



All About Portugal - Website dedicado ao turismo externo em Portugal

RUI MANUEL DA SILVA MARTINS

Outubro de 2015

All About Portugal

Website dedicado ao turismo externo em Portugal

Rui Manuel da Silva Martins

**Dissertação para obtenção do Grau de Mestre em
Engenharia Informática, Área de Especialização em
Sistemas Gráficos e Multimédia**

Orientador: Doutor Carlos Vaz de Carvalho, DEI/ISEP

Coorientador: Alexandre Gomes, InfoPortugal

Júri:

Presidente:

[Nome do Presidente, Categoria, Escola]

Vogais:

[Nome do Vogal1, Categoria, Escola]

[Nome do Vogal2, Categoria, Escola] (até 4 vogais)

Porto, Outubro de 2015

Resumo

Grande parte do tráfego *online* tem origem em páginas de resultados de motores de pesquisa. Estes constituem hoje uma ferramenta fundamental de que os turistas se socorrem para pesquisar e filtrar a informação necessária ao planeamento das suas viagens, sendo, por isso, bastante tidos em conta pelas entidades ligadas ao turismo no momento da definição das suas estratégias de marketing.

No presente documento é descrita a investigação feita em torno do modo de funcionamento do motor de pesquisa *Google* e das métricas que utiliza para avaliação de *websites* e páginas *web*. Desta investigação resultou a implementação de um *website* de conteúdos afetos ao mercado de turismo e viagens em Portugal, focado no mercado do turismo externo – *All About Portugal*.

A implementação do website pretende provar, sustentando-se em orientações da área do *SEO*, que a propagação de conteúdos baseada unicamente nos motores de pesquisa é viável, confirmando, deste modo, a sua importância. Os dados de utilização desse mesmo *website* introduzem novos elementos que poderão servir de base a novos estudos.

Palavras-chave: Turismo, motores de pesquisa, *Google*, *SEO*

Abstract

Most of online traffic comes from search engine results pages. They are nowadays an essential tool for travelers in the moment of searching and filtering information in order to plan their trips. For this reason, tourism related entities recognize search engines as a crucial factor for marketing strategies planning.

In the following document it is described an investigation about how Google search engine works and about the metrics that it uses to evaluate websites and webpages. This investigation resulted in the development of a tourism and travel related website – All About Portugal.

The website development aims to prove that the propagation of content, based only on search engine power, is viable. That demonstrate their importance nowadays.

All About Portugal traffic data introduces new elements that can be used in further investigations.

Keywords: Tourism, search engines, Google, SEO

Agradecimentos

Esta tese representa o cumprimento de um objetivo importante para a minha construção pessoal e profissional. Como tal, reservo este espaço para agradecer a todas as pessoas com que me cruzei durante este período e que, de alguma forma, contribuíram para a sua realização.

Aos meus amigos e família por todo o apoio incondicional.

A toda a equipa da InfoPortugal envolvida neste projeto pela amabilidade, disponibilidade, oportunidade e reconhecimento.

Ao Eng. Alexandre Gomes, diretor técnico da InfoPortugal e coorientador deste projeto.

Ao Doutor Carlos Vaz de Carvalho pela orientação, disponibilidade e precioso contributo em todo o processo.

A todo o corpo docente do ISEP com que me fui cruzando ao longo dos dois anos em que frequentei este mestrado, num ciclo que culmina nesta tese.

Índice

1	Introdução.....	1
1.1	Estrutura do documento.....	2
2	Estado da arte	5
2.1	Otimização orientada aos motores de pesquisa	6
2.1.1	Página de resultados de pesquisa	7
2.1.2	Importância da posição nos <i>rankings</i> dos motores de pesquisa	10
2.1.3	Fatores de avaliação	12
2.1.4	Orientações de desenvolvimento	14
2.2	Caso particular do turismo	17
2.3	Casos de estudo.....	20
2.3.1	TripAdvisor	21
2.3.2	Visit Portugal.....	29
2.3.3	Tripomatic	36
2.3.4	YourTour	39
2.4	Conclusões.....	42
3	Análise e implementação da solução	45
3.1	Metodologia de desenvolvimento	45
3.2	Levantamento de requisitos	46
3.3	Conteúdos editoriais e ponderações ao nível das palavras-chave a usar	49
3.4	Tecnologia	56
3.4.1	Backend	57
3.4.2	Frontend	59
3.5	Sincronização de conteúdos	59
3.6	Hierarquia de páginas.....	62
3.6.1	Mapa do site.....	62
3.6.2	Estrutura de <i>URLs</i>	63
3.7	Estrutura de dados	64
3.8	Metadados <i>HTML</i>	65
3.9	Interface e principais funcionalidades	66
3.9.1	Layout	67
3.9.2	Principais funcionalidades	68
3.10	Detalhes de implementação do sistema de recomendação de planos de viagem	74
4	Resultados.....	79
4.1	Definição das métricas consideradas	80
4.2	Análise de dados das características dos visitantes	82
4.2.1	Dados gerais	82
4.2.2	Dados demográficos.....	84
4.2.3	Dados geográficos	85
4.3	Análise de dados de pesquisa.....	87

4.4	Análise de dados comportamentais.....	92
4.4.1	Páginas de entrada no portal	92
4.4.2	Performance.....	95
5	Conclusões e trabalho futuro.....	97
5.1	Conclusões.....	97
5.2	Trabalho futuro	99

Lista de Figuras

Figura 1 - Exemplo de página de resultados de uma dada pesquisa no <i>Google</i>	8
Figura 2 - Menu de navegação para conteúdos relativos à página de detalhe da cidade Lisboa (<i>TripAdvisor</i>)	26
Figura 3 - Menu de navegação global (<i>TripAdvisor</i>)	26
Figura 4 - Bloco com itens visitados recentemente (<i>TripAdvisor</i>)	28
Figura 5 - Bloco de listagem dos conteúdos destaque na página de destino (<i>TripAdvisor</i>)	28
Figura 6 - Bloco relativo às análises dos utilizadores ao ponto de interesse (<i>TripAdvisor</i>)	29
Figura 7 - URL correspondente a uma página de detalhe de um restaurante (<i>Visit Portugal</i>)	34
Figura 8 - bloco de clipping (<i>Visit Portugal</i>)	34
Figura 9 - Página do mapa interativo com rota e indicações (<i>Visit Portugal</i>)	36
Figura 10 - Menu "Places" (<i>Tripomatic</i>)	37
Figura 11 - Ecrã inicial do portal <i>Tripomatic</i>	37
Figura 12 - Passo 1 e 2 da criação de um plano de viagem (<i>Tripomatic</i>)	38
Figura 13 - Exemplo de rota proposta para um dos dias da viagem (<i>Tripomatic</i>)	39
Figura 14 - Formulário de planeamento de viagem do <i>YourTour</i>	40
Figura 15 - À esquerda visualização detalhada de um dos dias da viagem e à direita o sumário de toda a viagem (<i>YourTour</i>)	41
Figura 16 - Resultado de pesquisa <i>Google</i> correspondente à página inicial	65
Figura 17 - Exemplo de partilha da página inicial no <i>Facebook</i>	66
Figura 18 - Exemplo de partilha da página inicial no <i>Twitter</i>	66
Figura 19 - Resultado de pesquisa <i>Google</i> correspondente à página de detalhe de um ponto de interesse	66
Figura 20 - Menu global do portal	69
Figura 21 - Caixa de pesquisa da página dedicada aos restaurantes	69
Figura 22 – Pontos de interesse relacionados com o restaurante "Casa Aleixo", Porto	70
Figura 23 - Sistema de comentários da página de detalhe de pontos de interesse	71
Figura 24 - Barra de topo da página de pesquisa	72
Figura 25 - Barra lateral da página de pesquisa	72
Figura 26 - Formulário de iniciação do sistema de recomendação de planos de viagem	73
Figura 27 - Exemplo de plano de viagem	73
Figura 28 – Representação em mapa da distribuição geográfica dos acessos ao portal de visitas	85

Lista de Ilustrações

Ilustração 1 – Relação entre <i>CTR</i> e posição no <i>ranking</i> da Google [Petrescu, 2014]	11
Ilustração 2 - Nível de influência das métricas de avaliação usadas para construção dos <i>rankings</i> [Moz, 2015]	13
Ilustração 3 - Tipos e frequências dos termos e palavras-chave utilizadas no contexto de turismo e viagens [Pan et al., 2006]	20
Ilustração 4 – Diagrama do modelo de desenvolvimento ágil	46
Ilustração 5 - Diagrama de Caso de Uso do portal	47
Ilustração 6 - Diagrama de Caso de Uso relativo à pesquisa avançada	48
Ilustração 7 - Diagrama de sequência relativo ao sistema de recomendação de plano de viagem	49
Ilustração 8 – Comparativo de popularidade entre pesquisas por termos relacionados com praia.....	52
Ilustração 9 - Comparativo de popularidade entre pesquisas por termos relacionados com natureza e meios rurais	52
Ilustração 10 - Comparativo de popularidade entre pesquisas por termos relacionados com restauração.....	53
Ilustração 11 - Comparativo de popularidade entre termos relacionados com hotelaria	53
Ilustração 12 - Comparativo de popularidade entre termos relacionados com património.....	54
Ilustração 13 - Comparativo de popularidade entre termos relacionados com lazer ...	55
Ilustração 14 - Comparativo de popularidade entre termos relacionados com arte e entretenimento	55
Ilustração 15 - Comparativo de popularidade entre termos relacionados com compras	56
Ilustração 16 – Diagrama de fluxo de dados <i>MVC</i> e <i>MTV</i>	58
Ilustração 17 - Mecanismo de sincronização de dados	60
Ilustração 18 - Diagrama de arquitetura	61
Ilustração 19 - Mapa do site	62
Ilustração 20 - Estrutura de <i>URLs</i>	63
Ilustração 21 - Relação de herança multitabela	64
Ilustração 22 - Ícones e cores relativos às categorias agregadoras.....	68
Ilustração 23 - Paleta de cores principal	68
Ilustração 24 - Exemplo de rota recomendada para três dias excluindo pontos de interesse da categoria agregadora “ <i>Food & Drink</i> ”	75
Ilustração 25 - Número de sessões por semana.....	82
Ilustração 26 - Número de sessões por semana.....	83
Ilustração 27 - Percentagem de sessões por género.....	84
Ilustração 28 - Percentagem de sessões por idade	85
Ilustração 29 - Sessões, novos utilizadores e taxa de rejeição para os 10 países mais visitantes	86
Ilustração 30 - Idioma dos utilizadores	87

Ilustração 31 - número de cliques versus número de impressões	88
Ilustração 32 - <i>CTR</i> média.....	89
Ilustração 33 - Resultados de pesquisa para "quermesse restaurante lisboa"	90
Ilustração 34 - Dez países com mais impressões origem	92
Ilustração 35 - Tempo médio de carregamento de páginas.....	95

Lista de Tabelas

Tabela 1 - Tipos de pesquisa mais comuns no âmbito do turismo [Pan et al., 2006] ...	18
Tabela 2 - Metadados HTML relativo às principais páginas do portal TripAdvisor	22
Tabela 3 - Metadados <i>HTML</i> relativo às principais páginas do portal <i>Visit Portugal</i>	30
Tabela 4 - Taxonomias de pontos de interesse disponíveis	50
Tabela 5 - Métricas gerais de sessão (entre 15 de Março de 2015 e 26 de Setembro de 2015)	84
Tabela 6 - Dez termos de pesquisa que resultaram em maior número de impressões	89
Tabela 7 - Termos de pesquisa que resultaram num maior número de cliques	90
Tabela 8 - Dez páginas com maior número de impressões.....	91
Tabela 9 - Páginas de entrada no portal mais visitadas	93
Tabela 10 – Número de sessões a tipos de páginas como páginas de entrada no portal	94
Tabela 11 - Atribuição das categorias e respetivos tipos às categorias agregadoras..	103

Acrónimos e Símbolos

Lista de Acrónimos

AJAX	<i>Asynchronous JavaScript and XML</i>
API	Application Programming Interface
CSS	<i>Cascading Style Sheets</i>
CTR	<i>Click-Through Rate</i>
GIS	<i>Geographic Information System</i>
H1	<i>Heading 1</i>
H2	<i>Heading 2</i>
HTML	<i>HyperText Markup Language</i>
HTTPS	<i>Hyper Text Transfer Protocol Secure</i>
JSON	<i>JavaScript Object Notation</i>
MVC	<i>Model-View-Controller</i>
ORM	<i>Object-Relational Mapping</i>
PIB	Produto Interno Bruto
POI	Ponto de Interesse
PR	<i>PageRank</i>
SEM	<i>Search Engine Marketing</i>
SEO	<i>Search Engine Optimization</i>
URL	<i>Uniform Resource Locator</i>
WGS84	<i>World Geodetic System</i>
XML	<i>Extensible Markup Language</i>

1 Introdução

“O turismo é um fenómeno social que consiste no deslocamento voluntário e temporário de indivíduos ou grupos de pessoas que, fundamentalmente por motivos de recreação, descanso, cultura ou saúde, saem do seu local de residência habitual para outro, no qual não exercem nenhuma atividade lucrativa nem remunerada, gerando múltiplas inter-relações de importância social, económica e cultural.”

Oscar de La Torre (México, 1992)

Segundo dados disponibilizados pelo *World Travel & Tourism Council* (WTTC), o sector económico relacionado com o turismo e viagens representou em 2013 cerca de 9.5% do produto interno bruto (PIB) mundial e conta com um registo onde se verifica um crescimento expressivo nos últimos anos. Este mesmo relatório, onde constam dados relativos ao impacto do sector Turismo e Viagens em Portugal para o ano 2014, conclui que a contribuição total (direta e indireta) deste sector no PIB português foi no mesmo ano de 15.6%, sendo que 67.3% desta contribuição teve origem no turismo externo – em 2013 o número de turistas estrangeiros que deram entrada em Portugal ultrapassou os 8 milhões [World Travel & Tourism Council, 2014]. O sector do turismo e viagens representa, portanto, uma porção significativa do sector financeiro e é, por isso, um apetecível alvo de investimento não só para o Estado, mas também, e sobretudo, para o sector privado.

A *Internet* alterou a forma como a informação relacionada com turismo é distribuída e como os turistas planeiam viagens [Buhalis & Law, 2008]. O portal *web* dedicado ao turismo e viagens mais visitado do mundo, *TripAdvisor*¹, divulgou um relatório que indica que as plataformas *web* dominam hoje entre os vários meios utilizados por

¹ Disponível em www.tripadvisor.com

turistas para planeamento de viagens. De acordo com os resultados apresentados, para um universo de 15595 viajantes e 19447 prestadores de serviços ligados à hotelaria 69% das pessoas preferem recorrer a plataformas web comunitárias de partilha e classificação de experiências turísticas, como o próprio *TripAdvisor* 57% preferem portais web de agências de viagens e apenas 18% recorrem aos tradicionais balcões das agências de viagem. [TripAdvisor & StrategyOne, 2013]

O trabalho descrito ao longo deste documento resulta de uma parceria com a empresa InfoPortugal. Esta detém uma das maiores bases de dados de conteúdos editoriais georreferenciados de Portugal - 149857 pontos de interesse devidamente categorizados e classificados - sendo que uma boa fatia destes pontos de interesse tem relevância no contexto da temática do turismo e viagens. Desta base de dados de conteúdos georreferenciados faz também parte uma seleção dos pontos de interesse mais relevantes para cada concelho, denominados “a não perder”, e que estão catalogados em dois níveis (grau 1 e grau 2). Esta classificação é dada através da apreciação dos colaboradores responsáveis pela manutenção da base de dados em questão.

Uma vez apresentado este moderno paradigma do sector de turismo e viagens enquadrado na *web*, surge a intenção de projetar os conteúdos referidos no mercado estrangeiro. É então proposto o desenvolvimento de um portal *web* cuja ambição é a de ser referência no consumo de conteúdos diretamente relacionados com o turismo externo em Portugal, com a premissa de que a conquista do público-alvo do portal deve ser feita através de uma boa presença nas páginas de resultados dos principais motores de pesquisa do mercado. No pretendido portal deve constar, além do diretório de pontos de interesse, uma ferramenta de recomendação de planos de viagem que seja simples do ponto de vista do utilizador.

Posto isto, constatou-se que seria necessário proceder a um estudo que permitisse aferir quais as práticas que poderiam resultar numa boa análise e indexação das páginas do *website* referido, em particular no contexto do turismo e viagens. Ao processo que conduz ao objetivo referido dá-se o nome de *SEO* (ou *Search Engine Optimization*) e a sua importância nos dias de hoje está bem patente na dimensão da indústria que se desenvolveu em torno dele.

Esta tese envolve duas áreas de interesse para o autor. Por um lado, o desenvolvimento de plataformas *web*, área em que o autor se inicia profissionalmente, por outro a construção e desenho de interfaces que é um tema que sempre lhe despertou o maior interesse.

1.1 Estrutura do documento

Este documento está dividido em cinco grandes capítulos.

O primeiro, e presente, capítulo faz uma introdução ao contexto em que se insere o tema de que trata esta tese e às motivações presentes na conceção deste projeto.

No segundo capítulo é feita uma descrição de todo o estudo feito em torno dos motores de pesquisa, da forma como estes avaliam as páginas *web* e quais as orientações de desenvolvimento que contribuem para uma presença eficaz de uma dada página *web* em pesquisas no *Google*. É este o motor de pesquisa em foco neste capítulo. Por fim é feita uma análise comparativa entre quatro plataformas *online* que se enquadram no mesmo contexto que o *website* desenvolvido no âmbito desta tese. Este *website* está dividido por duas vertentes: por um lado o desenvolvimento de um diretório de pontos de interesse, e por outro o desenvolvimento de uma ferramenta automatizada de recomendação de planos de viagem. Finalmente analisaram-se as seguintes plataformas, escolhidas tendo em conta as duas vertentes atrás referidas:

- *TripAdvisor*² e *Visit Portugal*³, no âmbito da vertente “diretório de pontos de interesse”;
- *Tripomatic*⁴ e *YourTour*⁵, no âmbito da vertente “sistema de recomendação de planos de viagem”;

No terceiro capítulo são descritas as várias fases de desenvolvimento da solução apresentada neste projeto. É primeiro descrita a metodologia de desenvolvimento. Segue-se uma análise das funcionalidades que o *website* contempla assim como descritos os diferentes casos de uso. É neste capítulo também que se descrevem os detalhes de implementação ao nível da arquitetura, páginas *web* que compõem o portal e respetivos *URLs*, metadados *HTML* bem como o algoritmo que serve o sistema de recomendação de planos de viagem.

O quarto capítulo diz respeito à recolha de dados de utilização do *website All About Portugal*, este que é hoje um produto disponível ao público desde meados de março de 2015.

No quinto e último capítulo apresentam-se as conclusões a retirar da investigação realizada bem como dos dados e conclusões apresentados no quarto capítulo.

² Disponível em www.tripadvisor.com

³ Disponível em www.visitportugal.com

⁴ Disponível em www.tripomatic.com

⁵ Disponível em www.yourtour.com

2 Estado da arte

“O utilizador quer a resposta e não milhões de páginas Web”

Google

A internet veio mudar para sempre o paradigma da comunicação e circulação de informação à escala global, e esta mudança está a ter um impacto sem precedentes na forma como as entidades comerciais e institucionais se posicionam para atingirem o seu público-alvo, também ele em processo acelerado de transformação ao nível das mentalidades e das tendências de consumo.

Entre óbvias vantagens das soluções *web* para difusão de informação estão o seu relativo baixo custo de produção e manutenção, informação personalizada, fácil cruzamento de dados e comparação entre produtos, interatividade, possibilidade de criação de comunidade alargada e disponibilidade de 24h [Wang et al., 2002].

Os motores de pesquisa tornaram-se um meio preponderante no processo da pesquisa e filtragem de informação no contexto do turismo e viagens. Constituem hoje um canal vital através do qual as entidades hoteleiras e outros prestadores de serviços ligados ao turismo chegam ao seu público-alvo [Pan, 2014].

O estudo de que trata o presente capítulo nasce da necessidade de compreender os processos de avaliação de documentos *web* usados pelos motores de pesquisa para construção dos seus *rankings*, por forma a compreender quais as melhores práticas de implementação que resultem num bom posicionamento das páginas do *website* descrito no capítulo 3.

2.1 Otimização orientada aos motores de pesquisa

Um relatório de 2013, disponibilizado pela *Shareaholic*⁶, que reúne dados relativos a uma rede de mais de 200 mil portais *web* recolhidos ao longo de mais de 13 meses, e que atingiu mais de 250 milhões de visitas únicas por mês, revela que mais de 40% do tráfego *online* tem origem em pesquisas orgânicas [Wong, 2013]. Ainda que o mesmo estudo sugira um decréscimo desta percentagem e um aumento de 110% no tráfego com origem em redes sociais, a verdade é que persiste uma clara distância entre os valores percentuais do tráfego oriundo de páginas de resultados de pesquisa e o de redes sociais (respetivamente 40% e 15%).

Como é de conhecimento geral, e comprovado por dados disponibilizados pelo *Alexa*⁷, o *Google* é o portal mais visitado de todo o mundo, por esta razão é nesse motor de pesquisa que se foca este capítulo.

Quando Sergey Brin e Lawrence Page criaram o primeiro protótipo daquele que viria a tornar-se no maior motor de pesquisa e uma das maiores empresas do mundo, este tinha como objetivo fazer uso da rede de ligações entre documentos *web* por forma a indexá-los e perceber qual a importância de cada um. A este grau de importância deram o nome de *PageRank*.

O cálculo deste atributo dos documentos *web* pode ser explicado pela equação seguinte, e que está presente no documento original que descreve o estudo académico que deu origem a este motor de pesquisa [Brin & Page, 1998]. O *PageRank* da página *A* é dado por:

- $T1$ e Tn - páginas que apontam para a página *A*;
- $PR(T1)$ ou $PR(Tn)$ – cada página tem noção do seu próprio *PageRank*;
- $C(T1)$ ou $C(Tn)$ – número de ligações que *A* contém para outras páginas;
- d – fator de amortecimento da influência das outras páginas. Este valor está compreendido entre 0 e 1 e é referido no documento que lhe é geralmente atribuído o valor 0.85.

$$PR(A) = (1-d) + d(PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn)) \quad (1)$$

Sendo este um valor de probabilidade distribuída entre todos os nodos do grafo que compõem a *web*, o valor do somatório de todos eles é 1. Compreende-se, a partir deste algoritmo, que o *PageRank* de uma determinada página *web* é definido pelo número de páginas que a referenciam.

⁶ Empresa detentora de uma plataforma que se centra na difusão de conteúdos *on-line* - <https://shareaholic.com/>

⁷ Serviço pertencente à *Amazon* que desde 1996 se dedica à análise estatística das visitas aos portais *web*.

Como é óbvio, a complexidade de uma pesquisa no motor *Google* ultrapassa o âmbito da comparação entre o *PageRank* de cada página. A construção dos *rankings* para uma dada pesquisa compreende outros atributos que não este, atributos estes discutidos no capítulo 2.1.3.

A tarefa de percorrer o grafo de documentos *web* é realizada por um pedaço de *software* a que se dá o nome de *crawlers*, ou motores de rastreio. Estes são responsáveis por descobrir novos nodos desse grafo e descarregar o respetivo conteúdo para os servidores do motor de pesquisa correspondente. São também estes os responsáveis por detetar ligações inativas ou alterações às páginas *web* já visitadas. São, portanto, responsáveis pelo total entendimento de toda a estrutura de documentos presentes na *internet*. É através da informação recolhida por estes que os motores de pesquisa constroem os seus índices de páginas disponíveis na *web* para, posteriormente, analisarem o seu conteúdo.

Estes *crawlers* apenas conseguiam entender texto. Por forma a contornar esta limitação, nos últimos anos foi introduzido um novo mecanismo que possibilita capturar informação semântica dos documentos *web* – *headless browsing*⁸. Através deste mecanismo os *crawlers* dos motores de pesquisa passaram a conseguir executar tarefas como como a *construção do DOM da página*, execução de código *JavaScript*, interpretação de conteúdo multimédia (como imagens ou ficheiros PDF), inserção de conteúdos em formulários, entre outras [Mavridis & Symeonidis, 2015]. Através destes, os motores de pesquisa passam a conseguir ver e interpretar as páginas *web* tal como os utilizadores. Um exemplo da utilização deste mecanismo está patente na forma como a *Google* consegue detetar se uma determinada página está ou não otimizada para dispositivos móveis, revelando que o *Googlebot* (*crawler* da *Google*) não só executa código *JavaScript* como também interpreta *CSS*.

2.1.1 Página de resultados de pesquisa

Na página de listagem de resultados de pesquisa no *Google* os resultados são apresentados em três zonas diferentes, zonas essas que dizem respeito a dois tipos de resultados distintos.

⁸ Pode ser definido como um *browser* em interface gráfica

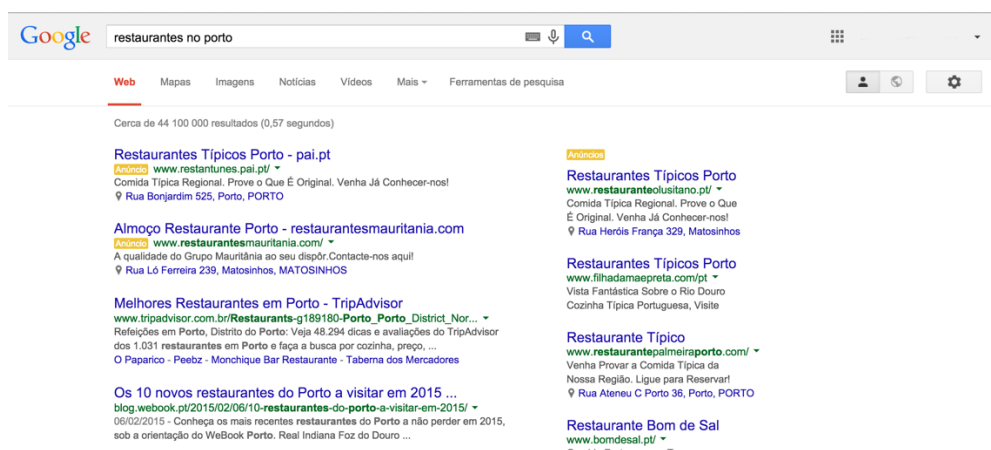


Figura 1 - Exemplo de página de resultados de uma dada pesquisa no *Google*

No caso particular da Figura 1, onde foi feita uma pesquisa por “restaurantes no porto” podem ver-se as três secções de listagem distintas referidas anteriormente: uma zona superior com dois resultados, uma imediatamente abaixo e uma coluna à direita.

Os dois resultados de topo e os da listagem da coluna da direita são resultados patrocinados, provenientes do programa *AdWords*⁹ da *Google*. Estes resultados podem ser diretamente influenciados pelo anunciante através da escolha de palavras-chave alvo e de um teto orçamental. O *Google* determina um *ranking* específico para resultados deste tipo baseado no limite de investimento monetário definido pelo anunciante, na *CTR* e na relevância do conteúdo da página a anunciar. Isto permite atrair mais tráfego para o portal e, dessa forma, potencialmente criar fidelização com novos utilizadores.

Os restantes resultados, são denominados resultados orgânicos. Conclui-se, portanto, que existem dois *rankings* distintos: o que diz respeito aos resultados patrocinados e o relativo aos resultados orgânicos. É neste último que se focam as ideias expostas neste documento.

O posicionamento de um determinado portal *web* nos resultados de pesquisa do *Google* depende daquilo que é o entendimento dado pela análise algorítmica relativamente à sua relevância no contexto de uma determinada pesquisa. Para esta análise a *Google* usa complexos algoritmos proprietários que estão em constante mudança, facto que dificulta a tarefa de compreender quais os fatores que influenciam diretamente a construção dos seus *rankings*. Consequentemente só com base nas guias disponibilizadas pela empresa, nos muitos estudos existentes e, sobretudo, na análise contínua dos dados estatísticos recolhidos de um determinado *website* se poderão criar orientações de desenvolvimento e manutenção que favoreçam um bom entendimento do motor de pesquisa em relação ao conteúdo

⁹ serviço de publicidade da *Google*: <http://www.google.pt/adwords/>

desse *website* e, deste modo, conseguir bons resultados relativamente ao posicionamento nas páginas de resultados de pesquisas.

Dada a importância que tem hoje o tráfego oriundo dos motores de pesquisa e ainda mais a necessidade de conseguir estreitar o foco de cada campanha a um público-alvo cada vez mais específico e exigente, o *SEM* (*Search Engine Marketing*) tornou-se uma indústria emergente nos países desenvolvidos [Pan et al., 2011]. Este pode, no geral, ser dividido em duas componentes essenciais: otimização dos portais *web* orientada aos motores de pesquisa (*SEO*) e a contratação de links patrocinados nas páginas de resultados dos motores de pesquisa [Pan, 2014]. Exemplo desta última é a plataforma *AdSense* da *Google* acima referida.

A contratação de resultados de pesquisa patrocinados não garante por si só o sucesso de um portal ou mesmo uma boa posição no *ranking* de resultados orgânicos. Embora a presença de resultados patrocinados influencie positivamente os orgânicos e vice-versa, o peso que os orgânicos têm nos patrocinados é maior do que o inverso [Yang & Ghose, 2010]. Além disso a maior percentagem de cliques vai mesmo para os resultados de pesquisa orgânicos [Jerath et al., 2014].

Uma das métricas mais relevantes para cálculo do retorno do investimento feito em *SEM*, tanto ao nível do *SEO* como do investimento em resultados de pesquisa patrocinados, é o *CTR* (*Click Through Rate*). Este é dado pela percentagem de utilizadores que clicam numa determinada ligação da página de resultados de pesquisa de entre todos os utilizadores que foram expostos à mesma [Pan, 2014].

SEO, ou *Search Engine Optimization*, é o nome dado ao processo de aproximação de um determinado conteúdo ao seu público-alvo no momento de uma pesquisa online. Este processo traduz-se num conjunto de guias de implementação e técnicas de análise que têm como objetivo guiar o desenho, desenvolvimento e manutenção de um portal *web* no sentido de proporcionarem melhores resultados em relação aos *rankings* orgânicos dos motores de pesquisa, o que resulta por si num tráfego mais volumoso e de maior relevância [Mavridis & Symeonidis, 2015]. Esta definição diz respeito apenas a uma primeira fase do processo, processo este que tem continuidade na análise de todos os fatores que estejam diretamente relacionados com a capacidade do portal de expandir o volume de tráfego e de fidelizar utilizadores/clientes.

A estratégia mais comum de *SEO* consiste em afinar o portal *web* por forma a aumentar o número de links orgânicos desse mesmo portal em páginas de resultados de pesquisa para um determinado conjunto de palavras-chave. Uma outra, e que corrobora as fronteiras da definição de *SEO* referidas acima, pode passar por tornar a navegação do portal mais amigável, providenciar conteúdo de qualidade, investir em publicidade e enriquecer o valor da marca através de processos da área do marketing.

Estas duas diferentes estratégias referidas no parágrafo anterior não são mutuamente exclusivas, na verdade complementam-se. Um investimento na força e

reconhecimento da marca pode influenciar indiretamente os rankings dos motores de pesquisa e vice-versa [Baye et al., 2014]. Embora este último não seja um fator diretamente relacionado com as pesquisas por palavras-chave específicas, o melhoramento deste fator traduz-se num aumento de resultados orgânicos para qualquer pesquisa, o qual, por sua vez, vai implicar um aumento de resultados em pesquisas por determinadas palavras-chave.

Naturalmente, quanto mais reconhecida a marca mais cliques irá obter, o que resulta em melhores posicionamentos nos rankings [Baye et al., 2014]. Os utilizadores tendem a clicar nas ligações que lhes são de alguma forma familiares e de referência. No entanto, é de frisar que não fazem parte do âmbito deste estudo quaisquer processos de marketing, nomeadamente onde se posicionam estratégias que envolvam o desenvolvimento de uma marca e a aproximação da mesma ao seu público-alvo por meios que não via motores de pesquisa.

