



Departamento de Ciências e Tecnologias da Informação

Pesquisa de Emprego na Web: Plataforma de agregação de ofertas de emprego e a aceitação destas ofertas no mercado de trabalho

Miguel António Tavares dos Santos Afonso

Dissertação submetida como requisito parcial para obtenção do grau de
Mestre em Informática e Gestão

Orientador:

Prof. Doutor Abílio Gaspar de Oliveira, Professor Auxiliar,

ISCTE-IUL

Coorientador:

Prof. Doutor Luís Filipe da Silva Rodrigues, Professor Auxiliar Convidado,

ISCTE-IUL

Dezembro, 2018

Agradecimentos

Quero agradecer e dedicar esta dissertação exclusivamente a uma pessoa, pessoa essa que teve um grande impacto no meu percurso acadêmico e profissional. Essa pessoa de forma espontânea, natural, ao exercer a sua atividade profissional deixa sempre uma mensagem de motivação, encorajamento, e desafia grande parte das vezes as pessoas que o rodeiam a serem uma melhor versão de si mesmas ou pelo menos a darem um pouco mais de si naquilo onde se encontram envolvidas.

Essa pessoa está presente no meu dia-a-dia deste o momento que tenho consciência de quem sou, e deu-me todas as ferramentas que necessito para ser uma pessoa integra, disciplinada, com objetivos a cumprir, e agradecido por tudo aquilo que tenho.

Essa pessoa nos últimos anos tem sido a minha motivação para que eu dê um pouco mais de mim todos os dias, para tentar ser mais resiliente mentalmente, têm-me mostrado mundos que eu nunca iria conhecer, mas principalmente nunca desistiu do projeto final que temos em comum.

E por fim essa pessoa foi escolhida por mim para eu poder partilhar os meus tempos livres, o meu local de trabalho, e as jornadas e desafios que vou tendo.

E para finalizar fica a faltar dizer quem é essa pessoa, e essa pessoa é o meu orientador. Cada parágrafo aqui descrito é um orientador diferente em áreas diferentes, como o Professor Abílio Oliveira, a minha Família, a minha esposa Ana Margarida, e os meus amigos, respetivamente.

Resumo

Estatísticas revelam que a empregabilidade está a aumentar, apesar de existir uma elevada taxa de desemprego jovem. Por detrás disto, há um claro desajuste entre as qualificações exigidas pelos empregadores e as apresentadas pelos candidatos. Apesar de um aumento ao acesso a todo o tipo de informação, a pesquisa por ofertas de emprego continua a ser um desafio. A requalificação de competências pode ser uma alternativa em uma situação de desemprego, mas só se essas competências forem requeridas pelo mercado de trabalho. Este tipo de situação pode ser controlada, caso a população tenha acesso aos requisitos mais valorizados pelo mercado de trabalho. O nosso objetivo é entender o peso que uma competência tem no mercado de trabalho. Atualmente já existem numerosos *websites* de agregação de ofertas de emprego, como por exemplo o Indeed, Monster ou Trovit. Contudo, nenhum destes analisa as ofertas de emprego a fim de perceber o impacto que uma competência profissional têm no mercado de trabalho. Desenvolvemos um motor de pesquisa de ofertas de emprego, baseado numa metodologia qualitativa, que lista ofertas de emprego, e apresenta qual o impacto que uma competência tem no mercado de trabalho. Técnicas de extração de dados, como *web scraping* ou consumindo interfaces de programação de aplicações (API) são usadas para listar ofertas de emprego. Contudo, a técnica de *web scraping* tem algumas vantagens em relação ao consumo de APIs, pois nem todos os *websites* disponibilizam APIs. Esta nova plataforma *web* desenvolvida pode contribuir para harmonizar a oferta e a procura de competências no mercado de trabalho.

Palavras-Chave: mercado de trabalho, desfasamento oferta-procura, ofertas de emprego, *web scraping*.

Abstract

Statistics have been showing an increase of employability, despite a high unemployment rate seen within young population. Behind this, there is a clear and increased imbalance between the skills employers seek and those that candidates hold. Although a general augmented access to all sorts of information, data screening from uncountable job search websites is still a challenge for a candidate. Requalification may be an alternative option in an unemployment situation, as long as this complements the requisites of the job market. This kind of situations may be controlled, or even avoided, if people have access to the most relevant job market requirements. It is thus our primary aim to understand the weight that a professional skill has in the labor market. Numerous job search websites are already developed and available, such as Indeed, Monster or Trovit. However, none has screened the data in search of understanding the impact of a professional skill has in the labor market. Herein, we developed a job search engine, based on a qualitative methodology such as focus group interviews, that lists the available job positions and highlights the relevance that a professional skill has in the labor market. Extracting data techniques, such as *web scraping* or consuming applications programming interfaces (API), have been used. However, the *web scraping* approach shows some advantages as some of the websites do not provide APIs to extract or share data. This new web platform may provide potential candidates a set of key professional skills demanded by employers, by efficiently selecting the proportion of job positions that request specific skills, which can be fulfilled by the candidate or eventually be further developed for future reference. This may contribute to harmonize supply and demand of skills in the labor market.

Keywords: job market, academia-employment mismatch, job offers, *web scraping*.

Índice

Departamento de Ciências e Tecnologias da Informação	i
Agradecimentos	i
Resumo	ii
Abstract	iii
Índice.....	iv
Índice de Abreviaturas	vi
Índice de Figuras.....	vii
Índice de Tabelas.....	ix
1. Introdução.....	1
Motivação e Delimitação do Tema	1
Questão de investigação e Objetivos	2
Abordagem Metodológica	3
Estrutura da Dissertação.....	4
2. Estado da Arte	7
Gestão de Recursos Humanos	7
Desfasamento Oferta-Procura.....	15
Pesquisa de Emprego em Portugal	19
Agregadores de ofertas de emprego	20
<i>Web Scraping</i>	20
Conclusão do Capítulo	26
3. Conceptualização e Modelação do Protótipo	27
Levantamento dos requisitos	27
Escolha dos <i>websites</i> de ofertas de emprego em Portugal	30
Estudo da árvore HTML do <i>website</i> de ofertas de emprego.....	34
Forma de estudo.....	34
Estudo da árvore HTML do <i>website</i> Net-Empregos	36
Construção do caminho XPath.....	36
Estudo do URL do <i>website</i> Net-Empregos	38
Esquema de Funcionamento da Aplicação de Agregação de Conteúdo	39
Descrição do funcionamento da aplicação de agregação de conteúdo	40
Conclusão do Capítulo.....	41
4. Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web	43
Construção da plataforma de agregação de ofertas de emprego	43
Framework .net Core 2.0.....	43
Camada de apresentação	43
View	43
Partial View.....	44
Controller.....	44

Bootstrap.....	45
Camada de negócio	45
Algoritmo de extração de dados	46
Propriedades	47
Variáveis	48
Método de extração das páginas (URLs) onde estão publicadas as ofertas alvo...	49
Método de extração de ofertas segundo uma página (url).....	51
Método de agregação de todas as ofertas alvo	54
Método de contagem das ofertas totais num distrito	55
Método de agregação de todas as ofertas	56
Conclusão do Capítulo	58
5. Avaliação e Teste do Protótipo	59
Metodologia adotada	59
Testes e Resultados	59
Focus Group	61
Conclusão do Capítulo.....	63
6. Discussão do trabalho desenvolvido	65
Solução Final	65
Dos objetivos à questão de investigação.....	65
Conclusão do Capítulo	66
7. Conclusões	67
Conclusões	67
Contribuição	67
Trabalho Futuro	68
Referências Bibliográficas.....	69
Anexos	75
Anexo A – Árvore HTML dos <i>websites</i> “indeed.pt” e “emprego.sapo.pt”	75
Árvore HTML "indeed.pt"	75
Árvore HTML " emprego.sapo.pt"	75
Anexo B – Algoritmo de extração da plataforma	76
Anexo C – Dados focus group	77
Focus Group – Levantamento dos requisitos	77
Focus Group – Avaliação da plataforma	77

Índice de Abreviaturas

API - Application programming interface

ASCII - American Standard Code for Information Interchange

CSS - Cascading Style Sheets

DOM - Document Object Model

HTML - Hypertext markup language

MVC - Model View Controller

RH - Recursos Humanos

RSS - Rich Site Summary

URL - Uniform Resource Locator

W3C - World Wide Web Consortium

WWW - World Wide Web

XML - Extensible markup language

XPath - XML path language

Índice de Figuras

Figura 1. As quatro atividades primárias e as suas componentes de Gestão de Recursos Humanos (DeCenzo, Robbins & Verhulst, 2013).....	8
Figura 2. Resumo visual de Anna Holm do processo de recrutamento tradicional (Holm, 2012)....	10
Figura 3. Cloud Tag resultante da questão “Quais os aspetos mais importantes que uma plataforma web de agregação de ofertas de emprego deve ter?”. Fonte: www.wordclouds.com. .	28
Figura 4. Cloud Tag resultante da questão "Quais os aspetos mais fracos que considera existirem nas plataformas web de agregação de ofertas de emprego?". Fonte: www.wordclouds.com.....	28
Figura 5. Cloud Tag resultante da questão "Quais considera serem os benefícios do uso das plataformas web de agregação de ofertas de emprego no auxílio de novas oportunidades de emprego?". Fonte: www.wordclouds.com.	29
Figura 6. Cloud Tag resultante da questão "Quais considera serem os inconvenientes do uso das plataformas web de agregação de ofertas de emprego no auxílio de novas oportunidades de emprego?". Fonte: www.wordclouds.com.	30
Figura 7. Opção Inspect do navegador web Google Chrome.	35
Figura 8. Figura com a árvore HTML: exemplo de uma oferta – tem data, descrição e link para a oferta.	35
Figura 9. Exemplo de um caminho XPath de uma oferta de emprego.	36
Figura 10. Apresentação de ofertas de emprego do website “net-empregos.com”.	37
Figura 11. Nó HTML de uma oferta presente no website “net-empregos.com”.	37
Figura 12. Elementos HTML a serem manipulados. Fonte: website “net-empregos.com”.....	38
Figura 13. Formulário HTML responsável por recolher os parâmetros de pesquisa definidos pelo utilizador. Fonte: Elaboração própria.	43
Figura 14. Tabela HTML responsável por conter as ofertas de emprego recolhidas.....	44
Figura 15. Controladores responsáveis pela alimentação das views com o model que contém os dados. Fonte: Elaboração própria.....	45
Figura 16. Criação de atributos do tipo class dentro dos elementos HTML div para utilizar o CSS pré-definido na biblioteca bootstrap. Fonte: Elaboração própria.....	45
Figura 17. Hierarquia de métodos presentes no algoritmo de extração de dados (extração net-empregos.com). Fonte: Elaboração própria.....	46
Figura 18. Propriedades e variáveis globais presentes no algoritmo de extração de dados (extração net-empregos.com). Fonte: Elaboração própria.	47
Figura 19. Propriedades presentes no algoritmo de extração de dados. Fonte: Elaboração própria.	47
Figura 20. Variáveis presentes no algoritmo de extração de dados (extração net-empregos.com). Fonte: Elaboração própria.	48
Figura 21. Assinatura do método “ExtrairUrlsNetEmpregos”. Fonte: Elaboração própria.	49
Figura 22. Objeto que contém os endereços (URLs) das páginas onde estão publicadas as ofertas a serem alvo de extração. Fonte: Elaboração própria.	49
Figura 23. Tratamento dos parâmetros de entrada descrição e distrito. Fonte: Elaboração própria.	50
Figura 24. Definição / Criação das urls das páginas alvo de extração.	50
Figura 25. Objeto do tipo HtmlWeb com a definição de uma atribuição de encoding à propriedade OverrideEncoding. Fonte: Elaboração própria.	50
Figura 26. Extração da página seguinte utilizando a biblioteca HTML Agility Pack. Fonte: Elaboração própria.....	51
Figura 27. Assinatura do método “ExtrairOfertasNetEmpregos”. Fonte: Elaboração própria.....	52
Figura 28. Objeto responsável por guardar todas as ofertas de emprego extraídas. Fonte: Elaboração própria.....	52
Figura 29. Objeto do tipo HtmlWeb com a definição de uma atribuição de encoding à propriedade OverrideEncoding. Fonte: Elaboração própria.	52
Figura 30. Extração da página alvo. Fonte: Elaboração própria.	53
Figura 31. Localização dos nós alvo na árvore HTML. Fonte: Elaboração própria.....	53

Figura 32. Validação para posterior extração dos dados associados a cada oferta de emprego. Fonte: Elaboração própria.....	53
Figura 33. Preenchimento das variáveis e propriedade necessárias para guardar o resultado final do método “ExtrairOfertasNetEmpregos”. Fonte: Elaboração própria.....	54
Figura 34. Assinatura do método “ExtrairTotalNetEmpregos”. Fonte: Elaboração própria.	54
Figura 35. Corpo do método “ExtrairTotalNetEmpregos”. Fonte: Elaboração própria.	55
Figura 36. Método “VerificarTotalOfertasNetEmpregos”. Fonte: Elaboração própria.	56
Figura 37. Assinatura do método “AgregaarOfertas”. Fonte: Elaboração própria.....	56
Figura 38. Criação das threads responsáveis pela aceleração do algoritmo de extração de dados. Fonte: Elaboração própria.....	57
Figura 39. Agregação de todas as ofertas alvo no mesmo objeto. Fonte: Elaboração própria.	58
Figura 40. Cálculo da percentagem de ofertas relevantes ao utilizador em relação ao total de ofertas presentes no distrito alvo e retorno do método. Fonte: Elaboração própria.....	58
Figura 41. Cloud Tag resultante da questão "Quais as mais-valias que considera existirem na plataforma web recém-criada em relação às suas congéneres?". Fonte: www.wordclouds.com . ..	61
Figura 42. Cloud Tag resultante da questão "Quais os aspetos mais fracos que considera existirem na plataforma web recém-criada em relação às suas congéneres?". Fonte: www.wordclouds.com	62
Figura 43. Cloud Tag resultante da questão "Quais considera serem os benefícios do uso da plataforma web recém-criada no auxílio de novas oportunidades de emprego?". Fonte: www.wordclouds.com	62
Figura 44. Cloud Tag resultante da questão "Quais considera serem os inconvenientes do uso da plataforma web recém-criada no auxílio de novas oportunidades de emprego?". Fonte: www.wordclouds.com	63

Índice de Tabelas

Tabela 1. Descrição Indeed. Fonte: Elaboração própria.....	20
Tabela 2. Descrição Monster. Fonte: Elaboração própria.....	20
Tabela 3. Descrição Simply Hired. Fonte: Elaboração própria.	20
Tabela 4. Descrição Trovit. Fonte: Elaboração própria.	20
Tabela 5. Questões legais e contratuais relacionadas com o processo de Web Scraping (Vanden & Baesens, 2018).....	26
Tabela 6. Palavra-chave "emprego". Fonte: Elaboração própria.....	31
Tabela 7. Palavra-chave "trabalho". Fonte: Elaboração própria.....	31
Tabela 8. Palavra-chave "ofertas de emprego". Fonte: Elaboração própria.....	31
Tabela 9. Palavra-chave "ofertas de trabalho". Fonte: Elaboração própria.....	32
Tabela 10. Websites resultantes da pesquisa. Fonte: Elaboração própria.....	33
Tabela 11. Número de ofertas de emprego, nos últimos 7 dias, nos websites "net-empregos.com", "indeed.pt" e "emprego.sapo.pt". Fonte: Elaboração própria.....	33
Tabela 12. Estrutura URL do website "net-empregos.com". Fonte: Elaboração própria.....	38
Tabela 13. Exemplos de URLs do website "net-empregos.com". Fonte: Elaboração própria.....	38
Tabela 14. Bateria de testes unitários. Fonte: Elaboração própria.....	60
Tabela 15. Bateria de testes de integração. Fonte: Elaboração própria.....	60
Tabela 16. Bateria de testes funcionais. Fonte: Elaboração própria.....	60
Tabela 17. Bateria de testes de aceitação. Fonte: Elaboração própria.....	61

1. Introdução

O aparecimento das tecnologias de informação no final do século XX permitiu uma explosão de conhecimento a nível mundial, acessível a quase todas as faixas sociais e etárias (Castells, 2011). Numa nova era, a 4ª revolução industrial, grande parte das nossas decisões são fundamentadas com o recurso às tecnologias em geral (World Economic Forum, 2016). Contudo, o crescente fluxo de informação torna-se muitas vezes difícil de filtrar, devido à sua extensão, por ser dispersa e, por vezes, de difícil leitura (Andrejevic, 2013). É, portanto, da responsabilidade das plataformas de informação *Web*, não só disponibilizar e potenciar o acesso a informação ao utilizador, como também ser capaz de o auxiliar na triagem dessa informação. Várias têm sido as plataformas criadas nesse sentido, incluindo os motores de busca *web* ou *websites* que auxiliam na tomada de decisões, compilando informação sobre um tema em específico, como o Booking.com, o Tripadvisor.com, para a pesquisa de hotéis, e o Indeed.com, na procura de emprego.

As plataformas *web* de auxílio à tomada de decisão, no que toca à procura de emprego, podem ajudar os seus utilizadores na procura de ofertas de emprego. Este trabalho pretende verificar se um agregador de ofertas pode contribuir para a harmonização das competências académicas existentes e as exigidas pelo mercado de trabalho. A motivação nasceu da problemática relacionada com o elevado desemprego jovem observado em toda a Europa atualmente. A taxa de desemprego jovem com idade inferior a 25 anos apresenta valores superiores a 20% e, em alguns países europeus, são mesmo superiores a 30% (International Labour Office, 2017). Vários estudos revelam ainda que muitos destes jovens possuem formação académica superior. Nesses estudos, é também perceptível a escassez de trabalhadores qualificados nas áreas da engenharia e das tecnologias da informação por toda a Europa, o que resulta num desfazamento entre o que o mercado precisa e o que oferece aos empregadores (Zimmermann, 2009; Eichhorst, Hinte & Rinne, 2013; Liu, Salvanes, & S & Snes 2016; Santos, 2016). Existe, pois, uma necessidade urgente de esclarecer a população, não só a camada jovem, sobre quais as necessidades do mercado de trabalho, na tentativa de harmonizar a procura com a oferta. Em Portugal, o desemprego atingiu o valor histórico de 16,2%, em 2013, e, apesar de em 2017 a taxa de desemprego apresentar o valor de 8,9%, continua a ser importante fazer as escolhas académicas ou de formação profissional certas para o futuro (OECD, 2018).

Motivação e Delimitação do Tema

Verifica-se que a satisfação e o rendimento de um trabalhador é superior se este estiver a desempenhar tarefas na sua área de competência (Allen & Van der Velden, 2001; Taris & Schaufeli, 2015; Plomp et al., 2016). A própria requalificação de um indivíduo que se encontre em situação de desemprego só é útil se as novas competências adquiridas derem resposta às necessidades do mercado de trabalho. Caso contrário, o esforço despendido, bem como os incentivos monetários do Estado, acabam por ser desperdiçados. Estes cenários poderão ser controlados caso a população

Introdução

tenha acesso, de uma forma clara e explícita, às exigências atuais do mercado de trabalho. Deste modo, cada indivíduo poderá adquirir e aprofundar determinadas competências profissionais, sociais e pessoais e adaptá-las às tendências de mercado existentes (Porter & Millar, 1985; Kalleberg, 2018).

A motivação deste trabalho relaciona-se com a tentativa de harmonizar o desfasamento atual que existe entre ensino-emprego, mostrando ao comum cidadão qual a aceitação por parte do mercado de trabalho de determinada formação. Isto é, um indivíduo que tenha ou pretenda vir a adquirir determinada formação académica ou profissional conseguirá verificar qual a aceitação dessa formação no universo das ofertas de emprego atuais disponíveis. Olhando para as plataformas atuais de agregação de ofertas de emprego, verifica-se que as mesmas se limitam a apresentar as ofertas de emprego ao utilizador, segundo os parâmetros por ele definidos. O que se pretende com este trabalho é agregar também as ofertas de emprego pertinentes ao utilizador, assim como mostrar a aceitação dos critérios de pesquisa do utilizador por parte do mercado de trabalho.

Questão de investigação e Objetivos

Nesta dissertação, a grande questão que se coloca é:

- Até que ponto um agregador de ofertas contribui para a harmonização das competências académicas existentes e as exigidas pelo mercado de trabalho?

O que passa por tentar desenvolver uma plataforma digital que auxilie na pesquisa *online* de ofertas de emprego, bem como na verificação da aceitação por parte do mercado de trabalho de uma determinada competência profissional e/ou académica. Em sequência, foram estabelecidos os seguintes objetivos:

Objetivos Teóricos:

- Compreender como surge uma oferta de emprego;
- Entender o conceito de *desfasamento ensino-emprego*;
- Verificar como se realiza o processo de extração de informação na web.

Objetivos Práticos:

- Fazer o levantamento de requisitos para a criação de uma plataforma de pesquisa e de agregação de ofertas de emprego;
- Fazer o estudo das árvores HTML dos *websites* alvo de extração para proceder à criação de um algoritmo de pesquisa e de agregação de ofertas de emprego;
- Conceptualizar a plataforma;
- Desenvolver um protótipo;
- Testar e avaliar esse protótipo, a fim de verificar se o mesmo consegue responder à questão de investigação.

Introdução

Abordagem Metodológica

De modo a concretizar estes objetivos, será criada uma plataforma Web que agregue ofertas de emprego contidas em *websites* de anúncios de emprego em Portugal e, simultaneamente, verifique a aceitação de uma competência profissional e/ou académica por parte do mercado de trabalho. O propósito final deste trabalho é orientar o comum Cidadão que procura emprego, fornecendo-lhe um único local na Web que agregue o maior número de ofertas de emprego em Portugal e forneça a percentagem de emprego para um determinado parâmetro de pesquisa. Uma plataforma web com estas características torna-se uma mais-valia para o Cidadão que esteja desempregado, empregado ou mesmo em processo de requalificação. As vantagens desta plataforma são várias, designadamente a velocidade de pesquisa das ofertas de emprego e posterior candidatura, uma vez que as probabilidades de uma candidatura ser analisada por um empregador são mais altas para aqueles que submetem as suas candidaturas mais cedo. Outra vantagem é a profundidade da pesquisa, isto é, visto que a plataforma de agregação de ofertas de emprego extrai as suas ofertas de vários *websites*, o utilizador acaba por ter acesso a mais ofertas de emprego, consultando apenas num único local que compara vários *websites* especializados, evitando que lhe escape algum anúncio ou desista da pesquisa. Um fator de inovação desta plataforma é calcular a percentagem de incidência de determinada função no universo das ofertas de emprego. Isto é, o utilizador ao fornecer uma descrição para uma pesquisa de emprego, a plataforma irá calcular a percentagem de ofertas de emprego devolvidas em relação ao total de ofertas. Desta forma o utilizador ficará a saber o impacto que determinada descrição (ex.: emprego, competência, curso superior) tem no mercado de trabalho.

Pretende-se que a aplicação web a desenvolver mostre, num único local, as ofertas de emprego publicadas pelos empregadores, num determinado período de tempo, bem como a percentagem de empregos que correspondem aos parâmetros de pesquisa definidos pelo utilizador. Esta ferramenta irá pesquisar por palavras-chave (descrição da função e localização) em *websites* de procura de emprego, tais como net-empregos.com, indeed.com, emprego.sapo.com, para posteriormente devolver ao utilizador,—todas as ofertas de emprego contidas nesses *websites*, respeitando os parâmetros de pesquisa fornecidos, bem como a percentagem de ofertas de emprego que respeitem esses parâmetros.

