversos aspetos do hotel (localização, limpeza, simpatia, preço, etc.).

4. PROTÓTIPO

A questão principal da experiência de utilização de uma visualização é: O que pensa o utilizador final da visualização proposta? [4, p. 10]. As seguintes a ter em conta, são:

- Será que os utilizadores interpretam detalhes da visualização de maneiras que não se esperava?
- Como é que a sua interpretação da visualização difere daquela de um histograma, por exemplo, mostrando os mesmos dados?
- Como gostaria de alterar ou manipular a vista?

Qualquer visualização para conseguir ser mais eficiente, exige um bom nível de interação com o utilizador final (tarefas de exploração) [3]. A avaliação da experiência de utilização procura compreender como as pessoas reagem a uma visualização quer num curto ou longo período. Uma visualização pode ser alternadamente um esboço inicial do projeto, como pode ser um protótipo de trabalho, bem como um produto acabado [4].

Tendo em conta todas as considerações e o levantamento efetuado que resultou no desenho do modelo conceptual (capítulo 3), foi elaborado um protótipo e colocado online no endereço www.iecivica.org/dataviz/.

A implementação foi realizada com base em HTML, CSS e JavaScript. Para garantir um melhor ajuste do conteúdo ao dispositivo em que se está a visualizar o modelo foi utilizada a *framework* web Bootstrap. Ao tema base foram adicionadas alterações nos estilos, através de CSS, para personalizar a estrutura do protótipo. Para facilitar a implementação da visualização utilizou-se a biblioteca de JavaScript D3.js, auxiliando na construção dos três gráficos existentes do protótipo.

O protótipo do modelo proposto foi posto à prova de potenciais utilizadores finais. O objetivo foi recolher resultados para concluir a eficácia ou não do modelo visual proposto.

Este capítulo apresenta o protótipo bem como os testes realizados com o mesmo.

4.1. INTERFACE VISUAL E FUNCIONALIDADES

Partindo de uma típica lista de resultados de uma pesquisa sobre hotéis, neste caso em Lisboa (a figura 4.1 ilustra a página inicial que é apresentada), é apresentada a opção ver opiniões, opção esta que dá acesso à visualização implementada, como a presente na figura 4.2.

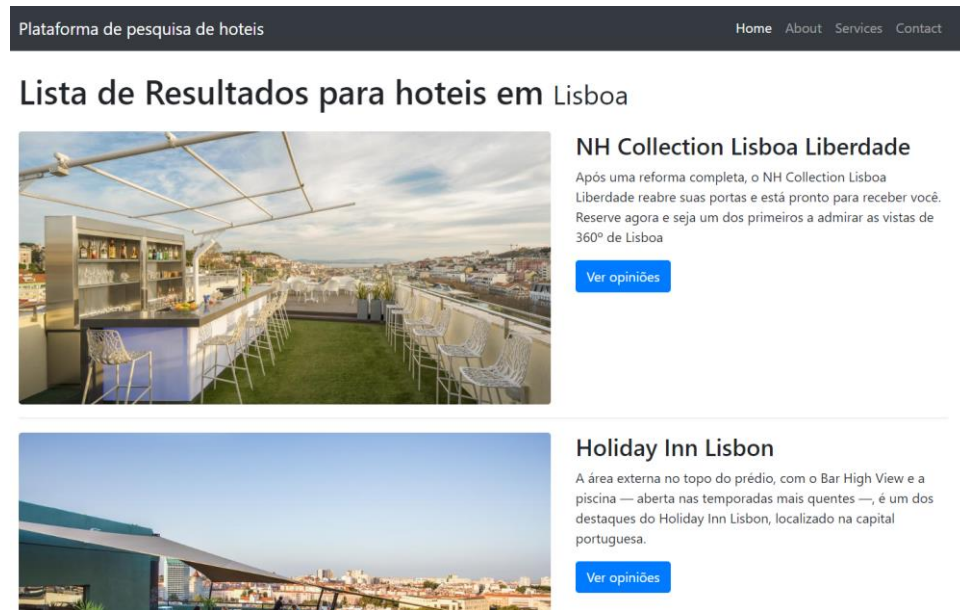


Figura 4.1 Captura do ecrã da página principal. Fonte: do próprio autor.

Esta visualização contém uma área visível com gráfico apresentando as emoções no geral e no topo um filtro relativamente aos anos. Na escolha de um ano, tem como efeito limitar a amplitude temporal dos dados visualizados. Existe ainda, uma ferramenta de *tooltip* que acompanha a navegação do rato pelas várias séries do gráfico, apresentando a informação sobre os detalhes dos dados abaixo da linha horizontal do gráfico, como se pode observar na figura 4.2.