O *SEO* é uma prática cuja atuação é progressiva e contínua, não podendo ser entendida apenas como uma fase prévia à implementação de determinado *website* nem ser esperados dela resultados imediatos.

2.1.2 Importância da posição nos *rankings* dos motores de pesquisa

Vários estudos sustentam a importância de uma boa presença nas páginas de resultados de pesquisa. Pan, Hembrooke e Joachims (2007) referem que os utilizadores prestam mais atenção aos primeiros resultados de uma página de pesquisa. Ghose, Ipeirotis e Li (2014) afirmam que os resultados mais bem posicionados têm taxas de *CTR* mais elevadas. Apurou-se ainda que as páginas *web* que não estejam listadas nas primeiras cinco páginas de resultados são alvo de menos 90% dos cliques que as restantes [Baye et al., 2014].

Um estudo realizado pela Moz¹⁰ corrobora esta ideia, usando como métrica a *CTR*.

¹⁰ Moz é uma empresa americana de consultoria de *SEO* que tem conquistado um expressivo valor no mercado - <https://moz.com/>

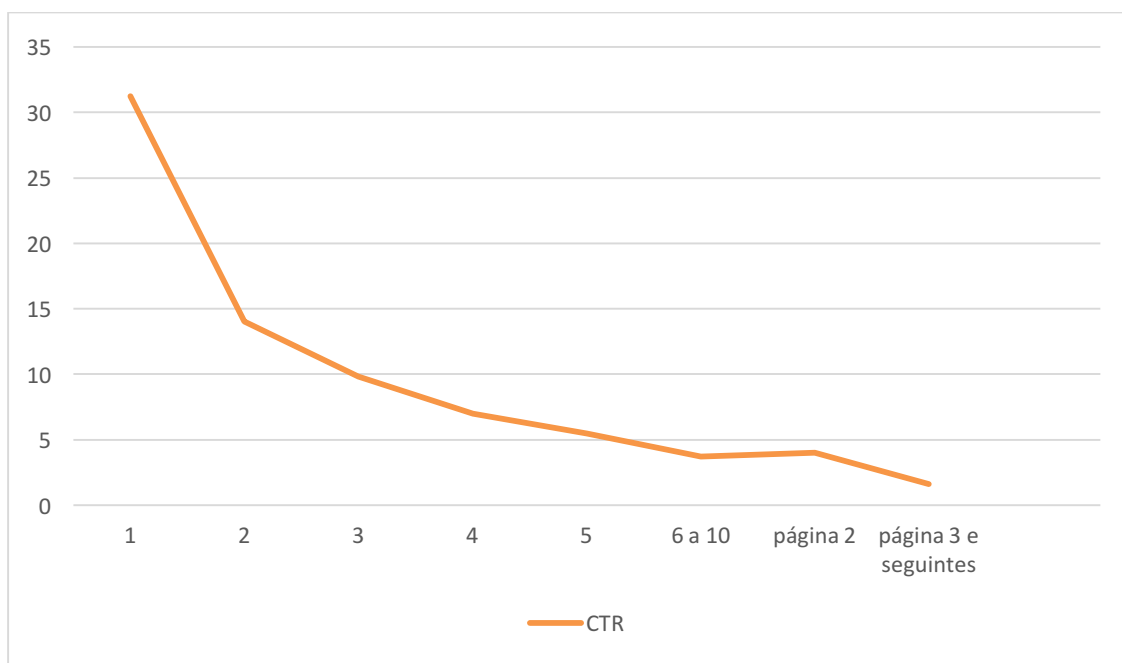


Ilustração 1 – Relação entre *CTR* e posição no *ranking* da Google [Petrescu, 2014]

A Ilustração 1 descreve a relação entre *CTR* e a posição no *ranking* de resultados orgânicos do motor de pesquisa *Google* com base em dados recolhidos da ferramenta *Google Webmaster Tools* relativamente ao portal *Advanced Web Ranking*¹¹, considerando apenas tráfego que não tenha origem em dispositivos móveis. Neste universo contam-se 7710 pesquisas, 1038335 impressões e 18662 cliques [Petrescu, 2014].

Note-se que, neste caso, a *CTR* para o resultado na primeira posição do ranking é significativamente mais expressiva do que para qualquer outro resultado, e que os primeiros cinco resultados da primeira página de pesquisa são aqueles em que os utilizadores mais clicam.

É também demonstrado no mesmo estudo que o comportamento da *CTR* ao longo das posições de *ranking* para o tráfego de origem em dispositivos móveis é bastante semelhante ao apresentado na Ilustração 1, divergindo apenas nos dados relativos à segunda página de resultados de pesquisa e seguintes onde a *CTR* é mais elevada para os dados oriundos de dispositivos móveis.

Pode, assim, concluir-se que uma posição alta no ranking é uma condição necessária para uma *CTR* elevada. No entanto, este atributo não é inteiramente suficiente. Várias páginas de ranking alto têm taxas de *CTR* baixas, o que indica que outros fatores entram em jogo – por exemplo, *snippets* que ajudam na representação do conteúdo das páginas de um determinado portal nas páginas resultados de pesquisa, o formato

¹¹ Ver: <http://www.advancedwebranking.com/>

dos *URLs*, título e descrição da página entre outros. Tudo isto deve ser tido em conta e trabalhando em função do público-alvo do portal [Pan, 2014].

2.1.3 Fatores de avaliação

A transformação e evolução da *web* tem acontecido de forma muito rápida ao longo das últimas duas décadas, desde que apareceu no contexto do consumo e que invadiu o quotidiano dos países desenvolvidos. Começou como universo de leitura, tornou-se, depois, um universo de leitura e escrita (*web 2.0*), posteriormente evoluiu para o atual universo de leitura, escrita e inferência (*web 3.0*). Esta transformação passou também por um acompanhamento dos motores de pesquisa na forma como compreendem a *Internet*, levando a que alguns dos fatores de avaliação dos documentos *web* se tenham tornado obsoletos enquanto novos fatores foram criados [Mavridis & Symeonidis, 2015].

Uma das mais importantes atualizações aos algoritmos de avaliação do motor de pesquisa *Google* foi introduzida a 23 de Fevereiro de 2011, com o nome “*Panda*”. Com esta atualização, entre outras especificações, a *Google* tornou as métricas do nível de domínio tão importantes quanto as de nível de página, e passou a considerar a qualidade do *layout* e usabilidade das páginas [Mavridis & Symeonidis, 2015]. Desde então que esta atualização tem sofrido melhoramentos de que resultaram implementações como o uso de dados relativos às plataformas *social media* para a avaliação de páginas e consequente determinação de posições de *ranking*.

Segue-se uma análise aos diferentes fatores de avaliação utilizados pela *Google* no momento da determinação da relevância dos conteúdos de um determinado *website* e respetivas páginas para uma dada pesquisa.

Anualmente a *Moz* realiza um inquérito acerca dos referidos fatores de avaliação. O alvo destes inquéritos é um conjunto de especialistas em *SEM* que expõem a sua opinião relativa à influência de cada uma das métricas apresentadas abaixo. Uma vez serem termos universais no contexto do *SEO*, serão usados os termos em inglês.

A Ilustração 2 demonstra os resultados obtidos pelo inquérito com uma escala de avaliação onde 1 significa pouco influente e 10 muito influente.

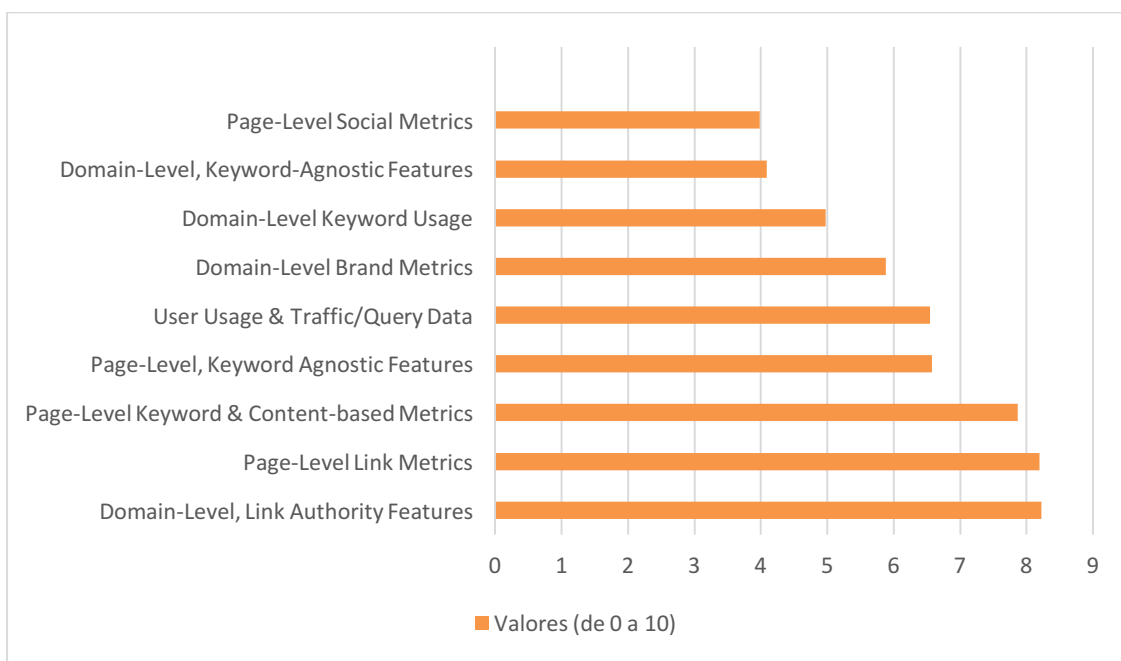


Ilustração 2 - Nível de influência das métricas de avaliação usadas para construção dos *rankings* [Moz, 2015]

Segue uma breve definição de cada um dos fatores listados na Ilustração 2:

- **Page-Level Social Metrics** – inclui a quantidade e qualidade de dados nas redes sociais relacionados com o portal *web* (“gostos”, “partilhas”, entre outros).
- **Domain-Level, Keyword-Agnostic Features** – especificações relacionadas com o domínio da página na sua forma literal. São considerados fatores como o número de caracteres do domínio, a quantidade de hífenes, ponderações ao nível do *top-level domain*¹², se o *website* contempla *layout* responsivo, ponderação ao nível do *CTR* agregado, entre outros;
- **Domain-Level, Keyword Usage** – ponderações ao nível das palavras usadas no domínio. As palavras pesquisadas pelos utilizadores podem ter correspondência com as palavras usadas no domínio do portal;
- **Domain-Level, Brand Metrics** – a este nível contam fatores como o valor e reconhecimento da marca referidos anteriormente, como o volume de pesquisas pelo nome do portal ou se este está ou não presente nas várias redes sociais;

¹² *Top-level domain* é a nomenclatura usada para definir a última parcela de um domínio (ex: .pt, .com, .org, etc). Este sufixo, por assim dizer, permite um enquadramento prévio do conteúdo do portal, como a sua origem geográfica ou mesmo o teor do seu conteúdo (ex: .pt remete para um portal de Portugal; .gov remete para um portal institucional pertencente a uma organização governamental)

- **User Usage & Traffic/Query Data** – dados diretamente relacionados com as páginas de resultados de pesquisa: tráfego de dados, taxa de rejeição, *CTR*, entre outros índices. Tanto ao nível do domínio como da página;
- **Page-level, Keyword-Agnostic Features** – a este nível são analisados parâmetros técnicos de implementação ao nível da página, como, por exemplo, a dimensão do seu conteúdo, a especificação do protocolo *Open Graph* ou de dados estruturados, desempenho de carregamento da página ou segurança de nível *HTTPS*;
- **Page-Level Keyword & Content-based Metrics** – a este nível são analisados fatores que estejam diretamente relacionados com o conteúdo da página, assim como a sua relevância e qualidade. É considerado também o conteúdo presente no código *HTML*, como *tags* título, descrição ou de imagem;
- **Page-Level Link Metrics** – baseia-se, sobretudo, na quantidade e qualidade de links de referência externos para uma dada página – estes dados servem o *PageRank*;
- **Domain-Level Link Authority Features** – o mesmo que *Page-Level Link Metrics* mas ao nível do domínio e não da página.

Rapidamente se verifica que os fatores considerados mais importantes neste inquérito, com uma distância significativa para os demais, são não só os que dizem respeito ao número de referências como também os que se referem à qualidade, autenticidade e singularidade do conteúdo.

Como é óbvio, existem também fatores que afetam negativamente o *ranking* de uma determinada página ou portal *web*. São altamente penalizadas páginas que alojem qualquer tipo de *software* malicioso, que contenham pedaços de conteúdo escondido do utilizador indevidamente ou que façam uso de esquemas de partilha ou venda de ligações. São estas óbvias más práticas que têm na sua origem a intenção de tentar enganar os motores de pesquisa [Enge et al., 2015].

2.1.4 Orientações de desenvolvimento

Com vista em melhorar os resultados das pesquisas feitas pelos seus utilizadores, os vários motores de pesquisa disponibilizam documentação e um conjunto de ferramentas que orientam os *webmasters* e *web developers* numa mais uniforme e eficiente estruturação dos seus conteúdos *web*. Conforme indicado anteriormente, o foco deste capítulo vai para o motor de pesquisa *Google*, no entanto é de notar que, apesar de existirem, entre *Google*, *Bing* e *Yahoo!*, pequenas divergências ao nível da implementação técnica, a maior parte das orientações são comuns [Mavridis & Symeonidis, 2015].

Seguem-se, além de uma recolha de orientações fornecidas pela Google [Google, 2015], guias que resultam do estudo descrito nos capítulos anteriores.

Orientações relacionadas com a estrutura de dados e conteúdo:

- A hierarquia de páginas e ligações entre elas deve ser apropriada, sendo capaz de facilitar o motor de pesquisa na tarefa de descobrir, navegar, e indexar devidamente o conteúdo do portal;
- Cada página deve ser acessível através de um link estático;
- É de extrema importância que o conteúdo das páginas seja claro, único e relevante;
- Devem ser incluídas palavras chave que estejam diretamente relacionadas com o conteúdo de que tratam as páginas, bem como ser tidas em conta as palavras que o seu público-alvo mais usará nas pesquisas que lhes dizem respeito;
- É também relevante a presença de palavras-chave sinónimas ou variantes próximas;
- É prioritário o uso de texto em detrimento de imagens no caso de elementos importantes, isto porque o *Googlebot* (e de outros importantes motores de pesquisa como o *Bing*) não é capaz de reconhecer texto em imagens;
- Os *websites* devem ser desenhados e estruturados em função do utilizador e não dos motores de pesquisa.

Orientações de cariz técnico:

- Os ficheiros estáticos - como imagens, ficheiros *CSS* ou *JavaScript* – devem ser acessíveis pelos *crawlers* uma vez que o sistema de indexação da Google vê as páginas em HTML com as alterações implicadas na importação dos seus ficheiros estáticos;
- Deve ser permitida, ao *crawler*, a navegação pelo *site* sem *session IDs* ou argumentos que o rastreiem;
- Deve ser habilitado o *header HTTP* “*If-Modified-Since*” no servidor que aloja o *site* em questão. Isto permite ao servidor *web* notificar o *crawler* se o seu conteúdo sofreu alterações desde a última visita;
- Deve ser garantido suporte alargado aos diferentes dispositivos e versões de *browsers*;

- A performance do portal – por exemplo, no que se refere ao tempo de carregamento das páginas – deve ser otimizada uma vez que este é um fator relevante no que à experiência do utilizador diz respeito, tornando-se, por isso, um fator preponderante para o posicionamento das páginas no *ranking* dos motores de pesquisa;
- O ficheiro *XML* descritivo do mapa de site, que pretende ajudar o *crawler* dos motores de pesquisa a analisarem a estrutura e conteúdo do website, não deve exceder os 10mb ou as 50 000 entradas. Caso tal aconteça, é necessário dividi-lo em vários ficheiros que cumpram estas restrições de dimensão e disponibilizar um índice para os mesmos;
- O layout das páginas deve ser responsivo, isto é, deve adaptar-se às várias dimensões de ecrã dos diversos dispositivos disponíveis no mercado, de forma a que seja passível de consulta em qualquer um deles. Este é fator bastante valorizado na última atualização da Google [Makino & Phan, 2015]. De notar que o impacto do não cumprimento deste requisito afeta maioritariamente as pesquisas feitas através de dispositivos móveis, e menos as pesquisas feitas através de *desktops*;
- O *design* do portal e a usabilidade são hoje fatores importantes;
- Ainda que pouco expressiva, a especificação de dados estruturados (Schema.org) e de mecanismos de comunicação com as plataformas de *social media* (*Twitter Cards* ou *Open Graph*) começam a ganhar cada vez mais importância;
- O título da página presente no código *HTML* e o *URL* devem conter palavras-chave relevantes;
- Por forma a satisfazer o requisito do tópico anterior as páginas devem ser servidas através de um *URL* semântico. Isto ajuda também a uma melhor perceção da estrutura do portal e do conteúdo da página por parte do utilizador, o que lhe transmite maior confiança em relação ao portal *web*;
- O tempo de resposta do servidor não deve exceder os 200ms.

Orientações de cariz estratégico:

- Uma elevada taxa de rejeição – percentagem de utilizadores que terminam a sua sessão num determinado *website* sem interagirem com o mesmo – pode ser prejudicial no momento da avaliação da relevância de uma página para uma dada pesquisa. No caso de uma percentagem muito elevada de novos utilizadores o mesmo efeito pode ser esperado;
- A quantidade e qualidade (reconhecimento que têm as referências) de ligações externas que apontem para o *website* em causa é muito relevante para a

avaliação da importância do mesmo. O número de links presentes (assim como outros dados como *likes* do Facebook) nas redes sociais é também relevante, embora não tenha ainda o mesmo peso que as ligações externas a partir de domínios reconhecidos.

Fatores altamente prejudiciais:

- A presença de links não naturais, ou seja, links que não tenham sido colocados com fins editoriais ou estruturais, mas apenas com vista a tentar “enganar” o *ranking*;
- Conteúdo duplicado;
- Páginas de fraco ou pouco conteúdo;
- Textos descritivos de ligações sobre-otimizados ou excessivo número de palavras-chave;
- Publicidade em demasia ou que impeça uma navegação natural;
- Páginas com metadados *HTML* relativos ao título ou descrição duplicados;
- Links não existentes ou erros de *HTML* são altamente prejudiciais.

2.2 Caso particular do turismo

A prática de otimização de um determinado portal para os motores de pesquisa compreende diversos âmbitos. É importante compreender como os utilizadores navegam e, sobretudo, como pesquisam dentro do contexto abrangido pelo o âmbito do *website* que se pretende desenvolver, isto é, quais as palavras-chave e termos mais utilizados, quais os fluxos de navegação que os levam a determinado tipo de páginas, e muitas outras questões que podem ser colocadas na hora da planificação e manutenção de um *website*.

Os processos de pesquisa online são divididos, maioritariamente, em três grupos:

- Transacional: o utilizador usa o motor de pesquisa com o objetivo de executar uma operação - por exemplo, uma pesquisa por “bilhetes avião Porto - Lisboa” sugere que o utilizador pretende comprar bilhetes para uma passagem aérea da companhia portuguesa TAP;
- Informacional: o utilizador procura informação acerca de um determinado assunto - por exemplo, quando o utilizador pesquisa por “História de Lisboa”;

- De navegação: o utilizador pesquisa por um portal web específico - por exemplo, pesquisa por “trip advisor” com vista a navegar para www.tripadvisor.com;

No caso particular do turismo na web existem estudos com recolhas de algumas orientações de referência que sustentam o desenvolvimento do projeto. Este capítulo, e o seguinte, focam-se naquilo que pode ajudar à estruturação do volume de dados disponível para o desenvolvimento e na sua possível representação em *frontend*.

Um cruzamento de dados dos estudos “Real Users, Real Trips, and Real Queries: An Analysis of Destination Search on a Search Engine” e “Real life, real users, and real needs: a study and analysis of user queries on the web” leva a conclusões interessantes acerca do comportamento dos utilizadores no âmbito da pesquisa online em torno do turismo e do planeamento de viagens:

1. No âmbito do turismo ou planeamento de viagens os utilizadores usam em média 3 termos para a pesquisa [Pan et al., 2006] enquanto que nas pesquisas de âmbito geral a média é de 2.6 [Jansen et al., 2000];
2. Na mesma sessão de pesquisa um utilizador que esteja à procura de conteúdos relacionados com turismo ou planeamento de viagem procede em média a 3 pesquisas [Pan et al., 2006] enquanto numa pesquisa de âmbito geral o utilizador procede em média a 2.3 pesquisas por sessão [Jansen et al., 2000];
3. O número de páginas de resultados de pesquisa visitadas em média por um utilizador no âmbito do turismo ou planeamento de viagens é de 2.1 páginas por pesquisa [Pan et al., 2006] enquanto que em pesquisas de âmbito geral esta média é de 1.7 [Jansen et al., 2000];

Estes três últimos tópicos permitem concluir que o processo de pesquisa no âmbito do turismo é mais extenso e detalhado do que no de um de âmbito geral, tem mais pesquisas por sessão e utiliza mais termos para essas mesmas pesquisas.

Tabela 1 - Tipos de pesquisa mais comuns no âmbito do turismo [Pan et al., 2006]

Tipo de pesquisa	Frequência	Percentagem
Cidade	188	10.5
Cidade + estado	135	7.6
Atração específica	99	5.5
Nome de cadeia de hotéis	57	3.2
País	49	2.7
Estado	49	2.7
Cidade + Nome de cadeia de hotéis	42	2.3
Região	36	2.0
Tipo de atividade	34	1.9

Cidade + hotel	31	1.7
Tipo de atividade + nome de cidade	31	1.7
Atividade específica	27	1.5
Atividade específica + cidade	25	1.4
Cidade + tipo de hotel	23	1.3
País + cidade	22	1.2
Tipo de atração + cidade	22	1.2
Cultura + estado	21	1.2
Rede de transportes	21	1.2
Tipo de atividade + estado	18	1.0
Estado + viagem	17	1.0
Cidade + viagem	15	0.8
Atração específica + estado	15	0.8
País + viagem	14	0.8
Cidade + atração específica	14	0.8
Hotel específico	14	0.8
Atração específica + cidade	14	0.8
País + atração específica	14	0.8
Tipo de atração + estado	14	0.8

Segundo [Pan et al., 2006], e como se pode ver na Tabela 1 - Tipos de pesquisa mais comuns no âmbito do turismo , a pesquisa mais frequente por utilizadores que estejam a planear uma viagem é por nome do destino (ex: Lisboa), sendo que, de um universo de 1788 pesquisas analisadas pelo estudo, 18.1% do total de palavras-chave são relativas ao nome da cidade destino, seguindo-se, por ordem de importância, o nome de uma atração específica ou cadeia de hotéis. Observa-se também que as pesquisas por pontos de interesse são quase sempre combinadas com uma localização.

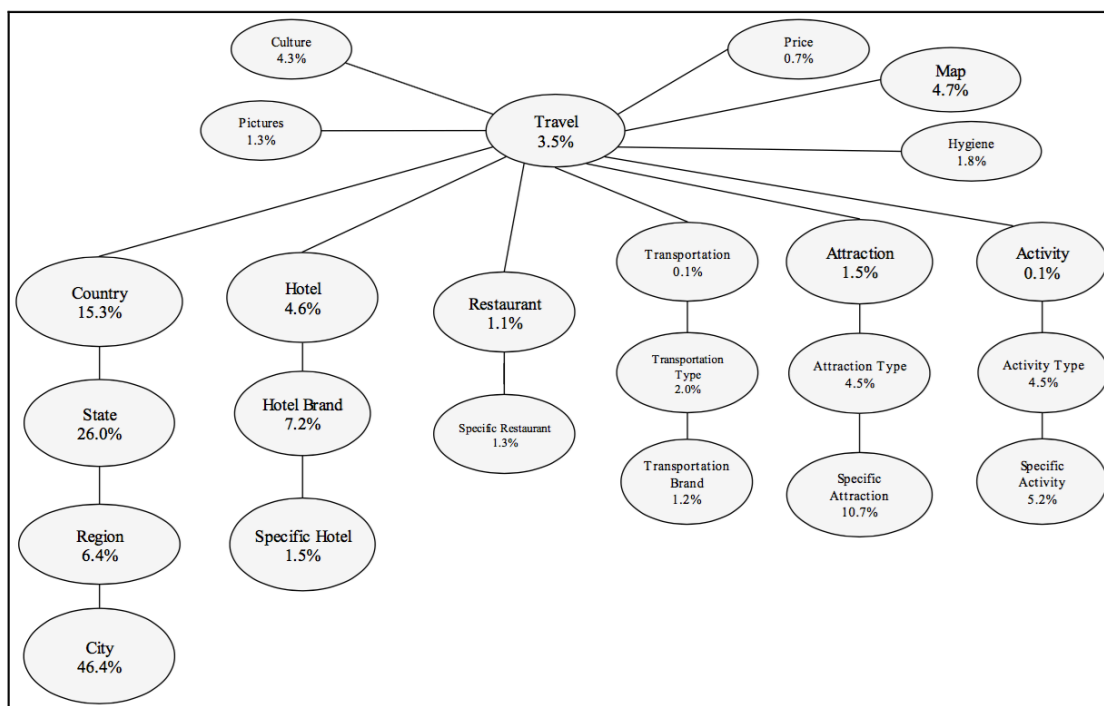


Ilustração 3 - Tipos e frequências dos termos e palavras-chave utilizadas no contexto de turismo e viagens [Pan et al., 2006]

2.3 Casos de estudo

Ao longo deste capítulo são analisados quatro casos distintos, escolhidos em função do seu enquadramento temático relativo às funcionalidades do portal *web* desenvolvido como consequência do estudo descrito até este capítulo.

No contexto do diretório de pontos de interesse afetos ao turismo de Portugal inteiro, com a particularidade de terem especial atenção o mercado de turistas estrangeiros, foram encontrados vários casos passíveis de análise, entre os quais estão o *TripAdvisor*¹³, *Visit Portugal*¹⁴, *Lonely Planet*¹⁵ e *LifeCooler*¹⁶, todos eles com boas posições em pesquisas relativas ao tema. Destes escolheram-se analisar os primeiros dois, o primeiro por ser líder mundial nesta temática e o segundo por ser o mais recente portal oficial de promoção turística em Portugal.

Escolheram-se analisar também dois portais que pudessem sustentar o desenvolvimento do sistema de recomendação de planos de viagem, funcionalidade contemplada pelo *website* desenvolvido - *Tripomatic* e *YourTour*.

¹³ <http://www.tripadvisor.com/>

¹⁴ <https://www.visitportugal.com/en/>

¹⁵ <http://www.lonelyplanet.com/>

¹⁶ <http://en.lifecooler.com/>

Nos casos *Visit Portugal* e *TripAdvisor* são analisados três tipos de páginas: a página inicial, uma página de detalhe de um destino turístico e uma página de detalhe de um ponto de interesse.

A análise a cada caso segue os tópicos apresentados na lista a seguir (apenas se a sua análise se justificar):

- Metadados *html* – títulos, descrições e cabeçalhos especificados;
- Densidade de palavras-chave – reafirmando as ideias a este respeito apresentadas no capítulo 2.1.3 é de extrema importância que as palavras-chave associadas a cada página sejam relevantes e orientadas para o público-alvo do portal. Assim, serão analisadas as palavras-chave mais frequentes de cada página dos portais que serão objeto de estudo neste capítulo por forma a compreender um padrão que possa servir o desenvolvimento do portal a que se dedica este projeto. Para recolha destas palavras-chave foi usada a ferramenta *web* gratuita para este fim disponibilizada pelo portal *SEO BOOK*¹⁷;
- Estrutura de dados – é crucial, tanto para a análise feita pelos motores de pesquisa como para a navegação do utilizador, que a estrutura de páginas seja intuitiva. Neste âmbito serão analisadas as estruturas de dados usadas pelos portais em estudo neste capítulo que digam respeito ao diretório de pontos de interesse;
- Análise de layout e funcionalidades das páginas principais;

2.3.1 TripAdvisor

O *TripAdvisor* é uma plataforma caracterizada por *online social travel networking*. Tem como grande foco o conteúdo gerado pelos utilizadores, estes comentam e classificam pontos de interesse e destinos afetos ao turismo. Este tipo de plataformas mudou a forma como os turistas planeiam as suas viagens e são hoje um importante meio de divulgação para os agentes de turismo. Talvez por isso estes têm hoje uma preocupação cada vez maior com a classificação dos seus estabelecimentos no *TripAdvisor*, pois dela depende uma fatia significativa do volume de clientes de cada um deles. Os dados disponibilizados pelo portal são demonstrativos da sua dimensão [TripAdvisor, 2015]:

- Traduzido em 48 línguas;
- Mais de 250 milhões de críticas e opiniões de utilizadores de todo o mundo;
- Mais de 5.2 milhões de pontos de interesse distribuídos por aproximadamente 123 mil destinos diferentes;

¹⁷ <http://tools.seobook.com/general/keyword-density/>;

- Uma média superior a 160 contribuições por minuto;
- Mais de 85% das perguntas colocadas nos fóruns de língua inglesa respondidas por outros membros num prazo de 24 horas;

Pela sua importância e dimensão (recorde-se que, na sua base, existe um diretório de pontos de interesse de todo o mundo), este portal é, sem dúvida, um objeto de estudo imprescindível tendo em vista o projeto a desenvolver no âmbito desta tese.

2.3.1.1 Metadados HTML

Tabela 2 - Metadados HTML relativo às principais páginas do portal TripAdvisor

	Título	Descrição	H1	H2
Página inicial¹⁸	<i>“Trip Advisor: Read Reviews, Compare Prices & Book”</i>	<i>“World's Largest Travel Site. 200 million+ unbiased traveler reviews. Search 200+ sites to find the best hotel prices.”</i>	<i>“Read Reviews, Compare Prices, Book your Perfect Trip”</i> <i>“Plan and Book Your Perfect Trip”</i>	<i>“Tips for your next vacation”</i> <i>“Follow us”</i> <i>“What travelers are talking about”</i>
Página de detalhe de uma localidade (Lisboa)¹⁹	<i>“Lisbon Tourism: Best of Lisbon, Portugal – TripAdvisor”</i>	<i>“Lisbon Tourism: TripAdvisor has 455,218 reviews of Lisbon Hotels, Attractions, and Restaurants making it your best Lisbon resource.”</i>	<i>“Lisbon Tourism: Best of Lisbon;”</i> <i>“Lisbon Portugal”</i>	<i>“Recently Reviewed Hotels around Lisbon”</i>
Página de detalhe de um POI (“Valverde Hotel”)²⁰	<i>“Valverde Hotel (Lisbon, Portugal) - Hotel Reviews – TripAdvisor”</i>	<i>“Book Valverde Hotel, Lisbon on TripAdvisor: See 175 traveler reviews, 81 candid photos, and great deals for Valverde Hotel, ranked #1 of 229 hotels in Lisbon and rated 5 of 5 at TripAdvisor.”</i>	<i>“Valverde Hotel, Lisbon”</i> <i>“Valverde Hotel”;</i>	<i>“Explore similar hotels”</i> <i>“Browse nearby”</i> <i>“Explore Lisbon”</i> <i>“You recently viewed...”</i>

Conforme se pode verificar na Tabela 2, o *TripAdvisor* é bastante rico no que diz respeito aos metadados *HTML*.

O nome da marca (*TripAdvisor*) está presente nos títulos de todas as páginas. O que lhe permite, por um lado reforçar o valor da marca, por outro, porque é muito forte e reconhecida a nível internacional, sobressair entre os vários resultados de pesquisa, chamando a si a atenção dos utilizadores.

Verifica-se o uso recorrente da palavra “*best*”, que é uma palavra muito usada em pesquisas de diferentes âmbitos. É possível também verificar o destaque dado à hotelaria na descrição da página inicial através da frase “*Search 200+ sites to find the best hotel prices*”. Isto comprova que este é um setor bastante valorizado, tal como já constatado anteriormente.