Este trabalho inicia-se com a revisão da literatura necessária para o enquadramento da temática e a posterior elaboração de um agregador de conteúdo web, atendendo aos temas de recrutamento, desfasamento ensino-emprego e extração de dados na *web*. A fase seguinte será de criação do algoritmo de extração de dados, como o pilar basilar da plataforma proposta nesta dissertação. Esse algoritmo será criado com recurso a uma das bibliotecas de extração de dados apresentada do capítulo de revisão bibliográfica. A metodologia utilizada para a criação desta plataforma/algoritmo é uma metodologia qualitativa, com recurso a um *focus group*. Na fase que se segue, será feita a implementação das fases anteriores, onde será criada a aplicação propriamente dita. De seguida,

Introdução

teremos a fase de testes, que servirá para realizar o controlo da qualidade da aplicação. Serão feitos testes unitários, de integração, funcionais e aceitação, bem como um segundo *focus group* de forma a avaliar a aceitação da plataforma recém-criada. O sucesso da plataforma é medido caso a mesma, de forma célere, consiga devolver ao utilizador todas as ofertas de emprego contidas nos *websites* que alimentam esta mesma plataforma, segundo os parâmetros de pesquisa definidos pelo utilizador, bem como calcular a percentagem de ofertas de emprego devolvidas em relação ao universo total de ofertas. Por fim, a plataforma será disponibilizada ao utilizador final. Esta plataforma Web será realizada com recurso a ferramentas/tecnologias de código aberto ou de acesso livre tais como: ASP.NET Core MVC2; Bootstrap 4; HTML Agility Pack.

Estrutura da Dissertação

A dissertação a apresentar será dividida em 7 capítulos:

- Introdução: Este capítulo contém uma breve introdução ao tema proposto a ser discutido, a delimitação desse mesmo tema, bem como a motivação e objetivos da dissertação.
- Estado da arte: Neste capítulo é realizado o levantamento bibliográfico sobre dois grandes temas: Gestão de recursos humanos com foco no recrutamento *online* e Extração de dados na Web.
- Conceptualização e modelação do protótipo: Capítulo dedicado ao levantamento dos requisitos necessários para a realização do algoritmo de extração de dados. É realizado um *focus group* a fim de fazer um levantamento de requisitos que facilite a adequação da futura plataforma à realidade existente no mercado de trabalho. São definidos quais os *websites* de ofertas de emprego a serem alvo de extração de dados e é estudado em detalhe a sua árvore *hypertext markup language* (HTML) para preparar o caminho onde o algoritmo de extração irá atuar.
- Desenvolvimento de um protótipo de extração de ofertas de emprego na *web*: Capítulo reservado à apresentação da construção do algoritmo que irá extrair todos os dados necessários dos *websites* de ofertas de emprego.
- Avaliação e teste do protótipo: Capítulo onde são apresentados os resultados dos testes de controlo da qualidade da plataforma recém-criada. Serão realizados testes unitários, de integração, funcionais e aceitação, bem como um *focus group* de forma a avaliar a aceitação da plataforma recém-criada. Espera-se obter um protótipo estável, ágil e que reflita as pesquisas segundo os *websites* que estão a ser consumidos.
- Discussão do trabalho desenvolvido: Capítulo destinado à discussão do trabalho desenvolvido e verificação do cumprimento dos objetivos propostos.

Introdução

- Conclusões: Capítulo dedicado à conclusão da dissertação, onde é dada a resposta à questão de partida definida no capítulo introdutório, bem como apresentadas críticas ao presente trabalho e sugestões para trabalhos futuros.

Introdução

2. Estado da Arte

Neste capítulo, realizar-se-á uma revisão bibliográfica assente em três temas. O primeiro tópico a ser abordado prende-se com o processo de recrutamento e recursos humanos. É durante o processo de recrutamento que nascem as ofertas de emprego publicadas na web. O segundo tópico está relacionado com o desfasamento ensino-emprego, tentando-se perceber em que consiste, as suas causas e as consequências do desfasamento entre a oferta e a procura no mercado de trabalho. O terceiro tópico a ser discutido está relacionado com a extração de dados na Web, mais especificamente, a extração de dados em *websites* de publicação de ofertas de emprego. Esta revisão da literatura servirá de base teórica para o desenvolvimento de uma plataforma de agregação de ofertas de emprego.

Gestão de Recursos Humanos

Gestão de Recursos Humanos (ou simplesmente RH) tem vários significados. Michael Armstrong propõe uma definição clara e objetiva sobre o que são RH (Armstrong, 2006). Afirmando que se trata de uma abordagem estratégica e coerente para a gestão dos ativos mais valiosos de uma organização – as pessoas que lá trabalham, que individual e coletivamente contribuem para a realização dos objetivos do negócio a que se dedicam. Por outro lado, RH refere-se a todas as atividades associadas à gestão do trabalho e às pessoas nas organizações (Boxall & Purcell, 2011).

Por vezes, a Gestão de Recursos Humanos é usada com o mesmo significado de “gestão de pessoal” (Armstrong, 2006). A gestão de pessoal, agora um nome obsoleto para definir gestão de recursos humanos, evoluiu com as novas exigências do mercado de trabalho por leis implementadas nos Estados Unidos da América na década de 60 (DeCenzo, Robbins & Verhulst, 2013). Quando os RH foram introduzidos nos Estados Unidos, através de um curso na Harvard Business School em 1981, a nova abordagem mudou o futuro da “gestão de pessoas” para sempre. A gestão de pessoal agora está restrita ao aspeto de pessoas apenas (Price, 2007).

O departamento de RH dentro de uma organização tem um papel extremamente importante e crucial para o desempenho da mesma. As atividades de RH concentram-se, aproximadamente, em quatro atividades gerais, conforme ilustra a Figura 1. As quatro atividades principais consistem em:

1. Pessoal
2. Treino e Desenvolvimento
3. Motivação
4. Manutenção

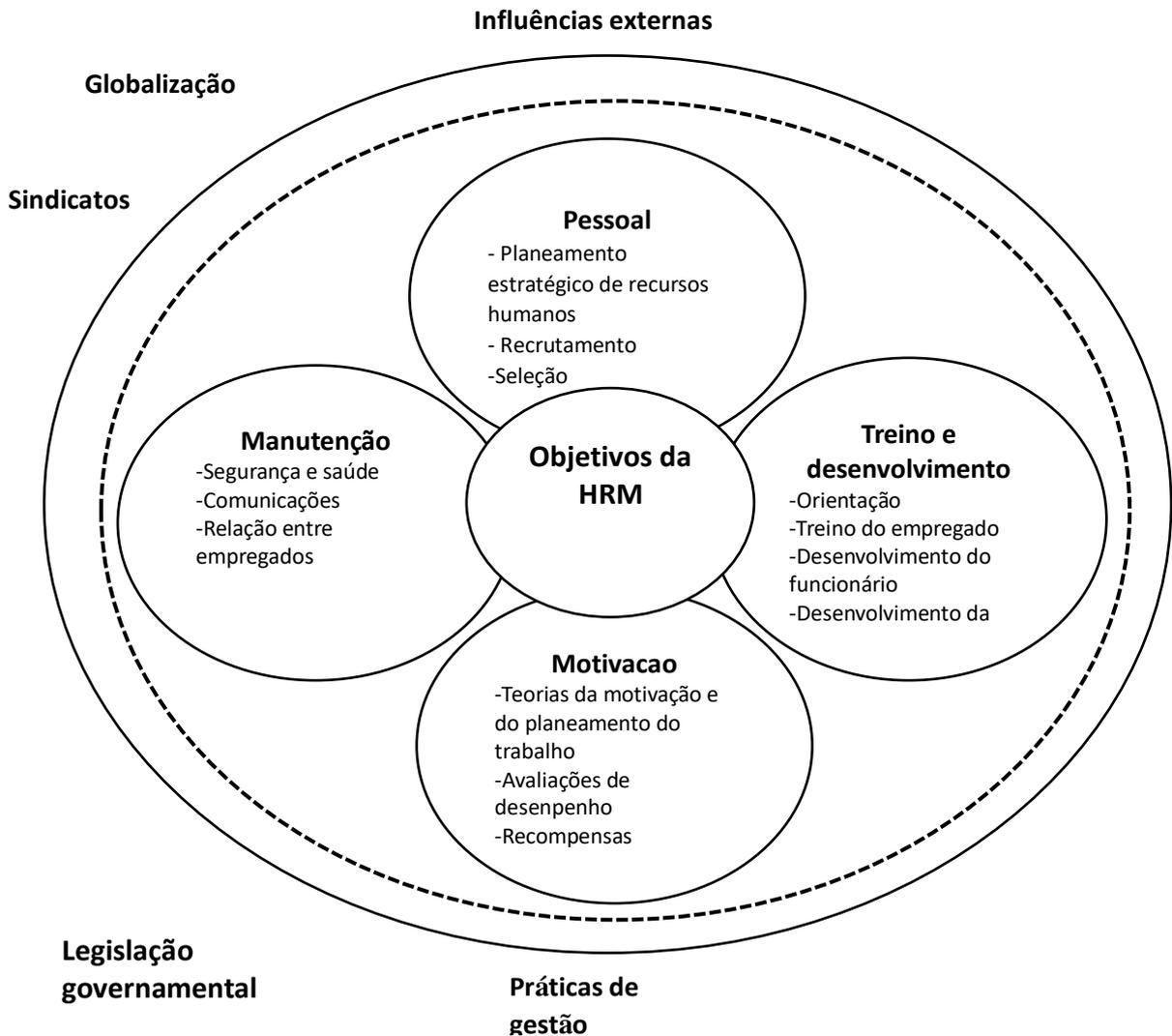


Figura 1. As quatro atividades primárias e as suas componentes de Gestão de Recursos Humanos (DeCenzo, Robbins & Verhulst, 2013).

Todas as atividades de RH precisam de ser integradas no processo de gestão estratégica das organizações. Cada função tem um peso para o sucesso final da organização. Para o propósito desta dissertação, só será explorada a atividade descrita anteriormente como “Pessoal”.

Pessoal

O setor de pessoal começa com o planeamento de recursos humanos. O planeamento de recursos humanos deve garantir que a organização tem as estratégias de longo prazo e de curto prazo no que toca à previsão do tipo e da quantidade certa de colaboradores para os cargos de apoio ao sucesso da organização. Um bom planeamento organizacional prepara os RH com uma visão das necessidades atuais e futuras de pessoal necessário dentro da organização (DeCenzo, Robbins & Verhulst, 2013).

Recrutamento e Seleção

Geralmente, “Recrutamento e Seleção” são reconhecidos como uma função conjunta. Como estas duas funções têm responsabilidades diferentes, é justo dar definições separadas.

O recrutamento não é um processo simples que possa ser descrito numa só frase. As estratégias de recrutamento diferem em todo o mundo, mas os elementos comuns são: atrair, encontrar e adquirir. O recrutamento é frequentemente relatado na literatura como o processo de análise dos requisitos de trabalho, reunindo uma rede de candidatos qualificados e contratando a pessoa mais adequada para o papel, a fim de obter uma vantagem competitiva (Boxall & Purcell, 2011).

A seleção é o processo de escolher a pessoa mais adequada para o papel, entre o grupo gerado de candidatos qualificados (Bratton & Gold, 2017). Estes dois processos complementam-se; sem recrutamento, não há seleção.

O Processo de Recrutamento

O recrutamento traz, essencialmente, talentos para a organização. O recrutamento pode fazer parte de uma das tarefas do departamento de RH ou pode ser realizado com recurso a processos de subcontratação, por meio de uma empresa externa à organização. Essas empresas externas substituem ou atuam de forma semelhante a um recrutamento interno, procurando e atraindo novos colaboradores para uma organização. Independentemente da abordagem, com recurso à subcontratação ou não, o processo de recrutamento segue um fluxo padronizado. Lee observou, no seu estudo sobre a evolução do recrutamento eletrónico (*E-recruiting*), que o processo tradicional de recrutamento segue um processo sequencial passo a passo (Lee, 2005). Na pesquisa de Anna Holm sobre *E-recruiting* e o processo de recrutamento, esta autora sintetizou (Figura 2) do processo tradicional de recrutamento em papel que foi adaptado de várias fontes de gestão de recursos humanos, incluindo a versão de Lee (Holm, 2012).

Estado da Arte

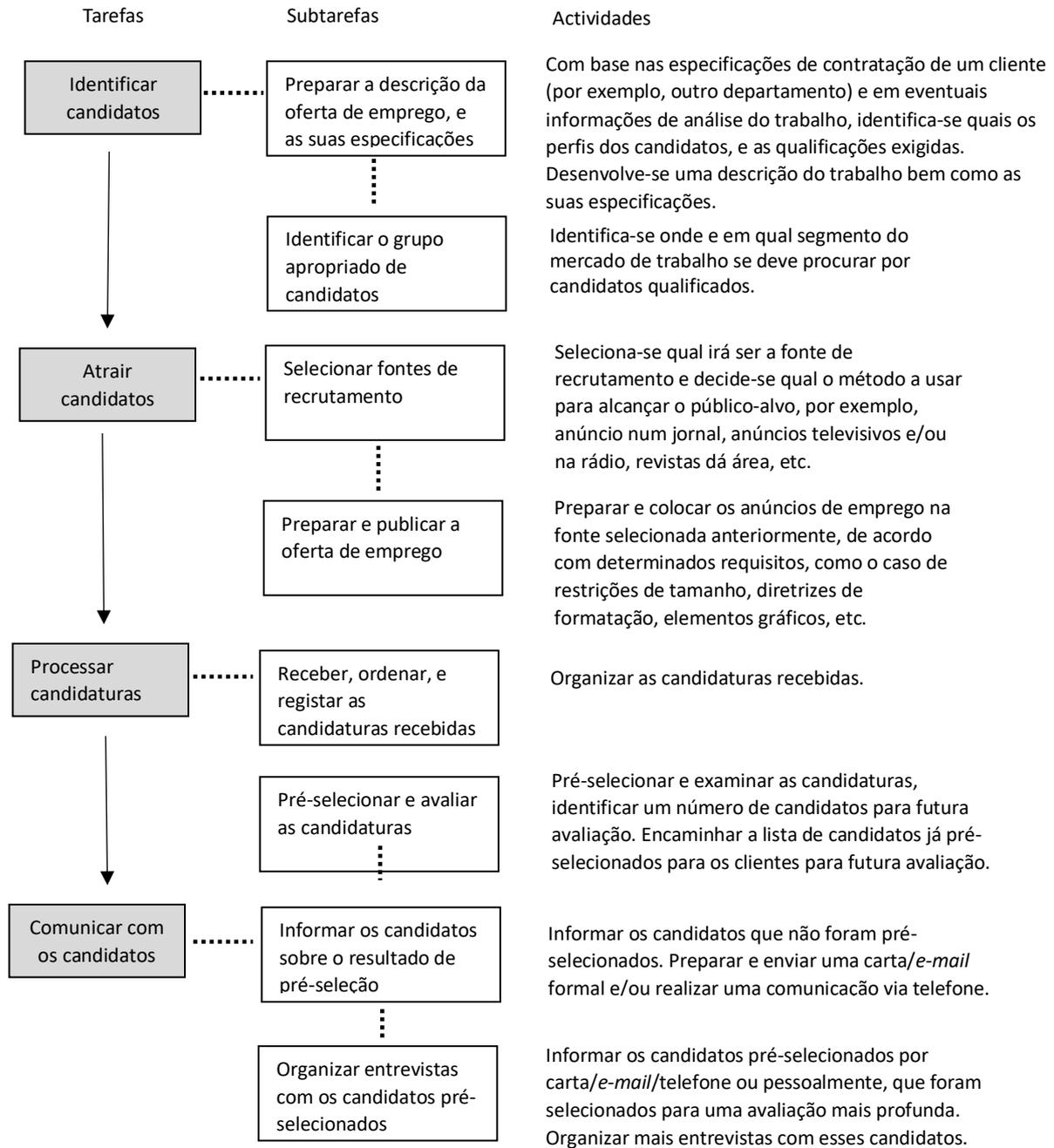


Figura 2. Resumo visual de Anna Holm do processo de recrutamento tradicional (Holm, 2012).

Holm relatou que a primeira tarefa do processo de recrutamento é identificar as necessidades de contratação (Holm, 2012). Uma vez identificadas as necessidades de contratação, é crucial realizar a subtarefa de criação de uma descrição do cargo, especificações do trabalho e identificação do grupo apropriado de candidatos. O responsável pela nova vaga de trabalho procede à revisão da descrição dos conhecimentos e habilitações essenciais para o cargo. Existem muitos métodos, por exemplo: a entrevista individual e métodos de entrevista em grupo, onde um indivíduo ou grupo de pessoas, em uma função similar, podem discutir com os Recursos Humanos as especificações e expectativas para o cargo. Quando uma função não existe, o método de conferência técnica é

valioso, pois permite que especialistas no assunto que possuem amplo conhecimento das principais funções do cargo possam contribuir para os Recursos Humanos (DeCenzo, Robbins & Verhulst, 2013).

A segunda etapa do processo de recrutamento de Holm é atrair potenciais candidatos através de um anúncio de emprego. Para a divulgação eficaz de um anúncio de trabalho, este deve ser atrativo, apelativo e claro. (Holm, 2012).

A terceira tarefa é o processamento das candidaturas recebidas, que consiste em receber, classificar, pré-selecionar e avaliar os candidatos. Durante esta tarefa, é importante que o recrutador comunique com o gestor responsável pela contratação. O gestor ajudará o recrutador a formular o próximo passo para a próxima tarefa do processo.

Por último, a comunicação com o candidato é fundamental. Há uma competição muito elevada para os melhores contratos e um recrutador que não tenha boa capacidade de comunicação pode perder candidatos valiosos. Após a pré-triagem, o recrutador discute a próxima etapa do processo de seleção com o gestor, o que pode resultar no encerramento do processo ou na pré-seleção do candidato para entrevistas.

A forma como a entrevista é conduzida depende da localização do candidato, da sua nacionalidade e do próprio cargo em questão. Por vezes, devido à localização física e restrições legais aí existentes, é benéfico organizar reuniões *online* primeiro. As entrevistas precisam de uma escala de avaliação. Se houver mais de um entrevistador, uma empresa deve ser capaz de classificar ou pontuar as entrevistas, recorrendo a critérios uniformes, para facilitar a tomada de decisão final. O recrutador e o gestor responsável pela contratação devem discutir em detalhe as próximas etapas da contratação do candidato (Holm, 2012).

Recrutamento Interno

O recrutamento interno refere-se à promoção de talentos dentro de uma organização. Normalmente, é iniciado por um colaborador que deseja mudar de funções ou cujo posto de trabalho foi extinto, necessitando de ser realocado. Há duas maneiras de recrutar dentro de uma organização: anúncio interno de trabalho, difundido através da intranet da empresa ou em espaço reservado para anúncios e comunicados. Segundo David DeCenzo, a contratação interna reduz os custos associados ao processo de recrutamento e melhora a probabilidade de uma boa seleção, pois existem informações detalhadas sobre o desempenho dos diferentes colaboradores agora candidatos a uma nova posição dentro da organização (DeCenzo, Robbins & Verhulst, 2013). Um estudo realizado por Matthew Bidwell concluiu que o recrutamento interno tem um custo menor, já que os colaboradores vindos de um processo de seleção externo tendem a exigir um salário maior (Bidwell, 2011). Regra geral, o esforço de adaptação de um colaborador que contacta com a organização pela primeira vez é maior quando comparado com um colaborador que aí trabalha e conhece a organização. Este simples facto resulta que o ordenado de um colaborador vindo de um processo de seleção externo à organização seja, em média, 17% mais elevado quando comparado

com uma promoção de um colaborador já pertencente à organização (Bidwell, 2011). Contudo, esta dissertação incide sobre o processo de recrutamento externo.

Recrutamento Externo

O recrutamento externo encontra talentos fora da organização. Tradicionalmente, o recrutamento externo era construído em torno da publicidade: anúncios à porta dos estabelecimentos, propaganda boca-a-boca, anúncios nos jornais e agências de emprego. Os anúncios de oferta de emprego através de jornais têm vindo a diminuir, uma vez que a tecnologia mudou a forma como as notícias são consumidas ou adquiridas (Baum & Kabst, 2014; Angelucci & Cagé, 2016). As agências de emprego têm sido um fator de apoio tanto a empregadores como a trabalhadores, de forma a cruzarem dois grandes universos. As feiras de emprego, que tiveram o seu papel na história do recrutamento, continuam a ser um método forte e adequado para atrair candidatos e aumentar ou publicitar a marca do empregador.

À medida que a competição por candidatos aumenta, as universidades e outras instituições de ensino desempenham grande parte do recrutamento externo. Como a maioria dos trabalhos estabelece algum tipo de requisito escolar (por exemplo, um diploma de ensino secundário, licenciatura, mestrado, entre outros etc.), essas instituições fornecem um agregado de recursos ideal.

Atualmente, *websites* de emprego são escolhidos para publicitar anúncios de oferta de emprego, juntamente com os *websites* das organizações. Estes são mais propensos a serem atualizados antes de qualquer outro canal, uma vez que são controlados de perto pelos recrutadores e altamente utilizado pelos candidatos (Coles & Mortensen, 2016).

E-recruiting

O mundo digital trouxe uma nova dimensão ao mundo do recrutamento. A *World Wide Web*, ou Web 1.0, encurtou o tempo de pesquisa, os custos e ofereceu um método transparente de informação aos candidatos (Salmen, 2012). No início dos anos 2000, as empresas só tinham os seus *websites* como fonte de comunicação das notícias da sua organização. O desenho e estrutura da Web 1.0 impôs um estilo de comunicação unidirecional. Os candidatos que encontravam as vagas de emprego no *website* de uma organização e tinham interesse enviavam um *e-mail* inicial com o seu currículo e uma carta de apresentação anexada. A comunicação bidirecional raramente ocorria, pois era uma estratégia pouco comum no recrutamento (Salmen, 2012). A Web dos dias de hoje, como a Web 2.0, alterou o processo de aquisição de talentos. Este fenómeno é conhecido como “E-recruiting” (Salmen, 2012). Quando dividido, o termo “E-recruiting” é composto por duas partes: o “E” significa “eletrónico” e “recruiting” significa “recrutamento”. Muitas vezes, o *E-recruiting* é conhecido como recrutamento *online*, recrutamento através de redes sociais, ou simplesmente recrutamento na internet. Para adicionar à definição anteriormente introduzida de recrutamento, o *E-recruiting* pode simplesmente ser descrito como o processo de publicitar uma oferta de emprego

Estado da Arte

via a internet. Segundo Holm, na sua tese, existe uma diferença entre o processo de recrutamento baseado em papel e o baseado na internet. No caso do *E-recruiting*, uma das principais diferenças é o poder de filtrar os candidatos mais adequados à vaga ou ao cargo em questão. Uma outra grande diferença é a velocidade de consulta e manipulação de dados extremamente elevada (Holm, 2012). O uso do *E-recruiting* está em pleno crescimento, embora os seus benefícios não tenham mudado muito ao longo dos anos. Carolien Handlogten listou esses benefícios no seu estudo sobre a implementação do *E-recruiting* numa empresa de aviação Holandesa (Handlogten, 2009):

- Diminuição de custos;
- Facilidade de uso por parte dos candidatos;
- Maior número de candidatos;
- Facilidade de uso por parte da organização;
- Aumento da velocidade de contratação;
- Sucesso na procura de candidatos;
- Capacidade de se manter à frente dos concorrentes.