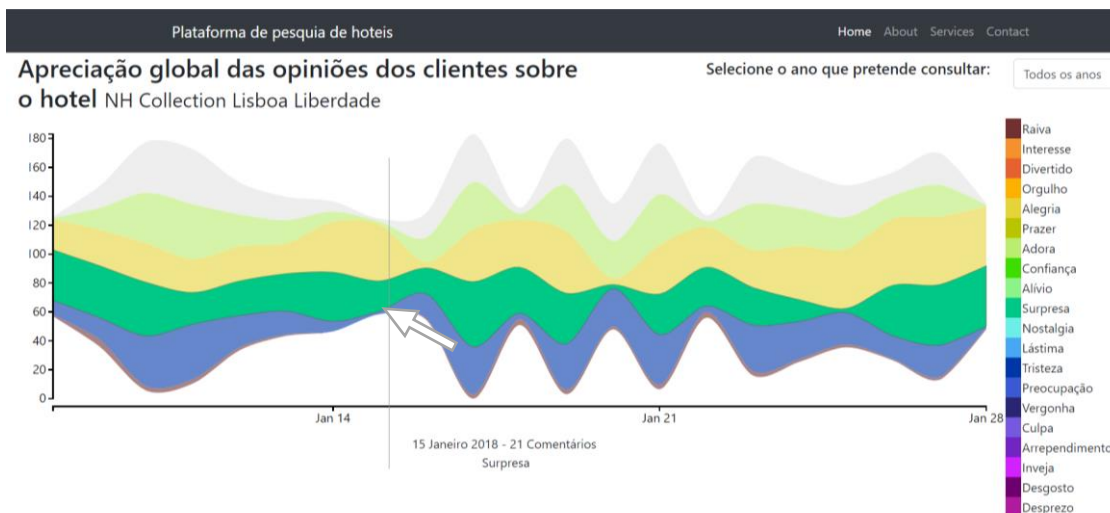


Figura 4.2 Captura do ecrã do gráfico principal. Fonte: do próprio autor.

De forma a facilitar a leitura do gráfico existe também uma escala, à direita do gráfico, com as cores e as emoções correspondentes. É possível ainda clicar em cada série e obter um quadro com os detalhes dos comentários, como mostra na figura 4.3.

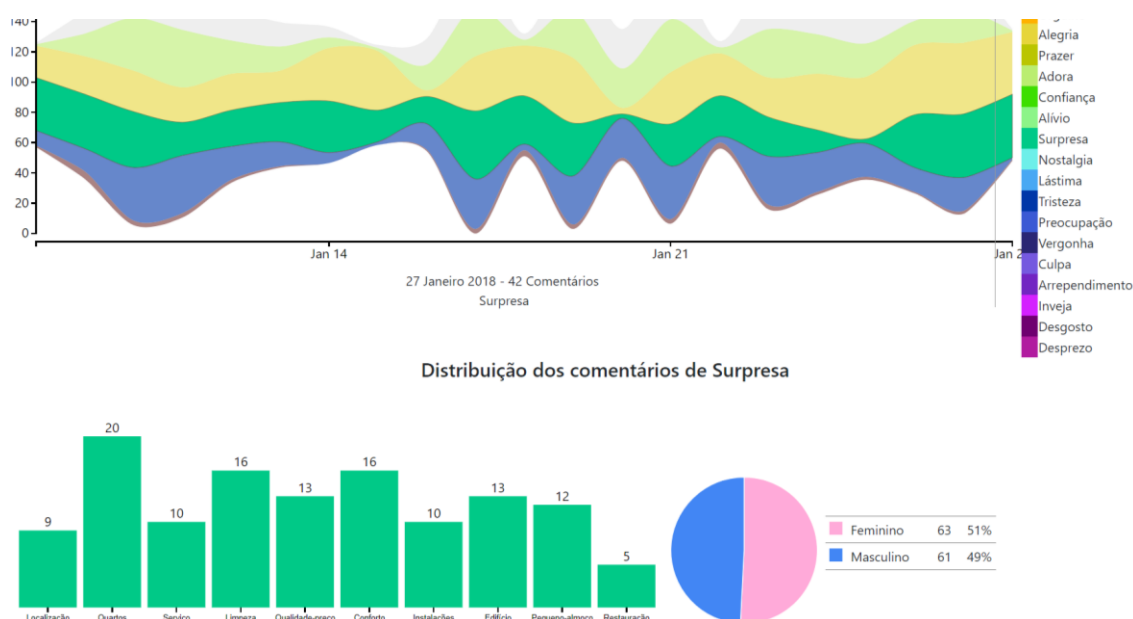


Figura 4.3 Captura do ecrã do quadro de detalhes. Fonte: do próprio autor.

Neste quadro existem três áreas, na área da esquerda está um gráfico de colunas que indica a quantidade de comentários identificados com essa emoção. Estes dados sempre que possível são agrupados pelos seguintes indicadores:

- Localização;
- Quartos;
- Serviço;
- Limpeza;
- Qualidade-preço;
- Conforto;
- Instalações;
- Edifício;
- Pequeno-almoço;
- Restauração;
- Outros.

Na área da direita está um gráfico circular que indica a distribuição do género dos indivíduos que fizeram os comentários, ao clicar em cada série de dados, esta funciona como um filtro que limita o gráfico da esquerda, mostrando a distribuição dos comentários pelo género seleccionado. Na área inferior são apresentados os comentários efetuados com a emoção, género e o indicador escolhido.

Os dados utilizados para gerar os gráficos foram simulados, conforme indicado na página 35 do capítulo 3. Os dados de entrada estão organizados segundo uma estrutura de dois níveis de detalhe um primeiro nível, este mais abrangente, constituído por tempo, emoção e quantidade. E um segundo nível, onde cada emoção está distribuída por parâmetros de avaliação e a respetiva quantidade e o género dos utilizadores que efetuaram os comentários.