2.3.1.2 Densidade de palavras-chave

De uma extração das palavras-chave da página inicial do portal *TripAdvisor* destacam-se as seguintes¹⁸:

- “*hotels*” - ocorrências: 135 - densidade: 10.95%;
- “*vacation*” - ocorrências: 17 - densidade: 1.38%;
- “*beach*” - ocorrências: 16 - densidade: 1.3%;
- “*reviews*”, “*travel*” - ocorrências: 15 - densidade: 1.22%;
- “*tripadvisor*” - ocorrências: 13 - densidade: 1.05%;
- “*book*” - ocorrências: 11 - densidade: 0.89%;
- “*rentals*” - ocorrências: 10 - densidade: 0.81%;
- “*restaurants*” - ocorrências: 5 - densidade: 0.41%;

Analisando os dados acima listados facilmente se conclui que os conteúdos relacionados com estadia são tidos como os mais importantes: não só “*hotels*” é a palavra-chave com maior predominância, e com uma distância significativa relativamente às restantes, como estão presentes também as palavras-chave “*book*” e “*rentals*”. Já a palavra “*restaurants*” aparece com menor densidade. As restantes palavras têm um carácter genérico no contexto do turismo. Uma vez mais se constata que o nome do *website* é bastante importante, aparecem treze vezes nesta página.

¹⁸ Dados retirados da página <http://www.tripadvisor.com/>.

De seguida são analisadas as palavras-chave extraídas das páginas de detalhe de destinos e de pontos de interesse.

No que à página de detalhe de um destino turístico diz respeito foi usado o caso de Lisboa como caso de estudo visto ser um dos destinos portugueses mais procurados por turistas estrangeiros. Desta análise resultaram as seguintes palavras-chave¹⁹:

- “lisbon” - ocorrências: 101 - densidade: 6.97%;
- “hotels” - ocorrências: 87 - densidade: 6%;
- “reviews” - ocorrências: 55 - densidade: 3.8%;
- “near” - ocorrências: 23 - densidade: 1.59%;
- “travel” - ocorrências: 17 - densidade: 1.17%;
- “best” - ocorrências: 11 - densidade: 0.76%;
- “rentals” - ocorrências: 10 - densidade: 0.69%;
- “restaurants”, “things” - ocorrências: 7 - densidade: 0.48%;

A palavra “lisbon” referente ao destino de que trata a página é, naturalmente, a palavra com maior frequência de todas. Verifica-se, uma vez mais, a importância dada à procura de alojamento pelas palavras “hotels”, “rentals”, entre outras. De seguida, mas não menos importante, constata-se que a palavra “near” é usada várias vezes, o que nos leva a concluir que a questão da geolocalização é também muito importante para os utilizadores. As palavras “restaurants” e “things” estão associadas aos restantes pontos de interesse localizados, neste caso, na zona de Lisboa.

As páginas de detalhe dos diferentes tipos de pontos de interesse são muito semelhantes no que à densidade de palavras-chave diz respeito. Por essa razão apresentam-se as palavras-chave mais relevantes do hotel “Valverde Hotel”²⁰:

- “hotel” - ocorrências: 126 - densidade: 3.55%;
- “lisbon” - ocorrências: 104 - densidade: 2.93%;
- “reviews” - ocorrências: 91 - densidade: 2.57%;
- “stars” - ocorrências: 61 - densidade: 1.72%;

¹⁹ Dados retirados da página retirados da página http://www.tripadvisor.com/Tourism-g189158-Lisbon_Lisbon_District_Central_Portugal-Vacations.html.

²⁰ Dados retirados da página http://www.tripadvisor.com/Hotel_Review-g189158-d6755430-Reviews-Valverde_Hotel-Lisbon_Lisbon_District_Central_Portugal.html

- “*valverde*” - ocorrências: 42 - densidade: 1.18%;
- “*near*” - ocorrências: 26 - densidade: 0.73%;
- “*prices*” - ocorrências: 16 - densidade: 0.45%;
- “*photos*” - ocorrências: 13 - densidade: 0.37%;
- “*restaurants*” - ocorrências: 10 - densidade: 0.28%;

Neste nível de navegação o grau de especificidade das palavras-chave é muito maior do que nos casos das que estão nos níveis superiores na hierarquia de páginas do portal. A confirmá-lo temos o caso das palavras “*starts*”, “*photos*” ou “*prices*”. É evidente também a relação com conteúdos relacionados, como é o caso de “*restaurants*” e a sua relação geográfica com o conteúdo a que a página diz respeito, evidenciada pela palavra “*near*”, que satisfará pesquisas como “*restaurants near Hotel Valverde*”.

2.3.1.3 Estrutura de dados

A distribuição dos pontos de interesse é feita com base na sua localização. Se olharmos para a página de detalhe de uma cidade, são apresentados, numa primeira secção, os grandes grupos de conteúdos que dizem respeito à cidade em causa (ver Figura 2). Verifica-se, então, que os conteúdos relativos a uma dada região geográfica são organizados em sete grandes grupos: “Hotels”, “Vacation Rentals”, “Flights”, “Things to Do”, “Restaurants” e “Forum”, sendo que apenas três destes - “Hotels”, “Restaurants” e “Things to Do” - dizem respeito a pontos de interesse. Daqui conclui-se, sem margem para dúvida, que a área da hotelaria e a restauração são as que o *TripAdvisor* considera mais relevantes para o seu público-alvo.



Figura 2 - Menu de navegação para conteúdos relativos à página de detalhe da cidade Lisboa (*TripAdvisor*)

Dentro da taxonomia de hotelaria os pontos de interesse são apresentados todos ao mesmo nível, sendo possível filtrar os resultados de acordo com a classe, estilo, cadeia, localização ou comodidades disponíveis.

No caso da taxonomia dedicada à restauração, esta divide-se em vários tipos – “restaurantes”, “dessert”, “coffee & tea”, “bakeries”, “ice cream”, etc. É possível também filtrar os dados pelas suas especificidades como o tipo de cozinha, localização ou gama de preços.

A taxonomia que tem maior diversidade a seu cargo é a “*Things to Do*”, que agrupa todos os restantes pontos de interesse relativos à região geográfica em causa. É subdividida em vários tipos como monumentos, parques naturais, museus, roteiros, atividades turísticas, entre muitos outros. Existe ainda um nível posterior a este, de maior detalhe, que agrupa, por exemplo, museus de arte ou museus de ciência.

2.3.1.4 Análise de layout e funcionalidades das páginas principais

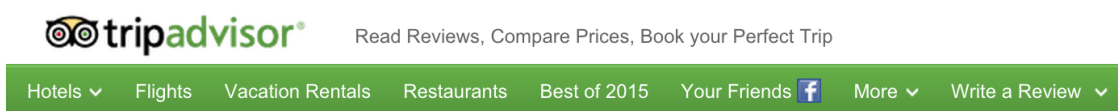


Figura 3 - Menu de navegação global (*TripAdvisor*)

Uma simples análise ao menu de navegação principal mostra-nos que, no âmbito dos pontos de interesse, as áreas da restauração e hotelaria são as de maior destaque. Isto pode levar-nos a concluir também que estas são as áreas mais procuradas pelos turistas e merecem, portanto, um lugar de acesso rápido e transversal a todas as páginas do portal - Figura 3.

Verifica-se, ao longo de todo o portal, a presença de ligações para as diversas redes sociais. Isto deve-se ao facto de *TripAdvisor* ser uma *virtual travel community* com forte componente social, mas também porque estas são ferramentas de difusão de informação em crescimento exponencial e são hoje importantíssimas na área do *marketing* e *SEO*.

Na página principal a primeira secção remete o utilizador para um formulário que lhe permita recolher informação relevante num contexto de viagem específico - é possível encontrar hotéis, restaurantes ou lugares recomendados para visitar, ou mesmo procurar voos. Imediatamente abaixo encontram-se diversas listas dos lugares mais recomendados pelos milhões de utilizadores do portal. Segue-se um conjunto de listagens e links de relevância para o utilizador e para os *crawlers* dos motores de pesquisa. Por outro lado, a página carece de ligações que reflitam a estrutura da informação de forma mais explícita. O foco principal vai para a pesquisa e para as listagens de conteúdos mais votados ou visitados, mas não facilita que o utilizador faça uma navegação estruturada pela hierarquia dos conteúdos.

Como referido anteriormente, na entrada do menu de navegação global encontram-se duas páginas de acesso rápido a conteúdos das áreas de restauração e hotelaria. A página dedicada aos restaurantes é composta apenas por um formulário de pesquisa de texto simples e por uma listagem de tipos de cozinha cujo acesso resulta nada mais do que numa pesquisa igual à anterior, mas com os resultados filtrados *à priori* com base no tipo de cozinha escolhido. No formulário de pesquisa são aceites dois tipos de conteúdo:

- Caso o utilizador pesquise por um destino, é reencaminhado para uma página de listagem de resultados de pesquisa com os devidos filtros – tipos de cozinha, preço e opções para refeições;
- Caso o utilizador pesquise pelo nome de restaurante é encaminhado para a página de detalhe do mesmo.

A página dedicada aos hotéis apresenta-se de forma similar à dos restaurantes: um formulário de pesquisa por hotel ou destino com a particularidade dos campos de data de início e de fim; e uma listagem dos destinos mais populares que funciona como filtro da pesquisa.

Ao longo da navegação pelo portal vai sendo apresentado nas diferentes páginas um bloco com os itens visitados recentemente pelo utilizador – Figura 4. Neste bloco surgem também recomendações baseadas no rasto de navegação de outros utilizadores.

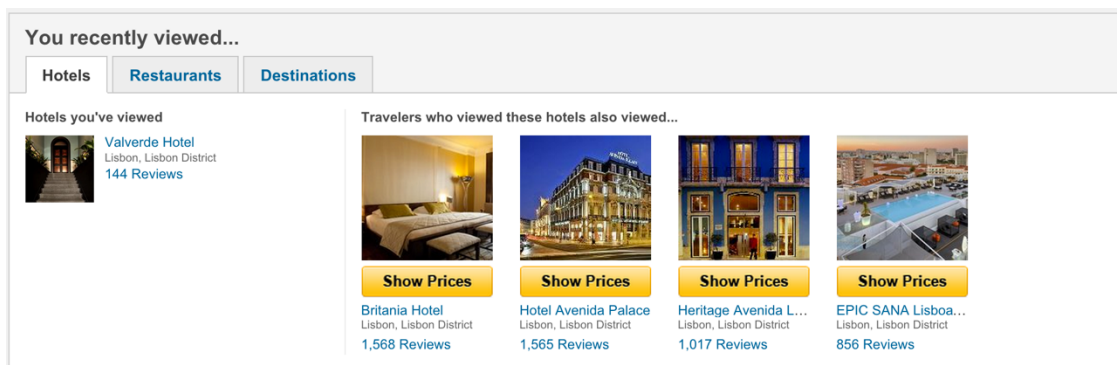


Figura 4 - Bloco com itens visitados recentemente (*TripAdvisor*)

As páginas de detalhe dos conteúdos, como pontos de interesse ou destinos, têm uma importância acrescida no mapa do site visto serem a porta de entrada de parte do tráfego do *website* vindo dos motores de pesquisa, e é por isso que, em muitos casos, é a partir destas páginas que novos utilizadores têm o seu primeiro contacto com o *website*.

Respeitando a hierarquia de páginas do portal, é analisada em primeiro lugar a página de detalhe de um destino. Para este fim foi usada a página de detalhe do destino “Lisboa”. Um dos blocos principais é o já falado bloco de acesso aos diferentes tipos de conteúdo que dizem respeito a esta localização e que está ilustrado na Figura 2. Este bloco, juntamente com um outro que contém uma galeria multimédia, que facilita o acesso rápido aos conteúdos que o utilizador procura. São estes dois blocos que se apresentam com maior destaque para o utilizador. De notar também a importância dada aos alojamentos e aos conteúdos destaque relativos aos pontos de interesse mais visitados ou mais bem classificados – Figura 5.

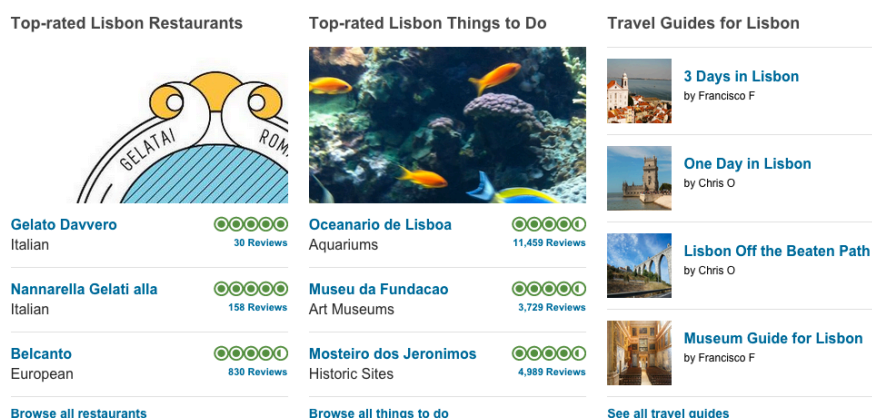


Figura 5 - Bloco de listagem dos conteúdos destaque na página de destino (*TripAdvisor*)

De notar que na página de detalhe dos pontos de interesse o primeiro destaque vai para a classificação e comentários ao ponto de interesse, em qualquer tipo - Figura 6. De forma a conduzir o utilizador para outras páginas do portal, são apresentados

outros pontos de interesse que possam de alguma forma ser interessantes dado o conteúdo da página atual. Este cruzamento de dados é feito com base em vários parâmetros:

- Os pontos de interesse mais visitados e comentados que partilhem a mesma região geográfica;
- Pontos de interesse mais próximos;
- Pontos de interesse mais próximos com os quais as ligações de *Facebook* do utilizador tenham interagido;
- Outras páginas que outros utilizadores tenham visitado a seguir a esta (isto é possível através da análise do rasto de navegação dos utilizadores);
- Pontos de interesse da mesma topologia e com características semelhantes.

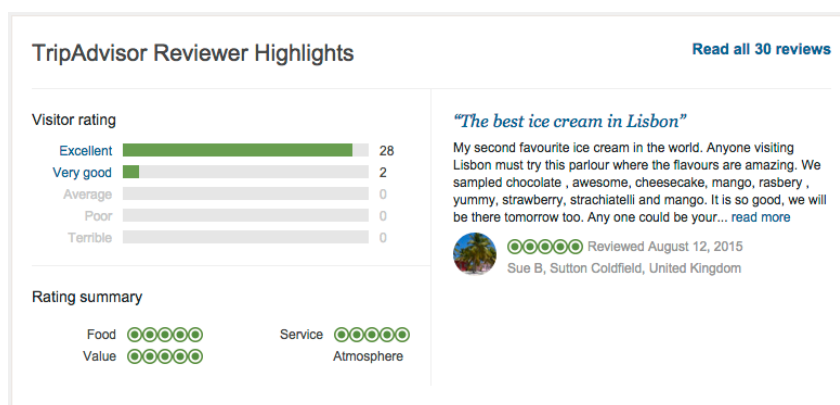


Figura 6 - Bloco relativo às análises dos utilizadores ao ponto de interesse (*TripAdvisor*)

2.3.2 Visit Portugal

O *Visit Portugal* é o portal oficial de promoção turística de Portugal criado por iniciativa do Instituto Público Turismo de Portugal e desenvolvido pela empresa InfoPortugal. Contém conteúdos traduzidos em 10 idiomas: português, inglês, espanhol, alemão, francês, italiano, neerlandês, russo, japonês e mandarim. O facto de este portal ser claramente orientado para o turismo externo em Portugal, componente igualmente central no projeto de que trata esta tese, determinou a sua escolha para análise neste documento.

Este é um portal que reúne um conjunto de informações úteis no contexto do turismo. Não se limita a ser apenas um diretório de pontos de interesse, incluindo igualmente dados sobre o país, cada região e localidades, bem como outras

informações úteis para o turista. Para efeitos de recolha de dados que ajudem ao desenvolvimento do projeto já referido será alvo de maior atenção a área que diz respeito aos pontos de interesse e sua catalogação.

2.3.2.1 Metadados *HTML*

	Título	Descrição	H1	H2
Página inicial ²¹	“www.visitportugal.com”		“www.visitportugal.com”	“Main menu” “Menu Destinos” “Pages” “Pages”
Página de detalhe de uma localidade (Lisboa) ²²	“Lisboa”	“On the right bank of the broad Tagus river estuary, the capital of Portugal graciously reclines over rolling hills. This is a spectacular geographical location”	“www.visitportugal.com”; “Lisboa”	“Main menu” “Your opinion” “Menu Destinos”
Página de detalhe de um POI (“Altis Avenida Hotel”) ²³	“Altis Avenida Hotel”	“The Altis Avenida Hotel, a five star Boutique Hotel, has an extraordinary location in the Lisbon’s centre, at Restauradores Square. The Hotel is a listed”	“www.visitportugal.com”; “Altis Avenida Hotel”	“Main menu” “Your opinion” “Menu Destinos”

Tabela 3 - Metadados *HTML* relativo às principais páginas do portal *Visit Portugal*

Verifica-se um descuido muito grande com a especificação de metadados *HTML* durante o desenvolvimento deste portal. A presente análise serve como exemplo de uma má prática relativamente ao SEO.

2.3.2.2 Densidade de palavras-chave

No que à página inicial diz respeito foram destacadas as seguintes palavras-chave de um universo de 451 únicas²¹:

- “portugal” - ocorrências: 26 - densidade: 1.88%;
- “tourism” - ocorrências: 14 - densidade: 1.01%;
- “events” - ocorrências: 12 - densidade: 0.87%;
- “suggestions” - ocorrências: 11 - densidade: 0.8%;
- “travel”, “azores”, “follow” - ocorrências: 7 - densidade: 0.51%;
- “share”, “porto”, “activities” - ocorrências: 6 - densidade: 0.43%;
- “sea” - ocorrências: 5 - densidade: 0.36%;
- “explore” - ocorrências: 4 - densidade: 0.29%;

À semelhança do que acontece com o portal *TripAdvisor*, também aqui se verifica a presença de palavras genéricas relacionadas com turismo, embora, desta vez, no contexto particular do turismo português. A presença da palavra-chave “porto” está de acordo com o crescimento exponencial de turismo estrangeiro na cidade Porto que se tem registado nos últimos anos. É importante que os conteúdos de destaque sejam atualizados com regularidade de forma a que estes estejam devidamente enquadrados com as tendências do consumo.

Para análise das páginas de detalhe de uma localidade foi usado o caso de Lisboa. Desta análise destacam-se as seguintes palavras-chave de um total de 409 palavras-chave únicas²²:

- “portugal” - ocorrências: 17 - densidade: 1.54%;
- “lisboa” - ocorrências: 10 - densidade: 0.9%;
- “tourism” - ocorrências: 8 - densidade: 0.72%;
- “share”, “follow” - ocorrências: 7 - densidade: 0.63%;
- “travel”, “transport” - ocorrências: 5 - densidade: 0.45%;
- “vídeos”, “maps”, “north” - ocorrências: 4 - densidade: 0.36%;

²¹ Dados retirados da página <https://www.visitportugal.com/en/>.

²² Dados retirados da página <https://www.visitportugal.com/en/NR/exeres/98BE771F-120C-4A83-A18B-DFC8C7907F5B>

- “alentejo”, “monuments” - ocorrências: 3 - densidade: 0.27%.

O que se pode retirar desta análise às palavras-chave da página de detalhe da localidade Lisboa é sobretudo a significância das palavras “share” e “follow”. Apesar do portal *Visit Portugal* não ser uma *virtual travel community*, como é o caso do *TripAdvisor*, isto é indicativo de que há uma aposta em *social marketing*, incentivando os utilizadores a partilharem a página com os seus contactos das diversas plataformas sociais. Ao contrário do que acontece no *TripAdvisor*, a presença de palavras relativas às especificidades do conteúdo em questão, embora aconteça, como é o caso de “monuments”, não é tão predominante quanto seria expectável. No entanto isto não quer dizer sequer que estas palavra-chave estejam relacionadas com o conteúdo Lisboa. Além disso, nesta análise aparecem com mais relevância palavras-chave relacionadas com conteúdos que podem ou não ter relação com a página em causa, como “alentejo” ou “north”, visto até que são palavras que se encontram no rodapé do portal e que, portanto, estão presentes em todas as páginas. Estas palavras serão omitidas nas próximas análises deste tipo.

Para análise de uma página de detalhe de um hotel foi usada a página relativa ao “Altis Avenida Hotel” em Lisboa²³. Desta análise destacam-se as seguintes palavras-chave de um total de 403 palavras-chave únicas:

- “hotel” - ocorrências: 19 - densidade: 1.77%;
- “altis” - ocorrências: 15 - densidade: 1.39%;
- “portugal” - ocorrências: 14 - densidade: 1.30%;
- “room” - ocorrências: 13 - densidade: 1.21%;
- “avenida” - ocorrências: 10 - densidade: 0.93%;
- “share”, “follow” - ocorrências: 7 - densidade: 0.65%;
- “accomodation” - ocorrências: 6 - densidade: 0.56%;
- “services”, “contacts” - ocorrências: 5 - densidade: 0.46%;

Nesta página encontram-se palavras especificamente relacionadas com o conteúdo e também com a tipologia, isto é, o nome e localização do hotel são as palavras-chave com maior densidade, mas seguem-se palavras como “services” ou “contacts” que são transversais a toda a tipologia da área de hotelaria.

²³ Dados retirados de <https://www.visitportugal.com/en/NR/exeres/F8EA5754-90D6-4089-8491-B8A059E4B22F>

Relativamente à página de detalhe de um restaurante foi usado o restaurante lisboeta “100 Maneiras”²⁴. De um total de 289 palavras-chave únicas destacam-se para o efeito as seguintes:

- “portugal” - ocorrências: 14 - densidade: 1.81%;
- “maneiras” - ocorrências: 7 - densidade: 0.91%;
- “lisboa” - ocorrências: 6 - densidade: 0.78%;
- “contacts” - ocorrências: 5 - densidade: 0.65%;

A página de detalhe de um restaurante não revela mais nenhum dado interessante dado aquilo que foi já dito relativamente às páginas anteriores.

2.3.2.3 Estrutura de dados

A partir do mapa do site disponibilizado em *frontend* verificamos que a estrutura do portal é composta por três grandes áreas relativas a conteúdos distintos:

- Conteúdo – nesta área está concentrada a maior parte dos conteúdos do portal, nomeadamente o diretório de pontos de interesses e localidades.
- Ferramentas – diz respeito a mecanismos de apoio ao funcionamento do portal, como o sistema de pesquisa, de registo e gestão de utilizadores, o planeador de viagem e o sistema de brochuras personalizadas dos utilizadores;
- Institucional – informação relativa ao Instituto Público dedicado ao turismo em Portugal que é também responsável pelo portal *Visit Portugal*, termos de utilização do portal e um formulário de contacto.

Como os conteúdos que mais interessam para o desenvolvimento do projeto a que esta tese diz respeito são os pontos de interesse e os destinos turísticos será neles que incidirá esta análise.

Os destinos turísticos, ou localidades, são distribuídos pelas seguintes regiões do país: Porto e Norte, Centro de Portugal, Região de Lisboa, Alentejo, Algarve, Açores e Madeira.

Os pontos de interesse estão alocados a 24 taxonomias: agências de viagens; jardins, parques e florestas; alojamento; localidades; áreas protegidas; marinas e portos; atividades; museus, monumentos e sítios; bares e discotecas; outros motivos de interesse; parques temáticos; casinos; praias; ciência e conhecimento; rent-a-car; circuitos e serviços de turismo; restaurantes e cafés; contactos no estrangeiro; termas,

²⁴ Dados retirados de <https://www.visitportugal.com/en/NR/exeres/39755A7C-84B9-48C4-904A-F7FE11BF15FD>

spas e talassoterapia; contactos úteis; terminais de transporte; transportes regulares; golfe; zoo e aquários.

À semelhança do que acontece com o *TripAdvisor*, neste portal também não são usados *URLs* semânticos em todas as páginas, tornando praticamente impossível ao utilizador conseguir decifrar o *URL* e perceber a que conteúdo ou tipo de conteúdo se refere, é considerado má prática pelas razões já referidas - Figura 7.

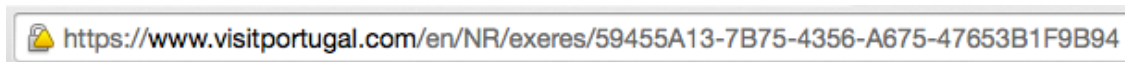


Figura 7 - *URL* correspondente a uma página de detalhe de um restaurante (*Visit Portugal*)

2.3.2.4 Análise de layout e funcionalidades das páginas principais

O propósito do *Visit Portugal* é bastante mais específico que o do *TripAdvisor*, uma vez que se apresenta sobretudo como uma ferramenta de divulgação dos produtos turísticos portugueses. Mais do que um diretório, é um cartão de visita.

A página inicial, como na maior parte dos portais, foca-se nos conteúdos destaque e na apresentação da temática.

O menu de navegação global, que dá acesso às diferentes páginas do portal de forma hierarquizada, está sempre fixo ao longo do do deslizamento vertical de qualquer página, por forma a que o utilizador consiga aceder ao mesmo sem ter que navegar para o topo das páginas. Esta é uma mais valia do ponto de vista da usabilidade e uma técnica cada vez mais usada pelos *websites* modernos.

A primeira secção a prender a atenção do utilizador num primeiro contacto é composta por um comum carrossel de imagens de alguns dos mais emblemáticos lugares do país. Apesar de não acrescentar informação relevante, este elemento cumpre muito bem o propósito de fazer uma primeira apresentação da riqueza do país. No entanto as fotografias deste carrossel não têm qualquer ligação para os respetivos conteúdos da página, ligação essa que seria uma mais valia pois permitiria ao utilizador navegar diretamente para as páginas de detalhe de destino relativas às fotografias, consultando, assim, a informação de que, à partida, não disporia sendo ele turista externo e, em princípio, desconhecedor dos lugares representados.



Figura 8 - bloco de clipping (*Visit Portugal*)

Imediatamente abaixo (Figura 8) pode ver-se uma pequena secção dedicada ao *clipping*, isto é, notícias relevantes relacionadas com o turismo em Portugal. A partir deste bloco o utilizador tem acesso a uma listagem de artigos que estejam relacionados com o produto em causa, que neste caso é o turismo em Portugal. Note-

se que esta é uma excelente forma de credibilizar e valorizar um determinado produto através de fontes que são já conhecidas pelo público geral.

As últimas secções são reservadas a quatro destaques, um mapa, sugestões e eventos. Considera-se importante a presença de conteúdos do tipo eventos, uma vez que não só são pertinentes no âmbito de uma estadia como também conferem dinâmica ao conteúdo do portal. O mesmo não acontece no caso dos pontos de interesse ou mesmo das páginas de detalhe das cidades que têm uma estrutura e conteúdo estáticos por um período de tempo alargado, implícito à sua natureza.

A página de detalhe de uma cidade é constituída pelos elementos padrão de uma página desta natureza – imagem, descrição e mapa. São cruzados dados de forma a obter eventos a decorrer, no momento ou no futuro, perto do destino de que trata a página. São ainda mostrados dois botões, “*Where to eat*” e “*Where to stay*”, que redirecionam o utilizador para a página de listagem de resultados de pesquisa filtrada por pontos de interesse da taxonomia de restauração e hotelaria, respetivamente, cuja localização seja a mesma da página em causa. No entanto, facilitaria mais a consulta se fossem listados alguns destes pontos de interesse mais relevantes sem que o utilizador tivesse que navegar para uma outra página.

Relativamente à página de detalhe de um ponto de interesse são apresentadas as especificações do conteúdo e, apesar de ela não conter um sistema de classificações, como é o caso do *TripAdvisor*, os utilizadores podem comentar nessa mesma página.

Uma vez que se trata de conteúdos georreferenciados, tanto na página de detalhe dos pontos de interesse como na das localidades é apresentado um formulário que permite ao utilizador consultar um trajeto de “como chegar”. Neste mesmo formulário é apenas possível indicar uma localidade como ponto de partida, mas seria interessante que este ponto de partida pudesse ser dado por coordenadas inseridas manualmente ou por uma morada ou ponto de interesse indicado pelo utilizador. Este formulário direciona o utilizador para a página de mapa interativo com o trajeto desenhado no mapa e as indicações listadas na barra lateral dessa mesma página, tal como pode ser visto na Figura 9.

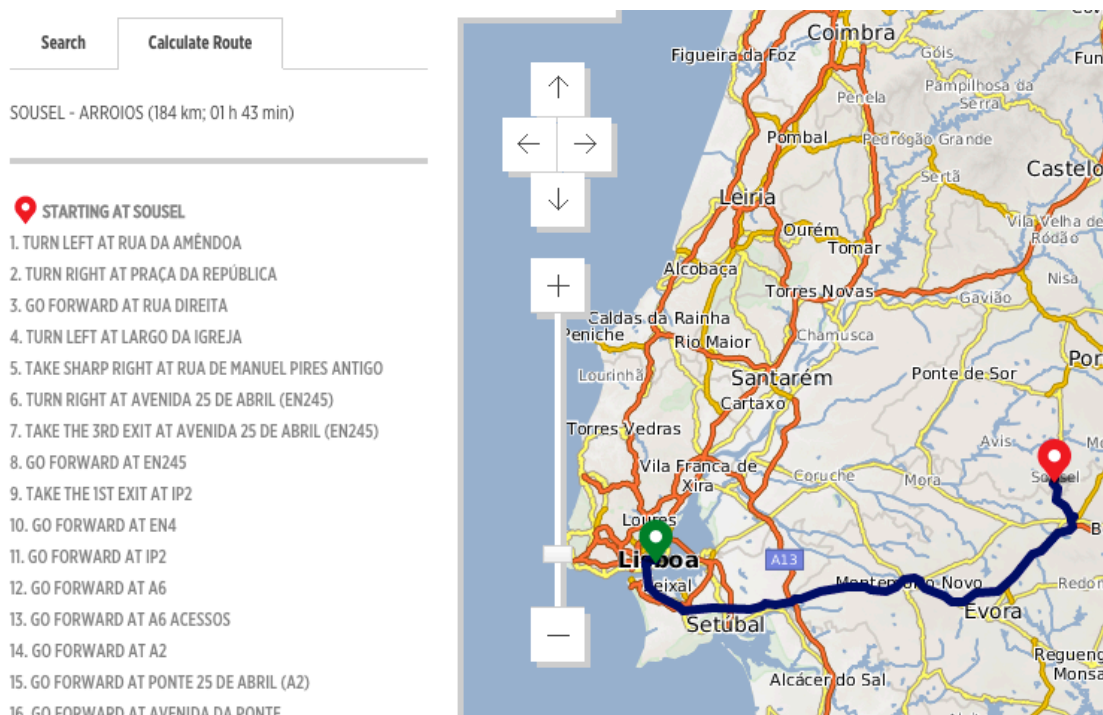


Figura 9 - Página do mapa interativo com rota e indicações (*Visit Portugal*)

2.3.3 Tripomatic

Tripomatic (<http://www.tripomatic.com/>) é uma plataforma que tem chamado a atenção junto da comunidade de viajantes por todo mundo. A empresa, sediada na República Checa, dedica-se ao desenvolvimento desta ferramenta de apoio ao planeamento de viagens, que está disponível gratuitamente tanto em formato *web* como para as plataformas *Android*, *iOS* e *Blackberry*. Apesar de existirem aplicações nativas para as plataformas *mobile* com maior quota de mercado (à exceção do *Windows Phone*) o portal *web* contempla um *layout* responsivo. Durante este capítulo será analisada apenas a versão *web*.

2.3.3.1 Estrutura de dados

O primeiro nível de distribuição dos conteúdos é a sua localização geográfica (destino turístico).

Na Figura 10 é possível ver quais as taxonomias aplicadas aos diferentes POIs da base de dados: arquitetura; arte; praia; café; entretenimento; família; comida; gratuito; golf; museu; natureza; vida noturna; restaurante; compras; paisagem; spa; desporto; tecnologia; transporte; inverno.