O retorno positivo que se observa com o uso do *E-recruiting* é comprovado pelo estudo de Jobvite, que realiza pesquisas anuais nos últimos seis anos para acompanhar os padrões de recrutamento nos Estados Unidos da América. Os resultados do estudo de 2013 revelam que 33% dos participantes afirmam que o tempo de contratação melhorou e 43% concordaram que a quantidade de candidatos também melhorou (Jobvite, 2018).

Desde o pioneiro *website* de pesquisa de emprego Monster.com lançado em 1995, as empresas começaram a perceber a magnitude da internet e o seu potencial na contratação de colaboradores. O Instituto de Economia da Universidade de Ciências da Computação em Frankfurt am Main (Alemanha) montou uma rede de pesquisa em 2002, em cooperação com a Monster.de, a contraparte alemã da Monster.com, e a Universidade Otto-Friedrich de Bamberg, para analisar as tendências do recrutamento. O estudo consiste nas mil maiores empresas da Alemanha (Beimborn, 2014).

Uma equipa de pessoas da rede de pesquisa CHRIS reuniu e analisou todos os relatórios sobre as tendências de recrutamento dos últimos 10 anos e compilou-os num único documento intitulado "Tendências de Recrutamento: Ein Rückblick auf 10 Jahre" (CHRIS, 2012). Este documento partilha informações pertinentes sobre como o recrutamento era, é e será realizado no futuro, pelo menos na Alemanha. Os papéis dos recrutadores e candidatos mudaram desde 1995. Anteriormente, procurava-se um colaborador generalista. Contudo, atualmente, as organizações procuram colaboradores especializados. Um colaborador generalista é muitas vezes conhecido, recorrendo à gíria, como uma pessoa que "faz um pouco de tudo, mas que não sabe muito de nada". Enquanto que um especialista é capaz de se concentrar num campo e ter pleno conhecimento desse campo.

Estado da Arte

Existem pontos positivos e negativos sobre esses dois títulos. Colaboradores generalistas são ainda hoje largamente contratados em sectores como o Marketing, Finanças, Recursos Humanos e Tecnologias de Informação. Um colaborador especialista é procurado em áreas como a engenharia de *software*, por exemplo. O CHRIS indica que, no recrutamento de hoje, deve haver diversos profissionais especializados em uma ou duas áreas de especialização (CHRIS, 2012). O papel do recrutador também mudou nos últimos 19 anos. O que começou como um administrador de operações, em que o recrutamento era visto como simplesmente realizar entrevistas. O recrutador atualmente é um consultor que medeia o gestor que contrata e o potencial candidato.

De acordo com o relatório de tendências do recrutamento, a maioria das mil maiores empresas da Alemanha respondeu que tem presença no Facebook e que avalia as redes sociais dos candidatos para adicionar um efeito positivo ao seu recrutamento (Sills, 2014). De acordo com a professora doutora Sonja Salmen no livro "Recruiting im Social Web", co-escrito por Bernd Rath (2012), o processo de recrutamento pode ser otimizado até 20% através das redes sociais e as empresas devem considerar essa estratégia. De acordo com a PewInternet.org, em janeiro de 2014, 74% dos adultos *online* usavam alguma forma de *website* de rede social. Este é um aumento de 66% face aos valores de 2005 (Duggan & Smith, 2014).

As redes sociais foram adotadas por 92% das empresas nos Estados Unidos da América. 93% delas estão no LinkedIn, 66% no FaceBook, e 54% no Twitter. Essas empresas procuram ativamente jovens talentos. As empresas competem por talentos de modo mais agressivo do que em anos anteriores e tomam medidas para reduzir os riscos de serem ultrapassadas pela sua concorrência. Registou-se um aumento de 85% nas atividades de recrutamento via redes sociais desde 2007 (Salmen, 2012).

O CHRIS descobriu que os três principais desafios externos para o recrutamento hoje são a mudança demográfica, a falta de trabalhadores qualificados e as redes sociais. Estas implicações afetam toda a indústria: engenharia, saúde, tecnologias de Informação, negócios, entre outros. As empresas alemãs têm vindo a tornar-se cada vez mais adeptas da aquisição de talentos com recurso a tecnologia, no entanto, não são rápidas o suficiente para as tendências atuais. Apenas duas em cada dez vagas de emprego são publicadas nos canais das redes sociais. No entanto, houve um aumento de 14,8% por parte das empresas alemãs a descobrir que o uso de redes sociais no recrutamento é positivo. Essas empresas entendem os desafios do recrutamento através das redes sociais. Um terço das principais empresas da Alemanha respondeu que possui estratégias e atividades nas redes sociais planeadas para o próximo trimestre. Num relatório conduzido por Eva Zils da SocialMedia-Recruiting.com, constatou-se que, das 335 empresas alemãs que participaram numa pesquisa em 2011, 45% não tinham orçamento de recrutamento através das redes sociais para 2012 (Zils, 2012).

Desfasamento Oferta-Procura

Depois de ter sido revisto como nasce e ocorre um processo de recrutamento, é essencial verificar as causas da taxa de desemprego e da escassez de profissionais em determinado setor onde existem vagas de emprego por preencher.

Pode-se assumir que um indivíduo escolhe uma área específica de estudos com a expectativa de trabalhar em um emprego relacionado com essa área. Dada a importância do ensino superior, espera-se que os indivíduos que possuam um diploma universitário tenham uma maior aceitação no mercado de trabalho e maiores regalias ou perspectivas profissionais, quando comparados com indivíduos que não possuam qualquer tipo de qualificação universitária. De acordo com a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO, 2012), os jovens que terminarem o seu percurso escolar devem ser capazes de fazer a transição da escola ou da faculdade para o mercado de trabalho com as competências e conhecimentos adquiridos. No entanto, esta relação linear entre educação e sucesso no mercado de trabalho nem sempre é uma realidade. A situação é mais aguda com licenciados dos campos das Humanidades. Muitos estudantes passam longos e dispendiosos anos a estudar temas como Geografia, História, Filosofia, Literatura e outras disciplinas relacionadas com as ciências Humanas, apenas para descobrir, após a licenciatura, que as suas novas competências não têm grande aceitação por parte do mercado de trabalho (Nunez & Livanos, 2010). O desfasamento entre as necessidades do mercado de trabalho e a oferta leva a que alguns indivíduos, motivados pela frustração e desespero, possam aceitar empregos que não combinam com o seu nível de formação, a fim de atenderem às suas necessidades de sobrevivência (Nunez & Livanos, 2010).

Definição de desfasamento educação-emprego

O levantamento da literatura sobre o desfasamento educação-emprego foi realizado a partir de diversos pontos de vista. O termo desfasamento entre educação e emprego tem sido usado de forma intercambiável com outros termos como educação excessiva, subutilização de habilitações, subeducação, incompatibilidade de habilitações, superqualificação e subemprego. Mahuteau et al. (McGuinness, Pouliakas & Redmond, 2017) definem o desfasamento educação-emprego como uma situação em que a qualificação de um empregado não corresponde à qualificação do trabalho que ele ou ela faz. Graham (Graham & Graham, 2013) definiu como uma situação em que um trabalhador está num trabalho que não corresponde ao seu nível de educação, experiência, ou interesses e sugeriu que tal incompatibilidade resulta da interação entre uma combinação de pessoas, necessidades, valores e expectativas, por um lado, e características e recompensas associadas aos seus empregos, por outro lado. Betti (Betti, D'Agostino, & Neri, 2007) indica que o desencontro educação-emprego se refere à falta de coerência entre o nível de escolaridade exigido e o oferecido para um trabalho. Existem dois tipos de desfasamento entre educação e trabalho: o desajuste vertical e o desajuste horizontal (Sloane et al., 2010; Kim, Ahn, & Kim, 2012). O desfasamento vertical refere-se ao descompasso entre o nível de educação e o emprego. O

desfasamento vertical pode ocorrer como excesso de educação ou como subeducação. A educação excessiva ocorre quando um indivíduo é recrutado para um trabalho que requer um nível mais baixo de educação, enquanto que a subeducação existe quando um indivíduo possui um nível de educação inferior ao exigido para um trabalho (Sloane et al., 2010).

O desfasamento horizontal, também conhecido como incompatibilidade de campo de estudo, ocorre quando um trabalhador, treinado num determinado campo de estudo, trabalha noutro campo (Montt, 2015) ou quando há divergência entre o campo de estudo e o trabalho (Kim, Ahn & Kim, 2012). Por exemplo, um licenciado em direito, gestão ou ciências sociais que trabalha no setor de serviços. A escolha do campo de estudos é geralmente, entre outras razões, impulsionada pela expectativa de adquirir um emprego onde o conhecimento adquirido durante a educação será aplicado e recompensado (Nordin, Persson & Rooth, 2010; Domadenik, Farcnik & Pastore, 2013).

A discrepância entre educação e emprego tem sido estudada especialmente nos países desenvolvidos e o foco principal tem sido medir a supereducação e os resultados negativos do mercado de trabalho (isto é, penalização salarial, insatisfação no trabalho e mobilidade profissional) (Allen & Van der Velden, 2001; Barone & Ortiz 2011). Além disso, a maioria das definições e o uso do termo desfasamento entre educação e trabalho têm sido mais consistentes com o desfasamento vertical do que o desfasamento horizontal (Betti, D'Agostino & Neri, 2007; McGuinness & Sloane, 2011). Por vezes, pode ocorrer a sobreposição do desfasamento vertical e horizontal. Um licenciado não só pode estar a trabalhar num emprego que ele é superqualificado (desfasamento vertical), mas também num trabalho onde o conhecimento que ele adquiriu através do ensino superior não tem relevância (desfasamento horizontal). Por este motivo, o desfasamento educação-emprego entre os licenciados é entendida num contexto em que o tipo de trabalho assegurado pelos licenciados não corresponde ao seu nível de escolaridade, ao campo de estudo ou a ambos.

Causas do desfasamento ensino-emprego

A questão do desfasamento entre a educação e o emprego pode ser explicada tanto a nível estrutural quanto a nível individual. O desajuste entre educação e emprego pode ser visto como um problema resultante de uma economia fraca e estagnada (Milana, 2009). Os desequilíbrios entre a procura e a oferta no mercado de trabalho (Robst, 2007; Montt 2015; Verhaest, Sellami & Van, 2017), a qualidade do currículo e o tipo de cursos oferecidos pelas universidades (Mbah, 2014), a desconexão entre as universidades, o mercado de trabalho e o campo de estudo (Robst, 2007).

De acordo com Montt (Montt, 2015), os trabalhadores desfasados, no que diz respeito ao binómio estudos-emprego, não escolhem voluntariamente ser incompatíveis, mas são levados a isso, porque não conseguem encontrar trabalho no seu campo de competências ou porque as suas competências são melhor recompensadas noutros grupos ocupacionais.

De acordo com Bhuvan e Marzo-Navarro (Bhuvan, 2018; Marzo-Navarro, Pedraja-Iglesias & Rivera-Torres, 2009), o desfasamento entre as competências ensinadas nas instituições de ensino superior

e as competências necessárias ao empregador explica por que é que os licenciados permanecem desempregados ou acham difícil ter acesso a empregos adequados no mercado de trabalho. Este desfasamento entre o ensino superior e o mercado de trabalho resulta em parte de uma falta de interação entre as instituições de ensino superior e o mercado de trabalho. Sloane (Sloane et al., 2010), também observou que, devido às disparidades existentes entre a educação e o mercado de trabalho, os jovens instruídos muitas vezes acham difícil conseguir empregos após a sua formação e, como tal, têm de esperar até encontrar um emprego adequado para a sua qualificação ou aceitar qualquer trabalho sem considerar as suas qualificações ou áreas de estudo.

Outra explicação para o desfasamento entre educação e emprego está relacionada com o excesso de licenciados numa determinada área de conhecimento, como consequência o mercado de trabalho para esses licenciados fica saturado (Montt, 2015). El-Khawas (El-Khawas, 2001) relata que o aumento ao acesso ao ensino superior está presente nas agendas políticas da maioria dos governos e organizações internacionais. Por exemplo, a Declaração Universal dos Direitos Humanos das Nações Unidas de 1948 realçou o acesso igual ao ensino superior para todos com base no mérito (UN, 1948). Em 1976, o Pacto Internacional das Nações Unidas sobre os Direitos Económicos, Sociais e Culturais reforçou este compromisso com a igualdade de acesso ao ensino superior para mulheres e homens. No entanto, a questão do acesso ao ensino superior ressurgiu nos objetivos de Desenvolvimento Sustentável Pós-2015 (UN, 2015). Isto mostra que há um maior foco na quantidade do que na qualidade do ensino superior e os governos na maioria dos países em desenvolvimento alcançam resultados bem-sucedidos na expansão do acesso ao ensino superior.

Apesar do aumento ao acesso ao ensino superior, a história de sucesso dos licenciados revela-se sombria quando se levanta a questão da utilidade de alguns graus académicos obtidos em algumas instituições de ensino superior (Humburg, De Grip & Van der Velde, 2017; Mbah, 2014; Neneh, 2014). Os governos em geral têm encontrado dificuldade para enfrentar os desafios da expansão não planeada, desemprego entre universitários, subemprego e falta de emprego, já que a maioria desses licenciados acaba por não encontrar emprego que utilize as suas competências adquiridas no ensino superior. Na opinião de Congregado (Congregado et al., 2016), as instituições de ensino superior estão a produzir mais licenciados a cada ano, a maioria dos quais não consegue encontrar emprego em seus respetivos campos de qualificação, e por frustração, e necessidade recorrem a qualquer tipo de trabalho que encontrarem. A qualidade do capital humano também é apontada para explicar o desfasamento entre os graduados (Balcerzak & Pietrzak, 2016). A qualidade do capital humano depende da natureza do currículo usado nas universidades, do tipo de competências adquiridas e da aplicabilidade dessas competências adquiridas. Alguns graduados deixam a universidade sem adquirir conhecimento de qualidade para os tornar empregáveis em áreas apropriadas no mercado de trabalho, acabando por aumentar o número de graduados desempregados (Andrews & Higson, 2008). De acordo com McGuinness e Hennemann (McGuinness, 2003; Hennemann & Liefner, 2010), o desfasamento entre os graduados

Estado da Arte

universitários está associado à educação de baixa qualidade devido a currículos e métodos de ensino desatualizados.

Os empregadores exigem que os candidatos a futuras vagas de emprego tenham uma formação acadêmica sólida e que lhes permitam adaptar-se rapidamente ao ambiente de trabalho (McGuinness, 2003; Hennemann & Liefner, 2010). No entanto, o foco do sistema educativo está na teoria e não no conhecimento prático (Wolbers, 2003; Paadi, 2014). Naong (Naong, 2011) observou que a aprendizagem universitária tem sido criticada por se concentrar muito em aumentar o “conhecimento declarativo” e ser insuficiente na compreensão profissional relevante ou conhecimento funcional. Como resultado, os jovens instruídos não conseguem acompanhar as mudanças do mercado de trabalho e economia. Da mesma forma, Paadi (Paadi 2014) apontou que os graduados universitários têm, na sua maioria, conhecimentos teóricos que não são apoiados por experiência suficiente de trabalho prático, em países onde a economia não está a crescer a uma velocidade que permita a criação de empregos qualificados. A falta de competências práticas e de experiência de trabalho leva a que os empregadores fiquem relutantes em empregar jovens graduados. A adoção de processos de produção tecnologicamente mais avançados fez com que os empregadores preferissem graduados mais qualificados e experientes, comparando com os graduados com conhecimentos mais teóricos e gerais (Branine & Avramenko, 2015).

Feiras de emprego, palestras e programas de estágio raramente são implementados na maioria das universidades de forma geral. Essas atividades que podem ajudar a trazer mais conhecimento do mundo exterior para as salas de aula são negligenciadas pelas universidades. Através de estágios, os estudantes podem aplicar o conhecimento adquirido na sala de aula, ganhar experiência de trabalho e aumentar a taxa de empregabilidade após a sua graduação (Callanan, 2003).

O desfasamento entre a educação e o mercado de trabalho também depende da área de estudo. Por exemplo, Robst (Robst, 2007) observou que a maior taxa de desfasamento é encontrada entre os graduados em artes e menor entre os graduados dos campos relacionados com a saúde. Isto ocorre porque, entre outras razões, os alunos ligados às áreas de saúde apresentam um maior conhecimento prático e mais adequado ao mercado de trabalho, o que reduz a probabilidade destes alunos virem a trabalhar num emprego fora da sua área de competência (Wolbers, 2003).

O desfasamento entre educação e trabalho entre graduados universitários também pode ser atribuído a restrições de procura de emprego (McGuinness & Bennett, 2007). Os indivíduos podem adquirir as competências certas, os trabalhos relacionados às suas competências podem estar disponíveis, mas esses indivíduos podem encontrar-se em empregos incompatíveis. Isto pode resultar da imperfeição da informação sobre as ofertas de emprego e as necessidades do mercado de trabalho (Bassegy & Atan, 2012). Graham (Graham & Graham 2013) observaram que a falta de emprego pode resultar de informações limitadas sobre a escolha dos empregos disponíveis. De acordo com Beck (Beck, 2002), o desfasamento entre ensino-emprego ocorre também, pois os indivíduos que se encontram desempregados ou desfasados não se esforçam o suficiente para procurar os trabalhos adequados às suas áreas de competência.

Consequências do desfasamento ensino-emprego

De acordo com Salinas-Jiménez (Salinas-Jiménez, Artés & Salinas-Jiménez, 2016), o desencontro entre educação e trabalho tem consequências negativas para o indivíduo, a empresa e a sociedade como um todo. Por exemplo, ao nível empresarial, o desencontro da educação leva a um baixo nível de produtividade, baixo nível de envolvimento no trabalho, altos índices de rotatividade de funcionários e as empresas podem ter que suportar custos extras na seleção, recrutamento e treino de novos funcionários repetidamente (Van & Van, 2000). A sociedade como um todo acaba por perder poder de produtividade quando um trabalhador está a exercer uma atividade fora da sua área de competência (Chevalier, 2003).

Segundo Nordin (Nordin, Persson & Rooth, 2010), indivíduos que trabalham numa ocupação que não corresponde à sua área de educação acabam por sofrer uma substancial penalização salarial. Robst (Robst, 2007) descobriu que o desfasamento entre educação e trabalho tem um efeito negativo sobre os salários e a categoria de trabalho. Os pais e responsáveis que lutaram para financiar a educação dos seus filhos muitas vezes ficam desapontados com esse investimento quando os educandos acabam em empregos que não exigem a educação que adquiriram.

A literatura também associa incompatibilidade entre educação e trabalho com uma menor satisfação no trabalho. A subutilização de competências afeta negativamente a satisfação no trabalho (Allen & Van der Velden, 2001). Indivíduos desfasados enfrentam níveis mais baixos de satisfação no trabalho (Groot & Van, 2000; Farooq, 2011). Quando a ocupação de um indivíduo não corresponde à sua área de educação escolhida, significa que a intenção original não foi alcançada e tais indivíduos enfrentam o desapontamento das expectativas não alcançadas (Montt, 2015)

Pesquisa de Emprego em Portugal

A literatura sobre recursos humanos, processos de recrutamento e desfasamento da procura-oferta no mercado de trabalho é vasta. Contudo, a literatura sobre pesquisa de emprego é reduzida e foca-se no mercado de trabalho dos Estados Unidos da América e do norte da Europa. Em Portugal, a literatura sobre o tema “pesquisa de emprego” é escassa, não havendo até ao momento nenhum estudo, artigo, tese ou documento que aborde quais são os principais *websites* de pesquisa de emprego em Portugal. Mais adiante, nesta dissertação, serão identificados quais os maiores *websites* de ofertas de emprego em Portugal, segundo o ranking Alexa¹, um ranking hierárquico do motor de busca mais popular do mundo, Google, e a quantidade de publicações de ofertas de emprego em determinado período em cada *website* (Ebizmba.com, 2018; Netmarketshare.com, 2018).

¹ Alexa é uma subsidiária da empresa Amazon e fornece dados de tráfego na internet.

Agregadores de ofertas de emprego

Os agregadores de ofertas de emprego, isto é, *websites* que agregam ofertas de emprego publicadas em outros *websites* de anúncios de emprego, não são uma novidade nem exclusivos a setores empresariais ou países. Contudo, estes mesmos agregadores são bastantes similares entre si, limitando-se, de grosso modo, a agregar ofertas de emprego provenientes de outros *websites*. Em Portugal, não existe nenhum agregador de ofertas com assinatura exclusivamente portuguesa, que consuma dados de *websites* que não forneçam uma *application programming interface* (API) – como o caso do site net-empregos.com, que é um dos maiores sites de ofertas de emprego em Portugal como será descrito no próximo capítulo, e que valide o peso de uma competência profissional no mercado de trabalho.

De seguida serão apresentados os maiores *websites* de agregação de ofertas de emprego:

Indeed	
Criação	2004 – Estados Unidos da América
Ranking Alexa	158
Presença Países	60 Países
Presença Portugal	Sim

Tabela 1. Descrição Indeed. Fonte: Elaboração própria.

Monster	
Criação	1994 – Estados Unidos da América
Ranking Alexa	1930
Presença Países	40 Países
Presença Portugal	Não

Tabela 2. Descrição Monster. Fonte: Elaboração própria.

Simply Hired	
Criação	2003 – Estados Unidos da América
Ranking Alexa	8582
Presença Países	24 Países
Presença Portugal	Sim

Tabela 3. Descrição Simply Hired. Fonte: Elaboração própria.

Trovit	
Criação	2006 – Espanha
Ranking Alexa	10422
Presença Países	38 Países
Presença Portugal	Sim

Tabela 4. Descrição Trovit. Fonte: Elaboração própria.

Web Scraping

Este tópico serve para descrever o que é o *Web Scraping*, as razões para o seu uso, as implicações legais e quais as técnicas mais comuns utilizadas associadas ao *Web Scraping*. A técnica de *Web Scraping* é uma técnica fundamental para construir a plataforma proposta nesta dissertação. O processo de *Web Scraping* é inteiramente explicado, através de exemplos, no capítulo que descreve a criação do algoritmo de extração de dados aqui utilizado.

O que é *Web Scraping*?

A *World Wide Web* consiste numa rede interligada de dados que são apresentados através de *websites* aos seus utilizadores. A *World Wide Web* mudou significativamente a maneira como partilhamos, recolhemos e publicamos dados. A quantidade de informação cresce constantemente.

A *International Data Corporation* prevê que, até 2025, a esfera de dados global aumentará para 163 zettabytes (isto é, um trilhão de gigabytes). São dez vezes os 16,1 ZB de dados gerados em 2016. Todos esses dados proporcionarão experiências únicas aos utilizadores e um novo mundo de oportunidades de negócios (Reinsel, Gantz & Rydning, 2017).

Com este crescimento de dados, por vezes desorganizados, não é possível mapear e registar manualmente todas as fontes disponíveis. É neste momento que evoluem as técnicas relacionadas com *Web Scraping*.

Várias definições do que é o *Web Scraping* surgiram durante a pesquisa bibliográfica. Watson (Watson, 2009) descreve que *Web Scraping* é a técnica de extração de dados de *websites*. Dados esses destinados a humanos e não a máquinas.