No caso de se tratar de um protótipo e para facilitar a tarefa de implementação do mesmo foram criados ficheiros de dados agrupados por ano.

4.2. AVALIAÇÃO

Para a avaliação deste modelo, foi utilizado o mesmo grupo de discussão, sendo o procedimento ligeiramente alterado. Existiram duas fases, onde na primeira o grupo utilizou o modelo quer na vertente de potencial turista, que na de gestor, e de seguida,

na segunda fase, seguiram-se as perguntas que ao contrário da sessão anterior as perguntas colocadas são fechadas, guião disponível no anexo VI.

Durante a sessão todas as respostas foram anotadas e compiladas posteriormente numa tabela de resultados disponível no anexo VII, e para facilitar a análise foram criados gráficos que serão expostos em seguida.

Após a utilização do protótipo pelo grupo foi colocada a questão se o uso da emoção melhorou a visibilidade das opiniões, tal como podemos constatar na figura 4.4 o grupo foi unanime na sua opinião que o uso das emoções melhorou a visibilidade das opiniões.



Figura 4.4 Distribuição das respostas para a pergunta sobre a melhoria na visibilidade das opiniões

O grupo quando questionado sobre o uso de uma escala temporal para distribuir as emoções, este também foi unanime na sua resposta, afirmando que facilitou a forma como avalia um hotel. Como pode ser observado no gráfico da figura 4.5.



Figura 4.5 Distribuição das respostas à pergunta sobre o uso de uma escala temporal para distribuir as emoções

Analisando a questão sobre a escala de cores para cada emoção o grupo, na sua maioria, respondeu que esta foi bem concebida, figura 4.6. Apenas 20% não concorda, questionando individualmente estes participantes sobre a razão desta discordância, verificou-se que emoções como a alegria, o orgulho e a tristeza geraram dúvidas sobre se a cor era de facto a mais adequada.

Na sua opinião a escala de cores a relação da cor com a emoção foi bem concebida?

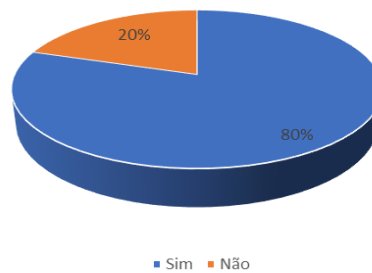


Figura 4.6 Distribuição das respostas sobre a relação da cor com a emoção

Em termos de avaliação do protótipo, avaliação dos detalhes e da interação com o sistema, a média situa-se, no geral, nos 17 valores numa escala de 0 a 20, tal como podemos observar no gráfico da distribuição das médias na figura 4.7. No caso da média da avaliação que o grupo fez à experiência de utilização foi de 16,4 valores, numa escala de 0 a 20.

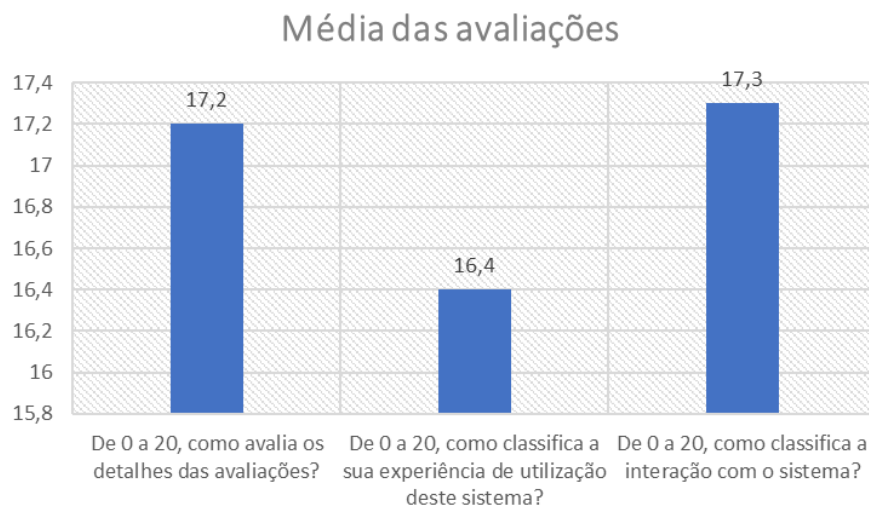


Figura 4.7 Média das avaliações do protótipo para os detalhes, experiência de utilização e interação

Numa apreciação global do protótipo destaca-se que 90% do grupo concorda com a melhoria deste face ao sistema tradicional como demonstra o gráfico da figura 4.8.

Comparando com o sistema tradicional, a visualização dos comentários melhorou?

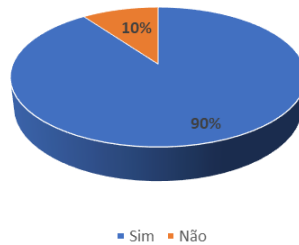


Figura 4.8 Captura do ecrã do quadro de detalhes

Analisando o gráfico da figura 4.9, destaca-se também, que 20% concorda que de alguma forma se devia de manter a avaliação clássica, ou seja, apresentar também os resultados dos questionários de avaliação dos visitantes do hotel.

Deveria de ser mantida a avaliação clássica dos hotéis?