Verifica-se também que a secção dedicada ao alojamento merece destaque em relação aos restantes pontos de interesse, como já verificado no caso do *TripAdvisor*.

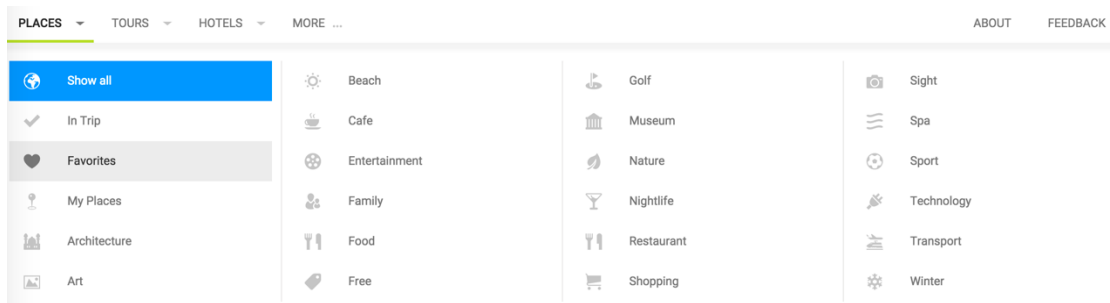


Figura 10 - Menu "Places" (Tripomatic)

2.3.3.2 Análise de layout e funcionalidades das páginas principais

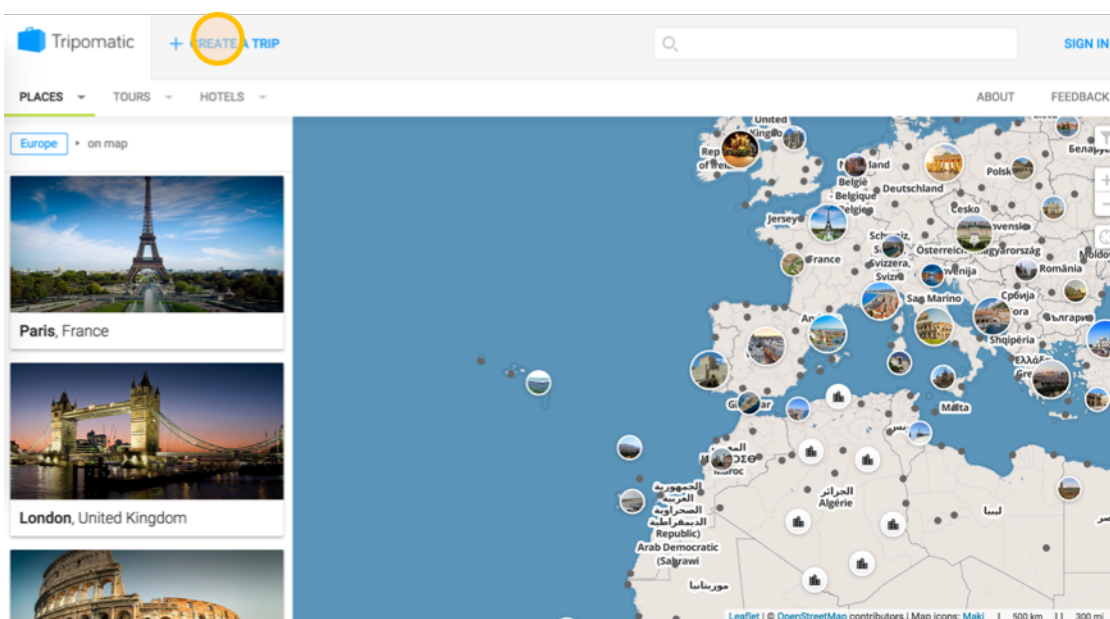


Figura 11 - Ecrã inicial do portal Tripomatic

A interface é dividida em dois painéis principais, à esquerda uma listagem dinâmica de destinos turísticos ou de pontos de interesse, à direita um mapa interativo com a representação geográfica respetiva.

O planeamento de uma viagem pressupõe o preenchimento prévio de três variáveis:

- Destino (ex: Lisboa, Porto ou Londres);
- Data de início e de fim da viagem;

O acesso a esta funcionalidade é feito através do botão de menu "+ create a trip" sinalizado na Figura 11, ação esta que remete o utilizador para o preenchimento das variáveis indicadas acima – Figura 12. O utilizador pode ainda especificar qual o local de chegada, indicando estações de comboio ou aeroportos.

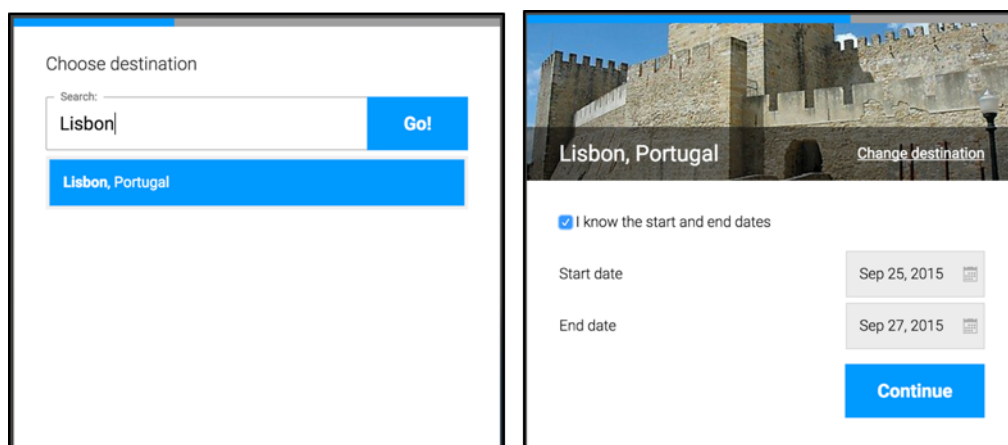


Figura 12 - Passo 1 e 2 da criação de um plano de viagem (*Tripomatic*)

Uma vez preenchido o formulário inicial, é apresentado ao utilizador um conjunto dos pontos de interesse mais populares próximos do destino por ele especificado. O detalhe de cada um destes pontos de interesse é bastante interessante, sendo constituído pelas informações comuns como contactos, descrição e listagem dos serviços disponíveis, mas também, nos casos em que se aplica, uma estimativa de tempo necessário para a visita ao ponto de interesse em causa. É com base nesta estimativa e nas distâncias entre os vários pontos e respetivo tempo médio que será necessário para as percorrer, que são preenchidos os vários dias da viagem, e são estes os dados que conferem o maior valor a esta plataforma, permitindo ao utilizador fazer uma estimativa daquilo que conseguirá fazer ou não em cada um dos dias. Apesar de serem mostrados com mais destaque os locais mais populares, a seleção dos dias de estadia com pontos a visitar é feita pelo utilizador de forma manual.

Uma vez terminado o planeamento é possível consultar a melhor rota para cada dia, alternando entre viagens de carro e a pé conforme a distância (ver Figura 13).

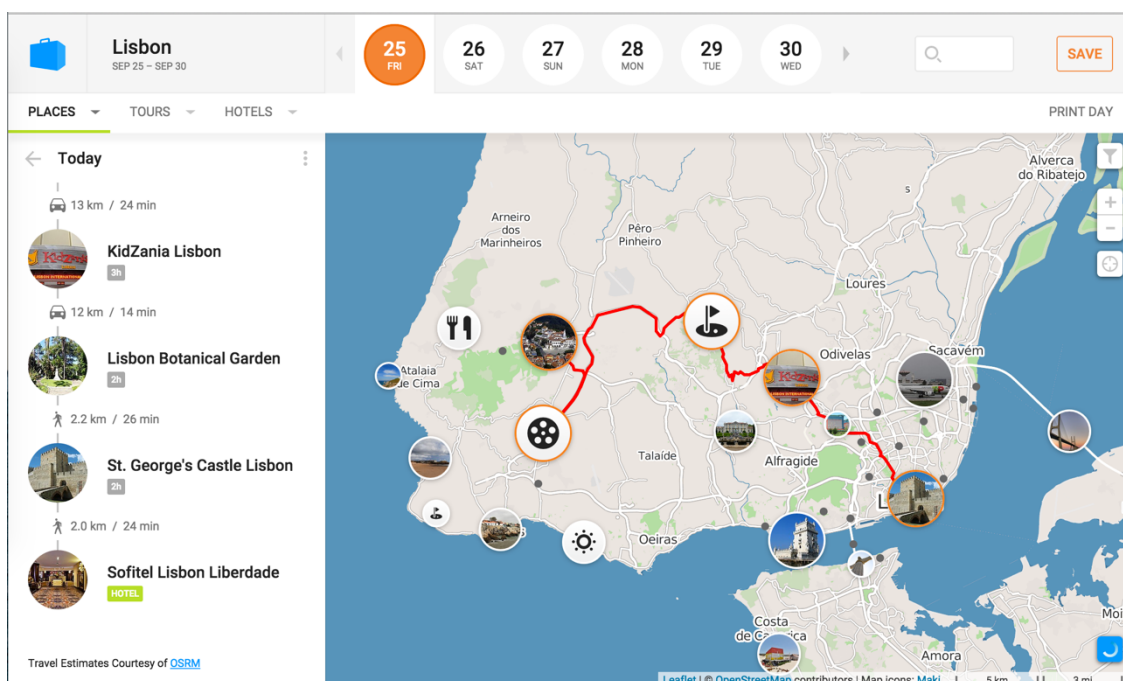


Figura 13 - Exemplo de rota proposta para um dos dias da viagem (*Tripomatic*)

2.3.4 YourTour

Embora a finalidade prática do *YourTour* seja idêntica à do *Tripomatic*, estes diferem muito na forma como o utilizador interage com a plataforma e na forma como são disponibilizados os resultados finais.

O *YourTour*, com empresa sediada na Bélgica, é uma plataforma que nasce de um projeto académico dedicado à automação de rotas personalizadas e automaticamente sugeridas através do perfil do utilizador [YourTour - Personal Tour Planner, n.d.]. O portal *web* é uma implementação ainda em fase de demonstração/protótipo, sendo que os destinos turísticos estão limitados ainda aos países Andorra, Espanha, França, Estados Unidos da América e Itália.

2.3.4.1 Análise de layout e funcionalidades das páginas principais

Plan your trip!

Destination Choose a country

Choose a region

No region selected

Themes [Add a theme to my tour](#)

Start point City [Specify](#)

Airport [Specify](#)

[End at a different place](#)

Dates **Exact dates** - [Flexible dates](#)

From to

Group Number of adults

Number of children (under age 12)

I agree to [the legal notice](#) [More options](#)

Build me a tour

Figura 14 - Formulário de planeamento de viagem do *YourTour*

O ponto inicial do planeamento de uma viagem é o preenchimento de um formulário que, por se tratar de uma plataforma baseada num mecanismo automatizado de recomendação de pontos de interesse, é ligeiramente mais extenso que o do *TripoMatic*. O utilizador tem então de responder às seguintes questões:

- País destino;
- Região do país para onde viaja, sendo possível adicionar mais do que uma;
- Filtragem opcional dos temas da viagem (ex: no planeamento de uma viagem a Barcelona são apresentados os temas “*Beaches*” e “*Wine & Spirits*”);
- O ponto de começo da viagem, que pode ser o centro de uma cidade ou um aeroporto. O ponto de fim da viagem pode diferir do ponto de começo;
- Data de início e de fim da viagem;
- Grupo de pessoas que farão a viagem, distinguidas entre crianças e adultos.

Ao contrário do que acontece no caso do *TripoMatic*, o plano de viagem é calculado automaticamente com base nos parâmetros definidos pelo utilizador, e é sobretudo, neste ponto que as plataformas divergem. São, então, sugeridas rotas para cada um dos dias com atividades e alojamento.

Uma das características interessantes desta plataforma é o cálculo de um orçamento aproximado para a viagem proposta, que tem como base os preços dos alojamentos e das atividades e deslocações entre cada um dos pontos. Tal como acontece no caso do *TripoMatic*, são tidos em conta os tempos de viagem.

O utilizador pode ver o sumário dos dias todos ou o detalhe de cada um, conforme ilustrado na Figura 15.

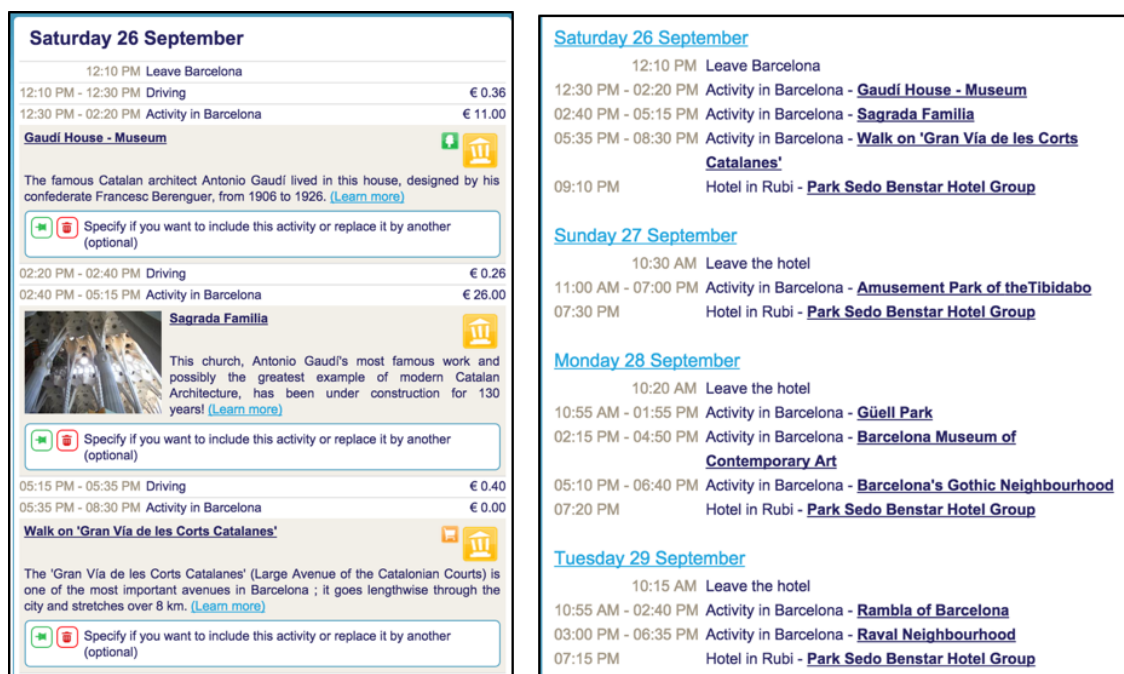


Figura 15 - À esquerda visualização detalhada de um dos dias da viagem e à direita o sumário de toda a viagem (*YourTour*)

A personalização da rota previamente calculada é possível através dos seguintes parâmetros:

- Adição e remoção de pontos de interesse específicos;
- Atribuição de níveis de preferência entre cada uma das categorias disponíveis – “Cultura e Descoberta”, “Desporto e Bem-estar”, “Compras e Restauração”, “Natureza e Ambiente”, “Divertimento e Entretenimento”;
- Parâmetros de filtragem dos hotéis como o número mínimo de classificação e serviços (exemplo: bar, piscina, spa, acesso à internet, entre outros);
- É possível limitar o orçamento por km, que condicionará, certamente, a liberdade de dispersão em termos geográficos permitida ao algoritmo para a seleção de pontos de interesse;

- São ainda disponibilizadas as seguintes opções avançadas de preferência do utilizador:
 - Serem ou não considerados mais que um ponto de alojamento;
 - Distâncias maiores ou menores entre os pontos;
 - Prioridade às características dos hotéis ou das atividades;
 - Redução custo das atividades;
 - Redução custo dos hotéis;

2.4 Conclusões

Com base no estudo descrito ao longo deste capítulo procedeu-se ao desenvolvimento de um portal *web* que dê resposta à procura de conteúdos diretamente relacionados com o turismo externo em Portugal. Esta procura pode ser representada, de forma geral, por destinos turísticos, hotelaria, restauração e atividades e atrações turísticas, e roteiros. No entanto, os conteúdos deste portal, tendo sido desenvolvido em parceria com a empresa InfoPortugal, estão limitados aos disponibilizados pela mesma.

É também objetivo deste sistema *online* demonstrar que é possível atrair utilizadores unicamente através dos motores de pesquisa. Isto tendo em conta que o projeto em causa foi tornado público sem qualquer suporte de *marketing*, estando inclusive dissociado da empresa em causa e por isso partiu, do ponto de vista do reconhecimento do público, completamente do zero.

De seguida são estabelecidas linhas de orientação para o desenho e implementação do portal referido, de acordo com as conclusões resultantes do estudo realizado acerca de *SEO* e turismo.

Através da análise dos portais *TripAdvisor* e *Visit Portugal*, assim como através da análise feita no capítulo 2.2, conclui-se que os três tipos de conteúdos mais importantes no âmbito do turismo e viagens *online* são os destinos turísticos, hotelaria e restauração. É, por isto, importante que o *layout* do portal e a sua estrutura de páginas sejam desenhados de forma a conferir destaque para estes três tipos de conteúdos.

O portal deve respeitar uma estrutura de informação intuitiva tanto para o utilizador como para os motores de pesquisa, espelhada na estrutura de páginas e *URLs*. Devem ser implementados *URLs* semânticos sendo que, como visto anteriormente, constituem vantagens tanto para o utilizador como para os motores de pesquisa. É, portanto, importante que todas as páginas do portal sejam servidas através de um

URL estático e que, para que os *crawlers* dos motores de pesquisa consigam recolher devidamente o conteúdo das mesmas.

Sendo hoje a qualidade do *layout* e usabilidade do portal fatores de avaliação utilizados pela *Google* para a construção dos seus *rankings*, e sendo estes também fatores importantes para a satisfação dos utilizadores, é crucial que o portal contemple um *layout* adaptável aos vários tamanhos de ecrã dos dispositivos mais utilizados do mercado e que seja também intuitivo e apelativo.

Por forma a tentar contornar taxas de rejeição altas nas páginas do novo portal devem ser relacionados os conteúdos com base em atributos comuns, isto visando um incentivo à navegação entre páginas do mesmo domínio.

Relativamente aos sistemas de planeamento de viagem estudados, ambos contemplam interfaces de interação simples e intuitivas, apesar de o *Tripomatic* se revelar claramente melhor neste atributo. De notar que ambos têm em conta tempos estimados tanto para as deslocações entre pontos de interesse como para a visita a esses mesmos pontos.

3 Análise e implementação da solução

O portal *web*, cujo desenvolvimento se encontra descrito neste capítulo, posiciona-se como resposta à procura cada vez maior de conteúdos diretamente relacionados com o turismo em Portugal, focado nos turistas estrangeiros. Este portal encontra-se em ambiente de produção desde a segunda quinzena de março de 2015 com o nome *All About Portugal* – acessível em <http://www.allaboutportugal.pt/>.

Pretende-se, após o estudo sobre a temática de *SEO*, turismo e análise de mercado, que a implementação deste portal resulte em resultados de sucesso ao nível da promoção de conteúdos via motores de pesquisa. Embora, como se pode ver no capítulo 2.3, o mercado seja bastante concorrido, o que faz com que a conquista de público-alvo unicamente pela via dos motores de pesquisa se torne mais difícil.

3.1 Metodologia de desenvolvimento

Sendo um dos objetivos desta tese o desenvolvimento de um portal *web* completamente funcional que, no final, deverá ser disponibilizado ao público, foi seguido um modelo de desenvolvimento do tipo “ágil” por forma a garantir o cumprimento deste objetivo. Este tipo de metodologia de desenvolvimento, que é muito usado hoje em dia, permite um planeamento adaptativo e um desenvolvimento contínuo que responde bem a alterações funcionais. Como ilustra a Ilustração 4, este processo caracteriza-se por iterações curtas de tarefas que traduzam um mínimo de requisitos possível, fazendo uso de *sprints* quinzenais com reuniões de ponto de situação que visaram a validação das funcionalidades e estrutura implementadas, assim como uma especificação progressiva de requisitos para as iterações seguintes.

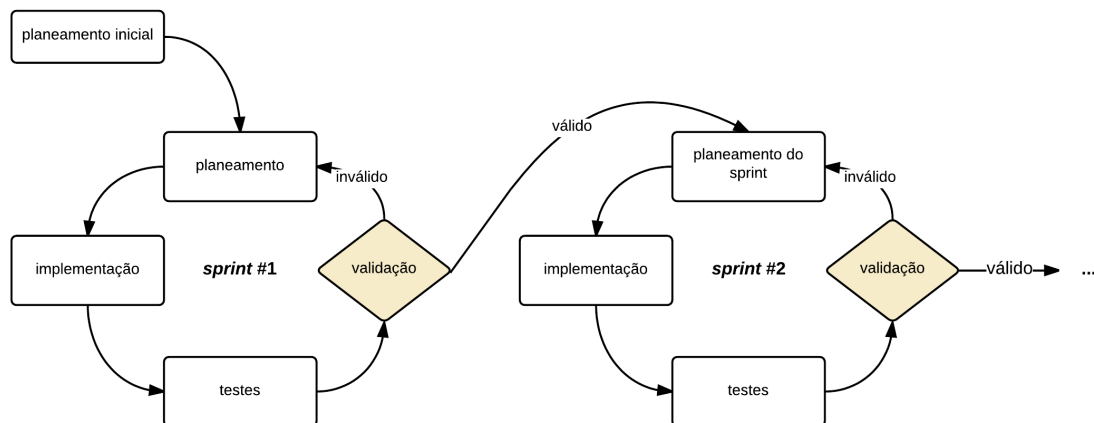


Ilustração 4 – Diagrama do modelo de desenvolvimento ágil

3.2 Levantamento de requisitos

O principal objetivo do referido portal é o de disponibilizar um diretório de pontos de interesse e destinos turísticos. Este deve ainda contemplar uma funcionalidade extra, um sistema de recomendação de planos de viagem.

Apesar de o projeto ter sido desenvolvido unicamente pelo autor deste documento, em consonância com o modelo de desenvolvimento descrito no capítulo 3.1, realizaram-se sessões de ponto de situação onde foram discutidas as funcionalidades implementadas e planeadas as iterações seguintes. Consoante o âmbito de cada uma, estas reuniões contaram com a participação do administrador da empresa, do diretor geral, do diretor técnico, da equipa de *design* e do líder da equipa de desenvolvimento de *software*.

Nas primeiras reuniões foram introduzidos, pelos membros da empresa envolvidos no projeto, as expectativas requisitos básicos iniciais para este projeto. Estes nasceram de prévias discussões em torno do tema e da experiência que foi sendo adquirida através do contacto com empresas parceiras que desenvolveram projetos com base nos conteúdos da InfoPortugal, nomeadamente o portal *iGoGo*²⁵.

O projeto arrancou de acordo com as seguintes premissas:

- O público-alvo do portal são os turistas estrangeiros que pretendem recolher informações acerca de Portugal;
- O desenvolvimento do portal deve contemplar características que visem favorecer a indexação dos conteúdos pelos motores de pesquisa;

²⁵ <http://www.igogo.pt/> - desenvolvido pela empresa 7Graus em parceria com a InfoPortugal

- O diretório deve conter uma ferramenta de pesquisa avançada que permita a navegação pelos diferentes tipos de pontos de interesse. Deve também ser possível fazer pesquisas por termos;
- Os conteúdos devem estar disponíveis em três línguas (inglês, espanhol e francês);
- Deve ser desenvolvido um layout atrativo e com suporte para dispositivos móveis;
- Inserir blocos publicitários da plataforma *Google AdSense*²⁶.

3.2.1.1 Navegação principal

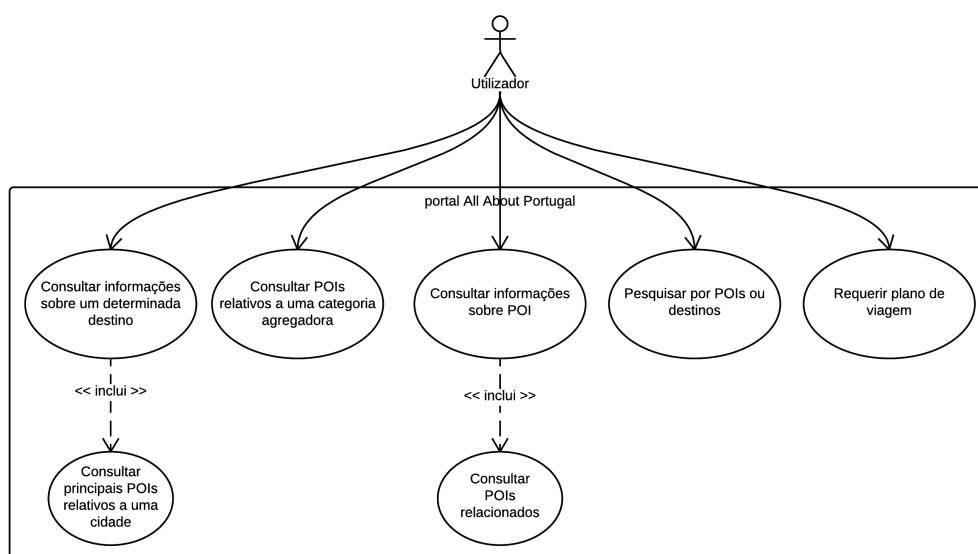


Ilustração 5 - Diagrama de Caso de Uso do portal

Tal como descrito na Ilustração 5, o portal é uma plataforma apenas de consulta que deve servir dois tipos de conteúdos diretamente relacionados – destinos turísticos e pontos de interesse.

Com vista a uma melhor organização dos dados, que favoreçam a usabilidade do portal enriquecendo as formas de acesso aos conteúdos, foi estipulado que os diferentes tipos de pontos de interesse seriam agregados em categorias de um nível hierárquico superior.

²⁶ <https://www.google.com/adsense/>

3.2.1.2 Pesquisa avançada

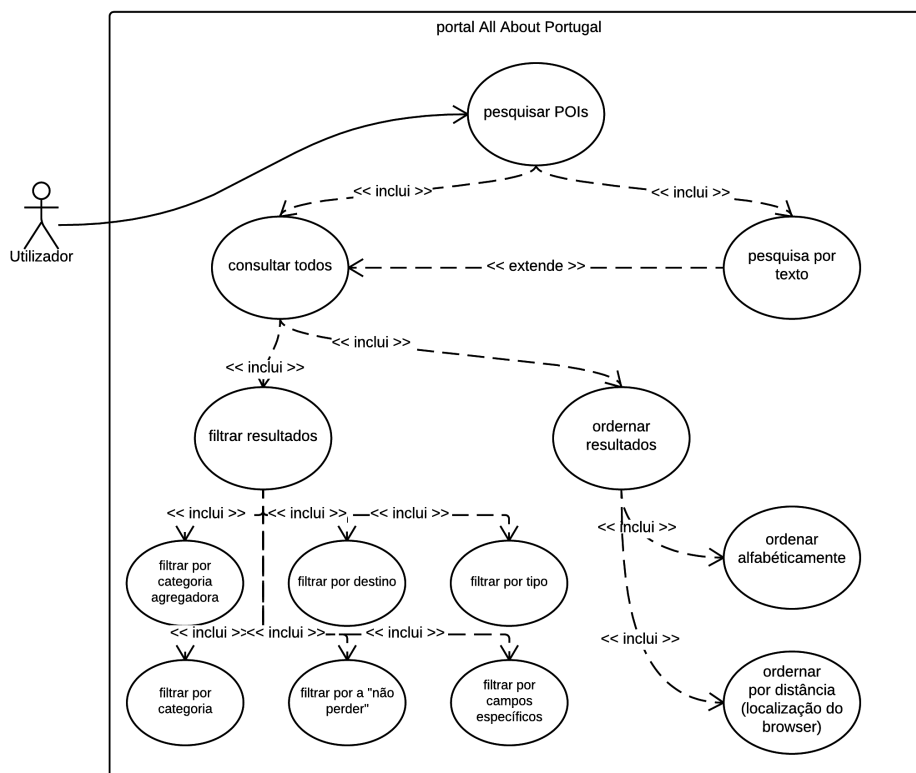


Ilustração 6 - Diagrama de Caso de Uso relativo à pesquisa avançada

A pesquisa avançada deve permitir que o utilizador filtre os resultados de acordo com as várias especificações dos pontos de interesse, que incluem:

- Categoria agregadora (ex: "Food & Drink");
- Categoria (ex: restaurantes);
- Tipo (ex: cafés típicos);
- Destino (ex: Porto);
- "A não perder";
- Campos específicos (ex: tem acesso à *Internet*);

A ferramenta tem de permitir também uma pesquisa por termos, assim como a ordenação por distância (através da localização do browser) ou alfabética.

3.2.1.3 Sistema de recomendação de plano de viagem

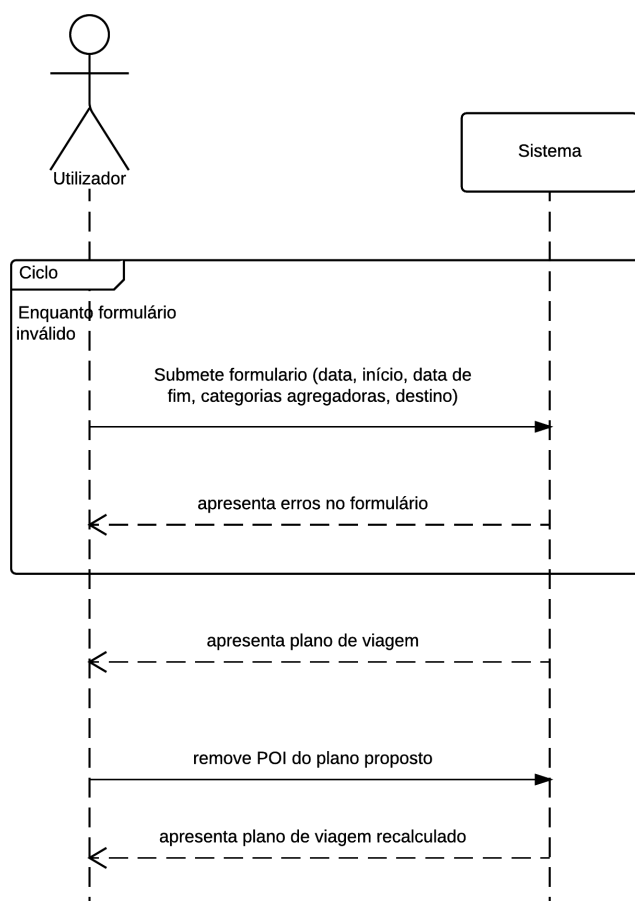


Ilustração 7 - Diagrama de sequência relativo ao sistema de recomendação de plano de viagem

Relativamente ao sistema de recomendação de plano de viagem, pretende-se que este seja simples e que exija um mínimo possível de *input* do utilizador, à semelhança do *YourTour*. Desta forma, definiu-se um conceito simples para a determinação dos planos de viagem. Ao utilizador são feitas três questões: “Para onde?”, “Quando?”, “O quê?”. As respostas do utilizador a estas três questões são usadas pelo sistema para calcular uma rota com um conjunto de pontos de interesse “a não perder” situados na zona do destino especificado.

O detalhe da implementação deste sistema é descrito mais à frente neste documento.

3.3 Conteúdos editoriais e ponderações ao nível das palavras-chave a usar

Um dos primeiros tópicos de discussão foi sobre quais os tipos de pontos de interesse da vasta base de dados da empresa que integrariam o referido portal *web*. Decidiu-se

que todos os pontos de interesse com fins utilitários (como bancos, escolas, correios, etc.) não fariam parte do conteúdo deste projeto, visto não serem da maior relevância para um turista estrangeiro. O conteúdo do portal baseia-se, portanto, em mais de 83 mil pontos de interesse e mais de 300 destinos turísticos. Na Tabela 4 estão listadas todas as categorias de pontos de interesse que constam da base de dados referida, sendo que as linhas a cinzento representam as taxonomias que não integram o portal *All About Portugal*.