A definição descrita acima indica que os dados extraídos se destinam a humanos. Tal definição provou estar desatualizada. Com a evolução das técnicas de automatização, provou-se que os humanos não são os únicos consumidores finais dos dados extraídos por técnicas de *Web Scraping*. As próprias máquinas (softwares) também podem consumir esses dados. No entanto, deve ter-se em conta que a data de publicação dessa primeira definição é de 2009 e o uso de *application programming interface* (API) era reduzido quando comparado com os dias de hoje (Gadre & Gupta, 2018).

Munzert (Munzert et al., 2014) define que *Web Scraping* é uma técnica para extrair dados presentes na *World Wide Web* (WWW) e o seu armazenamento numa base de dados, para posteriormente serem analisados.

Técnicas de *Web Scraping*

As técnicas de *Web Scraping* evoluíram em conjunto com a *World Wide Web*. De seguida serão apresentadas as técnicas atuais de *Web Scraping* relevantes para esta dissertação.

Scraping Manual

A técnica de *scraping* manual está relacionada com a recolha de dados por um humano sem recurso a qualquer tipo de ferramentas tecnológicas de automatização. Isto é, o programador localiza visualmente os dados a serem recolhidos e manualmente insere esses dados na sua base de dados para futura análise (Polidoro et al., 2015). Esta técnica ainda é uma opção em situações específicas como:

Estado da Arte

- Quando a quantidade de dados é mínima;
- Quando a extração de dados não necessita de ser feita de forma repetida;
- Quando a construção de um algoritmo de extração de dados levaria mais tempo que a própria extração de dados de forma manual;
- Quando por medidas de segurança, legais ou impossibilidade técnica não é possível extrair os dados de forma automática.

Análise do HTML

A análise da estrutura da árvore HTML de um *website* é uma das técnicas de *Web Scraping* onde o programador informático procura localizar os dados alvo a serem extraídos, dentro dos elementos HTML, identifica um padrão na árvore HTML, onde esses dados se encontram, e constrói um algoritmo de extração de dados ou utiliza ferramentas para esse efeito (Gupta et al., 2003).

Análise do DOM

A análise do *document object model* (DOM) é uma evolução da técnica anterior, em que o programador informático segue os mesmos passos da técnica de análise do HTML para extrair os dados que necessita de um *website*. Contudo, esta técnica permite que sejam utilizadas ferramentas de extração de dados mais rápidas, modernas e eficientes. (Bei-jie, Shuang & Bei-hong, 2007)

Extração via XPath

A extração de dados via *extensible markup language* (XML) *path language* (XPath) é uma técnica que acaba por estar dependente das últimas duas técnicas anteriores, mas merece um destaque isolado, pois esta linguagem de consulta pode tornar-se bastante complexa e tem uma curva de aprendizagem maior que as duas últimas técnicas. Após ser analisada a árvore HTML ou o DOM de um *website*, o programador informático desenha o caminho até aos dados que pretende extrair, recorrendo à linguagem XPath e/ou com auxílio de ferramentas que usem esta linguagem. Este tipo de linguagem é muito sensível a qualquer tipo de alteração na estrutura da árvore HTML ou DOM de um *website*, como tal o programador informático tem de estar constantemente atento a qualquer tipo de alteração, para poder atualizar o novo caminho XPath que leva o algoritmo ou ferramenta de extração até aos dados alvo (Munzert et al., 2014).

Consumo de APIs

Esta última técnica de extração de dados de um *website* é a técnica mais rápida, barata, legal, apesar de dependente de terceiros, que existe para extrair ou consumir dados na web. Não se trata de uma técnica de *Web Scraping* pura, pois a única atividade realizada é consumir uma API de um *website* que, como resposta aos nossos pedidos, nos fornece os dados que necessitamos (Ferrara et al., 2014).

Por vezes é cada vez mais, os *websites* fornecem APIs aos utilizadores, numa tentativa de disponibilizar os dados aí contidos de uma forma simples e rápida. Uma API é um *software* desenhado para comunicar com outros softwares e serve de ligação entre a base de dados de um *website* e um *software* (utilizador) que realiza uma consulta (Scheller & Kühn, 2015).

Ferramentas disponíveis

Existem várias ferramentas para realizar *Web Scraping*. De seguida, serão apresentadas as quatro categorias em que essas ferramentas se dividem:

- Cloud;
- Desktop;
- Bibliotecas de programação;
- Extensões *Web Browser*.

Só serão apresentados exemplos de soluções do tipo “bibliotecas de programação”, visto serem as soluções utilizadas nesta dissertação e as restantes soluções não se adequam aqui.

Cloud

As soluções Cloud fornecem uma interface ao utilizador por meio de um *web browser*, enquanto o *backend* da solução reside num servidor Cloud. Esta arquitetura limita os requisitos de *hardware* do utilizador ao mínimo. Podem ser extraídas grandes quantidades de dados sem ter o próprio computador ligado. Podem ser criados projetos em larga escala sem haver necessidade de ter uma ligação à internet veloz e/ou um sistema de *hardware* de última geração (Chaulagain, 2017).

Em algumas situações, é muito importante definir de forma correta onde se localiza o serviço Cloud. Por exemplo, por vezes, alguns *websites* podem não estar disponíveis na Europa, caso o serviço Cloud se encontre na Ásia (Chaulagain, 2017).

As soluções Cloud geralmente são financiadas por subscrições por parte dos utilizadores. O preço depende dos recursos alocados. Os fornecedores deste tipo de solução oferecem subscrições mensais ou anuais (Chaulagain, 2017).

Desktop

Os dados presentes num *website* são extraídos, analisados e armazenados localmente. Este tipo de aplicações de desktop exigem uma ligação à internet de banda larga. Em comparação com a solução Cloud é necessário um *hardware* mais avançado (Meschenmoser, 2016).

Bibliotecas de programação

Os tipos de *software* de *Web Scraping*, prontos a usar, como as anteriores soluções apresentadas, são mais fáceis de usar em comparação com a utilização de uma biblioteca de programação. Estas bibliotecas terão de ser inseridas no código desenvolvido pelos programadores informáticos, a fim de realizarem a extração dos dados pretendidos (Mitchell, 2018). Estas bibliotecas são classes que terão de ser instanciadas, a fim de serem manipuladas na forma de objetos. Essas classes possuem algoritmos que realizam a extração dos dados pretendidos. O funcionamento neste tipo de solução será descrito nesta dissertação em pormenor, no capítulo reservado à apresentação da construção do algoritmo de extração de dados.

Para alguns fins, as soluções apresentadas anteriormente não são as mais adequadas. Quando se deseja ter uma solução mais personalizada, específica e independente tem de se recorrer a este tipo de técnica. Estas bibliotecas permitem os programadores informáticos não começarem os seus algoritmos de extração do zero, isto é, estas classes na forma de bibliotecas abstraem já grande parte do processo de extração de dados (Mitchell, 2018).

De seguida, serão apresentadas as três bibliotecas mais pertinentes no que toca ao processo de *Web Scraping*.

Iron Web Scraper

A Iron Web Scraper é uma biblioteca avançada de *Web Scraping* escrita em *c#*. Esta biblioteca permite aos programadores informáticos simular e automatizar o comportamento de navegação humana, num *website*, para extrair conteúdo, arquivos e imagens. Esta biblioteca permite utilizar a opção de multi-threading e pode ser usada para migrar conteúdo de *websites* existentes ou ser usado como ferramenta de controlo da qualidade.

HTML Agility Pack

Esta biblioteca permite extrair e analisar a árvore HTML ou DOM de um *website* ou documento “fora da web”. Suporta a linguagem XPath, bem como HTML mal formatado ou não formatado de forma padrão, segundo as boas práticas do *World Wide Web Consortium* (W3C).

AngleSharp

A biblioteca Angle Sharp é uma biblioteca de análise de conteúdo HTML. Permite analisar documentos HTML5, CSS3 e XML, segundo as especificações oficiais do W3C.

Existem várias bibliotecas que auxiliam no processo de *Web Scraping*. Contudo, não serão apresentadas nesta dissertação por apresentarem uma ou mais das seguintes limitações:

- Tratar-se meramente de uma biblioteca que reescreve uma das bibliotecas apresentadas anteriormente;

Estado da Arte

- Não existir atualmente suporte ou manutenção;
- Estarem desenhadas unicamente para funcionar num só *web browser*;
- Terem elevada curva de aprendizagem sem trazer nenhum benefício quando comparadas com as três bibliotecas já apresentadas.

Extensões *Web Browser*

Este tipo de solução serve para uso recreativo e trata-se de um complemento que pode ser instalado no *web browser* do utilizador de forma a adquirir a funcionalidade de extração de dados. Estas ferramentas são as menos poderosas de todas apresentadas e servem para extrair um número limitado de dados inseridos em projetos pequenos (Chavan et al., 2018).

Aspetos legais da extração de dados

O uso de dados resultantes de técnicas de *Web Scraping* é um tema pouco definido a nível legal. Indivíduos ou empresas ativas na extração de dados via *Web Scraping* têm uma estratégia de defesa legal simples, indicando que os *websites*, no geral, têm dados que foram publicados com a clara intenção de serem vistos por terceiros. (Mahto & Singh, 2016).

O processo de *Web Scraping* pode violar várias leis ou termos contratuais. A tabela seguinte resume alguns pontos onde o processo de *Web Scraping* pode levantar algumas questões de legitimidade (Vanden & Baesens, 2018).

<i>Web Scraping</i> Vs Questões Legais e contratuais	
Termos e condições do website	Muitas empresas proíbem expressamente a extração de dados dos seus <i>websites</i> , e deixam bem claro essa informação nos termos e condições dos seus <i>websites</i> . Se os <i>websites</i> podem impor tais termos ainda não é claro, mas dependendo das circunstâncias, uma reivindicação por quebra contratual é possível.
Direitos de autor, de propriedade e intelectual	Como o processo de <i>Web Scraping</i> envolve a cópia de dados, isso pode levar a uma reivindicação por violação de direitos de autor. Se tal alegação tem algum mérito dependerá das circunstâncias, porque nem todos os dados extraídos se qualificam ao abrigo da lei de direitos de autor.
Proteção jurídica da base de dados	A proteção jurídica da base de dados é violada quando toda ou parte substancial de uma base de dados é extraída ou reutilizada sem o consentimento do proprietário. A extração

Estado da Arte

	repetida pode influenciar a performance da própria base de dados alvo. O proprietário da base de dados alvo de extração pode ter custos adicionais para melhorar a <i>performance</i> da base de dados devido ao processo contínuo de <i>Web Scraping</i> que é alvo e imputar esses custos à empresa responsável pela diminuição da performance da base de dados.
Marcas registadas	Se a empresa ou indivíduo que realiza o método de <i>Web Scraping</i> se fizer passar por uma marca sem o consentimento do proprietário, esse mesmo proprietário poderá tomar uma medida legal para travar esse processo de <i>Web Scraping</i> .
Proteção de dados	As empresas que utilizam técnicas de <i>Web Scraping</i> automatizadas podem recolher informação sobre indivíduos, e dessa forma violar a lei de proteção de dados. A grande questão central é se os indivíduos consentiram ou não a recolha desses dados pessoais.

Tabela 5. Questões legais e contratuais relacionadas com o processo de *Web Scraping* (Vanden & Baesens, 2018).

Conclusão do Capítulo

Neste capítulo, consideraram-se os conceitos teóricos mais importantes que deram suporte à criação de uma plataforma de agregação de ofertas de emprego em Portugal, pois avalia o peso de uma determinada competência profissional no mercado de trabalho. Plataforma essa que não depende do consumo de nenhuma API externa ou Rich Site Summary (RSS) Feeds, tornando esta plataforma mais independente, sendo uma mais-valia para o utilizador. Neste capítulo, foi revisto também como ocorre o processo de recrutamento, especialmente via *online*, bem como o conceito de desfasamento ensino-emprego.

3. Conceptualização e Modelação do Protótipo

O presente capítulo serve para descrever e identificar os principais *websites* de ofertas de emprego em Portugal, rever a sua estrutura HTML e quais os nós da árvore HTML a considerar para realizar a extração do conteúdo alvo. Todas as etapas serão orientadas à **construção de um algoritmo de extração de dados**, com o objetivo de poder aplicar esse algoritmo a um maior número de *websites* de ofertas de emprego, poder agregar o maior número de ofertas e devolver ao utilizador final a aceitação por parte do mercado trabalho de uma determinada competência profissional. Este capítulo aborda também a metodologia adotada para a criação da plataforma proposta. Trata-se de uma metodologia qualitativa, fazendo recurso a um *focus group*, que nos deu os dados necessários para adequar o mais possível a plataforma às necessidades do utilizador final.

Levantamento dos requisitos

Segundo Silverman (Silverman, 2016) o recurso a uma metodologia qualitativa, como o caso de um *focus group*, é algo que é utilizado para guiar uma pesquisa de um produto ou serviço em uma determinada área. E é essa metodologia que foi utilizada para fazer o levantamento dos requisitos para a criação da plataforma tecnológica, aqui proposta (Silverman, 2016).

Foram elaboradas quatro questões que foram apresentadas a um grupo de 5 pessoas. A discussão proveniente dessas quatro questões gerou um conjunto de informação, condensada e disponível no Anexo E deste documento. Segundo o autor Bernard uma das formas de estudo dos resultados provenientes de um *focus group* é fazer recurso a ferramentas como “www.wordclouds.com”, que permitem realizar uma *cloud tag* com as palavras mais relevantes e relacionadas entre si extraídas durante a discussão anteriormente realizada (Bernard, Wutich & Ryan, 2016). Esse estudo será apresentado se seguida:

Após a questão “Quais os aspetos mais importantes que uma plataforma web de agregação de ofertas de emprego deve ter?” pode-se verificar na Figura 3 a *cloud tag* resultante.

Na Figura 4 observa-se que os aspetos mais fracos que existem nas plataformas web de agregação de ofertas de emprego são: ofertas (emprego) repetidas, a descrição das ofertas, saber se as ofertas estão ativas, e fracas análises.

O estudo da questão “Quais considera serem os benefícios do uso das plataformas web de agregação de ofertas de emprego no auxílio de novas oportunidades de emprego?” pode-se verificar na Figura 5 através de uma *cloud tag*.



Figura 5. *Cloud Tag* resultante da questão “Quais considera serem os benefícios do uso das plataformas web de agregação de ofertas de emprego no auxílio de novas oportunidades de emprego?”. Fonte: www.wordclouds.com.

Na Figura 5 observa-se os benefícios do uso das plataformas web de agregação de ofertas de emprego no auxílio de novas oportunidades de emprego são: acesso a diferentes ofertas em um menor período de tempo.

O resultado após a questão “Quais considera serem os inconvenientes do uso das plataformas web de agregação de ofertas de emprego no auxílio de novas oportunidades de emprego?” é observado na Figura 6 através de uma *cloud tag*.



Figura 6. Cloud Tag resultante da questão "Quais considera serem os inconvenientes do uso das plataformas web de agregação de ofertas de emprego no auxílio de novas oportunidades de emprego?". Fonte: www.wordclouds.com.

Na Figura 6 observa-se que os inconvenientes do uso das plataformas web de agregação de ofertas de emprego no auxílio de novas oportunidades de emprego são: muitas ofertas numa só pesquisa, a plataforma só acede a *websites* de ofertas, e saber quais as ofertas que estão propriamente disponíveis.

Após análise dos resultados obtidos pode-se observar que um utilizador espera mais, de uma plataforma de agregação de ofertas de emprego, do que somente a apresentação das ofertas. O utilizador, fundamentalmente, espera ter mais informação que o guie na busca de uma nova oferta de emprego.

Escolha dos *websites* de ofertas de emprego em Portugal

A etapa de identificação dos *websites* de ofertas de emprego em Portugal é uma tarefa dificultada pela falta de literatura que aborde esse tema. Até ao momento, não foi identificado nenhum estudo que caracterize os principais *websites* de ofertas de emprego em Portugal. Como tal, terão de ser criados critérios específicos para a sua identificação.

Esses critérios são:

Problema 1: Quais são os *websites* de ofertas de emprego em Portugal?

Critério 1: Pesquisa no maior motor de busca mundial – Google – pelas seguintes palavras-chave da área vocabular de emprego, tais como emprego, trabalho, ofertas de emprego, ofertas de trabalho, a fim de identificar os principais *websites* de ofertas de emprego em Portugal. Verifica-se que quanto mais baixo o ranking de pesquisa de um *website*, isto é, a posição ocupada por um *website* após uma pesquisa num motor de busca, maior é a influência desse *website*. Intenda-se

Conceptualização e Modelação do Protótipo

influência como penetração na web, ligação a outros *websites* e vice-versa (Langville, Meyer & Fernández, 2008).

Problema 2: Quais são os maiores *websites* de ofertas de emprego em Portugal?

Critério 2: Verificar qual o número total de ofertas de emprego que cada *website* conteve nos últimos 7 dias.

Ao responder ao problema 1, segundo o critério 1, isto é, quais os *websites* de ofertas de emprego em Portugal encontrados no maior motor de busca – Google –, segundo as palavras-chaves “emprego”, “trabalho”, “ofertas de emprego” e “ofertas de trabalho”, obtiveram-se os seguintes resultados:

Palavra-chave: “emprego”	
Website	Ranking no web browser Google.pt
net-empregos.com	1
indeed.pt	3
emprego.sapo.pt	5
expressoemprego.pt	7
alertaempregp.pt	9

Tabela 6. Palavra-chave “emprego”. Fonte: Elaboração própria

Palavra-chave: “trabalho”	
Website	Ranking no web browser Google.pt
net-empregos.com	1
emprego.sapo.pt	2
indeed.pt	3
trabalhar.pt	4
babysits.pt	5

Tabela 7. Palavra-chave “trabalho”. Fonte: Elaboração própria

Palavra-chave: “ofertas de emprego”	
Website	Ranking no web browser Google.pt
net-empregos.com	1
indeed.pt	3
emprego.sapo.pt	5
pt.jooble.org	7
expressoemprego.pt	8

Tabela 8. Palavra-chave “ofertas de emprego”. Fonte: Elaboração própria

Palavra-chave: “ofertas de trabalho”	
Website	Ranking no web browser Google.pt

Conceptualização e Modelação do Protótipo

indeed.pt	1
emprego.sapo.pt	3
net-empregos.com	4
pt.jooble.org	6
expressoemprego.pt	7

Tabela 9. Palavra-chave "ofertas de trabalho". Fonte: Elaboração própria

Os resultados obtidos para responder ao problema 1 resultaram da pesquisa no motor de busca Google.pt pelas palavras-chave "emprego", "trabalho", "ofertas de emprego", e "ofertas de trabalho" no dia 24 de abril de 2018. Desta pesquisa, verifica-se a ocorrência de dez *websites* de ofertas de emprego distintos.

Resultados	
Website	Soma de todos os rankings
net-empregos.com	Este <i>website</i> ocorre 4 vezes em 4 pesquisas distintas nos primeiros 10 resultados de pesquisa, e tem um ranking total, somando todos os rankings das quatro pesquisas, de 7 pontos .
indeed.pt	Este <i>website</i> ocorre 4 vezes em 4 pesquisas distintas nos primeiros 10 resultados de pesquisa e tem um ranking total, somando todos os rankings das quatro pesquisas, de 10 pontos .
emprego.sapo.pt	Este <i>website</i> ocorre 4 vezes em 4 pesquisas distintas nos primeiros 10 resultados de pesquisa e tem um ranking total, somando todos os rankings das quatro pesquisas, de 15 pontos .
expressoemprego.pt	Este <i>website</i> ocorre 3 vezes em 4 pesquisas distintas nos primeiros 10 resultados de pesquisa e tem um ranking total, somando todos os rankings das três pesquisas, de 22 pontos .
trabalhar.pt	Este <i>website</i> ocorre uma única vez em 4 pesquisas distintas nos primeiros 10 resultados de pesquisa e tem um ranking total de 4 pontos . Apesar de ter um valor de ranking total baixo, esse valor não é de relevo, pois o <i>website</i> só ocorre em uma das 4 pesquisas realizadas.
babysits.pt	Este <i>website</i> ocorre uma única vez em 4 pesquisas distintas nos primeiros 10 resultados de pesquisa e tem um ranking total de 5 pontos . Apesar de ter um valor de ranking total baixo, esse valor não

Conceptualização e Modelação do Protótipo

	é de relevo, pois o <i>website</i> só ocorre em uma das 4 pesquisas realizadas.
alertaempregp.pt	Este <i>website</i> ocorre uma única vez em 4 pesquisas distintas nos primeiros 10 resultados de pesquisa e tem um ranking total de 9 pontos . Apesar de ter um valor de ranking total baixo, esse valor não é de relevo, pois o <i>website</i> só ocorre em uma das 4 pesquisas realizadas.
pt.jooble.org	Este <i>website</i> ocorre duas vezes em 4 pesquisas distintas nos primeiros 10 resultados de pesquisa e tem um ranking total, somando todos os rankings das duas pesquisas, de 13 pontos .
ofertasdeemprego.pt	Este <i>website</i> ocorre duas vezes em 4 pesquisas distintas nos primeiros 10 resultados de pesquisa e tem um ranking total, somando todos os rankings das duas pesquisas, de 19 pontos .
expressoemprego.pt	Este <i>website</i> ocorre 3 vezes em 4 pesquisas distintas nos primeiros 10 resultados de pesquisa e tem um ranking total, somando todos os rankings das três pesquisas, de 22 pontos .

Tabela 10. Websites resultantes da pesquisa. Fonte: Elaboração própria

Destes dez *websites*, verifica-se que somente três se repetem e com um baixo ranking total. Podendo-se afirmar que os *websites* “net-empregos.com”, “indeed.pt” e “emprego.sapo.pt” são os *websites* com maior relevo no que toca à pesquisa de ofertas de emprego em Portugal, segundo os critérios de pesquisa definidos anteriormente.

Após identificar os três principais *websites* portugueses de ofertas de emprego, identificou-se o número de ofertas total que cada *website* continha.

Desta forma, obteve-se o seguinte resultado:

Resultados	
<i>Website</i>	Número de ofertas
net-empregos.com	300 Ofertas nos últimos 7 dias
indeed.pt	11592 Ofertas nos últimos 7 dias
emprego.sapo.pt	11692 Ofertas nos últimos 7 dias

Tabela 11. Número de ofertas de emprego, nos últimos 7 dias, nos *websites* “net-empregos.com”, “indeed.pt” e “emprego.sapo.pt”. Fonte: Elaboração própria

Conceptualização e Modelação do Protótipo

Já os resultados obtidos para responder ao problema 2 resultam da consulta aos *websites* “net-empregos.com”, “indeed.pt” e “emprego.sapo.pt” no dia 24 de abril de 2018.

Verifica-se que os *websites* “indeed.pt” e “emprego.sapo.pt” têm um número de ofertas de emprego quase similar nos últimos 7 dias. Já o *website* “net-empregos.com” apresenta uma característica técnica muito própria, não permitindo visualizar mais de 300 ofertas numa única consulta, isto é, 15 ofertas de emprego por página, num limite máximo de 20 páginas.

Estudo da árvore HTML do *website* de ofertas de emprego

Identificados os maiores *websites* de ofertas de emprego portugueses, é altura de realizar um estudo da estrutura da sua árvore HTML, com o objetivo de localizar a informação a ser extraída, de forma a construir os caminhos XPath necessários para alimentar o algoritmo de extração de dados descrito no capítulo seguinte.