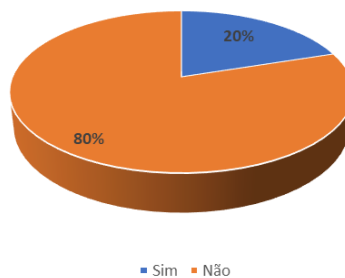


Figura 4.9 Captura do ecrã do quadro de detalhes

4.3. CONCLUSÕES

Refletindo sobre os dados obtidos nesta fase de avaliação é seguro afirmar que os resultados são positivos e que o protótipo convenceu os utilizadores. Isto deve-se não só às respostas dadas pelos participantes, como também ao feedback positivo transmitido aquando da primeira fase do grupo de discussão.

O uso de emoções para dispor, qualificar e quantificar os comentários foi aceite e indicado este fator como ponto positivo deste protótipo. O mesmo se sucedeu com a

distribuição destas emoções numa linha temporal, que transmitiram não só continuidade na obtenção de feedback como uma contextualização temporal dos dados.

Também da análise às respostas denota-se que para alguns utilizadores o antigo paradigma encontra-se vigente, como é o caso dos participantes que concordaram com a inclusão da avaliação clássica num desenvolvimento futuro. Ou o caso do participante que não concordou que este protótipo representa uma melhoria na visualização de opiniões de hotéis.

Dos parâmetros avaliados pode-se concluir que existe espaço para melhoria numa implementação futura. De todos os parâmetros avaliados, o que carece de mais atenção é a experiência de utilização, que avaliado com 16 cria espaço para melhoria numa implementação futura, o que nesse caso, obrigaria a mais intervenções entre quem desenvolve a solução e quem a utiliza.

Quando se inicia uma idealização para uma concretização existe um longo caminho em que conceitos intrínsecos a cada utilizador se perdem. Neste caso não foi exceção, apesar do grupo ter aceite a solução avaliando-a de forma positiva, há algum caminho ainda a percorrer para encontrar o ponto de equilíbrio entre o utilizador e a solução.

CONCLUSÃO E TRABALHO FUTURO

Este trabalho foi impulsionado pela resposta a cinco questões inicialmente colocadas, sendo a primeira relacionada com a representação visual de opiniões online de forma a apoiar a tomada de decisões. A extração de emoções dos comentários efetuados pelos utilizadores das plataformas, revelou-se fundamental na categorização, quantificação e agrupamento dos comentários. Sendo o tempo uma dimensão com importância para os utilizadores a aplicação desta na visualização foi um requisito que pesou na opção da técnica escolhida, *theme river*. Apesar desta técnica não ser muito considerada em visualizações, neste caso revelou-se com um passo para o sucesso, claro deixando a ressalva que a metáfora não é compreendida da mesma forma por todos os utilizadores.

A técnica *Theme River* aplicada como visualização principal, funciona como o ponto de partida para um utilizador fazer a sua análise dos dados, este dispõe de todas as emoções extraídas das opiniões, ao longo do tempo, tempo este que pode ser filtrado, facilitando a análise de um período de tempo. Num ponto escolhido pelo utilizador levar a mostrar detalhes sobre um tipo de emoções escolhido num determinado dia. A interação implementada sobre esta técnica catapultou-a de uma visualização a filtro de informação mais detalhada de cada emoção numa determinada data.

Uma das limitações deste projeto foi a ausência de dados reais para serem utilizados na implementação do protótipo, que para execução deste existiu a necessidade de simular estes. Outra limitação foi a própria biblioteca D3.js, utilizada para implementar as visualizações, que nas interações, transições e outras animações, revelou a inexistência nativa de processos assíncronos de programação, dificultando assim todos estes processos importantes para indicar ao utilizador, mudanças nos dados apresentados.

A execução deste projeto cumpre um papel essencial na formulação não só desta resposta como respostas futuras. Abrindo caminho a novas formas de visualizar os dados obtidos a partir de opiniões sobre hotéis, assim como o desenvolvimento de formas distintas de responder ao problema formulado. Do mesmo modo, o modelo desenvolvido

neste trabalho cumpre um papel essencial na formulação de novas opções de desenvolvimento futuro. Futuro este que a acontecer terá de se ter em conta duas situações que pesaram neste trabalho, a ausência de dados reais e as limitações tecnológicas.

No caso dos dados uma solução passará por sondar a plataformas tradicionais sobre a utilização dos dados que possuem para serem submetidos a uma análise emocional. Trabalho este que por si dá espaço a uma dissertação sobre o tema. No que toca a limitação tecnológica, no momento em que concluo esta dissertação está a ser testado um upgrade à biblioteca D3.js onde inclui um padrão de desenvolvimento conhecido como *Observables*, que permite sincronizar todos os eventos, chamadas e existências de dados num determinado âmbito. Esta atualização será o facilitador para o melhoramento de questões relacionadas com a experiência de utilização, interação e performance no carregamento de grandes quantidades de dados. As técnicas visuais utilizadas apesar de consideradas válidas pelo grupo de utilizadores que testaram o protótipo, o leque apresentado pode ser alargado para combinar e melhorar os resultados da observação de dados. Aqui também fica o trabalho futuro de uma recolha massiva das opiniões dos utilizadores sobre esta nova forma de visualizar os dados.