Tabela 4 - Taxonomias de pontos de interesse disponíveis

Nome da categoria	Qtd.	Descrição
Aldeias	415	Aglomerados rurais, aldeias e vilas históricas.
Atividades de passeio	1072	Rotas, percursos, trilhos, etc.
Atrações turísticas	155	Jardins zoológicos, parques temáticos, etc.
Bancos	7809	Balcões bancários.
Bares e similares	4771	Bares, casas de fado, cervejarias, etc.
Bem estar	987	Termas, SPAs, ginásios, etc.
Bombas de gasolina	2957	Postos de abastecimento.
Centros comerciais	896	Centros e galerias comerciais.
Correios	1216	Postos de correio.
Desportos	6310	Campos de futebol, golfe, piscinas, etc.
Educação	14112	Escolas, universidades, etc.
Espaços culturais	3838	Bibliotecas, salas de espetáculo, museus, etc.
Grande retalho	3312	Hipermercados, <i>cash & carry</i> , etc.
Hotéis	5646	Aldeamentos, hotéis, pousadas, etc.
Jardins	6470	Jardins, miradouros, parques, etc.
Mercados e feiras	813	Mercados e feiras.
Monumentos	13117	Estátuas, muralhas, palácios, castelos, etc.
Negócios	477	Parques industriais, empresariais, etc.
Parques de campismo	270	Parques de campismo.
Parques de estacionamento	2352	Parques de estacionamento.
Praias e marinas	1108	Praias fluviais, marítimas e marinas.
Recursos naturais	821	Grutas, montanhas, áreas protegidas, etc.
Religião	18328	Templos, cemitérios, simbolismos religiosos, etc.
Resorts	2	Resorts
Restaurantes	18017	Restaurantes
Saúde	5538	Hospitais, farmácias, centros de saúde, etc.
Segurança e justiça	1985	Polícia, estabelecimentos prisionais, etc.
Transportes	2493	Estações de metro, de comboio, etc.
Turismo	3743	Agências de viagem, empresas de animação, etc.

Na base de dados do gestor de conteúdos da empresa InfoPortugal consta também uma seleção de pontos de interesse considerados mais importante, denominados “a não perder”. Estes pontos de interesse “a não perder” estão agrupados por concelho, cada um destes concelhos tem uma descrição e fotografia associados, representando assim cada um uma página do portal, páginas estas correspondentes a um dos primeiros níveis do mapa do site – destinos turísticos. Este modelo apresenta uma falha do ponto de vista editorial, que é a inexistência de páginas dedicadas a destinos

com forte componente turística que não são sedes de concelho, como é o caso de Estoril ou Vilamoura que são freguesias (não sede) dos concelhos Cascais e Loulé respetivamente. De forma a contornar este problema foi adotado neste projeto um modelo diferente que permita numa fase posterior inserir conteúdos relativos a estas localidades em falta. Para tal os concelhos passam a constar no portal como destinos turísticos e está disponível em *backoffice* uma zona para inserção de novas entradas deste tipo, sendo possível desta forma manter intacta a estrutura original da base de dados fonte e expandir a informação editorial apenas no âmbito deste *website*.

Com vista a uma melhor organização dos dados e, conseqüentemente, a um melhor e mais intuitivo acesso aos conteúdos, foram criadas categorias agregadoras das várias categorias de pontos de interesse disponíveis na base de dados do *All About Portugal*. Relativamente a este tema, e quando comparado com os casos de estudo *TripAdvisor* e *Visit Portugal*, este portal permite a navegação por um nível adicional.

Com base nos dados da Tabela 4, dividiu-se o conjunto de dados disponíveis nos 8 âmbitos seguintes:

- Praias, parques aquáticos, piscinas, SPAs, entre outros;
- Natureza e meios rurais;
- Hotelaria;
- Restauração;
- Património arquitetónico e cultural;
- Arte e entretenimento;
- Lazer e vida noturna;
- Compras;

Uma vez definido o âmbito de cada uma destas categorias agregadoras, e com o objetivo de direcionar o portal para o maior volume de público-alvo possível, foi feito um estudo acerca de quais os nomes mais apropriados para cada uma. Sendo este um portal orientado ao turismo externo, foram usados nesta pesquisa termos em inglês. Para tal foi usada a ferramenta *Google Trends*²⁷ que permite perceber a popularidade de termos ao longo do tempo, assim como obter pesquisas relacionadas com estes termos.

²⁷ <https://www.google.pt/trends/>

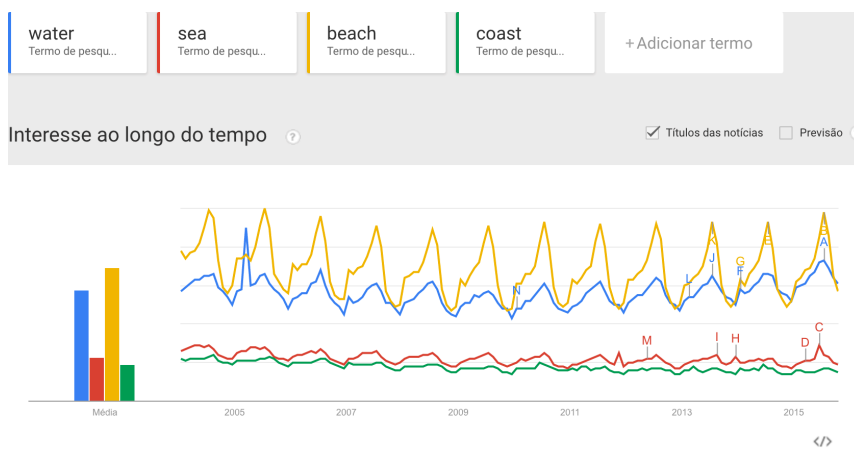


Ilustração 8 – Comparativo de popularidade entre pesquisas por termos relacionados com praia

Relativamente aos conteúdos relacionados com praia, é verifica-se na ilustração acima que o termo mais pesquisado é “*beach*”. Nas pesquisas relacionadas constata-se que é frequente a pesquisa pela junção de termos que indiquem a localização da praia – por exemplo “*miami beach*”. A presença do termo “*hotels*” nas pesquisas relacionadas com “*beach*” serve um propósito também ele relevante para o projeto, uma vez que remete para um tipo de conteúdos também ele contemplado no portal *All About Portugal*.

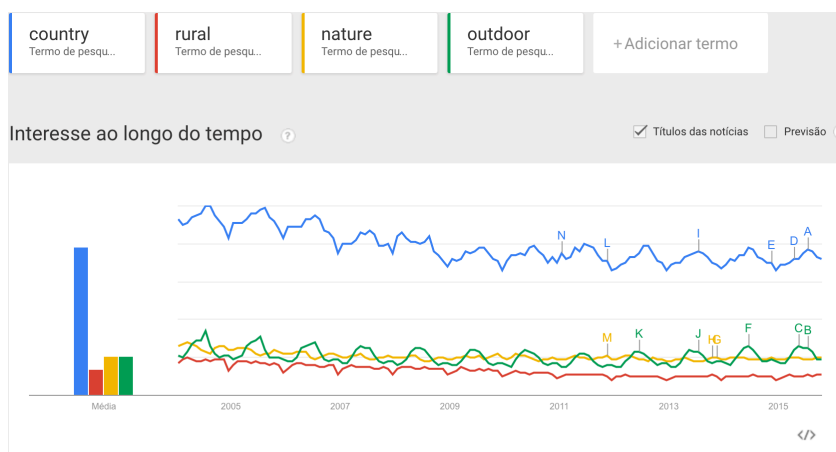


Ilustração 9 - Comparativo de popularidade entre pesquisas por termos relacionados com natureza e meios rurais

A Ilustração 9 livra-nos de quaisquer dúvidas acerca do termo a usar para os conteúdos relativos a natureza e ar livre, sendo claramente o termo “*country*” o mais frequente em pesquisas no *Google*.

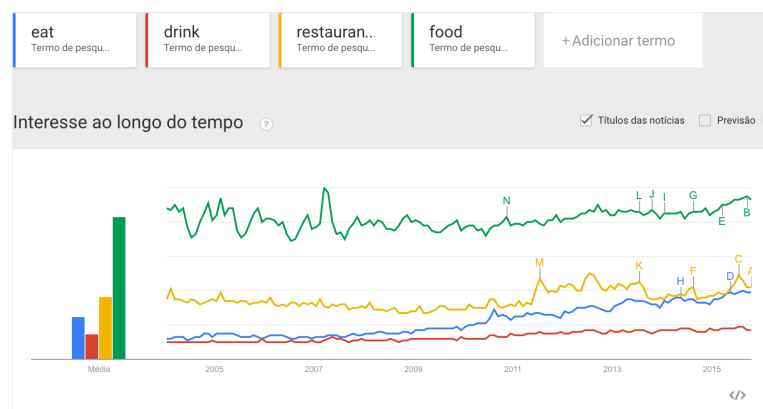


Ilustração 10 - Comparativo de popularidade entre pesquisas por termos relacionados com restauração

No que à categoria agregadora de conteúdos relacionados com restauração em geral diz respeito, a escolha recai para uma junção de dois termos bastante pesquisados – “*food*” e “*drink*” – por serem também os mais representativos do conteúdo a que se destina esta categoria agregadora.

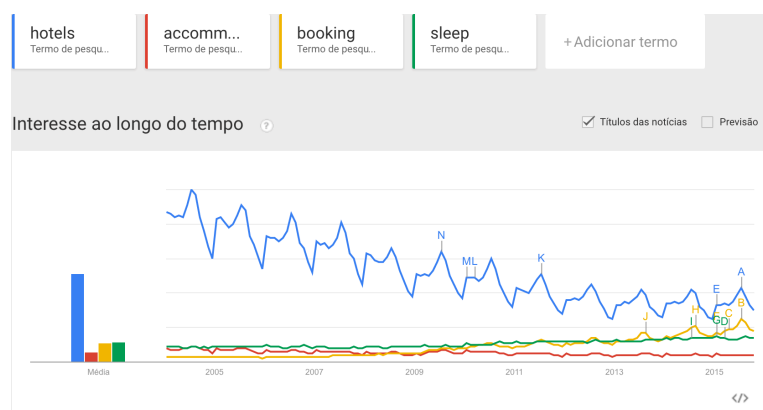


Ilustração 11 - Comparativo de popularidade entre termos relacionados com hotelaria

No caso particular da hotelaria, onde cabem os vários tipos de modalidades de dormida (tanto hotéis, como pensões, parques de campismo, entre outros), “*accomodation*” seria o termo mais indicado para a descrição destes conteúdos. No entanto, e dada a diferença tão grande de popularidade entre este termo e “*hotels*”, foi escolhido “*hotels*” como nome desta categoria agregadora (isto verifica-se igualmente no portal *TripAdvisor*).

Um dado interessante a retirar desta análise tem que ver com o decréscimo de popularidade do termo “*hotels*” e um crescimento de popularidade do termo “*booking*”. O aumento de popularidade do termo “*booking*” pode ser influenciado, em parte, pelo crescimento exponencial da plataforma que tem este mesmo nome²⁸. No

²⁸ <http://www.booking.com/>

entanto, a inversa proporcionalidade da popularidade entre este termo e “hotels” sugere um aumento do tipo de pesquisas transacionais – de que pode ser exemplo uma pesquisa como “hotel room booking in Lisbon”.

Retiram-se, das pesquisas relacionadas, palavras-chave interessantes do ponto de vista do *SEO*, e que devem, portanto, ser tidas em conta no momento da construção dos conteúdos das páginas:

- São frequentes pesquisas por “cheap hotels” ou “cheap accommodation”;
- É também muito usual a pesquisa composta pela junção do termo “hotel” com o nome de um destino turístico, que corrobora o que foi dito no capítulo 2.2. Disto são exemplo “London hotels”;
- De acordo como já referido, a pesquisa “hotel booking” é também bastante frequente, de onde se pode concluir que as plataformas de reservas de alojamento *online* são bastante procuradas hoje em dia.



Ilustração 12 - Comparativo de popularidade entre termos relacionados com património

Relativamente aos pontos de interesse relativos património cultural edificado (onde contam monumentos, museus ou zonas históricas) foi escolhido o termo “heritage” visto que, apesar de não ser o mais popular, é também bastante pesquisado e é o que representa com mais rigor o conteúdo da categoria agregadora a que se destina.



Ilustração 13 - Comparativo de popularidade entre termos relacionados com lazer

A escolha do nome para a categoria agregadora relacionada com lazer foi para “*leisure*”, uma vez que é o termo que apresenta maior popularidade.

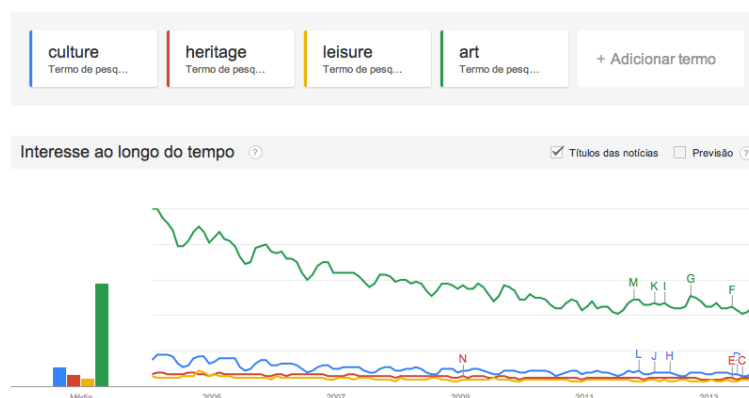


Ilustração 14 - Comparativo de popularidade entre termos relacionados com arte e entretenimento

Relativamente a pontos de interesse de tipos como galerias de arte, museus, salas de espetáculo, e outros tipos desta natureza, não foi possível apurar um termo que melhor o descrevesse que “*art*”. Com efeito, na Ilustração 14 demonstra-se o grande nível de popularidade deste termo quando comparado com outros termos usados.



Ilustração 15 - Comparativo de popularidade entre termos relacionados com compras

Concluindo, foram atribuídos os seguintes nomes a cada uma das oito categorias agregadoras de pontos de interesse:

- “Art” – arte e entretenimento;
- “Beach” – praias, parques aquáticos, piscinas, SPAs, entre outros;
- “Country” - natureza e meios rurais;
- “Food & Drink” – restauração;
- “Heritage” – património arquitetónico e cultural;
- “Hotels” – hotelaria;
- “Leisure” – lazer e vida noturna;
- “Shopping” – compras;

Após esta definição de categorização dos pontos de interesse e respetiva atribuição dos seus nomes, foi necessário estabelecer uma relação destas categorias com as categorias de pontos de interesse listadas na Tabela 4. Esta relação é de muitos-para-muitos, na medida em que vários pontos de interesse podem ser associados a várias categorias agregadoras. A especificação desta relação está detalhada no ANEXO A.

3.4 Tecnologia

Para execução do desenvolvimento do portal foi necessário escolher as ferramentas a utilizar. Foi necessário, portanto, escolher uma *framework* de desenvolvimento de *web*, um motor de base de dados, um motor de indexação de suporte à pesquisa, mecanismo de cache e ferramentas de suporte ao desenvolvimento de *frontend*. De notar que, tendo sido este projeto desenvolvido no âmbito de um estágio na empresa

InfoPortugal, foi tido em conta a preferência pelas tecnologias já utilizadas pela mesma.

Os sistemas *web* da empresa são, na sua maioria, baseados nas seguintes tecnologias:

- *PostgreSQL*²⁹
- *Django Framework*³⁰

3.4.1 Backend

Uma vez que o sistema proposto se baseia em tarefas de leitura (consulta de dados) não há, portanto, necessidade de suportar alta performance em tarefas de leitura/escrita de grandes volumes de dados simultâneas. Por isso se descartou a possibilidade de implementação do modelo de dados num motor de base de dados não relacional e optou-se pelo *PostgreSQL* sobre o qual são bem conhecidas as suas vantagens - reconhecida boa performance com objetos *GIS* (geográficos e espaciais, através da extensão *PostGIS*³¹), comunidade e documentação alargada, fiabilidade, robustez e escalabilidade.

Uma boa *Framework web* pode ser determinada por fatores como capacidade de gerir grande volume de tráfego, controlo de bugs, atualizações frequentes, reutilização e clareza de código, ferramentas de suporte à implementação de funcionalidades, entre outras. Há hoje um sem número de possibilidades no mercado, baseadas em diferentes linguagens de programação, sendo que todas elas tendem a convergir nos pontos acima referidos, portanto a escolha desta não é absolutamente determinante para o sucesso de um portal.

Django Framework é constituída por um conjunto de ferramentas de suporte ao desenvolvimento *web* baseadas em *Python*³², e como tal dispõe de todas as vantagens e potencialidades desta linguagem. Esta, *Python*, é uma linguagem de programação orientada a objetos na qual imperam a legibilidade do código e performance, também para a qual existe um conjunto alargado de bibliotecas de suporte às mais variadas funcionalidades. Eis alguns fatores que determinaram a escolha desta *Framework*:

- Com vista na abstração da implementação relativa ao modelo de dados e na forma de interagir com a base de dados, *Django* dispõe de uma camada de acesso à base de dados (ORM) compatível um alargado número de sistemas de base de dados, possibilitando, por exemplo, que uma alteração de sistema de base de dados envolva um esforço muito reduzido;

²⁹ <http://www.postgresql.org/>

³⁰ <https://www.djangoproject.com/>

³¹ <http://postgis.net/>

³² <https://www.python.org/>

- Baseada no paradigma de desenvolvimento *MVC* (ver Ilustração 16). À semelhança desta, existem três componentes de implementação que dizem respeito à lógica de três fases distintas da aplicação *web*:
 - *Models* – camada de acesso à base de dados;
 - *Template* – camada de apresentação que contém a lógica de como os blocos de informação devem ser apresentados nas páginas *web*;
 - *View* – equivalente ao controlador na arquitetura *MVC*. Contém toda a lógica que serve de comunicação entre os dados e a apresentação.
- O sistema de roteamento é baseado em expressões regulares e permite, por isto, a especificação de qualquer tipo de *URL*, sendo possível e intuitiva a implementação de *URLs* semânticos;
- Sistema modular baseado em aplicações. A filosofia para construção destes módulos baseia-se no conceito UNIX “*Do one thing and do it well*”. Isto permite não só uma boa estruturação de código assim como facilmente estender as funcionalidades do portal.
- *Backoffice* incorporado com acesso a todo o tipo de manipulação de objetos da base de dados.

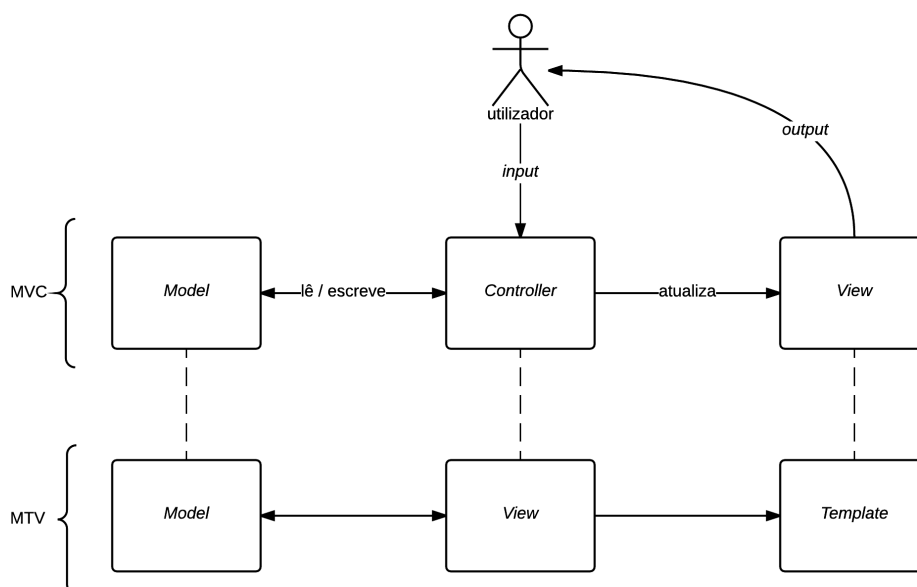


Ilustração 16 – Diagrama de fluxo de dados *MVC* e *MTV*

Como tecnologia de suporte à pesquisa, indexação de conteúdos, foi escolhida a ferramenta *ElasticSearch*³³. Esta é uma ferramenta desenvolvida em *Java* desenhada para operações em tempo real e para ser usada em sistemas distribuídos por forma a garantir flexibilidade e escalabilidade. Uma das interessantes funcionalidades desta ferramenta é a possibilidade de utilização de agregações, estas permitem facilmente desenvolver pesquisas avançadas com filtros encadeados.

O mecanismo de *cache* escolhido foi o *Memcached*³⁴. Este é um sistema de cache de objetos em memória. Tal permite reduzir, substancialmente, o número de acessos à base de dados preservando os seus resultados em memória, o que resulta, tal como desejado, num aumento significativo de desempenho do portal.

3.4.2 Frontend

De forma a auxiliar o desenvolvimento da componente *frontend* do portal foram escolhidas as seguintes ferramentas:

- *Bootstrap*³⁵ – *Framework frontend (HTML, CSS, JavaScript)* baseada num sistema de grelha de 12 colunas que pretende auxiliar a implementação de layouts responsivos;
- *jQuery*³⁶ – *Framework JavaScript*;
- *SASS*³⁷ – pré-compilador *CSS*;
- *Grunt*³⁸ – gestor de tarefas *frontend*. Este foi utilizado em tarefas como compilar dos ficheiros *SASS* e minificar os ficheiros *CSS* e *JavaScript*;

Todo o *layout* e respetiva linha gráfica foram desenvolvidos pelo autor deste documento, para tal foram utilizadas, essencialmente, as ferramentas *Adobe Photoshop*³⁹ e *Adobe Illustrator*⁴⁰.

3.5 Sincronização de conteúdos

Como já referido o conteúdo do portal baseia-se nos conteúdos editoriais georreferenciados da empresa InfoPortugal. Estes conteúdos são geridos numa

³³ <https://www.elastic.co/>

³⁴ <http://memcached.org/>

³⁵ <http://getbootstrap.com/>

³⁶ <https://jquery.com/>

³⁷ <http://sass-lang.com/>

³⁸ <http://gruntjs.com/>

³⁹ <http://www.adobe.com/pt/products/photoshop.html>

⁴⁰ <http://www.adobe.com/pt/products/illustrator.html>

ferramenta interna de gestão de conteúdos, também ela desenvolvida em *Python* (*Django Framework*), cujo acesso à base de dados é, por razões de segurança, restrito à sub-rede da empresa. Desta forma, e estando o portal *All About Portugal* alojado numa estrutura externa à rede da empresa, foi necessário criar um mecanismo de sincronização de dados indireta, via transmissão de ficheiros.

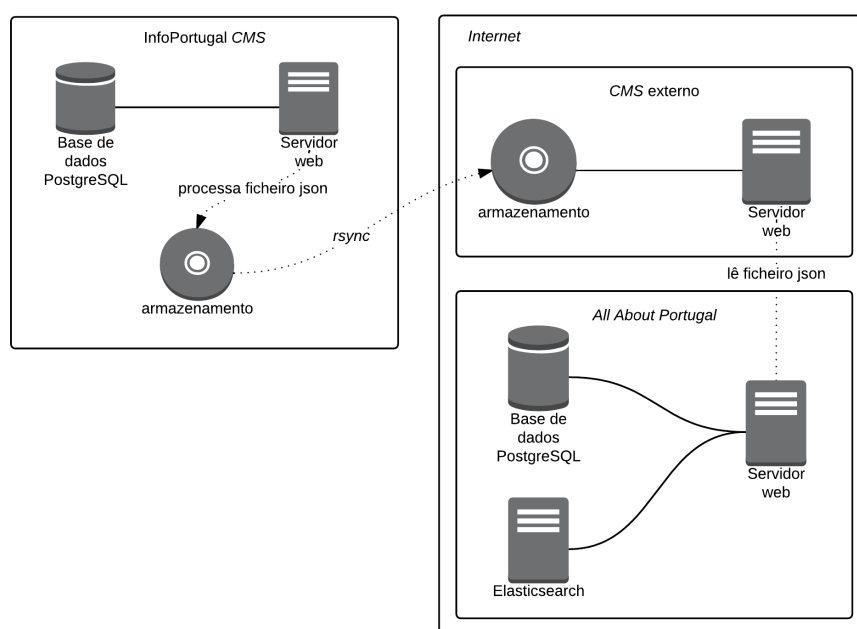


Ilustração 17 - Mecanismo de sincronização de dados

Conforme demonstrado na Ilustração 17, pode dizer-se que a sincronização de dados entre o gestor de conteúdos da InfoPortugal e o portal *All About Portugal* é feita em três passos.

O primeiro passo consiste em exportar os dados necessários a partir da base de dados do gestor de conteúdos. Para isto foi criado um *script* em *Python* que recolhesse esses mesmo dados e os armazenasse num ficheiro em formato *JSON*. Foi dada a instrução ao servidor *web* do gestor de conteúdos que executasse este *script* uma vez por semana. Sendo o servidor *web* em causa baseado em *Unix* esta instrução é dada através do ficheiro que especifica as instruções, e respetiva periodicidade, a executar pelo *cron*. A data e hora de cada execução deste processo de exportação é armazenada numa tabela da base de dados deste mesmo projeto por forma a que seja possível que na exportação seguinte sejam considerados apenas os conteúdos com datas de modificação posterior à mesma.

Uma vez terminada esta exportação, o ficheiro em formato *JSON* resultante da mesma é transferido via *rsync* para uma máquina que externa à sub-rede da empresa e que aloja um serviço que tem como finalidade disponibilizar este ficheiro ao portal *All About Portugal* via *HTTP* com autenticação.

Por último, um *script*, também ele escrito em *Python*, acede ao ficheiro alojado na segunda máquina (via *HTTP*) e trata de sincronizar a nova informação com a já existente na base de dados do *website All About Portugal*. Por sua vez, é feito uso da ferramenta de sinais da *Framework* que permite notificar o *Elasticsearch* de alterações na base de dados e conseqüentemente atualizar também os registos do mesmo.

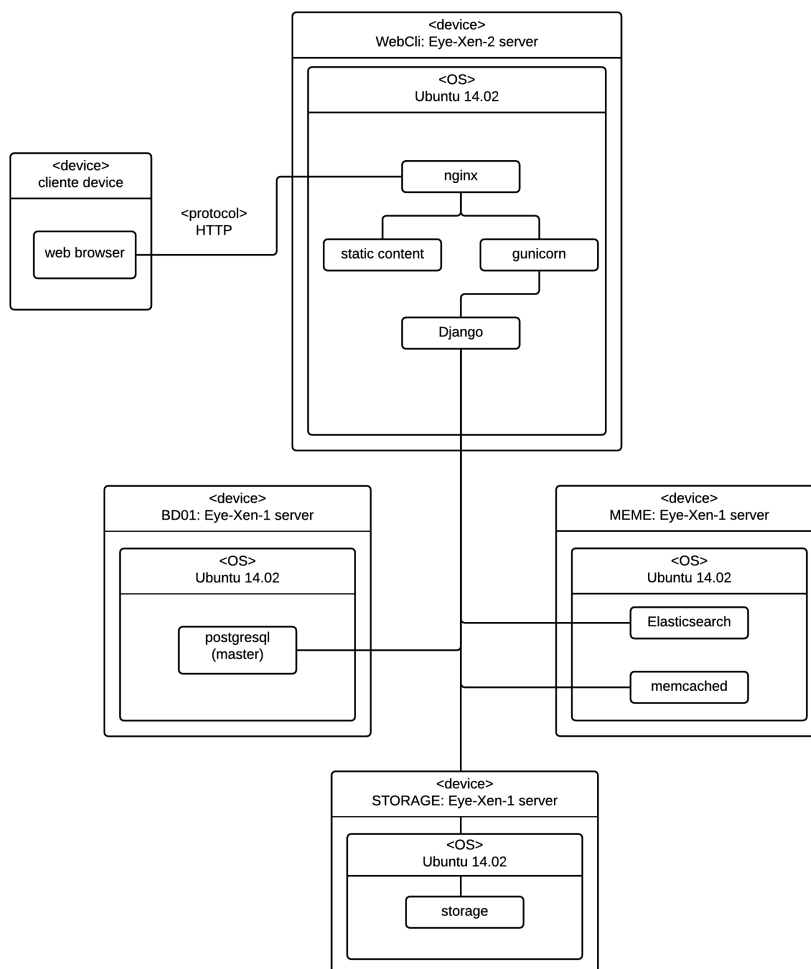


Ilustração 18 - Diagrama de arquitetura

3.6 Hierarquia de páginas

3.6.1 Mapa do site

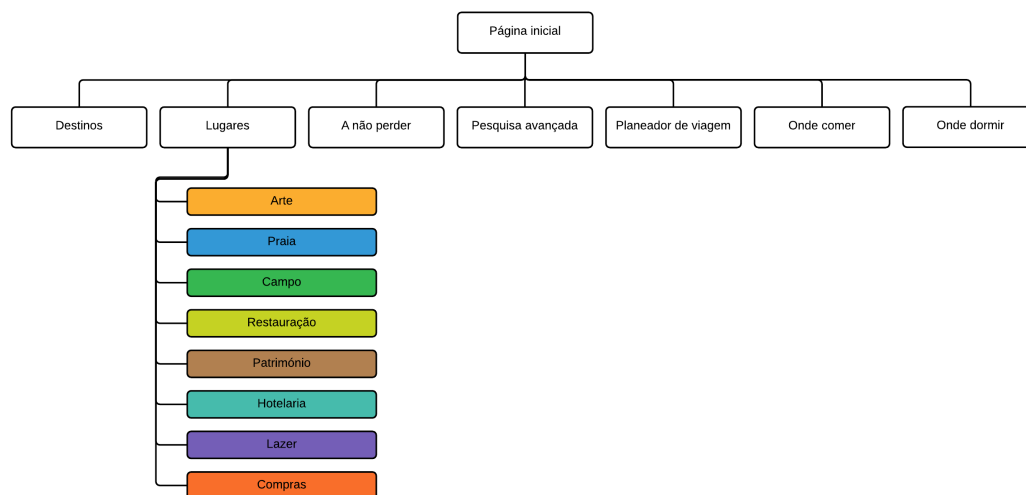


Ilustração 19 - Mapa do site

A estrutura do portal, constituindo um diretório de conteúdos com características muito semelhantes, pretende-se simples, pouco profunda e que permita um acesso rápido e intuitivo aos conteúdos.

A página “Destinos” pretende ser uma listagem de todos os destinos turísticos disponíveis. Nesta página deve ser possível pesquisar pelo nome de destino.

“Lugares” constitui um acesso rápido a listagens de pontos de interesse de acordo com cada uma das categorias agregadoras. E “A não perder” uma listagem de pontos de interesse marcados com essa mesma característica.

A página “pesquisa avançada” pretende ser uma ferramenta completa de filtragem e pesquisa por texto relativamente aos todos os pontos de interesse da base de dados do portal, o comportamento e funcionalidades desta página está descrito no capítulo 3.2.1.2.

“Planeador de viagem” dá acesso à funcionalidade descrita no capítulo 3.2.1.3.

As páginas “onde comer” e “onde dormir” servem o propósito de dar destaque, e acesso rápido, aos conteúdos relacionados com as áreas de hotelaria e restauração, visto ter sido concluído serem áreas de grande procura às quais os restantes portais desta natureza dão também destaque.

3.6.2 Estrutura de URLs

Como já referido anteriormente, é importante que um portal *web* seja servido com uma hierarquia de páginas intuitiva que reflita a navegação do portal e a hierarquia páginas.

A utilização de *URLs* semânticos permite que o utilizador consiga facilmente decifrar qual o conteúdo da página e qual o nível na hierarquia de páginas que lhe corresponde, além disso inspira maior confiança ao utilizador quando confrontado com um determinado *URL*. Posto isto foi especificado um conjunto de expressões regulares que definem a estrutura de páginas do portal, estas expressões regulares são interpretadas por um mecanismo da *framework Django* e usadas para reencaminhamento de pedidos *HTTP*.