Serão estudadas três árvores HTML correspondentes aos *websites* “net-empregos.com”, “indeed.pt” e “emprego.sapo.pt”. Contudo, como o tipo de estudo é similar entre *websites*, só será apresentado o estudo da árvore HTML do *website* “net-empregos.com”. As árvores HTML dos *websites* “indeed.pt” e “emprego.sapo.pt” estarão disponíveis para consulta nos Anexos desta dissertação.

Forma de estudo

A Figura 8 representa a árvore HTML de uma oferta de emprego presente no *website* “net-empregos.com”. Este pedaço da árvore HTML contém os dados necessários a serem extraídos, isto é, os dados alvo. Esses dados são: a descrição da oferta de emprego, a data da publicação da oferta e o *Uniform Resource Locator* (URL) da oferta.

Esse estudo é realizado com auxílio da funcionalidade *Inspect* do *web browser* Google Chrome – Figura 7, que permite visualizar toda a árvore HTML do *website* alvo e estudar quais os elementos HTML que serão manipulados através do algoritmo apresentado no próximo capítulo.

Conceptualização e Modelação do Protótipo

Back	Alt+Left Arrow
Forward	Alt+Right Arrow
Reload	Ctrl+R
Save as...	Ctrl+S
Print...	Ctrl+P
Cast...	
Translate to English	
View page source	Ctrl+U
Inspect	Ctrl+Shift+I

Figura 7. Opção Inspect do navegador web Google Chrome.

```
<tbody>
  <tr>
    <td align="left" width="531" height="30" valign="bottom">
      <strong>
        <small>
          <font face="Arial" style="FONT-SIZE: 15px; LINE-HEIGHT: 14px">
            <a href="/5182847/licenciado-area-alimentar-m-f-lisboa-e-cascais-restauracao/">Licenciado área alimentar (m/f) - Lisboa e Cascais- Restauração</a>
          </font>
        </small>
      </strong>
    </td>
  </tr>
  <tr>
    <td align="left" valign="top" width="100%" height="1">
      <table align="left" border="0" cellpadding="0" cellspacing="0" width="738" height="35">
        <tbody>
          <tr>
            <td align="left" valign="top" width="40"></td>
            <td align="left" valign="top" width="140">
              <font face="Arial" style="FONT-SIZE: 11px; LINE-HEIGHT: 14px">&nbsp;   25-7-2018 </font>
            </td>
            <td align="left" valign="top" width="59"></td>
            <td align="left" valign="top" width="495"></td>
          </tr>
        </tbody>
      </table>
    </td>
  </tr>
</tbody>
```

Figura 8. Figura com a árvore HTML: exemplo de uma oferta – tem data, descrição e link para a oferta.

A manipulação da árvore HTML é conseguida através da linguagem XPath, que é suportada pelo *web browser* Google Chrome. O caminho XPath correto para o elemento HTML pretendido torna o processo de extração de dados mais fácil de ocorrer. O grande desafio prende-se com a construção do caminho XPath correto e que possa ser replicado em cada sessão / ligação ao *website* de ofertas de emprego. Na Figura XPath, verifica-se o caminho XPath para um nó ou elemento HTML onde se situam os dados correspondentes a uma oferta de emprego, mais concretamente a descrição de uma oferta de emprego.

Conceptualização e Modelação do Protótipo

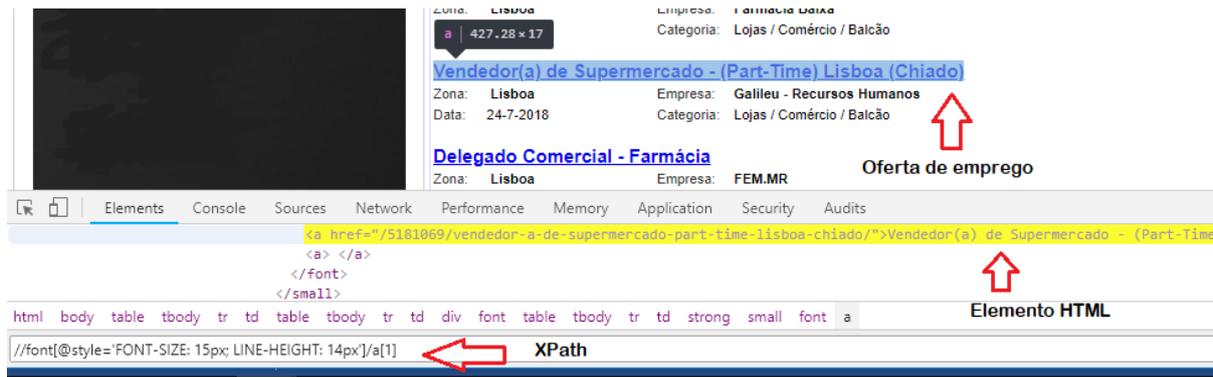


Figura 9. Exemplo de um caminho XPath de uma oferta de emprego.

A alteração da árvore HTML do *website* de ofertas de emprego acaba por interferir com os caminhos XPath utilizados no algoritmo de extração de dados, impedindo a extração correta dos mesmos. Contudo, os *websites* de ofertas de emprego, alvo de extração por parte desta dissertação, têm alguma maturidade, estabilidade e poucas alterações estruturais ao longo dos meses, tornando a extração de dados via HTML- XPath uma abordagem perfeitamente adequada e fiável.

Estudo da árvore HTML do *website* Net-Empregos

O objetivo do estudo da árvore HTML prende-se com a identificação das regiões onde se encontram os dados a extrair, para posteriormente desenhar os caminhos XPath necessários que irão indicar o caminho dos dados alvo ao algoritmo de extração. Isto é, o algoritmo de extração, com recurso à biblioteca de análise HTML, HTML Agility Pack, irá usar estes identificadores XPath para extrair somente os dados necessários.

O *website* “net-empregos.com” tem uma árvore HTML bastante padronizada, em que todo o conteúdo do *website* está contido dentro de tabelas – elemento HTML “table”, o que facilita a extração dos dados pretendidos, pois percorrer os diferentes elementos HTML de uma tabela, tais como “tr” e/ou “td” é uma tarefa bastante padronizada e facilitada com recurso à linguagem XPath.

Construção do caminho XPath

A construção de um caminho XPath começa pela identificação, de forma genérica, de uma oferta de emprego, como mostra a Figura 10.

<u>Licenciado área alimentar (m/f) - Lisboa e Cascais- Restauração</u>	
Zona: Lisboa	Empresa: Pluritema
Data: 25-7-2018	Categoria: Farmácia / Biotecnologia
<u>Operador Contact Center – Planos Nutricao - Lisboa</u>	
Zona: Lisboa	Empresa: Vertente Humana
Data: 25-7-2018	Categoria: Call Center / Help Desk
<u>Gestor de Cliente - LISBOA</u>	
Zona: Lisboa	Empresa: nutribalance
Data: 24-7-2018	Categoria: Comercial / Vendas
<u>Responsável loja - Ervanária em Lisboa</u>	
Zona: Lisboa	Empresa: Farmácia Baixa
Data: 24-7-2018	Categoria: Lojas / Comércio / Balcão
<u>Vendedor(a) de Supermercado - (Part-Time) Lisboa (Chiado)</u>	
Zona: Lisboa	Empresa: Galileu - Recursos Humanos
Data: 24-7-2018	Categoria: Lojas / Comércio / Balcão

Figura 10. Apresentação de ofertas de emprego do website “net-empregos.com”.

Depois de apresentadas as ofertas de emprego ao utilizador, é feita a inspeção, através da opção “inspect” anteriormente referida, para poder analisar a árvore HTML que está por detrás da parte gráfica apresentada ao utilizador. Na Figura 11, verifica-se o nó HTML de uma oferta.

```

<tbody>
  <tr>
    <td align="left" width="531" height="30" valign="bottom">
      <strong>
        <small>
          <font face="Arial" style="FONT-SIZE: 15px; LINE-HEIGHT: 14px">
            <a href="/5182847/licenciado-area-alimentar-m-f-lisboa-e-cascais-restauracao/">Licenciado área alimentar (m/f) - Lisboa e Cascais- Restauração</a>
          </font>
        </small>
      </strong>
    </td>
  </tr>
  <tr>
    <td align="left" valign="top" width="100%" height="1">
      <table align="left" border="0" cellpadding="0" cellspacing="0" width="738" height="35">
        <tbody>
          <tr>
            <td align="left" valign="top" width="40"></td>
            <td align="left" valign="top" width="140">
              <font face="Arial" style="FONT-SIZE: 11px; LINE-HEIGHT: 14px">&nbsp;   25-7-2018 </font>
            </td>
            <td align="left" valign="top" width="59"></td>
            <td align="left" valign="top" width="495"></td>
          </tr>
        </tbody>
      </table>
    </td>
  </tr>
</tbody>

```

Figura 11. Nó HTML de uma oferta presente no website “net-empregos.com”.

O nó HTML apresentado na Figura 11 apresenta toda a informação correspondente a uma oferta de emprego, isto é, apresenta a descrição, a data de publicação e o caminho URL para a oferta. Estes são os dados que se pretendem extrair, mas para isso é necessário ter o caminho XPath para os mesmos, que irá alimentar o algoritmo de extração.

Conceptualização e Modelação do Protótipo

Na Figura 12, é apresentado o caminho XPath para o nó HTML que contém a informação correspondente à data de publicação de uma oferta de emprego. De igual forma, consegue-se desenhar os caminhos XPaths para localizar o URL de uma oferta, bem como a sua descrição.

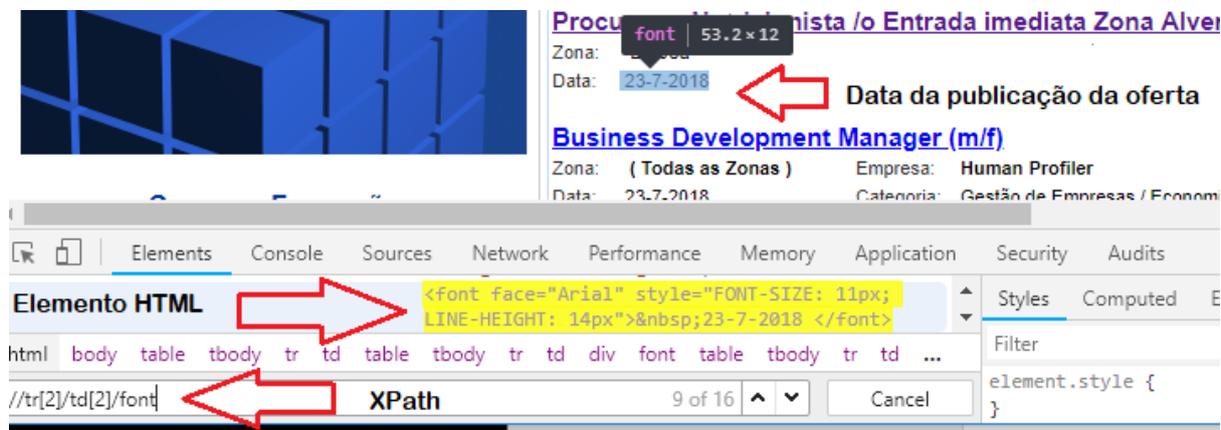


Figura 12. Elementos HTML a serem manipulados. Fonte: website “net-empregos.com”.

Estudo do URL do website Net-Empregos

O estudo do URL do website “net-empregos.com”, após uma pesquisa de emprego, é fundamental, uma vez que é através do de um caminho URL que a biblioteca HTML Agility Pack consegue analisar / extrair conteúdo de um website. O URL do website “net-empregos.com”, após uma pesquisa de emprego, é constituído da seguinte forma:

Estrutura URL		
Domínio	Parâmetro de pesquisa: “O Quê?”	Parâmetro de pesquisa: “Onde?”
http://www.net-empregos.com/	/nutricao-	lisboa/
http://www.net-empregos.com/	/ciencias-da-nutricao-	vila-real/

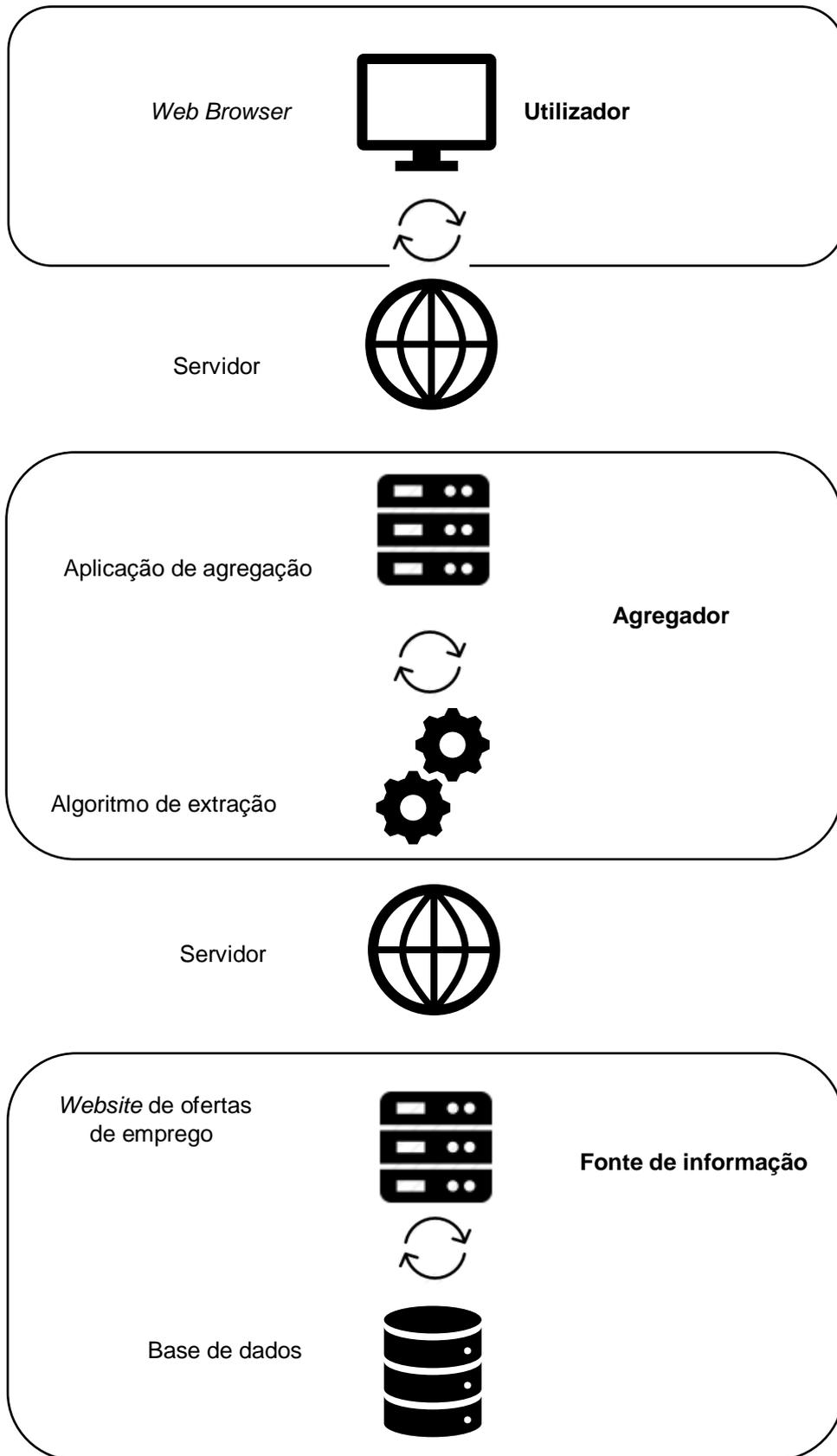
Tabela 12. Estrutura URL do website “net-empregos.com”. Fonte: Elaboração própria.

Exemplos de URLs		
Parâmetro de pesquisa: “O Quê?”	Parâmetro de pesquisa: “Onde?”	URL
Nutrição	Lisboa	http://www.net-empregos.com/nutricao-lisboa/
Ciências da Nutrição	Vila Real	http://www.net-empregos.com/ciencias-da-nutricao-vila-real/

Tabela 13. Exemplos de URLs do website “net-empregos.com”. Fonte: Elaboração própria.

Verifica-se que os parâmetros de pesquisa “O Quê?” e “Onde?” disponíveis no website são normalizados para poderem formar um URL com o caminho para a pesquisa desejada. Observa-se que todas as letras maiúsculas são transformadas em letras minúsculas, os caracteres que não sejam alfanuméricos são eliminados, os espaços entre as palavras são substituídos por hífen e, após o domínio do website, o primeiro valor a ser apresentado é o valor correspondente ao campo “O Quê?” e, em seguida, o valor correspondente ao campo “Onde?”.

Esquema de Funcionamento da Aplicação de Agregação de Conteúdo



Descrição do funcionamento da aplicação de agregação de conteúdo

Como representado no esquema de funcionamento da aplicação de agregação de conteúdo, pode observar-se que esse esquema é dividido em 3 blocos:

Utilizador – *web browser*;

Agregador – aplicação de agregação;

Fonte de informação – *website* de ofertas de emprego.

O presente tópico serve para descrever o bloco “Agregador”, onde se encontra a aplicação de agregação de conteúdo com o seu algoritmo de extração de dados. Contudo, é importante abordar o bloco “Utilizador” e “Fonte de informação”.

O bloco “**Utilizador**” diz respeito a todos os utilizadores que estejam a utilizar a internet com o objetivo de encontrar ofertas de emprego em Portugal. O utilizador pode aceder à aplicação de agregação de conteúdo diretamente através do seu URL correspondente ou pesquisando num *web browser*. Utilizando um *web browser*, o utilizador tem acesso à interface gráfica do *website* de agregação de ofertas de emprego e pode realizar uma pesquisa por ofertas de emprego, definindo uma descrição alvo, bem como o distrito desejado. Tanto a descrição como o distrito serão os dados recolhidos pelo agregador de conteúdo para orientar a pesquisa de ofertas de emprego, dentro dos *websites* alvo, que serão a fonte de informação.

O bloco “**Fonte de informação**” diz respeito aos *websites* de ofertas de emprego portugueses, onde estão publicados os dados das ofertas de emprego que os empregadores definiram. Estes *websites* são o substrato da aplicação de agregação de ofertas de emprego e serão consultados pelo algoritmo de extração de dados da aplicação de agregação sempre que um utilizador realizar uma nova pesquisa.

O bloco “**Agregador**” diz respeito à aplicação de agregação de ofertas de emprego propriamente dita. Esta aplicação realiza a ponte entre os empregadores e os potenciais candidatos a uma vaga de emprego. A aplicação aceita dois campos de pesquisa: a “descrição” e o “distrito”. A “descrição” está relacionada com as palavras-chave de pesquisa que o utilizador define para descrever a sua oferta de emprego alvo, enquanto que o distrito diz respeito à área geográfica onde o utilizador deseja pesquisar por uma oferta de emprego. Após o utilizador submeter o formulário de pesquisa, o algoritmo presente na aplicação de agregação é responsável por aceder aos *websites* de emprego pré-definidos – “net-empregos.com”, “indeed.pt”, “emprego.sapo.pt” – e recolher as ofertas de pesquisa dentro desses *websites*. Após o conteúdo ter sido extraído, é altura de devolver esses dados ao utilizador. Desta forma, o utilizador tem acesso a todas as ofertas de emprego que deseja visualizar, segundo a sua pesquisa, num único local sem ter de as pesquisar isoladamente em cada um dos *websites* de ofertas de emprego. Paralelamente a aplicação de agregação de ofertas de

Conceptualização e Modelação do Protótipo

emprego, verifica o número de ofertas de emprego disponíveis no distrito escolhido e o número de ofertas que correspondem à descrição pretendida. Desta forma, o utilizador tem também acesso à percentagem das ofertas de emprego existentes para uma determinada descrição, em relação ao total de ofertas existentes num determinado distrito.

De referir que a aplicação de agregação não realiza qualquer persistência de dados, isto é, não possui nenhuma base de dados que guarde qualquer tipo de oferta de emprego. O que é devolvido ao utilizador é a data da publicação da oferta de emprego, a descrição e o URL para a oferta em questão.

Conclusão do Capítulo

Neste capítulo, foram identificados os principais requisitos, e fontes de informação que irão alimentar o *website* de agregação de ofertas de emprego, isto é, foram identificados os principais *websites* de ofertas de emprego em Portugal e estudada a sua árvore HTML, bem como o seu URL após pesquisa, de forma a preparar o caminho XPath e URL para a extração dos dados alvos.

4. Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

Este capítulo pretende apresentar a construção da plataforma *web* de agregação de ofertas de emprego. É dado grande ênfase à construção e estrutura do algoritmo de extração de dados, pois é esse algoritmo que é o responsável pela extração de dados de diferentes *websites* e posterior apresentação desses dados num único local, de forma rápida e fiável, sem recorrer a APIs ou RSS feeds dos *websites* de ofertas de emprego, visto não possuírem esses instrumentos ou, quando os possuem, estarem desatualizados.

Construção da plataforma de agregação de ofertas de emprego

Framework .net Core 2.0

A *framework* .net Core 2.0 permite implementar o padrão de desenho *Model-View-Controller* (MVC), permitindo separar a lógica da aplicação em duas grandes camadas. A camada de apresentação e a camada de negócio. Esta *framework* foi eleita para a construção da plataforma de agregação de ofertas de emprego, pois segue o conceito *open source*, permite a aplicação do padrão de desenho MVC e toda a documentação de suporte à aplicação é organizada, revista e compilada pela empresa transnacional Microsoft.

Camada de apresentação

A camada de apresentação desta aplicação segue um fluxo padrão, mas ao mesmo tempo minimalista, de maneira a facilitar e a melhorar a experiência sentida pelo utilizador ao aceder à aplicação. Esta camada diz respeito à letra “V”, de *view*, do padrão de desenho MVC.

View

A *view* é o conjunto de todos os ficheiros do tipo *.cshtml*, onde, recorrendo à sintaxe Razor, se consegue embeber código servidor, *c#*, numa página tipicamente HTML. A *view* principal “Index” contém o HTML necessário para recolher os parâmetros de entrada do utilizador. Esta *view* tem pouco mais do que um formulário, onde o utilizador pode definir quais os seus critérios de pesquisa.

```
<form action="/" id="my-form" method="post">
  <div class="form-row align-items-center">
    <div class="col-auto">
      <input type="text" class="form-control" placeholder="Descrição" id="Descricao" name="Descricao" value="">
    </div>
    <div class="col-auto">
      <select id="Distrito" class="form-control">...</select>
    </div>
    <div class="col-auto">
      <button type="submit" class="btn btn-primary btn-block">Submeter</button>
    </div>
  </div>
</form>
```

Figura 13. Formulário HTML responsável por recolher os parâmetros de pesquisa definidos pelo utilizador. Fonte: Elaboração própria.

Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

Partial View

A *partial view* é uma *view* parcial que contém a tabela que irá apresentar os dados após a extração e a percentagem de ofertas relevantes em relação ao total de ofertas de um determinado distrito. Nesta camada, recorrendo à linguagem Razor, é possível iterar a Lista, presente no *model* desta aplicação, no próprio HTML. Pode observar-se na Figura 14 que será construído um elemento HTML "td" por cada parâmetro do tuplo - criado na camada de negócio - contido na lista que contém os dados das ofertas de emprego já extraídas.

```
<table class="table table-hover">
  <thead>
    <tr>
      <th scope="col">#</th>
      <th scope="col">Data</th>
      <th scope="col">Link</th>
      <th scope="col">Descrição</th>
    </tr>
  </thead>
  <tbody>
    @var i = 1;
    @foreach (var element in Model.Ofertas)
    {
      <tr>
        <th scope="row">@i</th>
        <td>@element.Item3</td>
        <td><a href="@element.Item1" target="_blank">Link da Oferta</a></td>
        <td>@element.Item2</td>
      </tr>
      @i++;
    }
  </tbody>
</table>
```

Figura 14. Tabela HTML responsável por conter as ofertas de emprego recolhidas.