Este modelo proposto, pode ser aplicado a outros contextos em que os utilizadores, clientes e outros que contribuam com a sua opinião para um produto ou serviço e cuja distribuição dos dados obtidos seja ao longo do tempo. Salvaguardando-se todos os ajustes necessários de contextualização do modelo à realidade do produto ou serviço.

Defendeu-se nesta tese que é possível propor um modelo visual para os dados das opiniões online sobre hotéis que proporcione uma melhor apreensão e experiência do utilizador final face ao que atualmente é trivialmente oferecido. A avaliação do protótipo assim o comprova com os seus resultados positivos e aceitação por parte dos utilizadores que o testaram.

5. BIBLIOGRAFIA

- [1] C. Ware, *Information Visualization Perception for Design*, 3ª ed., Elsevier, 2012.
- [2] E. S. Carvalho e M. S. Chaves, "Detecting End-User's Visual Model to Build a Visualization Tool Based on Online Reviews," *PARSONS JOURNAL FOR INFORMATION MAPPING*, vol. V, pp. 1-11, 2013.
- [3] P. Simon, *The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions*, New Jersey: Willey, 2014.
- [4] H. Lam, E. Bertini, P. Isenberg, C. Plaisant e S. Carpendale, "Seven Guiding Scenarios for Information Visualization Evaluation," UNIVERSITY OF CALGARY, 2011.
- [5] D. Keim, G. Grinstein e M. O. Ward, *Interactive Data Visualization*, A K Peters, Ltd, 2010.
- [6] A. Kirk, *Data Visualization: a successful design process*, Packt Publishing, 2012.
- [7] E. Marrese-Taylor, J. D. Velasquez e F. Bravo-Marquez, "A Novel Deterministic Approach for Aspect-Based Opinion Mining in Tourism Products Reviews," *Expert Systems with Applications*, vol. 41, pp. 7764-7775, 2014.
- [8] Y. Wu, F. Wei, S. Liu, N. Au, W. Cui, H. Zhou e H. Qu, "OpinionSeer: Interactive Visualization of Hotel Customer Feedback," *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, VOL. 16, NO. 6, NOVEMBER/DECEMBER 2010, pp. 1109-1118, 210.
- [9] D. M. Wong, *THE WALL STREET JOURNAL - Guide to Information Graphics*, W.W. Norton & Company, 2013.
- [10] E. S. Carvalho, M. S. Chaves e N. Sacramento, "Visualização de Informação de Opiniões Online sobre Restaurantes: Uso de Técnicas Orientadas à Visualização de Grafos," *Revista Eletrônica de Sistemas de Informação (RESI)*, 2014.
- [11] A. Kerren, C. Paradis e K. Kucher, "The State of the Art in Sentiment Visualization," *COMPUTER GRAPHICS forum*, vol. 37, n.º 1, pp. 71-96, 2018.
- [12] D. P. C. C. H. B. P. a. O. B. Milne, "We Feel: Taking the emotional pulse of the world," em *Proceedings of the 19th Triennial Congress of the International Ergonomics Association (IEA 2015)*, Melbourne, Victoria, Australia, 2015.
- [13] L. a. S. C. t. i. t. D. E. P. o. CSIRO, "We feel," CSIRO and Black Dog Institute, with support from Amazon AWS and GNIP, 2015. [Online]. Available: wefeel.csiro.au. [Acedido em 10 02 2016].
- [14] W. G. Parrott, *Emotions in Social Psychology: Essential Readings*, Georgetown : Psychology Press , 2001.
- [15] B. Alper, H. Yang, E. Haber e E. Kandogan, "OpinionBlocks: Visualizing Consumer Reviews," em *In Proceedings of the IEEE VisWeek Workshop on Interactive Text Analytics for Decision Making*, 2011.
- [16] L. Di Caro e M. Grella, "Sentiment Analysis via Dependency Parsing," *Computer Standards & Interfaces*, vol. 35, n.º 5, p. 442-453, 2013.
- [17] C. Görg, Z. Liu, J. Kihm, J. Choo, H. Park e J. Stasko, "Combining Computational