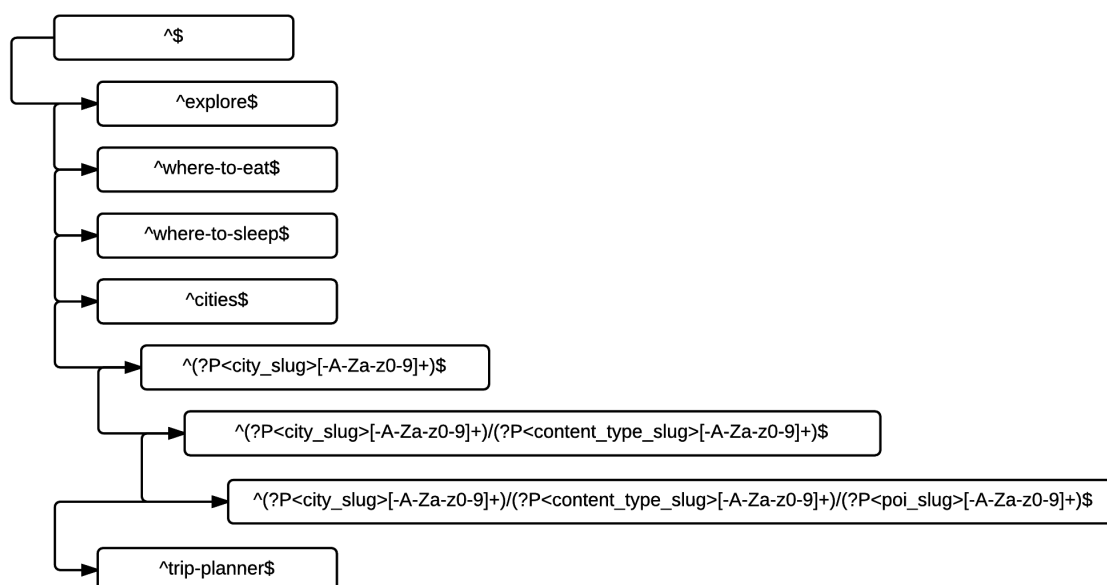


Ilustração 20 - Estrutura de URLs

A Ilustração 20 lista os *URLs* das páginas principais que constituem o portal, foram omitidos *URLs* que não tenham tanta relevância do ponto de vista do *frontend* como *endpoints* destinados a pedidos *AJAX*, de acesso ao painel de administração do portal ou de redireccionamento. A ordem com que estão apresentados na ilustração é a mesma pela qual as expressões regulares são testadas pelo sistema, sistema esta que ao receber um pedido *HTTP* testa a *string* correspondente ao *URL* contra cada uma das expressões regulares definidas até que seja encontrada uma correspondência.

Para aceder a um ponto de interesse o *URL* obedece à sequência hierárquica: cidade, tipo de conteúdo e nome do POI. Dado por exemplo o *URL* “/porto/restaurants/cafe-guarany”, o utilizador facilmente compreende que a página diz respeito a um restaurante com o nome “Café Guarany” situado na zona do Porto.

3.7 Estrutura de dados

Após análise dos requisitos e enquadramento descritos anteriormente conclui-se que o modelo de dados deste projeto deve espelhar parcialmente o modelo de dados especificado na base de dados de pontos de interesse do gestor de conteúdos que armazena o conteúdo em que se baseia o *website* desenvolvido. Foram desenhadas novas especificações de base de dados de forma a otimizar a sua utilização, assim como enquadrá-la melhor no âmbito restrito em que se insere este projeto.

Uma das especificações que não estão presentes na base de dados de origem diz respeito à implementação de relações de herança entre as tabelas relativas aos pontos de interesse. Como tal, foi criado um modelo “POI” com os campos comuns a todos os tipos de pontos de interesse e subclasses desse mesmo modelo que estendem o número de campos a especificidades do tipo de ponto de interesse em causa. Este tipo de herança estabelece uma relação entre a classe pai “POI” e os seus descendentes (ex: “Hotel”, “Restaurant”, “Village”) o que permite agregar todos os tipos de conteúdos semelhantes numa só tabela, neste caso a tabela “POI”.

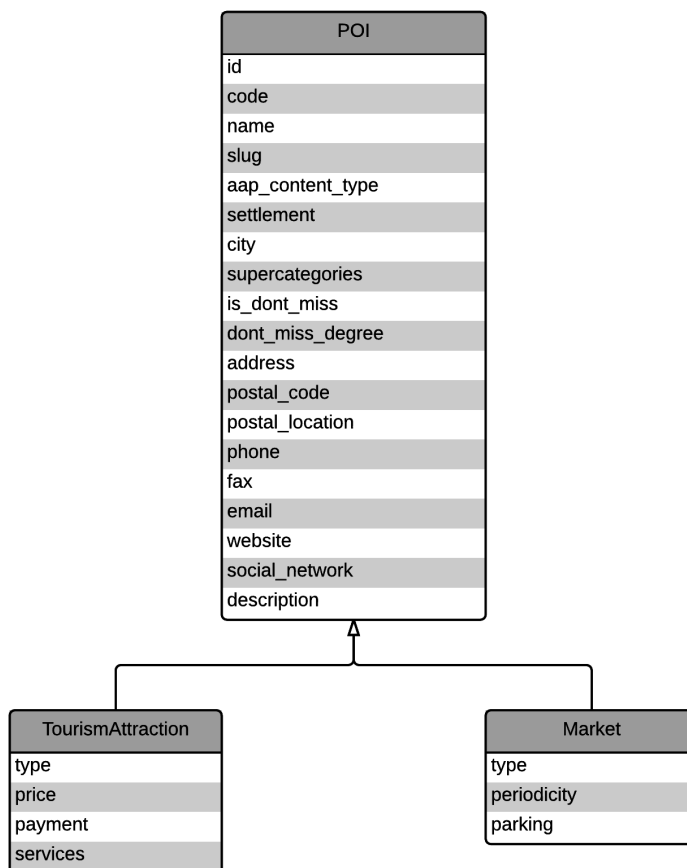


Ilustração 21 - Relação de herança multitableta

O modelo de dados, especificado no anexo B, é, como referido anteriormente, composto por tabelas capazes de armazenar dados já existentes, mas também por tabelas de fim utilitário que sirvam determinado propósito ao portal. Foram excluídas todas as tabelas que dizem respeito à *Framework Django*.

As entradas na tabela “*aap_content_type*” armazenam o nome de cada tipo de ponto de interesse presente na base de dados, tipos estes representados, cada um, por uma tabela. Esta tabela permite não só listar rapidamente todos os tipos de pontos de interesse existentes como também traduzir os nomes dados a cada tipo para as diferentes línguas. Relativamente à tabela “*supercategory*”, esta constitui entradas para cada uma das categorias agregadoras.

3.8 Metadados *HTML*

Como concluído ao longo do capítulo 2.1, o conteúdo de um portal é o fator mais importante para uma boa posição nos *rankings* dos motores de pesquisa. Este é o motor de todas as métricas, uma vez que bom conteúdo é referenciado e partilhado, tem grande volume de dados de relevância e isso faz com que o portal tenha boas pontuações em todas as métricas de avaliação.

No entanto, é necessário que este chegue ao seu público-alvo, e para isto, tendo em conta o enquadramento desta tese, é preciso que os utilizadores sejam levados a clicarem nos resultados de pesquisa referentes ao portal.

Um dos aspetos mais preponderantes para que tal acontecimento tenha lugar tem que ver com a forma como o conteúdo do portal é representado no pequeno espaço que lhe é dedicado nas páginas de resultados de pesquisa. Esta representação pode, e deve, ser manipulada pelos gestores do portal através de metadados especificados nos ficheiros *HTML* correspondentes às suas páginas.

Desta forma, em todas as páginas foram implementadas metadados *HTML* relativos ao título e descrição da mesma, assim como foram também especificados metadados para *Twitter Cards*, *Open Graph* e *Schema.org*.

Tal como verificado nos casos de estudo, o título e descrição da página inicial refletem um apanhado geral de todo o conteúdo do portal.

All About Portugal - Find the best places in Portugal

www.allaboutportugal.pt/ ▼ Traduzir esta página

All About Portugal is a selection of the best places in Portugal. Here you can find the best restaurants, hotels, beaches, museums and others.

Figura 16 - Resultado de pesquisa *Google* correspondente à página inicial



Figura 17 - Exemplo de partilha da página inicial no *Facebook*

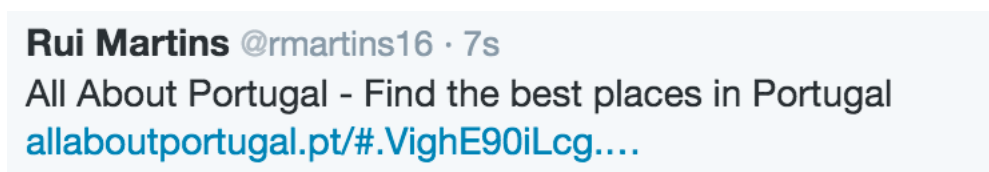


Figura 18 - Exemplo de partilha da página inicial no *Twitter*

Especificando marcação semântica *HTML*, também denominado dados estruturados, é possível informar os motores de pesquisa sobre qual o significado de cada bloco *HTML*. No caso ilustrado na figura a seguir é possível ver-se a interpretação do código *Schema.org* relativo ao bloco que mostra o rasto de navegação do portal (*breadcrumbs*).

[Golden Tulip Porto-Gaia Hotel & SPA - Vila Nova de Gaia ...](#)

www.allaboutportugal.pt > [Vila Nova de Gaia](#) > [Hotels](#) ▼

Golden Tulip Porto-Gaia Hotel & SPA. Porto > Vila Nova de Gaia > Lavadores. Located at Lavadores Beach, a few minutes away from Oporto, it has 56 rooms ...

Figura 19 - Resultado de pesquisa Google correspondente à página de detalhe de um ponto de interesse

3.9 Interface e principais funcionalidades

Há dois momentos chave que ditam o sucesso do portal descrito neste documento: o primeiro momento é o de garantir que o portal é encontrado pelos utilizadores em páginas de resultados de motores de pesquisa e o segundo, após os utilizadores visitarem o portal pelas primeiras vezes, está em conseguir converter estas visitas

esporádicas em visitas periódicas. Para que a segunda condição se verifique a sua experiência de utilização tem de ser satisfatória. Isto é garantido através de uma navegação intuitiva, conteúdo relevante e um *layout* apelativo, moderno e devidamente contextualizado.

3.9.1 Layout

O desenho da interface *frontend* deste portal resulta de *benchmarking* e, sobretudo, de sessões de *brainstorming* e validação das propostas realizadas em conjunto com a equipa de *design* e diretores técnico e geral da empresa.

Tendo em conta as funcionalidades especificadas anteriormente, concluiu-se que seriam contempladas nesta fase de desenvolvimento do *layout* e sucessora fase de implementação de *frontend* as seguintes páginas:

- Página inicial;
- Página de listagem de destinos turísticos;
- Página de detalhe de destinos turísticos;
- Página de detalhe de pontos de interesse;
- Página de pesquisa avançada;
- Página de restaurantes;
- Página de hotéis;
- Página de resultados do sistema de recomendação de plano de viagem;

Uma vez definida a estrutura de cada página foi necessário criar uma linha gráfica para o portal. Na Ilustração 22 estão figurados o logotipo do portal e a atribuição de cores e iconografia a cada uma das oito categorias agregadoras, cores estas que são usadas ao longo de todo o *website*.



Ilustração 22 - Ícones e cores relativos às categorias agregadoras

Além das cores de cada uma das categorias agregadoras foi também definida uma paleta de cores que servisse os restantes elementos de interface do portal - Ilustração 23.

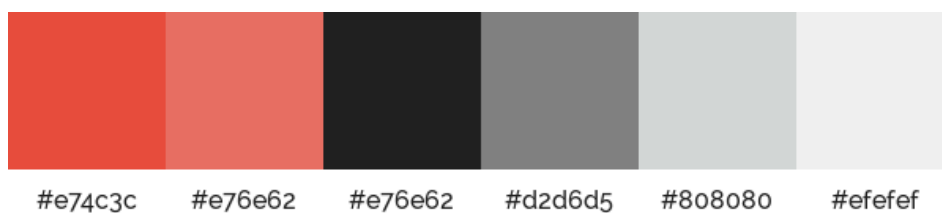


Ilustração 23 - Paleta de cores principal

3.9.2 Principais funcionalidades

Numa tentativa de conseguir referências ao portal nas plataformas *social media*, incentivando à partilha, foi implementado um menu lateral fixo presente em todas as páginas. Para este efeito recorreu-se à plataforma *AddThis*⁴¹.

Respeitando a ideia de que os conteúdos relacionados com a hotelaria e restauração são os mais procurados pelos consumidores de informação turística foram criadas

⁴¹ <http://www.addthis.com/>

páginas específicas para cada um deles, constando como entradas no menu global do portal.

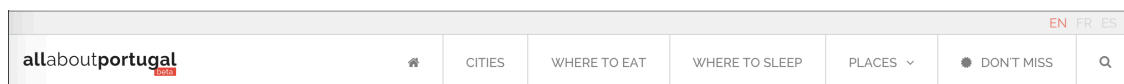


Figura 20 - Menu global do portal

Estas páginas, tanto a dedicada aos hotéis como restaurantes, são bastante idênticas tanto do ponto de vista funcional como de *layout*. Ambas contêm uma caixa de pesquisa por texto que permite ao utilizador encontrar restaurantes por nome ou ainda listar os restaurantes “a não perder” de um determinado destino turístico (ver Figura 21). Esta funcionalidade de pesquisa recorre ao *Elasticsearch* para obter sugestões de acordo com o está a ser digitado pelo utilizador.

São ainda apresentados ao utilizador os seis pontos de interesse “a não perder” mais próximos deste. Estes pontos de interesse são determinados com base na localização do *browser* cliente, obtida através da *API geolocation* suportada nos *browsers* modernos via *HTML5*.



Figura 21 - Caixa de pesquisa da página dedicada aos restaurantes

Na página de detalhe dos pontos de interesse são apresentados outros pontos de interesse conteúdos relacionados. Esta relação baseia-se em três parâmetros:

- Destino turístico – pontos de interesse “a não perder” pertencentes ao mesmo destino turístico;
- Distância – pontos de interesse “a não perder” mais próximos;
- Categorias agregadoras – outros pontos de interesse das mesmas categorias agregadoras;

São então apresentados ao utilizador pontos de interesse relacionados com o da página que visita, com o intuito de melhorar a sua experiência de utilização na medida em que poderão nesta lista constar conteúdos do seu interesse.

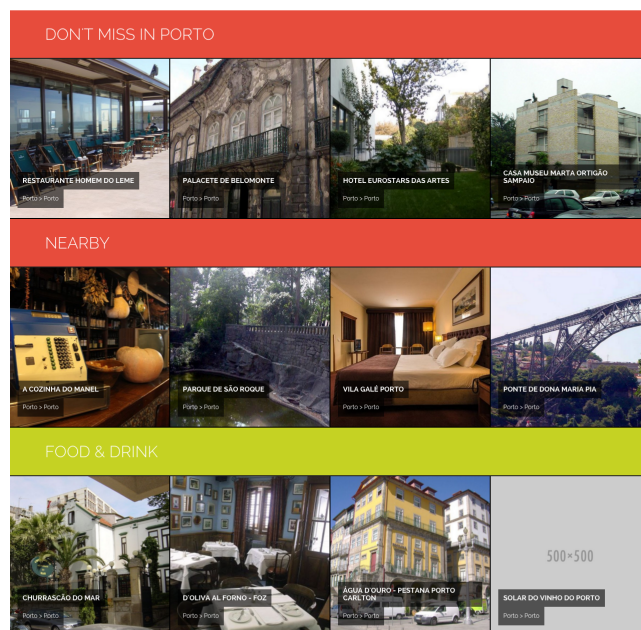


Figura 22 – Pontos de interesse relacionados com o restaurante "Casa Aleixo", Porto

Nesta mesma página foi criada uma zona para comentários feitos pelos utilizadores. Um dos problemas levantados pela empresa estava relacionado com a inexistência de recursos humanos que se pudessem dedicar a tempo inteiro à moderação destes comentários. Como tal, optou-se pela utilização do serviço *online Disqus*⁴² (ver Figura 23). Esta zona de comentários é fonte de conteúdo de interesse para os motores de pesquisa, sendo que o conteúdo gerado pelos utilizadores pode ser bastante relevante para algumas pesquisas.

⁴² <https://disqus.com/>

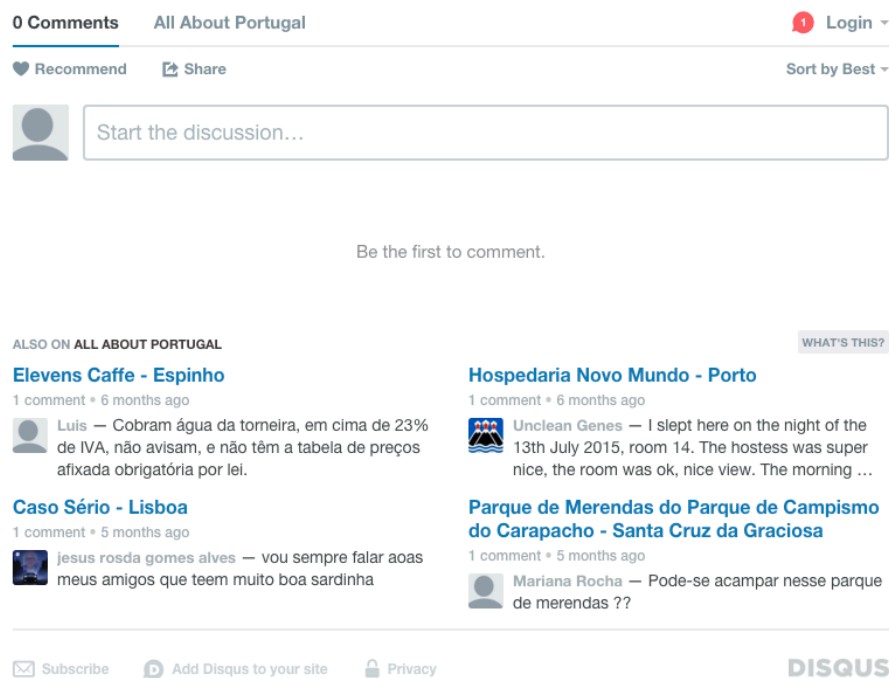


Figura 23 - Sistema de comentários da página de detalhe de pontos de interesse

Um dos requisitos deste portal é o da inclusão de uma ferramenta de pesquisa avançada que possibilitasse ao utilizador realizar vários tipos de filtragem e cruzamento de dados entre os diferentes pontos de interesse e as suas especificidades, tal como descrito no caso de uso do capítulo 3.2.1.2. Todas as ações de filtragem podem ser encadeadas, não se anulando em momento algum à exceção dos seguintes casos:

- Os filtros relativos aos atributos dos pontos de interesse não podem ser encadeados com os das categorias, uma vez que estes são dependentes;
- Se o utilizador refinou os resultados de pesquisa de acordo com uma categoria de ponto de interesse, assim que alterada a filtragem relativa às categorias agregadoras esta primeira filtragem é desativada;
- O mesmo comportamento do tópico anterior se verifica se o utilizador estiver num nível de filtragem correspondente aos atributos dos pontos de interesse;

Para a implementação desta funcionalidade recorreu-se às APIs do *Elasticsearch*, nomeadamente às que dizem respeito à agregação de conteúdos e pesquisa por termos.

Na barra de topo desta página é possível executar três das ações possíveis da pesquisa avançada.

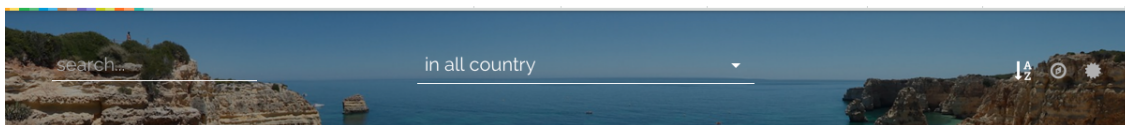


Figura 24 - Barra de topo da página de pesquisa

Do lado esquerdo desta barra de topo o utilizador pode digitar termos para uma pesquisa *full-text*, ao centro pode filtrar os pontos de interesse pelo destino turístico a que pertencem e do lado direito é possível tanto ordenar alfabeticamente como por distância. No caso de o utilizador não permitir que o browser disponibilize a sua localização o botão de ordenação por distância é desativado. O último botão à direita permite listar apenas os pontos de interesse “a não perder”.

A figura seguinte ilustra a zona da página de pesquisa dedicada ao filtro por categorias agregadoras, categorias e atributos dos pontos de interesse.

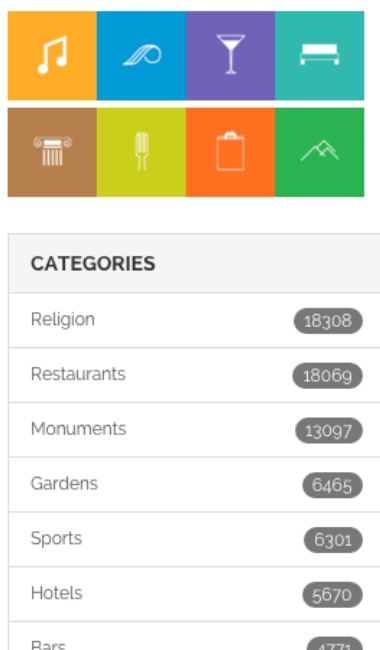


Figura 25 - Barra lateral da página de pesquisa

O processo de criação do plano de viagem é iniciado pelo utilizador através do preenchimento de um formulário presente na página inicial. Este é apresentado ao utilizador através de um formulário de linguagem natural, este tipo de formulário é uma tendência recente do *design* de interfaces. A implementação deste tipo de portais pode ser encontrada, por exemplo, no *website Betterment*⁴³.

⁴³ <https://www.betterment.com/>

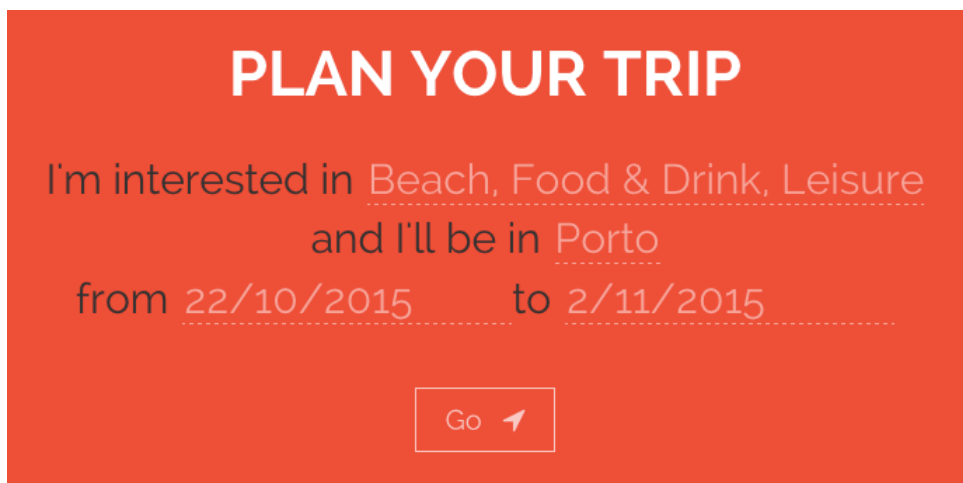


Figura 26 - Formulário de iniciação do sistema de recomendação de planos de viagem

Todos os campos do formulário são de preenchimento obrigatório. No clique de cada espaço vazio da frase é mostrado ao utilizador um ecrã contextual relativo ao tipo de *input* esperado para o campo correspondente. Se o formulário for válido, isto é, se todos os campos estiverem preenchidos, o sistema calcula o plano de viagem referido (ver Figura 27). Os detalhes de implementação deste sistema estão descritos no capítulo seguinte.

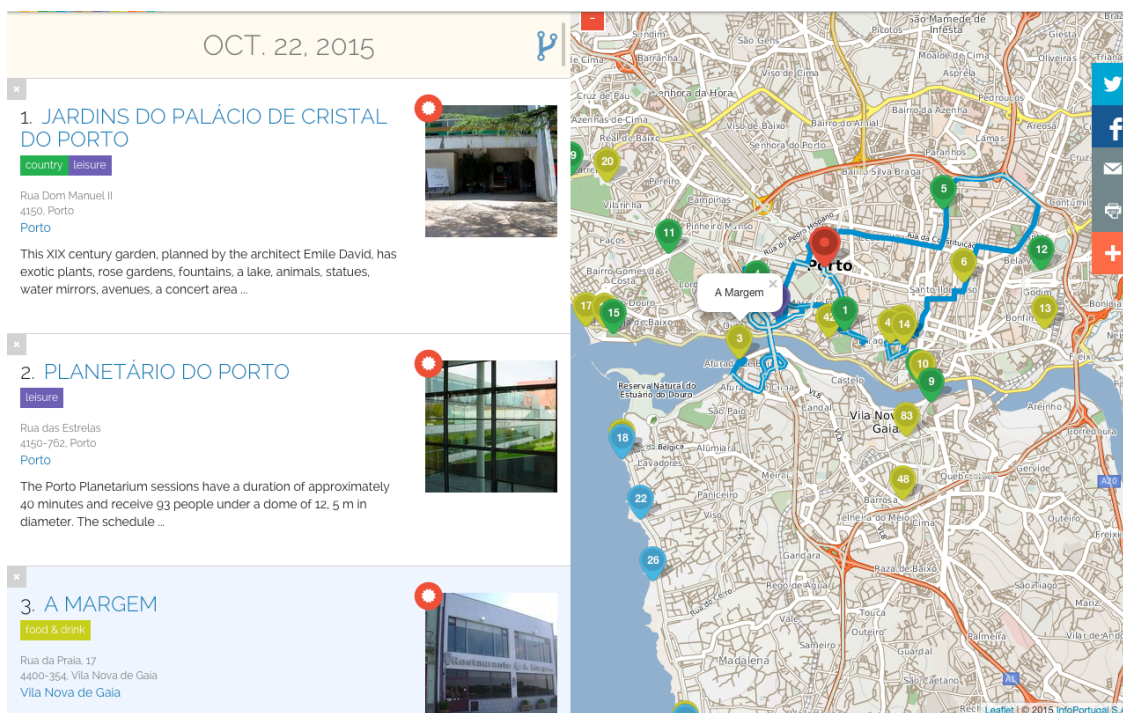


Figura 27 - Exemplo de plano de viagem

3.10 Detalhes de implementação do sistema de recomendação de planos de viagem

A ferramenta descrita neste capítulo determina um conjunto de pontos de interesse, e respetiva rota de navegação entre eles, dado um conjunto de condições definidas pelo utilizador. Ao contrário de outros sistemas de planeamento de viagem como o *Tripomatic*, o conceito que sustenta esta ferramenta assenta na ideia de que o plano de viagem é proposto ao utilizador e não criado pelo mesmo, tal como o *YourTour*.

Foi pensado um conceito simples que definisse este tipo de interação. Por forma a requisitar uma proposta de rota o utilizador é confrontado apenas com três perguntas simples:

- Para onde? – o utilizador deve definir um destino geográfico válido para o território português, este pode ser o centro de uma cidade, um hotel ou um endereço postal;
- Quando? – intervalo de datas da sua potencial visita;
- O quê? – escolha entre as oito categorias agregadoras descritas anteriormente.

Uma particularidade deste sistema de recomendação é o foco no ponto geográfico definido como origem, isto é, o pressuposto é que para cada dia é sugerido um conjunto de pontos a visitar e no final desse mesmo dia o turista retorna ao seu local de partida (na maior parte dos casos será o local de estadia). Desta forma a proposta difere do usual, em vez de se centrar unicamente nos pontos de interesse mais populares centra-se na descoberta de território em torno de um determinado ponto.

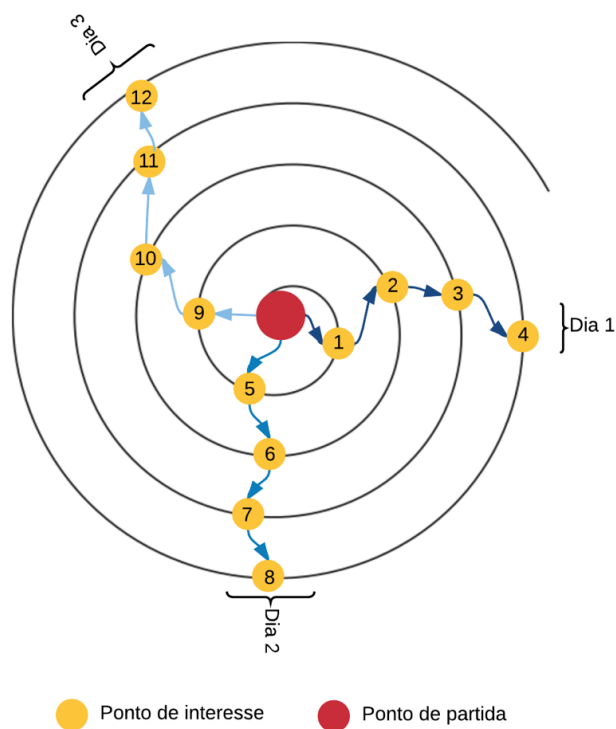


Ilustração 24 - Exemplo de rota recomendada para três dias excluindo pontos de interesse da categoria agregadora “Food & Drink”

No caso de o utilizador seleccionar a categoria agregadora “Food & Drink” o sistema calcula três pontos adicionais, um que corresponderá ao local de recomendação para almoço, outro para jantar e por fim um local de diversão noturna, passando a serem considerados 7 pontos para cada dia.

O algoritmo responsável pelo cálculo dos pontos de interesse que compõem o plano de viagem requisitado pelo utilizador é implementado no método “fetch_pois”, que por sua vez é invocado pela *view* correspondente a este pedido. Este método recebe os seguintes parâmetros:

- *lat, lon* – posição geográfica (latitude e longitude) do ponto inicial definido pelo utilizador. Estas coordenadas são dadas segundo o sistema WSG84;
- *deleted_pois* – lista de identificadores únicos dos pontos de interesse requisitados pelo utilizador para exclusão;
- *date_begin, date_end* – datas de início e de fim da viagem, estabelecidas pelo utilizador;
- *supercategories* – lista de identificadores únicos das categorias agregadoras a serem consideradas;

Este método baseia-se no seguinte algoritmo demonstrado em pseudocódigo:

```

Entrada: categoriasAgregadoras, poisExcluidos, dataInicio, dataFim,
pontoOrigem
início
  se "Food & Drink" em categoriasAgregadoras
    numPoisPorDia ← 7
  senão
    numPoisPorDia ← 4
  fim-se

  numDias ← dataFim - dataInicio + 1
  countPois ← 0
  vetor poisUsados(numPoisPorDia * numDias)
  vetor planoViagem(numDias, numPoisPorDia)

  para i ← 1; i ≤ numDias; i ← i+1 fazer
    para x ← 0; x < numPoisPorDia; x ← x+1 fazer
      se x = 0 então
        poisUsados[countPois] ←
          obterPoiMaisProximo(
            origem=pontoOrigem,
            cat=categoriasAgregadoras,
            exclui=poisExcluidos,
            tipos=[])
      senão se "Food & Drink" em categoriasAgregadoras então
        se x = 2 ou x = 5) então
          poisUsados[countPois] ←
            obterPoiMaisProximo(
              origem=poisUsados[countPois -
                1],
              cat=categoriasAgregadoras,
              exclui=poisExcluidos,
              tipos=["restaurantes"])
        senão se x = 6 então
          poisUsados[countPois] ←
            obterPoiMaisProximo(
              origem=poisUsados[countPois -
                1],
              cat=categoriasAgregadoras,
              exclui=poisExcluidos,
              tipos=["bares"])
        fim-se
        countPois ← countPois + 1
      senão
        poisUsados[countPois] ←
          obterPoiMaisProximo(
            origem= poisUsados[countPois - 1],
            cat=categoriasAgregadoras,
            exclui=poisExcluidos,
            tipos=[])
        fim-se
        planoViagem[i][x] ← poisUsados[countPois]
      fim-para
    fim-para
  fim

```

Código 1 - Algoritmo de cálculo de planos de viagem

A função “obterPoiMaisProximo” que figura no algoritmo acima refere-se ao pedido à base de dados, pedido este que faz uso de funções geográficas para calcular o ponto mais próximo relativo ao especificado no argumento “origem” (ex: 41.1621376, -8.6568725). O argumento “cat” será uma lista de identificadores das categorias agregadoras usadas para filtrar os pontos de interesse a considerar (ex: “Food & Drink”), o argumento “poisExcluidos” será uma lista de identificadores únicos dos pontos de interesse a não considerar e “tipos” serão o filtro de último nível relativo aos tipos de pontos de interesse a considerar (ex: “restaurantes”).