Controller

A classe HomeController não faz parte da camada de apresentação ou de negócio, pois trata-se somente de um controlador, que transporta os dados e contacta as camadas/classes/objetos pré-definidos. É considerada nesta dissertação como uma classe pertencente à camada de apresentação unicamente para facilitar a leitura deste capítulo.

```
[HttpGet]
public virtual IActionResult Index()
{
    return View();
}

[HttpPost]
public virtual IActionResult Index(ExtractModel model)
{
    model.Ofertas = model.AgregarOfertas(model.Descricao, model.Distrito);
    return PartialView("_Partial", model);
}
```

Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

Figura 15. Controladores responsáveis pela alimentação das views com o model que contém os dados. Fonte: Elaboração própria.

Nesta classe, foram criados dois métodos com o objetivo de devolver ao utilizador a *view* principal que contém o formulário de pesquisa, bem como a *partial view* com a tabela de resultados de pesquisa, como apresentado na Figura 15.

O método “Index”, que não recebe nenhum argumento, devolve a *view* principal “Index”. A sua única função é apresentar ao utilizador o formulário de pesquisa presente na *view* principal, onde o utilizador irá colocar os seus parâmetros de pesquisa. Já o outro método “Index” recebe os parâmetros de pesquisa do utilizador – descrição e distrito -, guarda esses dados de pesquisa nas propriedades “Descricao” e “Distrito” do objeto *model* e preenche a propriedade Ofertas com a lista vinda do método AgregadorOfertas (algoritmo discutido mais adiante) e passando como argumentos desse método, as propriedades Descricao e Distrito referidas anteriormente.

Bootstrap

Bootstrap é a biblioteca responsável pela parte gráfica desta aplicação. Esta biblioteca *open source* possibilitou utilizar *Cascading Style Sheets* (CSS) já pré-definido, recorrendo às classes *container*, *jumbotron*, entre outras, para definir qual o CSS a adotar, como representado na Figura 16.

```
<div class="container">  
  <div class="jumbotron">
```

Figura 16. Criação de atributos do tipo *class* dentro dos elementos HTML *div* para utilizar o CSS pré-definido na biblioteca *bootstrap*. Fonte: Elaboração própria.

Camada de negócio

É na camada de negócio onde reside o código de extração de dados e, como tal, o “cérebro” da aplicação. Esta aplicação não possui base de dados, pois o objetivo não é armazená-los, mas sim utilizar dados de outros *websites* / fontes e apresentá-los num único local de forma estruturada. Com uma base de dados, duplicar-se-iam dados em vão, o que iria tornar a aplicação mais lenta, visto os dados extraídos terem de percorrer mais um fluxo – persistência e consulta de dados na base de dados. A letra “M” do padrão de desenho MVC acaba por não representar o modelo de dados da aplicação, pois esse modelo não existe por não existir base de dados, mas representa ~~sim~~ antes o algoritmo de extração de dados. É a classe *ExtractModel* que contém o algoritmo de extração de dados. Esta classe contém seis propriedades públicas, nove variáveis (três por cada *website* alvo de extração) privadas e treze métodos, dos quais doze métodos privados (quatro por *website* alvo de extração) e um público. De seguida, será apresentada a estrutura e funcionamento do algoritmo de extração de dados.

Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

Algoritmo de extração de dados

O algoritmo de extração de dados tem como objetivo receber os parâmetros de pesquisa do utilizador e devolver as ofertas de emprego correspondentes a esses parâmetros de pesquisa. De seguida, será apresentado o algoritmo utilizado para extrair dados no *website* net-empregos.com. Na Figura 17, e 18, pode observar-se a hierarquia de métodos, bem como as propriedades e variáveis globais presentes no algoritmo de extração de dados, respetivamente. O algoritmo foi programado para primeiro extrair os endereços (URLs) das páginas onde as ofertas de emprego alvo estão; de seguida, extrair as ofertas de cada uma dessas páginas, recolher o número de ofertas totais do distrito definido pelo utilizador; e, por fim, agregar todas as ofertas recolhidas nos três *websites* referidos e calcular a percentagem de ofertas relevantes ao utilizador, no universo total de ofertas presentes no distrito definido pelo mesmo.

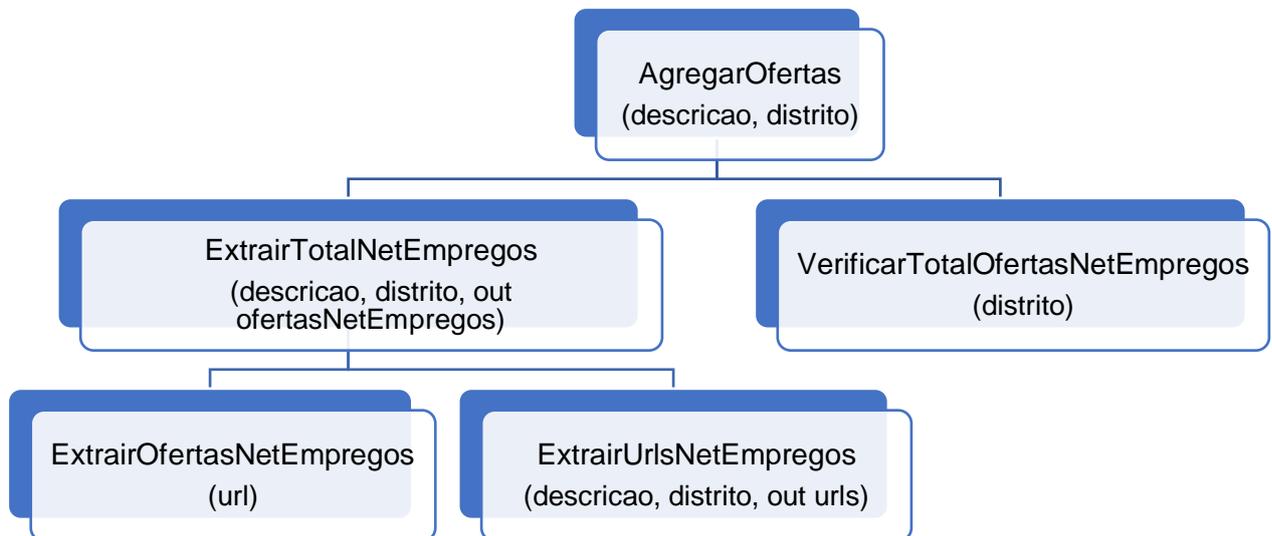


Figura 17. Hierarquia de métodos presentes no algoritmo de extração de dados (extração net-empregos.com). Fonte: Elaboração própria.

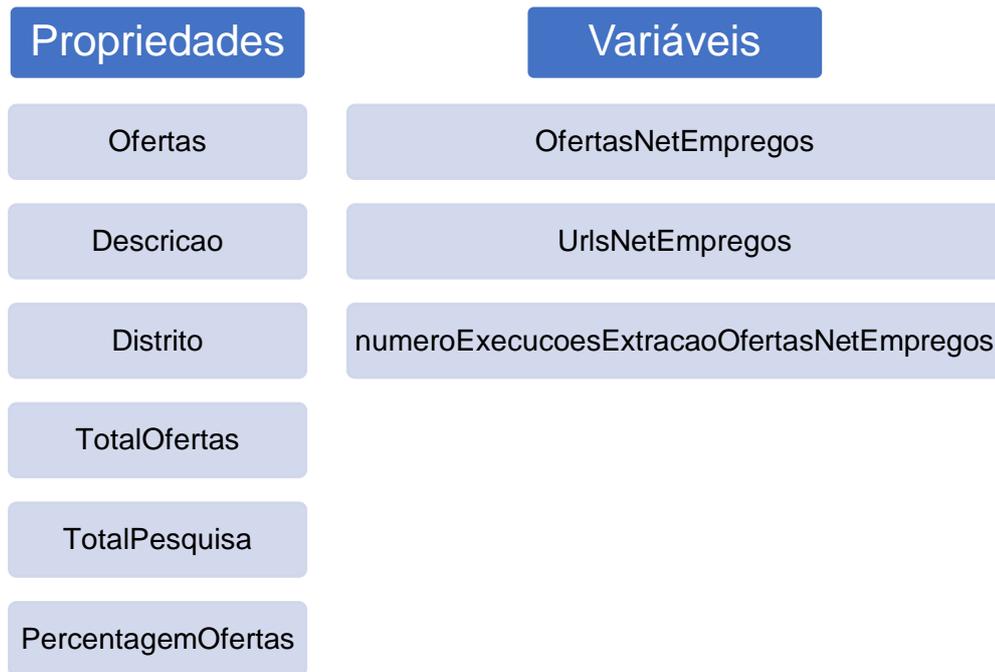


Figura 18. Propriedades e variáveis globais presentes no algoritmo de extração de dados (extração net-empregos.com).
Fonte: Elaboração própria.

Propriedades

A classe que contém o algoritmo de extração de dados possui seis propriedades globais, como se observa na Figura 18.

Estas propriedades (figura 19) permitem a troca de dados entre o utilizador, o algoritmo de extração de dados e a *partial view* devolvida ao utilizador com os resultados finais.

```
// Propriedades
// lista que agrega todas as ofertas relevantes ao utilizador de todos os websites
public List<Tuple<String, String, String>> Ofertas { get; set; }
// descricao da oferta definida pelo utilizador
public String Descricao { get; set; }
// distrito da oferta definido pelo utilizador
public String Distrito { get; set; }
// total de ofertas por distrito definido pelo utilizador
public double TotalOfertas { get; set; }
// total de ofertas relevantes ao utilizador
public double TotalPesquisa { get; set; }
// percentagem de ofertas relevantes ao utilizador em relação ao número total de ofertas em um determinado distrito
public double PercentagemOfertas { get; set; }
```

Figura 19. Propriedades presentes no algoritmo de extração de dados. Fonte: Elaboração própria.

A propriedade “**Ofertas**” é uma lista de tuplos com três itens do tipo de dados *string*. Esta propriedade guarda todas as ofertas de emprego a serem mostradas ao utilizador. O primeiro item de cada tuplo guarda o URL que dá acesso à oferta alvo; o segundo item guarda a descrição da oferta; e o terceiro guarda a data de publicação dessa oferta. Foi escolhida uma estrutura do tipo lista e não um *array* de *strings*, pois as listas são mais fáceis, intuitivas e rápidas de manipular.

Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

A propriedade “**Descricao**” é do tipo de dados *string* e representa a descrição desejada para uma oferta de emprego por parte do utilizador. Esta propriedade vai servir como parâmetro de entrada para o método de extração de dados “AgregaOfertas”.

A propriedade “**Distrito**” é do tipo de dados *string*, à imagem da propriedade anterior, e representa o distrito onde o utilizador deseja procurar por ofertas de emprego. Em conjunto com a anterior, esta propriedade também é um dos parâmetros de entrada para o método de extração de dados “AgregaOfertas”.

A propriedade “**TotalOfertas**” é do tipo de dados *double* e guarda o valor total de ofertas num distrito definido pelo utilizador. Esta propriedade serve para calcular a percentagem de ofertas relevantes ao utilizador, em relação ao total de ofertas presentes no distrito eleito para pesquisa.

A propriedade “**TotalPesquisa**” é do tipo de dados *double* e guarda o valor total de ofertas de acordo com os parâmetros de pesquisa do utilizador – descrição e distrito. Esta propriedade, em conjunto com a anterior, serve para calcular a percentagem de ofertas relevantes ao utilizador em relação ao total de ofertas presentes no distrito eleito para pesquisa.

A propriedade “**PercentagemOfertas**” é do tipo *double* e guarda o valor resultante do cálculo da percentagem de ofertas relevantes ao utilizador em relação ao total de ofertas presentes no distrito eleito para pesquisa.

Variáveis

No que toca a variáveis globais e relacionadas com a extração de informação do *website* “net-empregos.com”, o algoritmo de extração de dados apresenta três variáveis globais todas privadas, como representado na Figura 20.

```
// NetEmpregos
// lista que guarda todas as ofertas relevantes ao utilizador
private List<Tuple<String, String, String>> OfertasNetEmpregos;
// lista que guarda os urls das páginas onde estão publicadas as ofertas
private List<String> UrlsNetEmpregos;
// verifica se o metodo "ExtrairOfertasNetEmpregos" já finalizou todas as extrações
private int numeroExecucoesExtracaoOfertasNetEmpregos = 0;
```

Figura 20. Variáveis presentes no algoritmo de extração de dados (extração net-empregos.com). Fonte: Elaboração própria.

A variável “**OfertasNetEmpregos**” é uma lista do mesmo tipo que a propriedade “Ofertas”, mas ao invés de guardar todas as ofertas, guarda somente as ofertas respeitantes à extração do *website* “net-empregos.com”.

A variável “**UrlsNetEmpregos**” é uma lista do tipo de dados *string* e guarda os endereços (url) das páginas onde estão publicadas as ofertas a serem alvo de extração. Esta variável vai ser alimentada pelo método “ExtrairUrlsNetEmpregos”.

Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

A variável “**numeroExecucoesExtracaoOfertasNetEmpregos**” é um inteiro, que serve para verificar se o método “**ExtrairOfertasNetEmpregos**” já finalizou a sua execução. Para agilizar o processo de extração de dados, foram usadas *threads* e, como tal, é necessário verificar, no método “**ExtrairTotalNetEmpregos**”, se todas as *threads* executaram o método que lhes estava agregado. Esta variável é a responsável por essa verificação.

Método de extração das páginas (URLs) onde estão publicadas as ofertas alvo

Este método é responsável por recolher os endereços (URLs) das páginas onde as ofertas de emprego alvo foram publicadas, para posteriormente recolher as ofertas presentes em cada uma das páginas.

- Assinatura do método

O método “**ExtrairUrlsNetEmpregos**” tem a seguinte assinatura:

```
private void ExtrairUrlsNetEmpregos(String descricao, String distrito, out List<String> urls)
```

Figura 21. Assinatura do método “**ExtrairUrlsNetEmpregos**”. Fonte: Elaboração própria.

Como parâmetro de entrada, este método recebe dois argumentos. Esses argumentos são definidos pelo utilizador, são do tipo de dados *string* e representam o parâmetro de pesquisa dentro de cada oferta de emprego (“*descricao*”), bem como a sua região geográfica (“*distrito*”).

- Corpo do método

O corpo do método inicia-se com a instanciação de uma lista de *Strings* com o objetivo de guardar nessa lista a(s) URL(s) ou páginas web onde se encontram publicadas as ofertas de emprego. Esta lista acaba por ser o resultado final deste método.

```
urls = new List<String>();
```

Figura 22. Objeto que contém os endereços (URLs) das páginas onde estão publicadas as ofertas a serem alvo de extração. Fonte: Elaboração própria.

- Tratamento e criação do URL de pesquisa

Em seguida, inicia-se o tratamento dos parâmetros de entrada deste método, definidos pelo utilizador, com o objetivo final de preencher a lista anterior (URLs) com o(s) caminho(s) para a(s) página(s) web onde se encontram publicadas as ofertas de emprego dentro de um único *website* de oferta de emprego. O tratamento do parâmetro de entrada *descricao* e *distrito* pode ser observado na Figura 23.

Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

```
descricao = Regex.Replace(descricao, " ", "-");
descricao = descricao.UniDecode();
descricao = Regex.Replace(descricao, "[^A-Za-z0-9-]", "");

distrito = Regex.Replace(distrito, " ", "-");
distrito = distrito.UniDecode();
distrito = Regex.Replace(distrito, "[^A-Za-z0-9-]", "");
```

Figura 23. Tratamento dos parâmetros de entrada descrição e distrito. Fonte: Elaboração própria.

O tratamento dos parâmetros de entrada ocorre da seguinte forma:

Os parâmetros de entrada descricao e distrito são tratados com o objetivo de substituir qualquer espaço por um “-” ou remover qualquer caractere que não seja alfanumérico, pois o URL não permite esse tipo de caracteres. O caractere “-” serve para separar cada palavra (de pesquisa) dentro do URL. O método “UniDecode” serve para representar o Unicode em caracteres do tipo *American Standard Code for Information Interchange* (ASCII).

De seguida, os parâmetros de entrada “descrição” e “distrito” são concatenados com o URL `http://www.net-empregos.com/` com o objetivo de fornecer um URL final que será adicionado a Lista “URLs”, como exemplificado na Figura 24.

```
urls.Add("http://www.net-empregos.com/" + descricao.ToLower() + "-" + distrito.ToLower() + "/");
```

Figura 24. Definição / Criação das urls das páginas alvo de extração.

- Extração da página alvo utilizando a classe HtmlWeb

É instanciado da class `HtmlWeb` um objeto com o nome “web”. Esse objeto (Figura 25) pertence à biblioteca HTML Agility Pack que permite extrair o HTML de uma página web, e posteriormente manipulá-lo. A biblioteca HTML Agility Pack foi eleita em relação às restantes bibliotecas de apoio ao *web scraping*, como apresentadas no capítulo do estado da arte, pois é uma biblioteca que suporta a linguagem XPath, facilitando a extração de dados, segue as boas práticas recomendadas pela W3C, e tem um bom desempenho a analisar HTML mal formato ou que não segue as boas práticas definidas pela W3C.

A propriedade “`OverrideEncoding`” do objeto web é preenchida com o objetivo de permitir tratar caracteres como “ç” ou “~” muito presentes na língua portuguesa.

```
HtmlWeb web = new HtmlWeb
{
    OverrideEncoding = Encoding.GetEncoding("ISO-8859-1")
};
```

Figura 25. Objeto do tipo `HtmlWeb` com a definição de uma atribuição de encoding à propriedade `OverrideEncoding`. Fonte: Elaboração própria.

Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

É criado um objeto do tipo `HtmlDocument` com o nome “doc”, que irá guardar a página alvo, segundo o URL adicionado à lista URLs anteriormente, onde serão extraídas as ofertas de emprego. Sempre que este objeto é instanciado, há obrigatoriamente uma ligação entre a aplicação e uma entidade externa (*website* de ofertas de emprego). Esta ligação é o ponto onde o algoritmo despende mais tempo, pois terá de se ligar ao URL definido e extrair ao seu HTML. Este tipo de ação é reduzida no algoritmo, pois compromete a performance da aplicação, visto que cada pedido troca de dados com entidades externas consome sempre mais tempo.

É realizado um ciclo do tipo `for` para poder iterar sobre a lista URLs. Embora nesta fase essa lista tenha somente um índice, um URL, durante o ciclo `for` o objeto URLs é alimentado – caso exista - com mais um URL. Este segundo URL representa a segunda página de ofertas de emprego dentro do *website* de ofertas de emprego, segundo os critérios de pesquisa definidos pelo utilizador. A lógica é verificar se existe, dentro de cada página, uma página seguinte e adicionar o seu URL à lista, como demonstra a Figura 26.

```
int pagina = 2;

HtmlDocument doc = new HtmlDocument();

doc = web.Load(urls[0]);
for (int i = 0; i < urls.Count; i++)
{
    var paginaSeguinte = doc.DocumentNode.SelectNodes("//a[text()=' " + pagina + "']");

    if (paginaSeguinte != null)
    {
        urls.Add("http://www.net-empregos.com/" + paginaSeguinte[0].Attributes["href"].Value);
    }
    pagina++;
}
```

Figura 26. Extração da página seguinte utilizando a biblioteca HTML Agility Pack. Fonte: Elaboração própria.

A variável “paginaSeguinte” guarda os nós da árvore HTML que representam os locais onde estão os endereços para as páginas seguintes.

- Retorno do método

Apesar de ser um método do tipo *void*, este método tem como retorno a lista de *strings* com os URLs extraídos anteriormente. Esta lista é retornada como argumento do método, através da palavra reservada `out` para se poder chamar este método numa *thread* criada no método “ExtrairTotalNetEmpregos”.

Método de extração de ofertas segundo uma página (url)

O método “ExtrairOfertasNetEmpregos” é responsável por extrair todas as ofertas de emprego de uma determinada página web. Isto é, este método recebe um URL como parâmetro de entrada e recolhe desse URL todas as ofertas de emprego que possam existir.

Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

- Assinatura do método

Este método tem uma assinatura bastante simples, como mostra a Figura 27.

```
private void ExtrairOfertasNetEmpregos(String url)
```

Figura 27. Assinatura do método “ExtrairOfertasNetEmpregos”. Fonte: Elaboração própria.

Como parâmetro de entrada, este método recebe somente um argumento. Esse argumento é definido no método descrito anteriormente e trata-se de uma *string*, que contém o endereço URL que possui as ofertas de emprego alvo de extração.

- Corpo do método

Este método inicia-se com a criação de uma lista de tuplos com três itens do tipo de dados *string* cada. Este objeto “ofertas” irá guardar os dados recolhidos no *website* de ofertas de emprego, segundo os parâmetros de pesquisa fornecidos pelo utilizador. Cada índice da Lista “ofertas” guarda um tuplo com três índices. O primeiro índice serve para guardar a descrição da oferta de emprego, o segundo item o *link* que dá acesso à oferta de emprego e, por fim, o último item guarda a data da publicação da oferta de emprego. A criação do objeto “ofertas” é apresentada na Figura 28.

```
List<Tuple<String, String, String>> ofertas = new List<Tuple<String, String, String>>();
```

Figura 28. Objeto responsável por guardar todas as ofertas de emprego extraídas. Fonte: Elaboração própria.

De seguida, é instanciado da class *HtmlWeb* um objeto com o nome “web”, como apresentado na Figura 29. Esse objeto pertence à biblioteca *HTML Agility Pack* que permite extrair o *HTML* de uma página web e, posteriormente, manipulá-lo. A propriedade “*OverrideEncoding*” do objeto *web* é preenchida com o objetivo de permitir apresentar caracteres como “ç” ou “~” muito presentes na língua portuguesa ao utilizador final.

```
HtmlWeb web = new HtmlWeb  
{  
    OverrideEncoding = Encoding.GetEncoding("ISO-8859-1")  
};
```

Figura 29. Objeto do tipo *HtmlWeb* com a definição de uma atribuição de *encoding* à propriedade *OverrideEncoding*.
Fonte: Elaboração própria.

Posteriormente, é criado um objeto do tipo *HtmlDocument* com o nome “doc”, que irá guardar a página alvo, segundo o URL fornecido como parâmetro de entrada neste método, onde serão extraídas as ofertas de emprego. Sempre que este objeto é instanciado, há obrigatoriamente uma ligação entre a aplicação e uma entidade externa (*website* de ofertas de emprego), como já referido anteriormente. Na Figura 30, pode ser observada a criação do objeto “doc”.

```
HtmlDocument doc = new HtmlDocument();  
doc = web.Load(url);
```

Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

Figura 30. Extração da página alvo. Fonte: Elaboração própria.