- Analyses and Interactive Visualization for Document Exploration and Sensemaking in Jigsaw,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 9, n.º 10, pp. 1646 - 1663, 2013.
- [18] E. S. Carvalho e M. S. Chaves, “New Findings on the User’s Preferences about Data,” *World Academy of Science, Engineering and Technology*, vol. 83, pp. 451-461, 2 10 2013.
- [19] “Booking.com,” Booking Holdings Inc., [Online]. Available: <https://www.booking.com>. [Acedido em 03 03 2016].
- [20] “Logitravel efetuado a 03/03/2016 10:11,” Logitravel S.L. - Agência de viagens on-line, [Online]. Available: <https://www.logitravel.pt>. [Acedido em 03 03 2016].
- [21] “Hoteis.com,” Hotels.com, L.P., [Online]. Available: <https://pt.hoteis.com>. [Acedido em 03 03 2016].
- [22] “Trivago,” trivago N.V., [Online]. Available: <https://www.trivago.pt>. [Acedido em 03 03 2016].
- [23] “Momondo,” momondo A/S, [Online]. Available: <https://www.momondo.pt>. [Acedido em 03 03 2016].
- [24] “TripAdvisor,” TripAdvisor LLC, [Online]. Available: <https://www.tripadvisor.pt>. [Acedido em 03 03 2016].
- [25] B. G. Glaser e A. L. Strauss, *THE DISCOVERY OF GROUNDED THEORY: Strategies for Qualitative Research*, Transaction Publishers, 1999.
- [26] W. Aigner, S. Miksch, H. Schumann e C. Tominski, *Visualization of Time-Oriented Data*, Springer London, 2016.
- [27] R. Plutchik, “The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice,” *American Scientist*, vol. 89, n.º 4, pp. 344-350, 2001.
- [28] R. Kempter, V. Sintsova, C. Musat e P. Pu, “EmotionWatch: Visualizing Fine-Grained Emotions in Event-Related Tweets,” em *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, University of Michigan, 2014.
- [29] I. S. Silva, A. L. Veloso e J. B. Keating, “Focus group: Considerações teóricas e metodológicas,” *Revista Lusófona de Educação*, n.º 26, pp. 175-190, 2014.
- [30] M. Bostock, “D3 Data-Driven Documents,” Mike Bostock, 2017. [Online]. Available: www.d3js.org. [Acedido em 4 06 2017].
- [31] F. Nelli, *Create Web Charts with D3*, Apress, 2014.
- [32] M. Dewar, *Getting Started with D3*, O'Reilly Media, 2012.
- [33] C. C. Gramazio, D. H. Laidlaw e K. B. Schloss, “Colorgical: Creating discriminable and preferable color palettes,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, pp. 521-530, 2016.
- [34] S. Liu, W. Cui, Y. Wu e M. Liu, “A survey on information visualization: recent advances,” *The Visual Computer*, vol. 30, pp. 1373-1393, 2014.
- [35] H. E. Chi, “A Taxonomy of Visualization Techniques using the Data State Reference Model,” *IEEE Symposium on Information Visualization 2000. INFOVIS 2000*, pp. 69-75,

2000.

6. ANEXOS

6.1. ANEXO I - GROUNDED THEORY - DETALHES

O *The Grounded Theory Institute* partilha uma descrição de Odis E. Simmons sobre as várias etapas que constituem a *Grounded Theory*, que são as seguintes:

1. Preparação

- ✓ Minimização de ideias preconcebidas;
- ✓ Sem bibliografia preliminar crítica;
- ✓ Tópico geral de investigação, mas sem problema predeterminado;

2. Recolha de dados

- ✓ Entrevistas intensivas, frequentemente combinada com observação participativa;
- ✓ Qualquer tipo de dados pode ser utilizado, inclusive dados quantitativos;
- ✓ Elaborar uma teoria com a amostra;
- ✓ A análise inicial determina para onde ir e o que procurar para a próxima na recolha de dados;

3. Análise

- ✓ Constante análise comparativa;
- ✓ Relacionar os dados com ideias, e depois, relacionar ideias com ideias;

4. Memorandos

- ✓ Anotações da teorização dos dados e as suas relações;
- ✓ Rever todas as anotações até que a hipótese formulada seja consistente;

5. Classificar e Teorizar

- ✓ Classificar as anotações;
- ✓ Salientar as relações entre estas;
- ✓ Equacionar mais recolha de dados ou novas anotações;
- ✓ Criar o esboço da teoria;

6. Escrever

- ✓ Passar do esboço para a teoria final;
- ✓ Afinar os últimos detalhes da teoria;

Resumindo o processo, o investigador aborda o assunto a ser investigado sem qualquer obrigação de testar uma teoria, este tem apenas a vontade de entender uma determinada situação, fenómeno ou ocorrência. Recorrendo a vários métodos de recolha de dados, reúne um volume de informações sobre o fenómeno observado. Comparando e codificando a informação recolhida, assim como detetando padrões nessas informações, seguindo métodos detalhados de extração de sentido destas informações, o investigador termina então, tirando as suas conclusões, com algumas teorias que emergiram desta análise rigorosa e sistemática, razão pela qual a metodologia intitula-se Teoria

Fundamentada nos Dados ("*grounded*" = apoiada, fundamentada, sustentada pelos dados).

Karl Popper debateu-se com o problema do método científico, o investigador ao formular uma hipótese quando parte para a fase experimental tem tendência a "contaminar" os dados recolhidos de forma a irem ao encontro da hipótese formulada. Pela *Grounded Theory*, o investigador após a observação do fenómeno, parte para a recolha e análise de dados e só depois formula as primeiras hipóteses, que mais tarde se comprovadas pelos dados são transformadas em teorias.

6.2. ANEXO II – QUESTIONÁRIO – 1ª PARTE

O questionário está localizado no endereço: https://docs.google.com/forms/d/e/1FAIpQLSeh6y-09KxVkgDaaa6Y_qDIANG8_AXSnhZe_WPf6RldP6i25g/viewform?usp=sf_link

Questionário Para Avaliação Dos Gráficos De Opiniões Sobre Hotéis

Responda ao seguinte questionário, posicionando-se no papel quer de um gestor de um hotel, quer de um turista que procura um hotel.

*Obrigatório

1. Indique a sua idade: *

2. Indique o seu género: *

Marcar apenas uma oval.