4 Resultados

O objetivo elementar de qualquer *website* é o de ser visitado. Mas, mais importante que isso, é que essas visitas sejam relevantes, ou seja, que a maior percentagem destas sejam de utilizadores que pertençam ao público-alvo a que se destinem os *websites*. É, pois, com base em dados acerca de utilizações que se aferem os sucessos e as limitações deste projeto. Além disso, considera-se que estes dados fazem também eles parte do estudo em torno do *SEO*.

Posto isto, imediatamente antes de proceder à disponibilização do *All About Portugal* em ambiente de produção (visível ao público) implementou-se no mesmo um canal de comunicação entre o referido *website* e a ferramenta *Google Analytics*. Esta é uma ferramenta *web* gratuita, disponibilizada pela *Google* que, uma vez ativa, recolhe dados relativos à atividade dos utilizadores. Posteriormente, com base nestes dados, disponibiliza um conjunto alargado de estatísticas que permitem perceber como é que os utilizadores descobrem, visitam e utilizam o *website* em causa, sendo possível até visualizar dados em tempo real. Esta é, possivelmente, a ferramenta mais utilizada em todo mundo para efeitos de recolha de dados deste tipo. É muito com base nos dados obtidos através desta que são tomadas decisões ao nível do *SEO*.

Os dados disponibilizados por esta ferramenta estão agrupados em seis grandes áreas: “Eventos de inteligência”, “Tempo real”, “Público-alvo”, “Aquisição”, “Comportamento” e “Conversões”. Para efeitos da análise presente neste capítulo são consideradas as áreas “Público-alvo”, “Aquisição” e “Comportamento”.

São considerados dados compreendidos entre o intervalo de 15 de Março de 2015 a 26 de Setembro de 2015.

4.1 Definição das métricas consideradas

Por forma a conseguir obter conclusões acerca dos registos de acessos disponibilizados pela ferramenta escolhida foi necessário compreender o significado das suas métricas e o seu enquadramento relativamente ao projeto de que trata este documento.

As métricas base do *Google Analytics*, em que se baseiam todas as outras, são dadas pelo número de sessões, número de utilizadores e número de visualizações de página, todas elas tendo em conta um dado intervalo de tempo. Uma sessão é definida pelo período de tempo que um determinado utilizador está ativo no *website*. O número de utilizadores inclui utilizadores que visitam o *website* pela primeira vez ou que o revisitam. Por último, o número de visualizações de página inclui visualizações de páginas repetidas.

Na área “Público-alvo” estão disponíveis dados estatísticos que dizem respeito às características dos visitantes do *website*. Destes destacam-se os seguintes:

- **Visualizações de páginas por sessão** – dado o número alargado de conteúdos de que dispõe o *website*, de maior ou menor relevância, é esperado que a maior parte dos utilizadores visitem o portal com uma intenção muito específica: consulta de uma determinada informação relativa a um ponto de interesse. Por esta razão, não é esperado um valor elevado nesta métrica, mesmo tendo em conta a listagem de conteúdos relacionados na página de detalhe do ponto de interesse ou da página de detalhe do destino turístico cujo objetivo é o de combater este comportamento;
- **Duração média da sessão** – com base nesta métrica pode perceber-se a quantidade de tempo que os utilizadores estão, em média, ativos no *website*. Será expectável, pelas mesmas razões apontadas no tópico anterior, que este valor não seja expressivo, uma vez que o comportamento natural de um visitante será o de obter a informação desejada e imediatamente sair do mesmo;
- **Taxa de rejeição** – percentagem de sessões que terminaram sem qualquer navegação interna. Esta métrica está diretamente relacionada com as anteriores;
- **Percentagem de novas sessões** – parcela de sessões cujos utilizadores visitam pela primeira vez o *website*. Esta métrica permite perceber o nível de fidelização dos utilizadores com o *website*. Uma percentagem muito elevada pode ser reveladora de fraco conteúdo ou de navegação pouco intuitiva, no entanto, num *website* recente esta percentagem será muito alta;
- **Demografia e Interesses** – percentagens estimadas relativas ao género e idade dos visitantes, assim como informações relativas à segmentação de mercado e

mercados relacionados. Através destas podem ser definidas estratégias de mudanças de público-alvo ou foco em nichos de mercado;

- **Geografia** – estes são também dados importantes para a análise deste portal, a partir destes pode-se concluir se a internacionalização dos conteúdos foi bem sucedida e se é vantajoso, por exemplo, o investimento em tradução dos conteúdos para mais idiomas. Sendo este *website* orientado ao turismo externo, é expectável verificar visitas de vários países, sobretudo, dos idiomas disponíveis;

Na secção “Aquisição” constam dados relativos à origem do tráfego do *website*. Este é outros dos aspetos importantes a analisar por forma a perceber se a indexação do *website All About Portugal* pela *Google* resultou em resultados satisfatórios, uma vez que esta é a única fonte divulgação do *website*. Desta secção destacam-se:

- **Canais** – permite visualizar atributos estatísticos das outras áreas, de acordo com cada tipo de origem de tráfego – tráfego orgânico, direto, de referência ou em redes sociais. Será expectável que grande parte do tráfego do *website* tenha origem nas páginas de resultados de pesquisa do *Google*;
- **Otimização do motor de pesquisa** – no caso deste tipo de métricas os dados tomados como base diferem das restantes. São fornecidos pela própria *Google* e baseiam-se na atividade registada nas suas páginas de pesquisa relativa a um determinado *website*. Este tipo de informação pode também ser consultada na ferramenta complementar *Google Webmaster Tools*. Através destes dados podem consultar-se, entre outros detalhes, os termos pesquisados que mais impressões ou cliques geraram, ou mesmo obter uma posição média de uma determinada página.

Da secção “Comportamento” são destacados os seguintes tipos de dados:

- **Fluxo de utilizadores** – estes dados permitem perceber quais os fluxos de navegação mais comuns. Desta é importante perceber quais os tipos de página que são mais usados como entrada no *website*. De acordo com o já referido, será expectável que as páginas de entrada mais utilizadas sejam as de detalhe dos pontos de interesse.
- **Tempo carregamento de uma página** – intervalo de tempo, em segundos, que decorre desde o momento em que é feito um pedido pelo cliente a uma determinada página até que esta é completamente carregada no seu *browser*. Como visto no capítulo 2.1.4, o desempenho de um determinado *website* é hoje fator de ponderação importante no momento da avaliação de uma determinada página *web* no contexto da construção de *rankings* do motor de pesquisa *Google*. Não existe, no entanto, nenhuma regra específica para este atributo, sendo que a *Google* apenas sugere aos *webmasters* que as páginas

devem ter o melhor desempenho possível. Assume-se que seria expectável que o tempo médio de carregamento de uma página ronda-se os 4 ou 5 segundos,;

- **Tempo de ligação ao servidor** – este é também um fator considerado importante pela *Google*. Foi indicado anteriormente que este não deveria ser superior a 200ms. Esta métrica é também conhecida por *time for first byte*, e representa o tempo decorrido desde que é feito o pedido *HTTP* pela máquina cliente até que é recebido o primeiro *byte* de informação do servidor.

Os tipos de dados/métricas descritos ao longo deste capítulo são os considerados mais importantes no momento de análise conclusiva deste projeto. Permitem não só verificar os fatores de sucesso da implementação de acordo com o desafio proposto como também identificar melhorias.

Os dados relativos ao *website* desenvolvido são demonstrados e analisados nos três capítulos que se seguem.

4.2 Análise de dados das características dos visitantes

4.2.1 Dados gerais

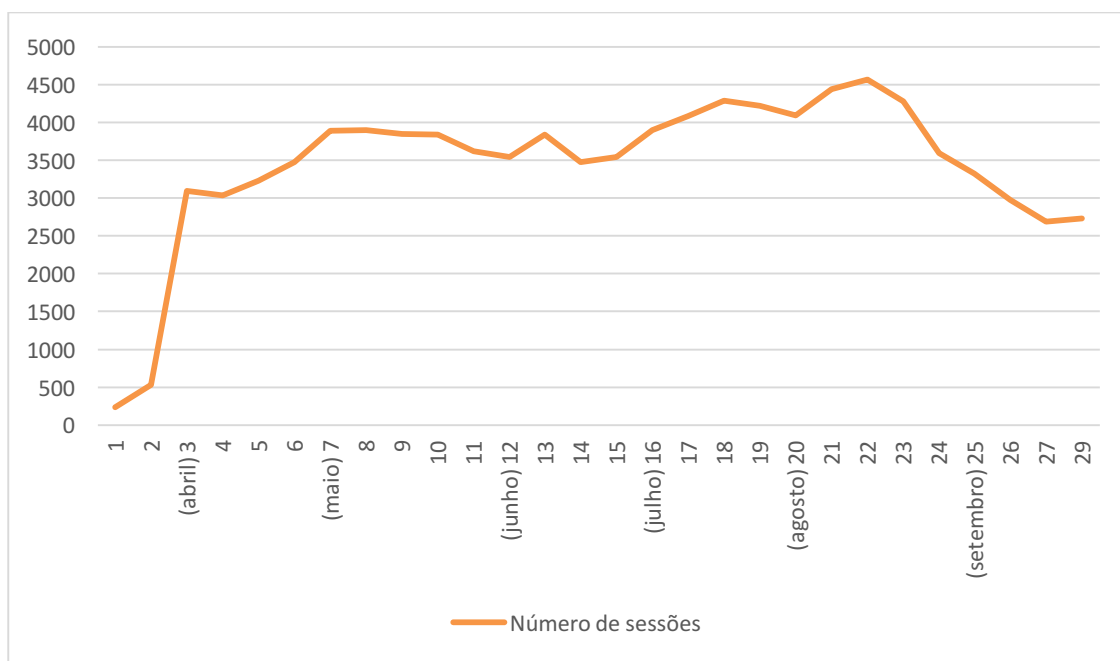


Ilustração 25 - Número de sessões por semana

A partir do momento em que o portal foi disponibilizado ao público verificou-se um rápido crescimento de visitas, o que sugere que o portal foi desenvolvido de forma a permitir uma boa indexação das suas páginas por parte dos motores de pesquisa. A Ilustração 25 demonstra exatamente o referido. Note-se, por exemplo, o crescimento

exponencial de sessões da segunda para a terceira semana – aumento de mais de 2500 sessões.

Atendendo a que o portal se foca essencialmente no turismo, e sendo os turistas o seu público-alvo, é natural que o ponto máximo de visitas semanais tenha ocorrido durante agosto, mês de maior movimento turístico no país. Isso poderá explicar, também, o decréscimo de visitas verificado no final do mês de agosto e início de setembro, já que este período coincide com o fim da chamada época alta.

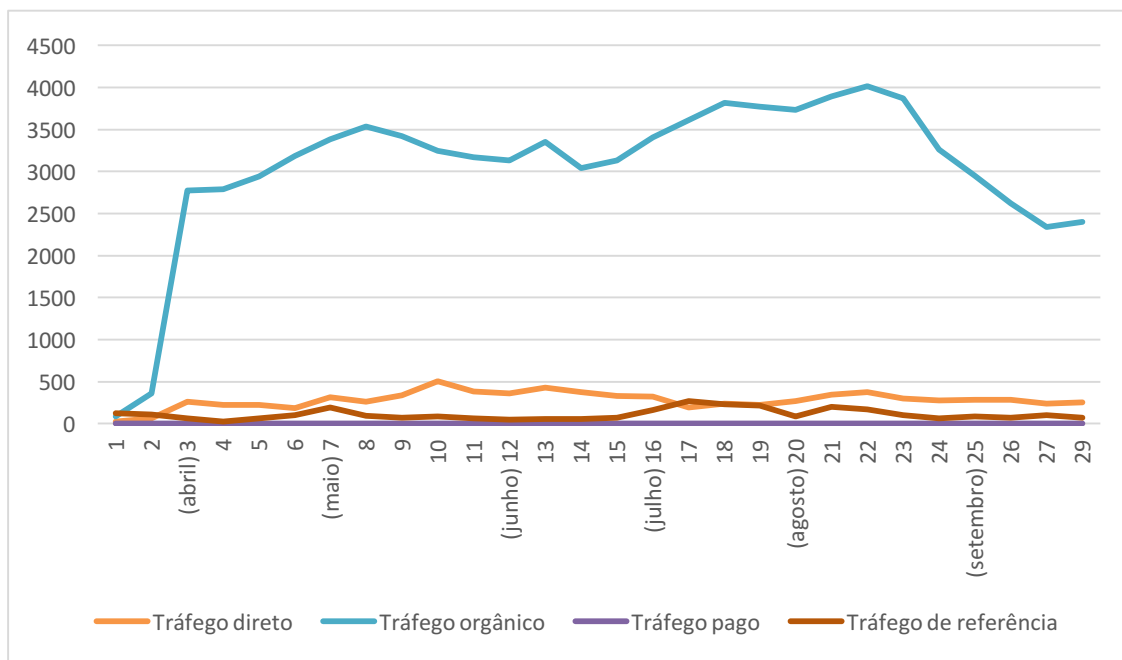


Ilustração 26 - Número de sessões por semana

Os dados da Ilustração 26 confirmam o cumprimento do objetivo de atrair tráfego através de pesquisas *online* (tráfego orgânico). Com efeito, na referida ilustração pode constatar-se que 88,6% das visitas tiveram origem em páginas de resultados de pesquisa. Isto significa que as pesquisas retornam páginas do portal *All About Portugal* numa posição relativamente significativa. No entanto, a baixa percentagem de tráfego direto e de referência revelam falha na fidelização dos utilizadores com o *website*, o que é demonstrado também pelas alta taxa de rejeição, elevada percentagem de novos utilizadores, um número baixo de páginas por sessão e de duração média por sessão reduzida (ver Tabela 5). Isto pode ser explicado por vários cenários possíveis. Como foi já referido, este seria um comportamento expectável dada a natureza do conteúdo do portal, no então, e uma vez sendo apresentados ao utilizador conteúdos relacionados, era desejável que se verificasse um número mais elevado de páginas por sessão.

Tabela 5 - Métricas gerais de sessão (entre 15 de Março de 2015 e 26 de Setembro de 2015)

Páginas por sessão	1.71
Duração média por sessão	00:01:01
Taxa de rejeição	77.09%
Percentagem de novos utilizadores	89.73%

4.2.2 Dados demográficos

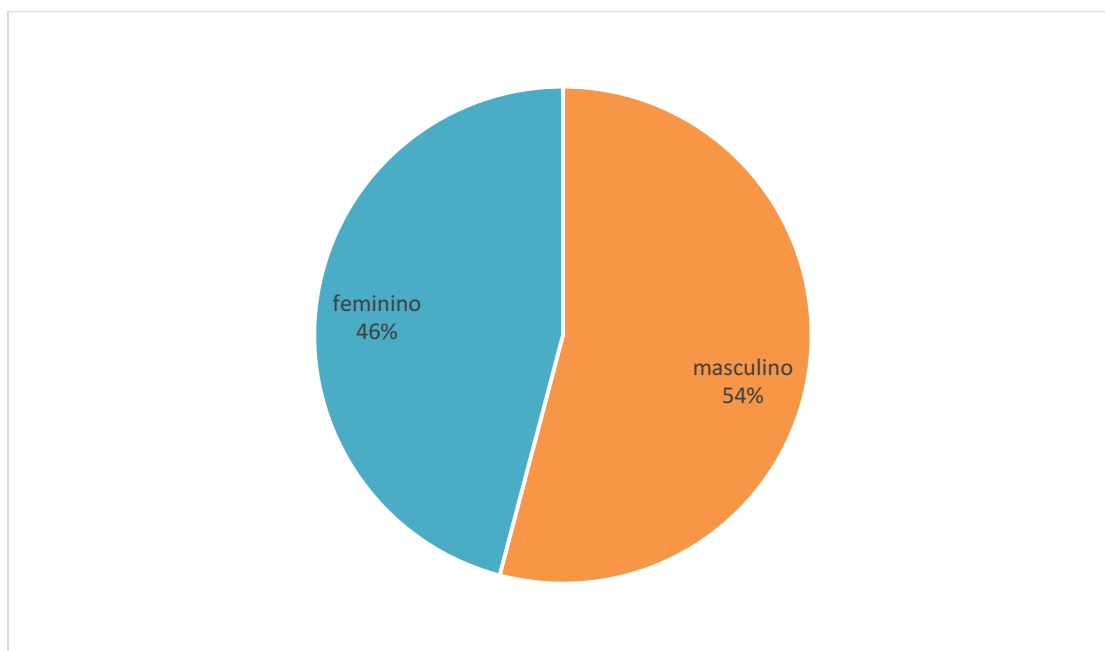


Ilustração 27 - Percentagem de sessões por género

As percentagens de género são muito equilibradas como seria expectável no contexto em que se inserem os conteúdos.

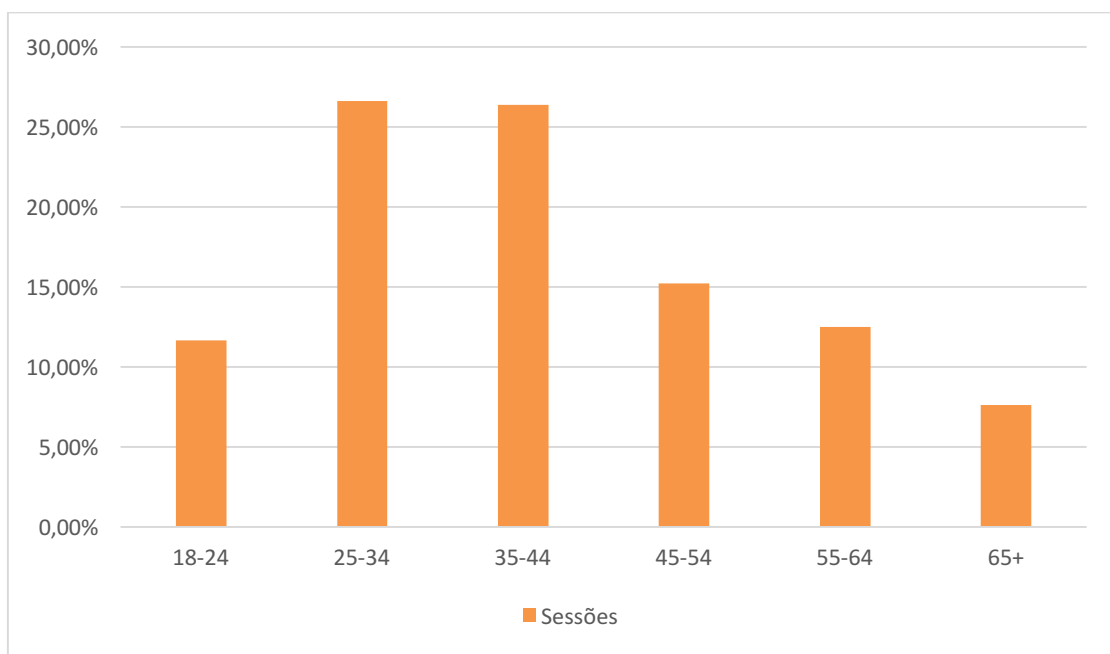


Ilustração 28 - Percentagem de sessões por idade

De notar que há uma dominância significativa dos utilizadores entre os 25 e os 44 anos. Conclui-se, portanto, que esta será uma faixa etária sobre a qual o processo editorial de desenvolvimento de conteúdos futuros se deve debruçar mais.

4.2.3 Dados geográficos

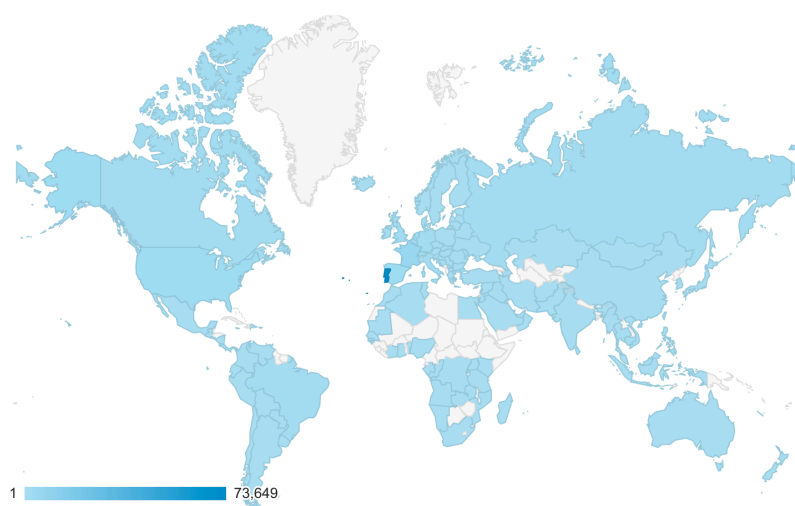


Figura 28 – Representação em mapa da distribuição geográfica dos acessos ao portal de visitas

A avaliar pela Figura 28, verifica-se que o portal foi visitado por utilizadores de, praticamente, todo mundo. A grande parte destes países não apresenta um número

de visitas elevado, porém, do ponto de vista da indexação pelos motores de pesquisa, pode-se assumir que uma primeira abordagem à internacionalização produziu resultados esperados, dado todo este amplo conjunto de países visitantes.

Pode ser visto no gráfico seguinte que há um desequilíbrio muito grande na distribuição das visitas pelos vários países, constatando-se, por exemplo, que 76.5% de todas as visitas ao portal têm origem em Portugal. Esta percentagem pode representar na sua maioria, e uma vez que o portal não está disponível em português, turistas estrangeiros que façam as suas pesquisas no momento da sua estadia em Portugal, e não em momento de planeamento ainda no seu país de origem. No entanto, os dados da Ilustração 30 levantam algumas dúvidas acerca desta conclusão uma vez que perto de 60% do tráfego vem de utilizadores de língua portuguesa de Portugal.

Como referido anteriormente, a taxa de rejeição e a percentagem de novos utilizadores são altas. No entanto, é de notar que a taxa de rejeição nos casos de França, Espanha e Suíça são ligeiramente mais baixas que as restantes, levando-nos a concluir que o tráfego oriundo destes países será de maior qualidade.

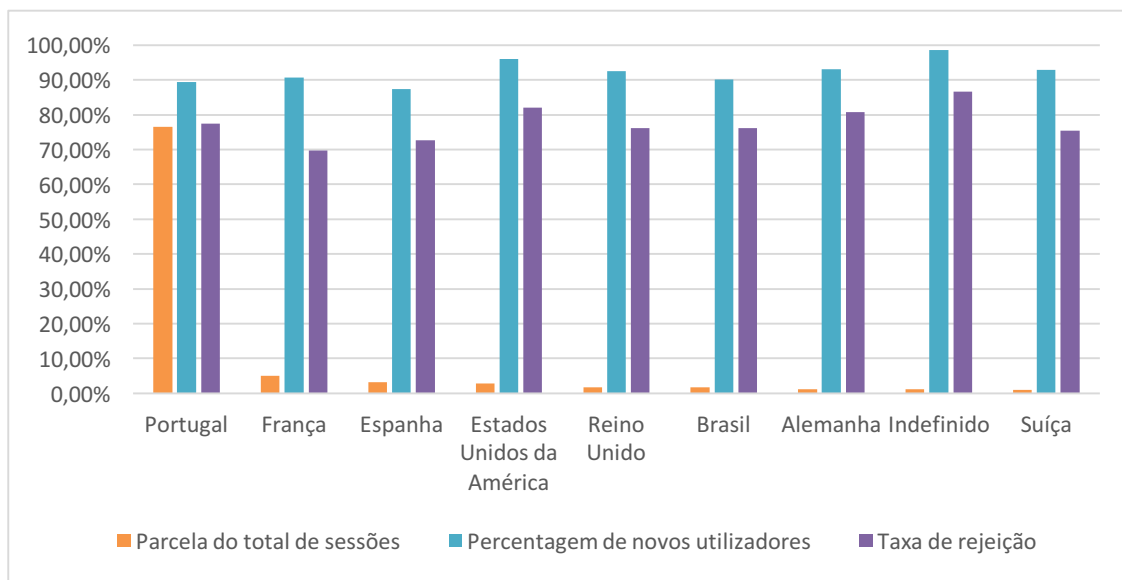


Ilustração 29 - Sessões, novos utilizadores e taxa de rejeição para os 10 países mais visitantes

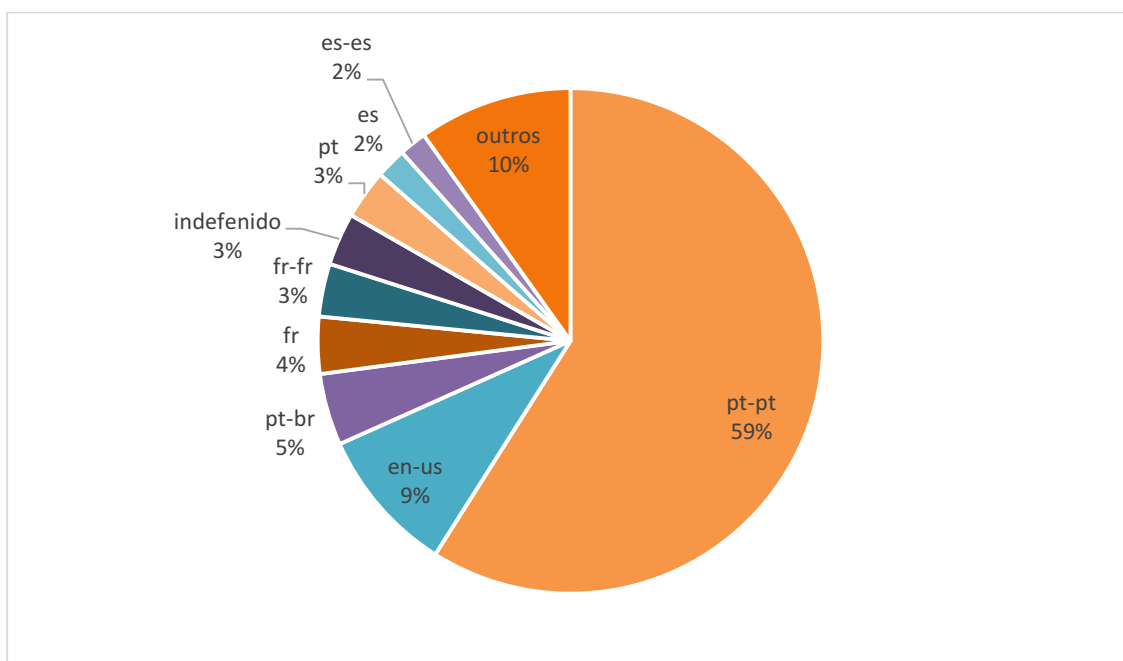


Ilustração 30 - Idioma dos utilizadores

4.3 Análise de dados de pesquisa

Tanto a ferramenta *Google Analytics* como *Google Webmaster Tools* guardam os dados apenas dos últimos três meses para este tipo de dados, pelo que, para efeitos desta análise, os dados são relativos ao intervalo de tempo de 2 de julho de 2015 a 1 de Outubro de 2015. Os dados da Ilustração 31 e Ilustração 32 foram recolhidos através da ferramenta *Google Webmaster Tools*.

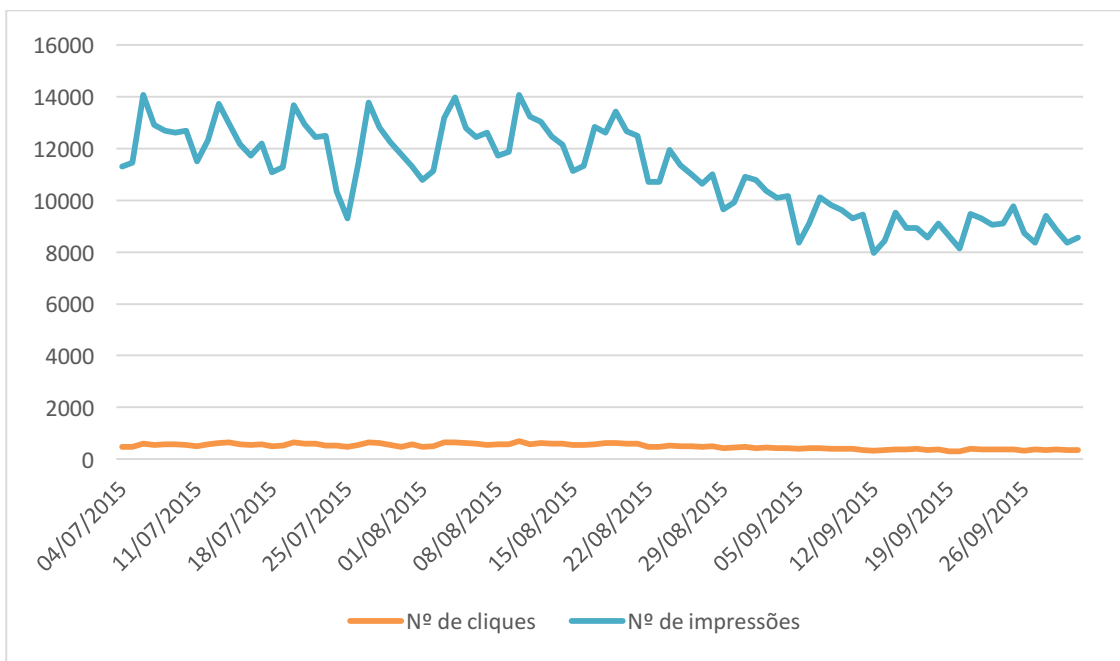


Ilustração 31 - número de cliques versus número de impressões

Apesar de haver um significativo número de impressões das páginas do portal de que trata este projeto em páginas de resultados de pesquisa, o número de cliques nestas mesmas páginas é muito reduzido, rácio este que resulta numa *CTR* baixa.

Como é demonstrado na Ilustração 32, a *CTR* não chega sequer, em momento algum, aos 6%, valor muito abaixo do desejável. Isto significa que, em média, menos de 5% dos utilizadores que são confrontados com uma ligação para o *All About Portugal* não clicam no mesmo. O que significa que será preciso rever estratégias diretamente relacionadas com o conteúdo que é apresentado nas páginas de resultados de pesquisa.