As três variáveis criadas de seguida, “nosUrl”, “nosDescricao”, e “nosData” pretendem armazenar os nós pertencentes à árvore HTML do *website* extraído, onde a informação pretendida se encontra. Faz-se recurso à linguagem XPath, como se observa na Figura 30, para identificar esses nós HTML. Sempre que a estrutura do *website* é alterada, há um grande risco dos caminhos XPath definidos deixarem de funcionar e terá de ser alvo de intervenção por parte do programador, para identificar as alterações e reescrever os novos caminhos XPath.

```
var nosUrl = doc.DocumentNode.SelectNodes("//font[@style='FONT-SIZE: 15px; LINE-HEIGHT: 14px']/..");
var nosDescricao = doc.DocumentNode.SelectNodes("//font[@style='FONT-SIZE: 15px; LINE-HEIGHT: 14px']");
var nosData = doc.DocumentNode.SelectNodes("//tr[2]/td[2]/font");
```

Figura 31. Localização dos nós alvo na árvore HTML. Fonte: Elaboração própria.

Verifica-se, de seguida, se a variável “nosDescricao” contém qualquer tipo de dados, isto é, se se verifica algum nó com o XPath “//font[@style='FONT-SIZE: 15px; LINE-HEIGHT: 14px]” no documento extraído. Caso a variável “nosDescricao” seja nula, significa que não existe nenhuma oferta de emprego no URL que alimenta o método. Caso a variável “nosDescricao” apresente valor diferente de nulo, passa-se à extração dos dados associados a cada oferta de emprego, como apresentado na Figura 32.

```
if (nosDescricao != null)
{
    for (int i = 0; i < nosUrl.Count; i++)
    {
        String valorUrl = nosUrl[i].OuterHtml.ToString();
        String valorDescricao = nosDescricao[i].InnerText;
        String valorData = (nosData[i + 1].InnerText.ToString()).Replace("&nbsp;", "");

        String valorUrlCut = valorUrl.Substring(valorUrl.IndexOf("/"));
        String valorUrlFinal = valorUrlCut.Substring(0, valorUrlCut.IndexOf(">") - 1);
        ofertas.Add(new Tuple<String, String, String>("http://www.net-empregos.com/" + valorUrlFinal, valorDescricao, valorData));
    }
}
```

Figura 32. Validação para posterior extração dos dados associados a cada oferta de emprego. Fonte: Elaboração própria.

Nesta fase do algoritmo, são extraídos os dados pretendidos do *website* de ofertas de emprego armazenados na lista “ofertas”, para, posteriormente, mostrar esses dados ao utilizador. Os dados a extrair são: URL da oferta de emprego, descrição dessa oferta e data de publicação. Para realizar a extração desses dados, é necessário percorrer as coleções “nosUrl”, “nosDescricao” e “nosData”, que correspondem aos nós da árvore HTML, onde estão publicadas as ofertas de emprego.

A variável “valorUrl” é do tipo de dados String e guarda o valor do atributo *href* presente no nó “nosUrl” no índice “i”. “i” representa o número inteiro de zero até ao número máximo de elementos presentes no nó “nosUrl”. Essa variável – “i” - é essencial para percorrer todas as ofertas de emprego publicadas no *website* de ofertas de emprego e extrair os dados alvo. O valor contido no atributo href do elemento “a” HTML, do nó a ser tratado, representa o URL da oferta de emprego que está a ser tratada e é armazenado na variável “valorUrl”.

Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

A descrição de cada oferta de emprego é armazenada na variável “valorDescricao” do tipo de dados String e corresponde ao texto/valor presente em cada índice “i” do nó “nosDescricao”.

Por fim, a variável “valorData” do tipo de dados String armazena a data correspondente à publicação da oferta de emprego. Essa data está contida em cada índice “i+1” do nó “nosData” e tem de ser tratada com o método Replace(), para remover a propriedade “non-breaking space” presente na tag HTML que contém o valor da data de publicação da oferta de emprego.

As variáveis “valorUrl”, “valorDescrição” e “valorData” são armazenadas num tuplo que corresponde a cada índice da lista “ofertas”.

Para finalizar o corpo do método, como apresentado na Figura 33, realiza-se o preenchimento das variáveis, “OfertasNetEmpregos”, “numeroExecucoesExtracaoOfertasNetEmpregos” e a propriedade “TotalPesquisa” descritas anteriormente.

```
TotalPesquisa += ofertas.Count();
for (int i = 0; i < ofertas.Count; i++)
{
    OfertasNetEmpregos.Add(ofertas[i]);
}
numeroExecucoesExtracaoOfertasNetEmpregos++;
```

Figura 33. Preenchimento das variáveis e propriedade necessárias para guardar o resultado final do método “ExtrairOfertasNetEmpregos”. Fonte: Elaboração própria.

Método de agregação de todas as ofertas alvo

O método “ExtrairTotalNetEmpregos” serve para agregar todas as ofertas de emprego do *website* alvo, segundo os parâmetros de pesquisa definidos pelo utilizador.

- Assinatura do método

A Figura 34 apresenta a assinatura do método “ExtrairTotalNetEmpregos”, que recebe dois argumentos de entrada e devolve um parâmetro de saída. Os parâmetros de entrada são os valores “descricao” e “distrito” definidos pelo utilizador. Já a lista “ofertasNetEmpregos” representa o resultado final do método, que é passado como argumento do método para que este possa ser usado e chamado por uma *thread* dedicada à melhoria da performance do algoritmo de extração.

```
private void ExtrairTotalNetEmpregos(String descricao, String distrito, out List<Tuple<String, String, String>> ofertasNetEmpregos)
```

Figura 34. Assinatura do método “ExtrairTotalNetEmpregos”. Fonte: Elaboração própria.

- Corpo do método

Este método nasce com a criação da lista que vai armazenar todas as ofertas relevantes ao utilizador. De seguida, o método “ExtrairUrlsNetEmpregos” é chamado para que sejam recolhidos os caminhos URLs, onde estão publicadas as ofertas a serem extraídas. Por fim, é criada uma lista de *threads* com o objetivo de criar uma *thread* por cada URL extraído, no método chamado

Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

anteriormente. Cada *thread* vai ser responsável por executar o método “*ExtrairOfertasNetEmpregos*”, que irá extrair todas as ofertas de emprego de cada URL definido anteriormente. A lista de *threads* é iniciada dentro de um ciclo *foreach* e, por fim, é realizado um ciclo *while* com o objetivo de esperar que todos os métodos executados pelas *threads* sejam executados. Todos estes passos podem ser observados na Figura 35.

```
ofertasNetEmpregos = new List<Tuple<String, String, String>>();

ExtrairUrlsNetEmpregos(descricao, distrito, out UrlsNetEmpregos);
List<Thread> executarThreads = new List<Thread>();

foreach (var url in UrlsNetEmpregos)
{
    executarThreads.Add(new Thread(new ThreadStart(() => ExtrairOfertasNetEmpregos(url))));
}

foreach (var thread in executarThreads)
{
    thread.Start();
    thread.Join(TimeSpan.Zero);
}

int d = 0;
while (numeroExecucoesExtracaoOfertasNetEmpregos < UrlsNetEmpregos.Count)
{
    d++;
}
```

Figura 35. Corpo do método “*ExtrairTotalNetEmpregos*”. Fonte: *Elaboração própria*.

Método de contagem das ofertas totais num distrito

O método “*VerificarTotalOfertasNetEmpregos*” é responsável por recolher o número total de ofertas num determinado distrito definido pelo utilizador. Esse valor aqui recolhido serve para calcular a percentagem de ofertas relevantes ao utilizador, em relação ao número total de ofertas disponíveis num determinado distrito. O código presente neste método é explicável por si só, como mostra a Figura 36, visto todos os objetos e variáveis criadas dele já terem sido explicadas anteriormente.

```
private void VerificarTotalOfertasNetEmpregos(String distrito)
{
    List<String> urls = new List<String>();

    distrito = Regex.Replace(distrito, " ", "-");
    distrito = distrito.UniDecode();
    distrito = Regex.Replace(distrito, "[^A-Za-z0-9--]", "");

    urls.Add("http://www.net-empregos.com/" + distrito.ToLower() + "/");

    HtmlWeb web = new HtmlWeb
    {
        OverrideEncoding = Encoding.GetEncoding("ISO-8859-1")
    };

    HtmlDocument doc = web.Load(urls[0]);

    var numeroPagina = 2;

    while (doc.DocumentNode.SelectNodes("//a[text()=' " + numeroPagina + "']") != null)
    {
        numeroPagina++;
    }

    if (doc.DocumentNode.SelectNodes("//a[text()=' " + 2 + "']") == null)
    {
        TotalOfertas += 15;
    }
    else
    {
        TotalOfertas = (numeroPagina - 1) * 15;
    }
}
```

Figura 36. Método “VerificarTotalOfertasNetEmpregos”. Fonte: Elaboração própria.

Método de agregação de todas as ofertas

Por fim, o método “AgregarOfertas” que representa o ponto onde todas as ofertas de emprego alvo, dos diferentes *websites*, são agregadas.

- Assinatura do método

Este último método, que será descrito e apresentado na Figura 37, tem como parâmetros de entrada dois argumentos. Estes argumentos correspondem aos critérios de pesquisa definidos pelo utilizador – “descricao” e “distrito”.

```
public List<Tuple<String, String, String>> AgregarOfertas(String descricao, String distrito)
```

Figura 37. Assinatura do método “AgregarOfertas”. Fonte: Elaboração própria.

- Corpo do método

O corpo do método inicia-se com a criação da lista “ofertas”, que segue a mesma filosofia da propriedade global “Ofertas” anteriormente descrita.

Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

De seguida, são criadas e inicializadas as *threads* responsáveis por executar os métodos que irão extrair as ofertas alvo de cada um dos *websites* de ofertas de emprego. Foi necessário usar *threads* para tornar a chamada dos diferentes métodos de forma assíncrona e, com isto, diminuir o tempo de extração dos dados alvo. A chamada do método “VerificarTotalOfertasNetEmpregos” não é feita com recurso a *threads*, pois trata-se de um método com um tempo de execução curto e inferior à execução das *threads* criadas anteriormente. É criado um ciclo *while* para que o algoritmo de extração espere pela execução de todos os métodos chamados pelas *threads*. A Figura 38 exibe o código descrito acima.

```
List<Tuple<String, String, String>> ofertas = new List<Tuple<String, String, String>>();

List<Thread> executarThreads = new List<Thread>
{
    new Thread(new ThreadStart(() => ExtrairTotalNetEmpregos(descricao, distrito, out OfertasNetEmpregos))),
    new Thread(new ThreadStart(() => ExtrairTotalIndeed(descricao, distrito, out OfertasIndeed))),
    new Thread(new ThreadStart(() => ExtrairTotalSapoEmpregos(descricao, distrito, out OfertasSapoEmprego))),
};

foreach (var thread in executarThreads)
{
    thread.Start();
    thread.Join(TimeSpan.Zero);
}

VerificarTotalOfertasNetEmpregos(distrito);
VerificarTotalOfertasIndeed(distrito);
VerificarTotalOfertasSapoEmpregos(distrito);

int d = 0;
while (executarThreads.Any(x => x.IsAlive))
{
    d=1;
}
```

Figura 38. Criação das *threads* responsáveis pela aceleração do algoritmo de extração de dados. Fonte: Elaboração própria.

O próximo passo tem o objetivo de agregar todas as ofertas recolhidas nos diferentes *websites* num único objeto, como mostra a Figura 39. Primeiro é feita uma validação e, caso os objetos que contêm as ofertas de cada um dos *websites* não estejam nulos, são iterados com o objetivo de alimentar a lista “ofertas” que agrega todas as ofertas recolhidas dos diferentes websites.

```
if (OfertasNetEmpregos != null)
{
    foreach (var elementNetEmpregos in OfertasNetEmpregos)
    {
        ofertas.Add(new Tuple<String, String, String>(elementNetEmpregos.Item1, elementNetEmpregos.Item2, elementNetEmpregos.Item3));
    }
}

if (OfertasIndeed != null)
{
    foreach (var elementIndeed in OfertasIndeed)
    {
        ofertas.Add(new Tuple<String, String, String>(elementIndeed.Item1, elementIndeed.Item2, elementIndeed.Item3));
    }
}

if (OfertasSapoEmprego != null)
{
    foreach (var elementSapoEmprego in OfertasSapoEmprego)
    {
        ofertas.Add(new Tuple<String, String, String>(elementSapoEmprego.Item1, elementSapoEmprego.Item2, elementSapoEmprego.Item3));
    }
}
```

Desenvolvimento de um Protótipo de Extração de Ofertas de Emprego na Web

Figura 39. Agregação de todas as ofertas alvo no mesmo objeto. Fonte: Elaboração própria.

Por fim, é calculada a percentagem de ofertas relevantes ao utilizador em relação ao total de ofertas presentes no distrito eleito e é feito o retorno do objeto “ofertas”, como representado na Figura 40.

```
if (TotalPesquisa != 0)
{
    PercentagemOfertas = Math.Round((TotalPesquisa * 100) / TotalOfertas, 2);
}
else
{
    PercentagemOfertas = 0;
}
return ofertas;
```

Figura 40. Cálculo da percentagem de ofertas relevantes ao utilizador em relação ao total de ofertas presentes no distrito alvo e retorno do método. Fonte: Elaboração própria.

- Retorno do método

Este método retorna uma lista de tuplos com três itens do tipo de dados *string*, à imagem da propriedade “Ofertas” descrita anteriormente. Esta lista representa o conjunto de ofertas de emprego que corresponde aos critérios de pesquisa definidos pelo utilizador. Está será a lista que vai alimentar a interface gráfica a ser devolvida ao utilizador.

Nota:

Os algoritmos de extração de dados dos sites emprego.sapo.pt e indeed.pt seguem a mesma lógica que o algoritmo de extração de dados apresentado neste capítulo, como tal esses dois algoritmos são descritos no capítulo reservado a Anexos.

Conclusão do Capítulo

Neste capítulo, foi apresentada a estrutura e construção da plataforma de extração de dados, com especial foco no algoritmo de extração de dados. Este algoritmo recorre à biblioteca HTML Agility Pack para interpretar a árvore HTML dos *websites* de ofertas de emprego e para extrair os dados requeridos, com recurso à linguagem XPath para identificar os nós alvo, da árvore HTML da página, que contêm os dados a serem extraídos. A biblioteca HTML Agility Pack foi a eleita, entre as suas rivais, pois trata-se de uma ferramenta *open source*, independente, veloz, e bastante versátil.

5. Avaliação e Teste do Protótipo

O presente capítulo serve para apresentar os testes unitários, de integração, funcionais e aceitação realizados à plataforma de agregação de dados proposta nesta dissertação. Estes testes servem para avaliar a plataforma e validar se a mesma está preparada para ser utilizada pelos utilizadores finais.

Metodologia adotada

Foram realizados quatro tipos de testes à plataforma recém-criada, com o objetivo de detetar alguma área de melhoria e, desta forma, avaliar se a plataforma está pronta para ser disponibilizada ao utilizador final. Os testes realizados foram:

- Testes unitários

Estes são os testes mais simples e de mais baixo nível. Têm o objetivo de testar os métodos do algoritmo de extração e agregação de dados de forma individual.

- Testes de integração

Os testes de integração pretendem combinar os testes unitários entre si e verificar se os seus resultados seguem o valor esperado.

- Testes funcionais

Os testes funcionais só verificam o resultado final das ações a serem testadas e confronta esse resultado com o resultado teórico esperado. Não há especial atenção à performance da aplicação nem à forma como as ações são executadas.

- Testes de aceitação

Por fim, os testes de aceitação são testes formais, que simulam o comportamento do utilizador final. Estes testes são realizados no mesmo ambiente que o utilizador final tem acesso.

Testes e Resultados

Os resultados dos testes realizados à plataforma, são apresentados nas tabelas seguintes:

Bateria de testes unitários	
Teste	Resultado – Negativo / Positivo
Testar a capacidade de codificação e normalização dos parâmetros de pesquisa definidos pelo utilizador, nos métodos que consomem esses dados	Positivo
Testar se a criação dos URLs para as páginas onde se encontram as ofertas alvo	Positivo

Avaliação e Teste do Protótipo

publicadas são URLs válidos e conforme o esperado	
Testar se os caminhos XPath's definidos nos métodos apontam para os nós HTML desejados	Positivo
Testar se a recolha e tratamento dos dados a apresentar ao utilizador são os esperados	Positivo

Tabela 14. Bateria de testes unitários. Fonte: Elaboração própria.

Bateria de testes de integração	
Teste	Resultado – Negativo / Positivo
Testar se os métodos de agregação de ofertas de cada website utilizam os dados de forma correta vindos dos métodos de extração de URLs e extração de ofertas de cada website, e criam os objetos de saída esperados	Positivo
Testar se todas as threads iniciadas chamam os métodos esperados, e na ordem esperada.	Positivo
Testar se a propriedade “Ofertas” transporta todos os dados extraídos dos websites alvo de extração	Positivo
Testar se a partial view comunica com a view principal	Positivo

Tabela 15. Bateria de testes de integração. Fonte: Elaboração própria.

Bateria de testes funcionais	
Teste	Resultado – Negativo / Positivo
Comparar o número de ofertas de emprego devolvidas pela plataforma de agregação com os websites que estão a ser consumidos pela mesma.	Positivo
Comparar o número de ofertas totais por distrito devolvidas pela plataforma de agregação com os websites que estão a ser consumidos pela mesma.	Positivo
Testar se o valor calculado para a percentagem de ofertas relevantes em relação ao total de ofertas em um distrito tem o valor esperado	Positivo
Testar se o caminho (url) para as ofertas de emprego alvo corresponde de facto à oferta publicada no website consumido pela plataforma	Positivo

Tabela 16. Bateria de testes funcionais. Fonte: Elaboração própria.

Bateria de testes de aceitação	
Teste	Resultado – Negativo / Positivo
Testar a responsividade da plataforma – <i>mobile, tablet e desktop</i>	Positivo
Testar a compatibilidade da plataforma em diferentes <i>web browsers</i> , e sistemas operativos	Positivo
Testar o tempo de pesquisa – tempo decorrido entre a plataforma receber os parâmetros de pesquisa, e devolver o resultado final ao utilizador	Positivo
Testar a resiliência da plataforma – teste de carga, onde várias pesquisas são	Negativo – teste dependente das características, configurações, e plano



Figura 42. Cloud Tag resultante da questão "Quais os aspectos mais fracos que considera existirem na plataforma web recém-criada em relação às suas congêneres?". Fonte: www.wordclouds.com.

Na Figura 42 observa-se os aspectos mais fracos que existirem na plataforma web recém-criada em relação às suas congêneres são: velocidade, e a falta de mais filtros de pesquisa.

O estudo da questão "Quais considera serem os benefícios do uso da plataforma web recém-criada no auxílio de novas oportunidades de emprego?" pode-se verificar na Figura 43 através de uma *cloud tag*.

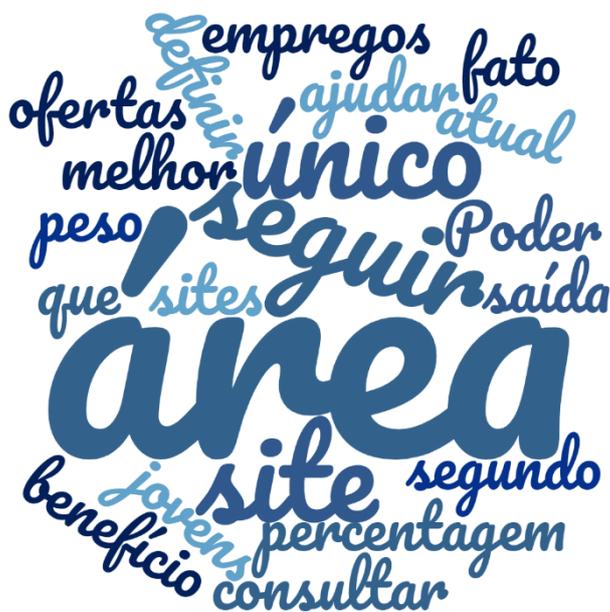


Figura 43. Cloud Tag resultante da questão "Quais considera serem os benefícios do uso da plataforma web recém-criada no auxílio de novas oportunidades de emprego?". Fonte: www.wordclouds.com.

Na Figura 43 observa-se os benefícios do uso da plataforma web recém-criada no auxílio de novas oportunidades de emprego são: consultar a percentagem de ofertas segundo uma área (profissional), verificar o peso de uma área (profissional), e ajudar a definir uma área (profissional)

O resultado após a questão “Quais considera serem os inconvenientes do uso da plataforma web recém-criada no auxílio de novas oportunidades de emprego?” é observado na Figura 44 através de uma *cloud tag*.

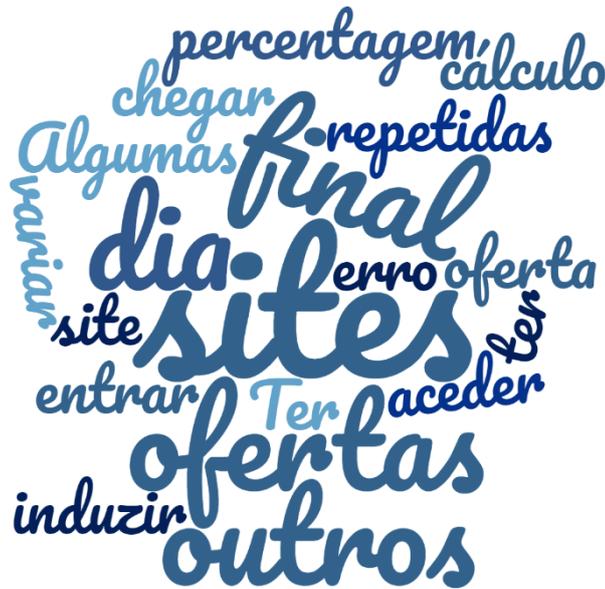


Figura 44. *Cloud Tag* resultante da questão “Quais considera serem os inconvenientes do uso da plataforma web recém-criada no auxílio de novas oportunidades de emprego?”. Fonte: www.wordclouds.com.

Na Figura 44 observa-se os inconvenientes do uso da plataforma web recém-criada no auxílio de novas oportunidades de emprego são: ter de entrar em outros *websites*, e ter algumas ofertas repetidas.

Conclusão do Capítulo

Este capítulo apresentou todos os testes realizados à plataforma de agregação e extração de dados recém-criada, observando-se que todos os testes propostos tiveram um resultado positivo, podendo afirmar-se que a plataforma está pronta para ser disponibilizada ao utilizador final.

Avaliação e Teste do Protótipo

6. Discussão do trabalho desenvolvido

Neste capítulo, será discutido o trabalho desenvolvido e analisado se os objetivos propostos foram atingidos, de forma a conseguir responder à questão de partida.

Solução Final

Como solução final desta dissertação, tem-se uma aplicação web de agregação de ofertas de emprego, que consegue com sucesso agregar ofertas de emprego publicadas em diferentes *websites* e, ao mesmo tempo, calcular a percentagem de ofertas de emprego que sejam relevantes ao utilizador da plataforma, segundo a descrição por ele definida, em relação ao número total de ofertas num determinado distrito. Desta forma, o utilizador consegue saber se as suas competências atuais ou futuras têm uma boa penetração no mercado de trabalho. Uma competência com uma boa aceitação por parte do mercado de trabalho é a que ocorre na descrição de um maior número de ofertas. Esta auscultação do mercado de trabalho ajuda o utilizador a situar as suas competências neste mercado. Isto é, de forma instantânea, consegue perceber a dinâmica oferta / procura existente.

Ao longo da construção da plataforma de agregação de ofertas de emprego, foi possível ir alcançando os vários objetivos teóricos, bem como práticos, desenhados na primeira fase desta dissertação. O atingir desses objetivos permitiu responder à questão de investigação.