Feminino

Masculino

3. É utilizador frequente de websites de opiniões sobre hotéis? *

Marcar apenas uma oval.

Sim

Não Após a última pergunta desta secção, passe para a pergunta 10.

4. Indique qual o que usa com maior frequência? *

5. Classifique a apresentação da avaliação dos hotéis nesse website *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	8	9	10	
Mau	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Excelente

6. Para si, a avaliação apresentada tem peso na escolha de um hotel? *

Marcar apenas uma oval.

Sim

Não

7. Quando escolhe um hotel, tem por hábito ler os comentários sobre os hotéis? *

Marcar apenas uma oval.

Sim

Não Após a última pergunta desta secção, passe para a pergunta 10.

8. Em média quantos lê?

9. Os comentários que lê têm influência na escolha de um hotel?

Marcar apenas uma oval.

Sim

Não

6.3. ANEXO III – QUESTIONÁRIO – 2ª PARTE

Elementos gráficos

Nesta secção pretende-se avaliar as representações gráficas das opiniões sobre hotéis nos 6 websites mais visitados sobre hotéis.

Gráfico 1



Gráfico 2

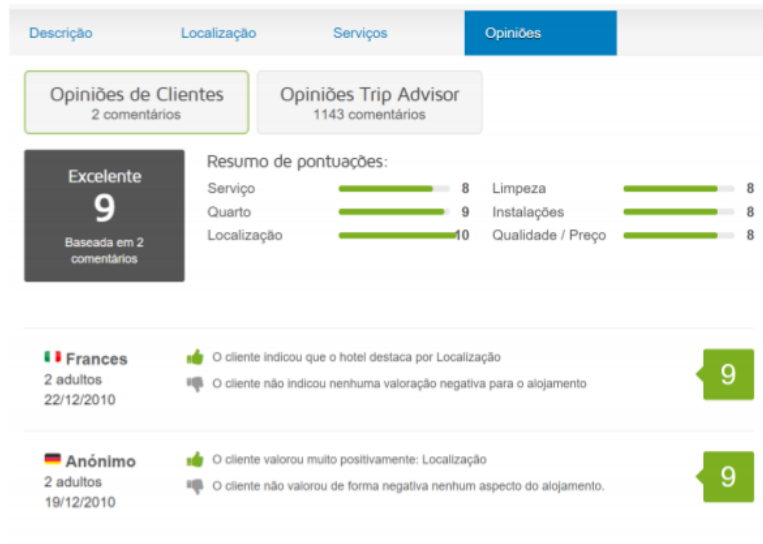


Gráfico 3



Gráfico 4

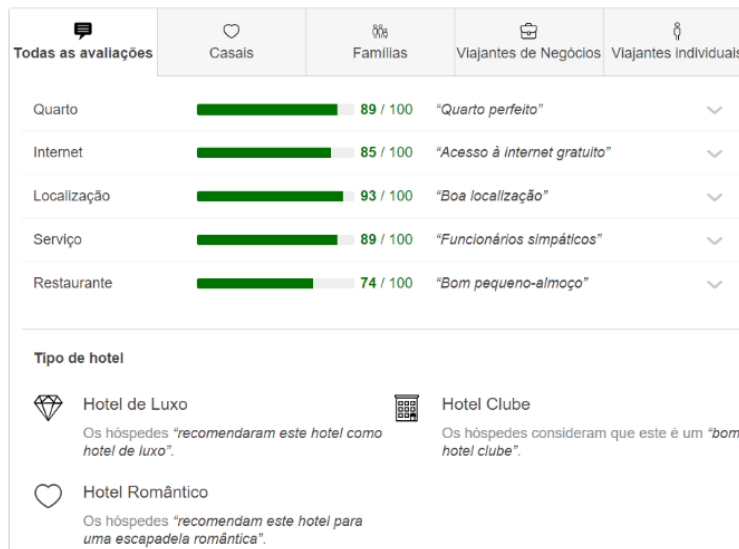
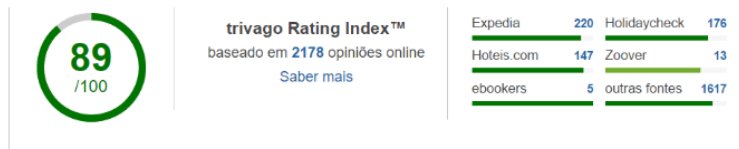


Gráfico 5

Resumo das 4 091 avaliações de hóspedes



Gráfico 6



10. Na sua opinião, em qual dos gráficos achou mais fácil consultar a informação?

Marcar apenas uma oval.

- Gráfico 1
- Gráfico 2
- Gráfico 3
- Gráfico 4
- Gráfico 5
- Gráfico 6
- Nenhum

11. E o gráfico mais difícil de consultar a informação?

Marcar apenas uma oval.

- Gráfico 1
- Gráfico 2
- Gráfico 3
- Gráfico 4
- Gráfico 5
- Gráfico 6
- Nenhum

12. Assinale em qual ou quais das avaliações é possível observar os aspectos mencionados de forma positiva ou negativa sobre o hotel.

Marcar tudo o que for aplicável.