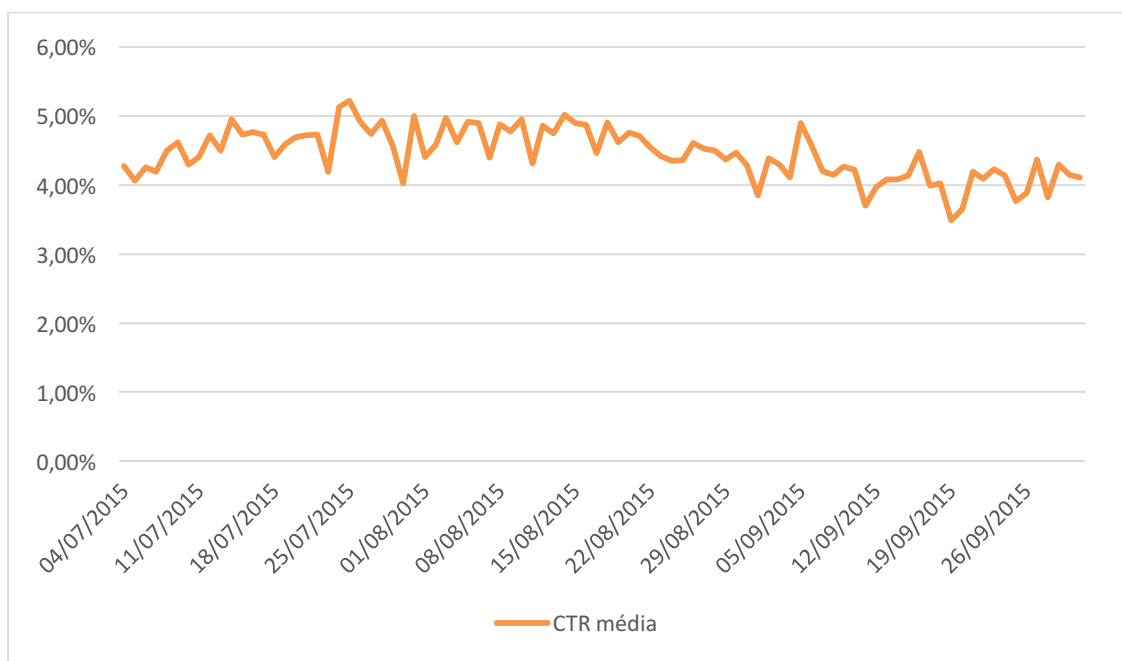


Ilustração 32 - CTR média

Termos de pesquisa	Impressões	Cliques	Posição média	CTR
motel da vim	2245	13	6.1	0.58%
areinho arouca	1163	16	6.3	1.38%
quermesse restaurante lisboa	995	2	10.8	0.2%
poço azul	984	10	8.5	1.02%
praia fluvial do vau	728	8	7.7	1.1%
poço azul de são pedro do sul	719	34	4.6	4.73%
motel vim	656	2	6.5	0.3%
praia fluvial do areinho	623	10	5	1.61%
praia fluvial do areinho arouca	602	7	5.5	1.16%
shape matosinhos	505	12	9.6	2.38%

Tabela 6 - Dez termos de pesquisa que resultaram em maior número de impressões

A Tabela 6 contém as primeiras dez entradas de uma listagem de termos de pesquisa que tenham resultado em impressões de páginas do *website All About Portugal*. Pode-se concluir que a natureza das pesquisas que estão a conduzir ao *website* é a correta, e que, apesar de se poder verificar uma baixa CTR, os valores da posição média revelam uma boa indexação e posicionamento do conteúdo das páginas. De notar que o facto da presença da pesquisa “areinho arouca” e relacionadas ter resultado num número elevado de impressões, e sabendo que esta está relacionada com um dos maiores fenómenos de popularidade de turismo no ano de 2015 (passadiços do rio

Paiva), confere-se uma boa utilização das palavras chave utilizadas nas paginas de detalhe dos pontos de interesse. A baixa *CTR* pode ser explicada pela existência de concorrência muito forte, uma vez que, tal como ilustrado na figura seguinte, dada uma pesquisa por “quermesse restaurante lisboa” o *website All About Portugal* é listado na posição 9, atrás de plataformas de popularidade mundial como *TripAdvisor*, *Zomato* ou *Facebook*.

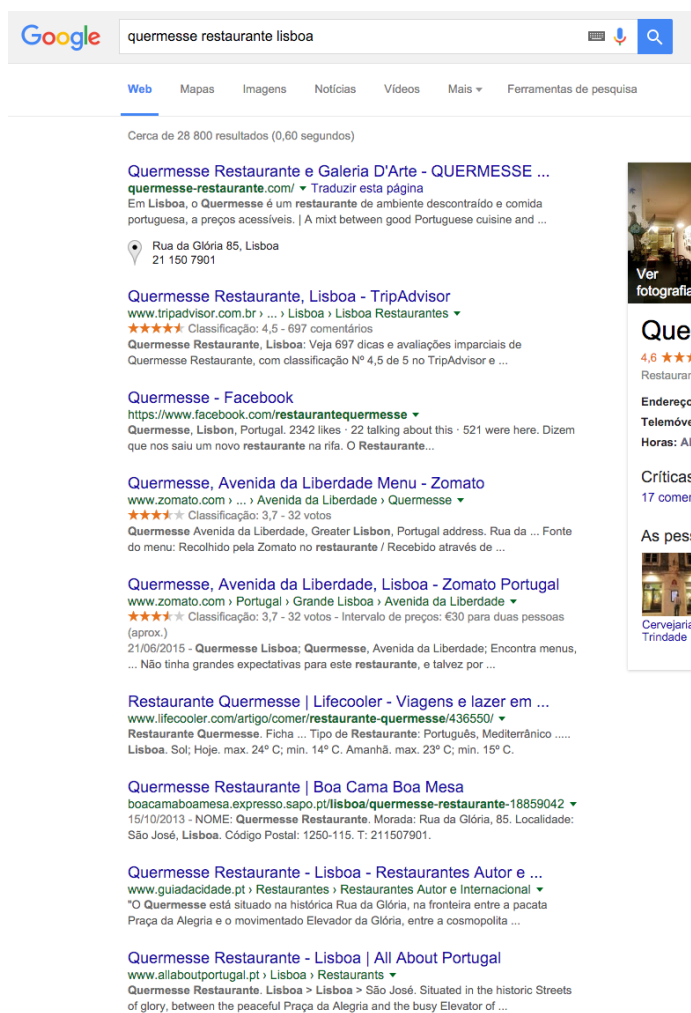


Ilustração 33 - Resultados de pesquisa para "quermesse restaurante lisboa"

Tabela 7 - Termos de pesquisa que resultaram num maior número de cliques

Termos de pesquisa	Cliques	Impressões	CTR	Posição
fintas espinho	68	326	20.86%	4.3
comboio turistico lagos	42	125	33.6%	2.6
poço azul sao pedro do sul	34	719	4.73%	4.6
poço azul viseu	32	479	6.68%	4.5
pensao astoria porto	26	331	7.85%	6.5
motel vila do conde	26	335	7.76%	6.6

praia fluvial odemira	25	71	35.21%	3.2
night star bustelo	24	35	68.57%	1.1
bowling lousada	23	132	17.42%	3.2
aqualine castro marim	23	62	37.1%	1.0

Tabela 8 - Dez páginas com maior número de impressões

URL da página	Impressões	Cliques	Posição média	CTR
http://www.allaboutportugal.pt/arouca/beach/praia-fluvial-do-areinho	5500	170	16,01	3,09
http://www.allaboutportugal.pt/es/sao-pedro-do-sul/recursos-naturales/poco-azul	3500	200	13,61	5,71
http://www.allaboutportugal.pt/setubal/beach/praia-do-galapinhos	3500	90	76,31	2,57
http://www.allaboutportugal.pt/aguada/gardens/parque-de-merendas-da-praia-fluvial-da-redonda	2500	170	13,01	6,80
http://www.allaboutportugal.pt/oliveira-de-frades/beach/praia-fluvial-do-vau	2500	90	35,95	3,60
http://www.allaboutportugal.pt/vila-real-de-santo-antonio/beach/praia-da-fabrica-cacela-velha	2500	30	48,08	1,20
http://www.allaboutportugal.pt/lisboa	1000	35	214,99	3,50
http://www.allaboutportugal.pt/santiago-do-cacem/restaurants/restaurante-do-badoca-safari-park	1000	22	93,58	2,20
http://www.allaboutportugal.pt/abranes/beach/praia-fluvial-da-aldeia-do-mato	900	30	87,29	3,33
http://www.allaboutportugal.pt/castelo-de-paiva/beach/praia-fluvial-da-varzea	900	110	21,72	12,22

As duas tabelas anteriores permitem aferir que tipo público-alvo consumiu mais o conteúdo do *website* ou simplesmente a qual foi possível chegar com esta primeira fase do projeto. Além disso verifica-se que a aposta nos nichos de mercado será mais válida e eficiente do que a aposta em conteúdos de grandes volumes de consumo já dominados e explorados por vários *websites* de larga dimensão e popularidade.

Relativamente aos países de origem das pesquisas, e como demonstra a Ilustração 34, confirmam-se as conclusões referidas no capítulo 4.2.2.

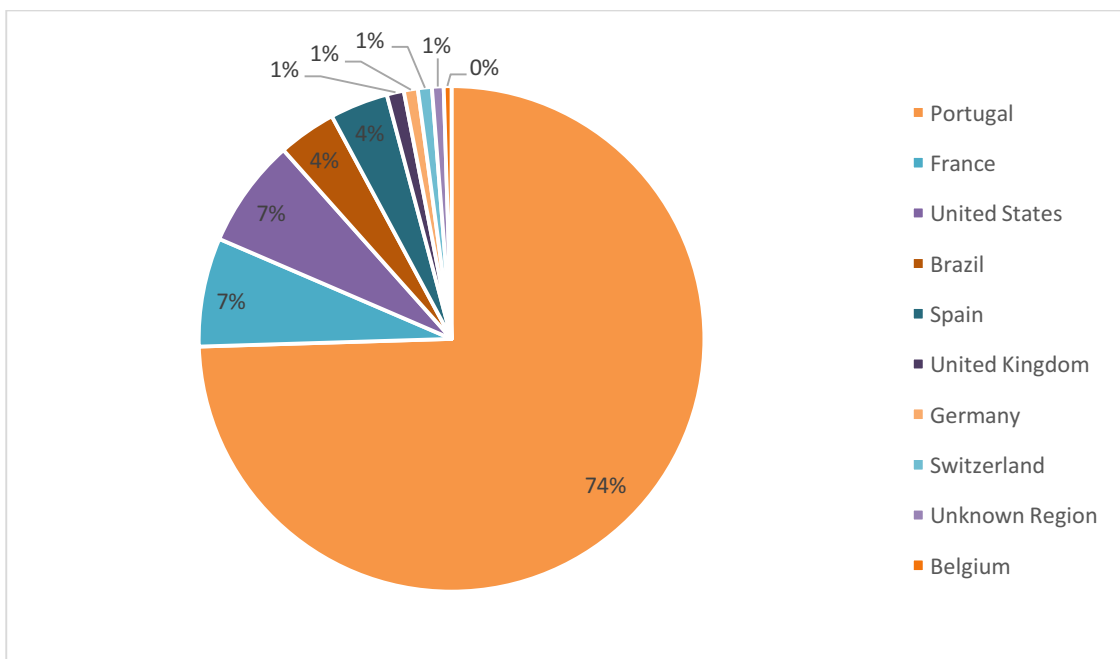


Ilustração 34 - Dez países com mais impressões origem

4.4 Análise de dados comportamentais

4.4.1 Páginas de entrada no portal

É muito importante perceber quais as páginas de entrada mais frequentes nas visitas ao *website*. Isto permite, por exemplo, distinguir as várias sessões de visita com base nos três tipos de pesquisa introduzidos no capítulo 2.2 – transacional, informacional e de navegação.

Na tabela seguinte podemos ver as dez páginas de entrada mais visitadas. Verifica-se que a página mais visitada é, objetivamente, a página inicial. No entanto, este tipo de análise não é relevante visto que os resultados para cada tipo de página têm de ser agrupados por forma a ser possível compreender-se qual o tipo de página mais visitado.

Página	Sessões	Taxa de rejeição
/	3824 (3.97%)	82,43%
/es/sao-pedro-do-sul/recursos-naturales/poco-azul	207 (0.22%)	84,06%
/agueda/gardens/parque-de-merendas-da-praia-fluvial-da-redonda	200 (0.21%)	81,50%
/castro-marim/tourist-attractions/aqualine-parque-aquatico	199 (0.21%)	80,40%

/espinho/tourist-attractions/fintas-parque-infantil-de-diversoes	183 (0.19%)	85,25%
/portimao/bars/exotica-strip-club	171 (0.18%)	85,38%
/arouca/beach/praias-fluvial-do-areinho	165 (0.17%)	84,85%
/lagos/outdoor-activities/comboio-turistico-de-lagos-3	136 (0.14%)	80,15%
/ponta-delgada/bars/showgirls-gentlemans-club	130 (0.14%)	71,54%
/coimbra/bars/impacto-club	127 (0.13%)	90,55%

Tabela 9 - Páginas de entrada no portal mais visitadas

Por forma a agrupar os *URLs* correspondentes a cada tipo de página recorreu-se à ferramenta de criação de relatórios personalizados do *Google Analytics*. Para identificar cada tipo de página através do *URL* foi usada uma expressão regular que definisse o padrão de cada uma:

- Página inicial
 - Exemplos: “/en/”, “/”;
 - Expressão regular: `^/([a-z]{2})?/?$`
- Página de detalhe de destino
 - Exemplos: “/porto”, “/es/porto”;
 - Expressão regular: `^/([a-z]{2})?/?([-A-Za-z0-9]+)$`
- Página de detalhe de ponto de interesse
 - Exemplos: “/es/sao-pedro-do-sul/recursos-naturales/poco-azul”, “/sao-pedro-do-sul/recursos-naturales/poco-azul”;
 - Expressão regular: `^/([-A-Za-z0-9]+)/([-A-Za-z0-9]+)/([-A-Za-z0-9]+)$`
- Página de listagem de restaurantes
 - Exemplos “/es/restaurantes”, “/restaurants”;
 - Expressão regular: `^/([a-z]{2})?/?(restaurants|restaurantes)$`
- Página de listagem de hotéis
 - Exemplos “/es/hoteles”, “/hotels”;
 - Expressão regular: `^/([a-z]{2})?/?(hotels|hoteles)$`

Tipo de página	Número de sessões	Taxa de rejeição
Página de detalhe de POI	88307	77.90%
Página inicial	3824	82.43%
Página de detalhe de destino	609	60.92%
Página de listagem de restaurantes	13	53.85%
Página de listagem de hotéis	14	71.43%

Tabela 10 – Número de sessões a tipos de páginas como páginas de entrada no portal

Há uma larga distância entre a página do tipo de detalhe de ponto de interesse e a página inicial, sendo a primeira, claramente, o tipo de página pelo qual os utilizadores fazem o primeiro contacto com o portal. Por isto pode-se concluir que o maior volume de tráfego é do tipo informacional - os utilizadores vão à procura de uma informação acerca de um determinado ponto de interesse. A taxa de rejeição, neste caso, é relativamente normal visto que o utilizador recolhe a informação que procura e dá como cumprida a tarefa da sua sessão neste *website*. No entanto, seria desejável conseguir que mais utilizadores navegassem pelos conteúdos relacionados.

Se, no caso da página de detalhe ponto de interesse, a elevada taxa de rejeição está de acordo com a natureza do seu conteúdo, visto ser uma página de último nível de navegação e usada muitas vezes para consulta de informação utilitária, no caso da página inicial a taxa de rejeição atual não é aceitável, uma vez que se trata de uma página distribuidora e de acesso aos conteúdos propriamente ditos.

Além disso, seria igualmente desejável que o número de sessões da página de detalhe de destino e das páginas de listagem de restaurantes e hotéis aumentasse. No entanto, verifica-se que uma boa parte dos utilizadores que visitaram estas páginas não terminaram a sua sessão sem interagirem com a mesma. Estas páginas não são muito bem cotadas pelos motores de pesquisa uma vez que têm pouco conteúdo.

4.4.2 Performance

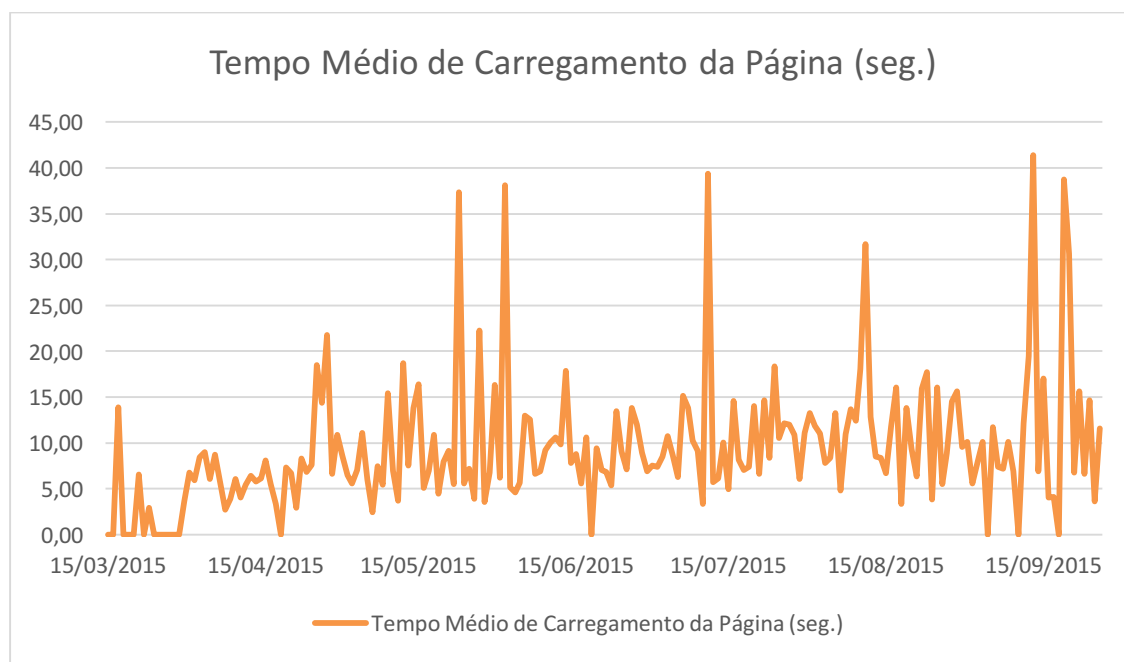


Ilustração 35 - Tempo médio de carregamento de páginas

Olhando para os dados da Ilustração 35, e sabendo que o cálculo do tempo médio de carregamento de página para este intervalo de datas é de 9.15 segundos, verifica-se que este é um dos atributos a melhorar no portal *All About Portugal*.

Em primeiro lugar, são notórios seis picos no gráfico que não estão de acordo com o padrão, momentos estes que podem ter coincidido com deficiências na estrutura de máquinas que sustentam o portal. Se não forem consideradas todas as medições acima de 30 segundos e for recalculada a média de carregamento de páginas esta reduz em 0.6 segundos. Apesar disto o tempo de resposta do servidor é respeitado, sendo, em média exatamente 200ms.

5 Conclusões e trabalho futuro

5.1 Conclusões

O presente estudo teve como intuito estudar o processo de otimização de páginas *web* por forma a afetar positivamente a visibilidade dessas mesmas páginas nos resultados orgânicos de motores de pesquisa, em particular no contexto do turismo e viagens. Sempre tendo em vista o desenvolvimento de um *website* sobre turismo em Portugal, focado no mercado de turismo externo.

O trabalho inicia-se com uma referência à expressão do turismo na economia mundial e à utilização da *Internet* como meio cada vez mais preponderante na procura e consumo de conteúdos relacionados com o planeamento de viagens.

Após uma descrição da motivação e enquadramento do tema, destacando a parceria com a empresa InfoPortugal na disponibilização de alguns dos recursos utilizados no desenvolvimento deste trabalho, começou por se tentar perceber a forma como os motores de pesquisa procuram e interpretam os documentos presentes na *web* e como, posteriormente, constroem os seus *rankings*. De seguida, foi feito um levantamento dos comportamentos mais comuns dos turistas no universo *on-line*, no que às pesquisas diz respeito.

Conclui-se que a posição de uma determinada página *web* numa página de resultados de pesquisa é absolutamente determinante para o seu sucesso na angariação de clientes por esta via, destacando que os resultados mais importantes para a maioria dos utilizadores estão nas primeiras cinco posições da primeira página de resultados de pesquisa.

Uma das fases mais importantes foi a da recolha de informação acerca das métricas usadas pelos motores de pesquisa para efeitos de avaliação de páginas *web*. Tal permitiu perceber quais os fatores mais influentes e, através disso, recolher alguns princípios e orientações de desenvolvimento cujo objetivo é o de conseguir ser

descoberto pelo público-alvo a que se destina o *website All About Portugal* no momento de pesquisas *on-line*.

Seguiu-se uma análise a quatro portais *web* que permitiu perceber, em parte, o mercado de atuação deste projeto. Destaca-se o *TripAdvisor* que é uma plataforma mundialmente reconhecida e líder na resposta à procura de conteúdos relacionados com turismo e viagens. Apesar de ser uma *virtual travel community*, assemelha-se ao *website* desenvolvido no âmbito deste estudo visto ser constituído, também, por um diretório de pontos de interesse. Concluiu-se que o portal *Visit Portugal* não cumpre com algumas das consideradas boas práticas de *SEO*, no entanto revela bons índices de presença em páginas de resultados de pesquisa. Tal deve-se, em grande parte, ao facto de pertencer a um órgão de governamental e por isto ser particularmente forte na métrica relativa ao valor da “marca” e conseguir facilmente um número alargado de referências externas de qualidade.

Uma vez feito o estudo acerca da otimização de páginas *web* orientada a motores de pesquisa, procedeu-se ao desenvolvimento do portal proposto.

Este processo de desenvolvimento envolveu esforços de várias naturezas, tendo sempre em conta o estudo prévio, bem como as conclusões obtidas através de discussão com elementos da empresa InfoPortugal.

Em primeiro lugar foi necessário compreender a estrutura dos conteúdos editoriais que servem o portal *All About Portugal*. Da convergência entre esta análise e as conclusões obtidas no âmbito do estudo prévio – tanto a nível de *SEO* como de mercado – surgem os requisitos do novo portal. Foi desenhada uma estrutura de informação, envolvendo, por exemplo, a criação de categorias agregadoras e respetiva atribuição dos conteúdos existentes e de seguida criada, pelo autor deste documento, a linha gráfica para aquele que é hoje o *All About Portugal*.

Sucederam-se as várias iterações de desenvolvimento que resultaram num produto considerado por todos os elementos envolvidos no processo como satisfatório e que está disponível ao público desde finais de março de 2015.

Os dados de acesso, recolhidos pelas ferramentas *Google Analytics* e *Google Webmaster Tools*, permitiram obter conclusões interessantes acerca do produto desenvolvido. Estes dados são, por si só, material de estudo. São, por isto, não só provas de sucesso da implementação até ao momento de que trata esta tese, como também uma base para trabalho futuro.

O facto de se ter conseguido um nível de visitas diárias relativamente considerável e estável ao longo do tempo, sendo este portal associado a uma marca criada completamente de raiz, dissociada inclusive do nome da empresa e, portanto, absolutamente desconhecida do público, é revelador do sucesso da implementação deste projeto. Destacam-se também os dados demográficos afetos às visitas do portal, uma vez que demonstram o cumprimento do desejado processo de internacionalização com relativo sucesso.

Apesar disto há índices que remetem para falhas ao nível da fidelização de utilizadores, nomeadamente elevadas taxas de rejeição e de novos utilizadores, assim como um número baixo de duração média de sessão e de páginas por sessão.

5.2 Trabalho futuro

Como foi concluído no capítulo 2, o processo de otimização de um determinado *website* aos motores de pesquisa é um processo progressivo, e que requer, portanto, um envolvimento continuado onde se analisem o impacto de cada especificidade do *website*, tanto na forma como os algoritmos dos motores de pesquisa interpretam o conteúdo contido nas páginas do referido portal, como na forma como os utilizadores interagem com o mesmo.

Revendo a análise feita no capítulo 4 é possível apontar uma conjunto de melhorias que contribuam para melhores posicionamentos nos *rankings* dos motores de pesquisa.

Um fator bastante considerado para efeitos de avaliação de páginas *web* é o seu tempo de carregamento, isto porque páginas com um tempo carregamento muito alto são pouco apreciadas pelos utilizadores, levando-os mesmo a desistir de a visitar, sendo por isto considerado até um problema de usabilidade. Posto isto, e reparando que o tempo médio de carregamento de páginas deste portal é de cerca de 9 segundos, será urgente identificar quais os fatores, e as páginas, que estão a contribuir para esta média elevada.

Como já referido anteriormente será importante melhorar o nível de compromisso dos utilizadores com o *website*, isto é, fazer com que naveguem mais entre páginas e que este seja uma referência na procura do tipo de conteúdos de que dispõe. Por forma a atingir este objetivo será necessário realizar testes aos *layouts* dos diferentes tipos de página com o objetivo de perceber quais as disposições dos blocos de informação que contribuam para um aumento da navegação entre páginas do portal.

Por forma a tornar o *All About Portugal* numa plataforma mais completa, seria interessante acrescentar funcionalidades como:

- Possibilidade de os utilizadores avaliarem cada ponto de interesse;
- Permitir uma maior flexibilidade de escolhas na ferramenta que diz respeito ao planeamento de viagem;
- Criação de zona destinada a artigos editoriais mais extensos relacionados com turismo. A qualidade do conteúdo é o fator mais importante de todos para o sucesso de um portal.

Referências

- [Baye et al., 2014] Baye, M.R., Santos, B.D.I. & Wildenbeest, M.R., 2014. 47405 *Search Engine Optimization: What Drives Organic Traffic to Retail Sites?* Indiana: Bloomington Indiana University.
- [Brin & Page, 1998] Brin, S. & Page, L., 1998. *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Tese de Doutorado. Stanford: Universidade de Stanford.
- [Buhalis & Law, 2008] Buhalis, D. & Law, R., 2008. Progress in information technology and tourism management: 20 years on and 10 years after the internet - The state of eTourism research. *Tourism Management*, 17 Janeiro. pp.609-23.
- [Enge et al., 2015] Enge, E., Spencer, S. & Stricchiola, J., 2015. *The Art of SEO: Mastering Search Engine Optimization*. Terceira Edição ed. O'Reilly Media, Inc.
- [Google, 2015] Google, 2015. *Webmaster Guidelines*. [Online] Available at: <https://support.google.com/webmasters/answer/35769?hl=en> [Accessed 30 Maio 2015].
- [Jansen et al., 2000] Jansen, B.J., Spink, A. & Saracevic, T., 2000. Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing & Management*, pp.207-27.
- [Jerath et al., 2014] Jerath, K., Ma, L. & Park, Y.-H., 2014. *Consumer Click Behaviour at a Search Engine: The Role of Keyword Popularity*. Estados Unidos da América: Journal of Marketing Research.
- [Makino & Phan, 2015] Makino, T. & Phan, D., 2015. *Official Google Webmasters Central Blog: Rolling out the mobile-friendly update*. [Online] Available at: <http://googlewebmastercentral.blogspot.pt/2015/04/rolling-out-mobile-friendly-update.html> [Accessed 24 Setembro 2015].
- [Mavridis & Symeonidis, 2015] Mavridis, T. & Symeonidis, A.L., 2015. Identifying valid search engine ranking factors in a Web 2.0 and Web 3.0 context for building efficient SEO mechanisms. *Engineering Applications of Artificial Intelligence*, pp.1-1.
- [Moz, 2015] Moz, 2015. *Ranking Factors Expert Survey - Moz*. [Online] Available at: <https://moz.com/search-ranking-factors/survey> [Accessed 20 Setembro 2015].
- [Pan, 2014] Pan, B., 2014. The power of search engine ranking for tourist destinations.
- [Pan et al., 2006] Pan, B., Litvin, S.W. & Goldman, H., 2006. Real Users, Real Trips, and Real Queries: An Analysis of Destination Search on a Search Engine. In *Annual Conference of Travel and Tourism Research Association*. Dublin, 2006.
- [Pan et al., 2011] Pan, B., Xiang, Z., Law, R. & Fesenmaier, D.R., 2011. The dynamics of search engine marketing for tourist destinations. *Journal of Travel Research*, pp.365-77.
- [Petrescu, 2014] Petrescu, P., 2014. *Google Organic Click-Through Rates in 2014*. [Online] Available at: <https://moz.com/blog/google-organic-click-through-rates-in-2014> [Accessed 23 Setembro 2015].
- [TripAdvisor & StrategyOne, 2013] TripAdvisor & StrategyOne, 2013. *TripBarometer*. Massachusetts: TripAdvisor.
- [TripAdvisor, 2015] TripAdvisor, 2015. *Fact Sheet*. [Online] Available at: http://www.tripadvisor.com/PressCenter-c4-Fact_Sheet.html

- [Wang et al., 2002] [Accessed 27 julho 2015].
Wang, Y., Head, M. & Arthur, N., 2002. E-tailing: an analysis of web impacts on the retail market. *Journal of Business Strategies*, pp.73-79.
- [Wong, 2013] Wong, D., 2013. *Search Traffic vs Scoail Referrals: How Fast Are They Growing?* [Online] Available at: <https://blog.shareaholic.com/search-traffic-social-referrals-12-2013/> [Accessed 15 Setembro 2015].
- [World Travel & Tourism Council, 2014] World Travel & Tourism Council, 2014. *Travel & Tourism Economic Impact 2014 Portugal*. London: World Travel & Tourism Council.
- [Yang & Ghose, 2010] Yang, S. & Ghose, A., 2010. Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependece? *Marketing Science*, pp.602-23.
- [YourTour - Personal Tour Planner, n.d.] YourTour - Personal Tour Planner, n.d. *About YourTour*. [Online] Available at: <http://www.yourtour.com/about-us> [Accessed 25 Setembro 2015].

Anexos

Anexo A – Relação entre categorias de pontos de interesse e categorias agregadoras

Tabela 11 - Atribuição das categorias e respetivos tipos às categorias agregadoras

Tabela	Tipo	Beach	Country	Hotels	Food & Drink	Heritage	Art	Leisure	Shopping
Atividades de passeio	Compras								x
	Conhecer					x			
	Histórico					x			
	Natureza		x						
Aldeias	TODOS		x			x			
Atrações turísticas	Lojas de artesanato					x			x
	Lojas de produtos regionais					x			x
	Artes tradicionais					x			
	Adegas e cooperativas				x	x			
	Caves de vinho do porto				x	x			x
	Lojas gourmet				x				x
	Doces regionais				x	x			x
	Pratos típicos				x	x			
	Aquários	x						x	
	Jardins zoológicos							x	
	Outras atrações							x	
	Parques aquáticos	x						x	
	Parques de diversões							x	
	Parques temáticos							x	
	Pavilhões temático							x	
Planetários							x		
Bem estar	SPAs	x		x					
	Termas	x							
	Talassoterapia								
	Healthclubs								
	Ginásios								
	Outros tratamentos								
Bares e similares	Bares de praia	x			x				
	Bares de strip				x			x	
	Bares de karaoke				x			x	

	Bares e pubs				x				
	Bares gay			x					
	Cafés típicos			x	x				
	Casas de chá			x					
	Casas de fado			x	x	x			
	Caves e lugares de petiscos			x					
	Cervejarias e marisqueiras	x		x					
	Clubes de jazz			x		x			
	Cyber cafés			x					
	Discotecas			x				x	
	Gelatarias típicas			x	x				
	Wine-bars			x					
Desportos	Bowling							x	
	Campos de futebol							x	
	Campos de golfe							x	
	Campos de ténis							x	
	Campos de tiro							x	
	Centros hípicos							x	
	Circuitos de manutenção							x	
	Desportos motorizados							x	
	Desportos náuticos	x							x
	Estádios de futebol								x
	Mini-golfe								x
	Pesca desportiva	x							x
	Piscinas								x
	Recintos polidesportivos								x
	Complexos desportivos								x
Espaços culturais	Bibliotecas					x	x		
	Casinos e bingos							x	
	Centros culturais						x		
	Cinemas						x	x	
	Fundações						x		
	Galerias de arte						x		
	Museus					x	x		
	Outros espaços culturais						x		
	Praças de touros					x			x
	Salas de teatro e espetáculos						x	x	
Hotéis	TODOS			x					
Jardins	TODOS		x					x	
Mercados e feiras	TODOS							x	
Monumentos	TODOS					x			
Parques de campismo	TODOS			x					
Praias e marinas	TODOS	x							
Recursos Naturais			x						

Religião	TODOS					x			
Restaurantes	TODOS				x				