- Gráfico 1
- Gráfico 2
- Gráfico 3
- Gráfico 4
- Gráfico 5
- Gráfico 6
- Nenhum

13. Em qual ou quais das avaliações é mais fácil obter uma síntese geral da informação

Marcar tudo o que for aplicável.

- Gráfico 1
- Gráfico 2
- Gráfico 3
- Gráfico 4
- Gráfico 5
- Gráfico 6
- Nenhum

14. Em qual ou quais gráficos a cor é o atributo visual mais relevante para mapear as informações?

Marcar tudo o que for aplicável.

- Gráfico 1
- Gráfico 2
- Gráfico 3
- Gráfico 4
- Gráfico 5
- Gráfico 6
- Nenhum

15. **Em qual ou quais gráficos a forma é o atributo visual mais relevante para mapear informações?**

Marcar tudo o que for aplicável.

- Gráfico 1
- Gráfico 2
- Gráfico 3
- Gráfico 4
- Gráfico 5
- Gráfico 6
- Nenhum

16. **Em qual ou quais gráficos é mais fácil quantificar o número de comentários sobre um determinado aspecto?**

Marcar tudo o que for aplicável.

- Gráfico 1
- Gráfico 2
- Gráfico 3
- Gráfico 4
- Gráfico 5
- Gráfico 6
- Nenhum

17. **Qual ou quais dos gráficos oferece a melhor visão a partir dos comentários sobre os diferentes aspectos do hotel?**

Marcar tudo o que for aplicável.

- Gráfico 1
- Gráfico 2
- Gráfico 3
- Gráfico 4
- Gráfico 5
- Gráfico 6
- Nenhum

18. **Considera importante a possibilidade de separação temporal das avaliações?**

Marcar apenas uma oval.

- Sim
- Não

Muito obrigado pela sua participação!

6.4. ANEXO IV – GUIÃO DE ENTREVISTA PARA O GRUPO DE DISCUSSÃO

Introdução

Este grupo de discussão tem como objetivo avaliar um primeiro esboço de uma visualização alternativa e com base nas opiniões online sobre hotéis. Esta avaliação deve de ser feita tendo em conta o ponto de vista do turista e do gestor do hotel.

Perguntas:

- 1) Face às visualizações anteriores como classifica a do esboço?
- 2) Pontos positivos do esboço face às outras visualizações?
- 3) Existem pontos a melhorar neste esboço?
- 4) Se sim, quais são os pontos a melhorar?

6.5. ANEXO V – RESULTADOS DA SESSÃO DE GRUPO DE DISCUSSÃO

Quadro de síntese das respostas dadas pelo grupo

Questões		
Face às visualizações anteriores como classifica a do esboço?		
Respostas		Contagem
Melhor		8
Igual		2
Pontos positivos do esboço face às outras visualizações?		
Respostas		Contagem
Escala temporal		10
Exposição emocional dos comentários		10
Interação com o gráfico		4
Existem pontos a melhorar neste esboço?		
Respostas		Contagem
Sim		7
Talvez		3
Se sim, quais são os pontos a melhorar?		
Respostas		Contagem
Rótulo de texto informativo de legenda no topo das cores e os sentimentos que representam		10
Uma vista detalhada onde pode ser possível filtrar por serviços ou comodidades do hotel		8
Comparar um ou mais detalhes do hotel		5

6.6. ANEXO VI – GUIÃO DE ENTREVISTA PARA O GRUPO DE DISCUSSÃO

Introdução

Este grupo de discussão tem como objetivo avaliar o protótipo de uma visualização alternativa e com base nas opiniões online sobre hotéis. Esta avaliação deve de ser feita tendo em conta o ponto de vista do turista e do gestor do hotel.

Perguntas:

O uso da emoção melhorou a visibilidade das opiniões?

O uso de uma escala temporal para distribuir as emoções, facilitou a forma como avalia um hotel?

Na sua opinião a escala de cores a relação da cor com a emoção foi bem concebida?

De 0 a 20, como avalia os detalhes das avaliações?

De 0 a 20, como classifica a sua experiência de utilização deste sistema?

De 0 a 20, como classifica a interação com o sistema?

Comparando com o sistema tradicional, a visualização dos comentários melhorou?

Deveria de ser mantida a avaliação clássica dos hotéis?

6.7. ANEXO VII – RESULTADOS DA SESSÃO DE GRUPO DE DISCUSSÃO

Resultados obtidos com o grupo de discussão aquando a avaliação do protótipo final.

Questões	Participantes									
	1	2	3	4	5	6	7	8	9	10
O uso da emoção melhorou a visibilidade das opiniões?	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim
O uso de uma escala temporal para distribuir as emoções, facilitou a forma como avalia um hotel?	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Na sua opinião a escala de cores a relação da cor com a emoção foi bem concebida?	Não	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim
De 0 a 20, como avalia os detalhes das avaliações?	15	19	16	18	17	14	19	20	18	16
De 0 a 20, como classifica a sua experiência de utilização deste sistema?	14	16	15	16	17	14	17	17	18	20
De 0 a 20, como classifica a interação com o sistema?	15	18	15	20	17	16	18	18	17	19
Comparando com o sistema tradicional, a visualização dos comentários melhorou?	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim
Deveria de ser mantida a avaliação clássica dos hotéis?	Não	Não	Não	Não	Não	Sim	Sim	Não	Não	Não