Dos objetivos à questão de investigação

Os primeiros objetivos a serem alcançados foram os objetivos teóricos, através de uma análise detalhada da literatura atual sobre três grandes temas:

- Construção de uma oferta de emprego:

Uma oferta de emprego nasce com a necessidade de preencher uma necessidade de uma empresa, no que toca a um recurso humano. O departamento de recursos humanos é responsável pelo recrutamento, seleção e contratação desse novo recurso. É na fase de recrutamento que nasce a oferta de emprego e é publicada a oferta nos *websites* de anúncios de ofertas de emprego, entre outros meios.

- Desfasamento ensino-emprego:

O desfasamento entre ensino-emprego ocorre quando as competências do capital humano disponível não correspondem às necessidades das oportunidades existentes no mercado de trabalho. Este desfasamento nasce de uma economia fraca e estagnada, de um desequilíbrio entre a procura e oferta no mercado de trabalho, da qualidade do currículo e do tipo de cursos oferecidos

Discussão do trabalho desenvolvido

pelas universidades, numa desconexão entre as universidades e o mercado de trabalho, e o excesso de graduados universitários num determinado campo de estudo.

- Processo de extração de dados na web

O processo de extração de dados na web ocorre de várias formas, O consumo de APIs e/ou a técnica de *web scraping* são dois dos maiores processos de extração de informação. A técnica de *web scraping* é dirigida para websites que não disponibilizam APIs de consumo de dados ou quando disponibilizam essas APIs essas não possuem todos os dados desejados. A análise da árvore HTML do website alvo, onde está a informação a ser extraída, pode ser feita através de bibliotecas de programação que facilitam o trabalho do programador, no que toca à extração de dados não estruturados ou semi-estruturados.

Os segundos objetivos alcançados foram os objetivos práticos. Estes objetivos foram atingidos através da análise dos *websites* alvo de extração de dados, da elaboração dos requisitos da plataforma proposta nesta dissertação, da construção da mesma e, por fim, através da validação de testes unitários, de integração, funcionais e de aceitação.

Com os objetivos teóricos e práticos atingidos foi possível responder à questão de investigação “Até que ponto um agregador de ofertas contribui para a harmonização das competências académicas existentes e as exigidas pelo mercado de trabalho?”. A resposta é afirmativa. É possível calcular a percentagem de procura de uma competência por parte do mercado de trabalho, através de técnicas de *web scraping* que cruzem – extraíndo – as ofertas de emprego relevantes a um utilizador e as ofertas totais disponíveis nos *websites* de publicação de ofertas de emprego.

Conclusão do Capítulo

Foi aqui discutido o produto final desenvolvido nesta dissertação cujos objetivos, tanto teóricos como práticos, ficaram alcançados, o que permitiu responder à questão de partida.

7. Conclusões

O presente capítulo é o último desta dissertação e serve para apresentar as conclusões retiradas deste trabalho e apresentar ideias ou sugestões para trabalhos futuros.

Conclusões

Esta dissertação abordou os temas de recrutamento, através de ofertas de emprego web, defasamento ensino-emprego e extração de dados na web. Estes foram os temas teóricos que permitiram sustentar a criação da plataforma de agregação de ofertas de emprego e cálculo da percentagem de ofertas de emprego existentes para uma dada competência definida pelo utilizador.

Como referido no capítulo de discussão do trabalho desenvolvido, os objetivos propostos inicialmente foram atingidos, permitindo desta forma responder à questão de investigação proposta.

É perceptível ao longo deste documento, e principalmente no capítulo que aborda a criação do algoritmo de extração de dados, a importância do conhecimento dos conceitos teóricos para poder realizar-se uma extração precisa dos dados alvo e, desta forma, construir uma plataforma de agregação de ofertas de emprego.

Ao longo de todo o desenvolvimento da plataforma tecnológica, surgiram dificuldades e limitações relativamente à escolha da biblioteca de programação mais indicada para analisar os documentos HTML dos *websites* alvo e posterior extração. A elevada curva de aprendizagem, a falta de suporte ou manutenção, o facto de algumas bibliotecas estarem desenhadas para funcionar apenas num tipo de *web browser* e o seu código reescrever outras bibliotecas limitaram o universo de escolha da biblioteca a ser utilizada nesta dissertação. Contudo, a biblioteca HTML Agility Pack eleita nesta dissertação cumpre bem a função a que se destina. No entanto, poderá ser feita uma análise futura relativamente a outras opções mais eficazes no que toca à velocidade de análise dos documentos HTML dos *websites* alvo.

Contribuição

Como principal contribuição desta dissertação, pode destacar-se que o fácil acesso à percentagem de ofertas de emprego para uma dada competência, que a plataforma de agregação de ofertas de emprego criada permite, capacita o utilizador com dados sobre a procura-oferta do mercado de trabalho de forma instantânea. Desta forma, o utilizador pode usar estes dados a seu favor e investir em áreas em que o mercado de trabalho esteja com uma maior aceitação de uma determinada competência. Esta ação quando ocorre permite contribuir para uma harmonização da oferta-procura no mercado de trabalho.

Todo o estudo e análise feitos ao longo desta dissertação serviram, não só para obter um maior nível de conhecimento, no que diz respeito ao processo de recrutamento, criação de uma oferta de

Conclusões

emprego, conceito de desfasamento ensino-emprego, *web scraping*, como também para adquirir mais e melhores competências técnicas, no respeito ao uso de bibliotecas de programação para realizar a extração de dados em *websites*.

Trabalho Futuro

Poder-se-á num futuro próximo explorar algumas áreas que acrescentem valor à plataforma de agregação de ofertas de emprego recém-criada, nomeadamente no que respeita a:

- Opção de escolha da fonte de dados:

Esta opção está relacionada com a possibilidade de escolher qual o *website* a ser alvo de extração. De uma lista disponível ao utilizador, dar a possibilidade de marcar qual ou quais os *websites* onde o utilizador deseja realizar as suas pesquisas. Com esta funcionalidade, dá-se mais controlo ao utilizador para escolher a fonte dos seus dados, criando independência e imparcialidade no ato da pesquisa.

- Registrar as competências mais valorizadas pelo mercado de trabalho

Com esta nova opção, pretende-se ir registando, persistindo na base de dados, qual a competência pesquisada pelos utilizadores que tenha tido um maior número de ofertas de emprego. Desta forma, é possível realizar um quadro de tendências das competências que devolvem um maior número de ofertas de emprego. O utilizador, com esta nova funcionalidade, fica com uma visão mais ampla das competências mais valorizadas pelo mercado de trabalho e não somente restrito às suas próprias pesquisas.

- Cruzar o número de ofertas de emprego com o número de recém-graduados:

Esta última sugestão pretende cruzar as ofertas de emprego disponíveis para uma determinada competência e o número de recém-graduados universitários desse ano ou ano anterior, relacionados com a área científica dessa competência. Desta forma, seria possível verificar se o mercado estaria de facto com excesso ou carência de recém-graduados universitários numa área / competência específica.

Referências Bibliográficas

- Allen, J., & Van der Velden, R. (2001). Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search. *Oxford economic papers*, 53(3), 434-452.
- Andrejevic, M. (2013). *Infoglut: How too much information is changing the way we think and know*. Routledge.
- Andrews, J., & Higson, H. (2008). Graduate employability, 'soft skills' versus 'hard' business knowledge: A European study. *Higher education in Europe*, 33(4), 411-422.
- Angelucci, C., & Cagé, J. (2016). Newspapers in times of low advertising revenues.
- Armstrong, M. (2006). *A handbook of human resource management practice*. Kogan Page Publishers.
- Assembly, U. G. (1948). Universal declaration of human rights. *UN General Assembly*.
- Assembly, U. G. (2015). Transforming our world: the 2030 Agenda for Sustainable Development. *New York: United Nations*, (1).
- Balcerzak, A. P., & Pietrzak, M. B. (2016). Quality Of Human Capital In The European Union In The Years 2004-2013. Application Of Structural Equation Modeling.
- Barone, C., & Ortiz, L. (2011). Overeducation among European University Graduates: a comparative analysis of its incidence and the importance of higher education differentiation. *Higher Education*, 61(3), 325-337.
- Bassey, G. E., & Atan, J. A. (2012). Labour market distortions and university graduate unemployment in Nigeria: Issues and remedies. *Current research journal of economic theory*, 4(3), 67-76.
- Baum, M., & Kabst, R. (2014). The effectiveness of recruitment advertisements and recruitment websites: Indirect and interactive effects on applicant attraction. *Human Resource Management*, 53 (3), 353-378.
- Beck, U. (2002). *Individualization: Institutionalized individualism and its social and political consequences* (Vol. 13). Sage.
- Bei-jie, D. A. I., Shuang, Y. U., & Bei-hong, J. I. N. (2007). XML editor based on efficient DOM parser [J]. *Computer Engineering and Design*, 22, 001.
- Beimborn, D. (2014). Centre of Human Resources Information Systems. Retrieved June, 29, 2014.
- Betti, G., D'Agostino, A., & Neri, L. (2007). *Educational Mismatch of Graduates: a Multidimensional and Fuzzy Approach*. Università di Siena, Dipartimento di metodi quantitativi.
- Bernard, H. R., Wutich, A., & Ryan, G. W. (2016). *Analyzing qualitative data: Systematic approaches*. SAGE publications.
- Bidwell, M. (2011). Paying more to get less: The effects of external hiring versus internal mobility. *Administrative Science Quarterly*, 56(3), 369-407.
- Bhuvan, K. C. (2018). Curriculum Development: Mismatch Between Supply and Demand. In *Pharmacy Education in the Twenty First Century and Beyond* (pp. 285-296).
- Boxall, P., & Purcell, J. (2011). *Strategy and human resource management*. Palgrave Macmillan.
- Branine, M., & Avramenko, A. (2015). A comparative analysis of graduate Employment prospects in european labour markets: A study of graduate recruitment in four countries. *Higher Education Quarterly*, 69(4), 342-365.
- Bratton, J., & Gold, J. (2017). *Human resource management: theory and practice*. Palgrave.

- Callanan, G. A. (2003). What price career success?. *Career Development International*, 8(3), 126-133.
- Castells, M. (2011). *The rise of the network society* (Vol. 12). John Wiley & Sons.
- Chaulagain, R. S., Pandey, S., Basnet, S. R., & Shakya, S. (2017, November). Cloud Based Web Scraping for Big Data Applications. In *Smart Cloud (SmartCloud), 2017 IEEE International Conference on* (pp. 138-143). IEEE.
- Chavan, H. A., Bhojane, H., Naik, R., & Priyadarshi, O. (2018). Commodity Search Using Web Browser Extension. *AADYA-Journal of Management and Technology (JMT)*, 8(1), 42-45.
- Chevalier, A. (2003). Measuring over-education. *Economica*, 70(279), 509-531.
- CHRIS. (2012). Recruiting Trends: Ein Rückblick auf 10 Jahre. Frankfurt am Main: Centre of Human Resources Information Systems.
- Coles, M. G., & Mortensen, D. T. (2016). Equilibrium labor turnover, firm growth, and unemployment. *Econometrica*, 84(1), 347-363.
- Congregado, E., Iglesias, J., Millán, J. M., & Román, C. (2016). Incidence, effects, dynamics and routes out of overqualification in Europe: A comprehensive analysis distinguishing by employment status. *Applied Economics*, 48(5), 411-445.
- DeCenzo, D. A., Robbins, S. P., & Verhulst, S. L. (2013). *Human Resource Management* (11 Ed.). Hoboken: Wiley.
- Domadenik, P., Farcnik, D., & Pastore, F. (2013). Horizontal mismatch in the labour market of graduates: The role of signalling.
- Duggan, M., & Smith, A. (2014). Social Media Use Over Time. Retrieved January 15, 2018, from PewInternet: <http://www.pewinternet.org/data-trend/social-media/social-media-use-all-users>.
- Ebizmba.com. (2018). Top 15 Most Popular Search Engines | May 2018. [Online] Available at: <http://www.ebizmba.com/articles/search-engines> [Accessed 20 Feb. 2018]
- Eichhorst, W., Hinte, H., & Rinne, U. (2013). *Youth unemployment in Europe: what to do about it?* (No. 65). IZA policy paper.
- El-Khawas, E. (2001). Today's universities: responsive, resilient, or rigid?. *Higher Education Policy*, 14(3), 241-248.
- Farooq, S. (2011). Mismatch between education and occupation: A case study of Pakistani graduates. *The Pakistan Development Review*, 531-552.
- Ferrara, E., De Meo, P., Fiumara, G., & Baumgartner, R. (2014). Web data extraction, applications and techniques: A survey. *Knowledge-based systems*, 70, 301-323.
- Gadre, D. V., & Gupta, S. (2018). Application Programming Interface (API). In *Getting Started with Tiva ARM Cortex M4 Microcontrollers* (pp. 95-100). Springer, New Delhi.
- Graham, J. and Graham, S. (2013). Literature Review Exploring Job Mismatch and Income, and Labour Market Outcomes for People with Disabilities. *Human Resources and Skills Development Canada (HRSDC)*.
- Groot, W., & Van Den Brink, H. M. (2000). Overeducation in the labor market: a meta-analysis. *Economics of education review*, 19(2), 149-158.
- Gupta, S., Kaiser, G., Neistadt, D., & Grimm, P. (2003, May). DOM-based content extraction of HTML documents. In *Proceedings of the 12th international conference on World Wide Web* (pp. 207-214). ACM.
- Handlogten, C. (2009). *Implementation of e-recruitment Enablers and success indicators from the*. (Master's thesis, University of Twente).

- Hennemann, S., & Liefner, I. (2010). Employability of German geography graduates: The mismatch between knowledge acquired and competences required. *Journal of Geography in Higher Education*, 34(2), 215-230.
- Holm, A. B. (2012). E-recruitment: towards an ubiquitous recruitment process and candidate relationship management. *German Journal of Human Resource Management*, 26(3), 241-259.
- Humburg, M., De Grip, A., & Van der Velden, R. (2017). Which skills protect graduates against a slack labour market?. *International Labour Review*, 156(1), 25-43.
- International Labour Office. (2017). *Global Employment Trends For Youth 2017: paths to a better working future*. International Labour Office.
- Jobvite. (2018). The 2013 Social Recruiting Survey Results Are Here! - Jobvite. [online] Available at: <https://www.jobvite.com/jobvite-news-and-reports/2013jobvitesocialrecruitingsurvey/> [Accessed 26 Feb. 2018].
- Kalleberg, A. L. (2018). Changing contexts of careers: Trends in labor market structures and some implications for labor force outcomes. In *Generating Social Stratification: Toward a New Research Agenda* (pp. 343-358). Taylor and Francis.
- Kim, H. K., Ahn, S. C., & Kim, J. (2012). *Vertical and horizontal education-job mismatches in the Korean youth labor market: a quantile regression approach* (No. 1201).
- Langville, A. N., Meyer, C. D., & Fernández, P. (2008). Google's pagerank and beyond: The science of search engine rankings. *The Mathematical Intelligencer*, 30(1), 68-69.
- Lee, I. (2005). The evolution of e-recruiting: A content analysis of Fortune 100 career web sites. *Journal of Electronic Commerce in Organizations (JECO)*, 3(3), 57-68.
- Liu, K., Salvanes, K. G., & Sørensen, E. Ø. (2016). Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review*, 84, 3-17.
- Mahto, D. K., & Singh, L. (2016, March). A dive into Web Scraper world. In *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on* (pp. 689-693). IEEE.
- Marzo-Navarro, M., Pedraja-Iglesias, M., & Rivera-Torres, P. (2009). Curricular profile of university graduates versus business demands: Is there a fit or mismatch in Spain?. *Education+ Training*, 51(1), 56-69.
- Mbah, M. F. (2014). The Dilemma of Graduate Unemployment within a Context of Poverty, Scarcity and Fragile Economy: Are there Lessons for the University?. *International Journal of Economics and Finance*, 6(12), 27.
- McGuinness, S. (2003). University quality and labour market outcomes. *Applied Economics*, 35(18), 1943-1955.
- McGuinness, S., & Bennett, J. (2007). Overeducation in the graduate labour market: A quantile regression approach. *Economics of Education Review*, 26(5), 521-531.
- McGuinness, S., & Sloane, P. J. (2011). Labour market mismatch among UK graduates: An analysis using REFLEX data. *Economics of Education Review*, 30(1), 130-145.
- McGuinness, S., Pouliakas, K., & Redmond, P. (2017). How Useful is the Concept of Skills Mismatch?.
- Meschenmoser, P., Meuschke, N., Hotz, M., & Gipp, B. (2016). Scraping Scientific Web Repositories: Challenges and Solutions for Automated Content Extraction. *D-Lib Magazine*, 22(9/10).
- Milana, M. (2009). The post-Lisbon discourse on skill mismatches and competence upgrading. *H. Holmarsdottir, Nordic Voices: Teaching and Researching Comparative and International Education in the Nordic Countries*, 9-24.

- Mitchell, R. (2018). *Web Scraping with Python: Collecting More Data from the Modern Web*. " O'Reilly Media, Inc."
- Montt, G. (2015), The causes and consequences of field-of-study mismatch: An analysis using PIAAC, OECD Social, *Employment and Migration Working Papers*, No. 167, OECD Publishing, Paris.
- Nunez, I., & Livanos, I. (2010). Higher education and unemployment in Europe: an analysis of the academic subject and national effects. *Higher Education*, 59(4), 475-487.
- Munzert, S., Rubba, C., Meißner, P., & Nyhuis, D. (2014). *Automated data collection with R: A practical guide to web scraping and text mining*. John Wiley & Sons.
- Naong, M. N. (2011). Promotion of Entrepreneurship Education-A Remedy to Graduates and Youth Unemployment—A Theoretical Perspective. *Journal of Social Sciences*, 28(3), 181-189.
- Neneh, B. N. (2014). An assessment of entrepreneurial intention among university students in Cameroon. *Mediterranean Journal of Social Sciences*, 5(20), 542.
- Netmarketshare.com. (2018). Search engine market share. [Online] Available at: <https://www.netmarketshare.com/search-engine-market-share.aspx?options=%7B%22filter%22%3A%7B%22%24and%22%3A%5B%7B%22deviceType%22%3A%7B%22%24in%22%3A%5B%22Desktop%2Flaptop%22%5D%7D%7D%5D%7D%2C%22dateLabel%22%3A%22Trend%22%2C%22attributes%22%3A%22share%22%2C%22group%22%3A%22searchEngine%22%2C%22sort%22%3A%7B%22share%22%3A-1%7D%2C%22id%22%3A%22searchEnginesDesktop%22%2C%22dateInterval%22%3A%22Monthly%22%2C%22dateStart%22%3A%222017-06%22%2C%22dateEnd%22%3A%222018-05%22%2C%22segments%22%3A%22-1000%22%7D> [Accessed 19 Feb. 2018].
- Nordin, M., Persson, I., & Rooth, D. O. (2010). Education–occupation mismatch: Is there an income penalty?. *Economics of education review*, 29(6), 1047-1059.
- OECD (2018), Unemployment rate (indicator). doi: 10.1787/997c8750-en (Accessed on 21 June 2018).
- Paadi, K. (2014). Perceptions on employability skills necessary to enhance human resource management graduates prospects of securing a relevant place in the labour market. *European Scientific Journal, ESJ*, 10(10).
- Plomp, J., Tims, M., Akkermans, J., Khapova, S. N., Jansen, P. G., & Bakker, A. B. (2016). Career competencies and job crafting: How proactive employees influence their well-being. *Career Development International*, 21(6), 587-602.
- Polidoro, F., Giannini, R., Conte, R. L., Mosca, S., & Rossetti, F. (2015). Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP compilation. *Statistical Journal of the IAOS*, 31(2), 165-176.
- Porter, M. E., & Millar, V. E. (1985). How information gives you competitive advantage.
- Price, A. (2007). *Human resource management in a business context*. Cengage Learning EMEA.
- Reinsel, D., Gantz, J., & Rydning, J. (2017). Data age 2025: The evolution of data to life-critical. *Don't Focus on Big Data*.
- Robst, J. (2007). Education and job match: The relatedness of college major and work. *Economics of Education Review*, 26(4), 397-407.
- Salinas-Jiménez, M. D. M., Artés, J., & Salinas-Jiménez, J. (2016). Educational mismatch and job aspirations: A subjective wellbeing analysis using quantile regression. *International Journal of Manpower*, 37(1), 115-134.
- Salmen, S. (2012). Einleitung-Die Suche von Top-Mitarbeitern im War for Talent. *B. Rath, & S. Salmen, Recruiting im Social Web*, 30-40.

- Santos, M. B. (2016). Beyond skill mismatch. Why there are so many unfilled vacancies and simultaneously high unemployment rates?.
- Scheller, T., & Kühn, E. (2015). Automated measurement of API usability: The API concepts framework. *Information and Software Technology*, 61, 145-162.
- Sills, M. (2014). E-recruitment: A comparison with traditional recruitment and the influences of social media: A qualitative and quantitative review.
- Silverman, D. (Ed.). (2016). *Qualitative research*. Sage.
- Sloane, P., Mavromaras, K., O'Leary, N., McGuinness, S., & O'Connell, P. J. (2010). *The skill matching challenge: Analysing skill mismatch and policy implications*. Publications office of the European Union.
- Taris, T. W., & Schaufeli, W. (2015). Individual well-being and performance at work: A conceptual and theoretical overview.
- United Nations Education, Scientific, and Cultural Organization. (2012). *Youth and skills: putting education to work*. New York: Unesco.
- Van Smoorenburg, M. S. M., & Van der Velden, R. K. (2000). The training of school-leavers: complementarity or substitution?. *Economics of education review*, 19(2), 207-217.
- Vanden Broucke, S., & Baesens, B. (2018). *Practical web scraping for data science: best practices and examples with Python*. Apress.
- Verhaest, D., Sellami, S., & Van der Velden, R. (2017). Differences in horizontal and vertical mismatches across countries and fields of study. *International Labour Review*, 156(1), 1-23.
- Watson, M. (2009). Using Web Scraping to Create Semantic Relations. *Scripting Intelligence: Web 3.0 Information Gathering and Processing*, 205-228.
- Wolbers, M. H. (2003). Job mismatches and their labour-market effects among school-leavers in Europe. *European Sociological Review*, 19(3), 249-266.
- World Economic Forum. (2016). *The future of jobs: Employment, skills and workforce strategy for the fourth industrial revolution*. World Economic Forum, Geneva, Switzerland.
- Zils, E. (2012). Social media Recruiting Studie 2012: Deutschland. *Strasbourg: SocialMedia-Recruiting.com*.
- Zimmermann, K. F. (2009). Labor mobility and the integration of European labor markets. DIW Berlin Discussion Paper No. 862.

Anexos

Anexo A – Árvore HTML dos websites “indeed.pt” e “emprego.sapo.pt”

Árvore HTML "indeed.pt"

```
▼ <tr>
  <td>&nbsp;</td>
  ▶ <td id="ctl100_ContentPlaceholderMain_RepJobs_ctl100_tdPartnerHighlight" class="col0">...</td>
  ▼ <td class="col1">
    <span>30-07-2018</span>
  </td>
  ▶ <td class="col2">...</td>
  ▼ <th class="col2" scope="row">
    <a href="/emprego/anuncio/5700916/operador-contact-center--planos-nutricao---lisboa.htm" title="Ver detalhes do anúncio Operador Contact Center - Planos Nutricao - Lisboa" onclick="ChangeSearchIndex(0);">Operador Contact Center - Planos Nutricao - Lisboa</a>
  </th>
  ▶ <th class="col2" scope="row">...</th>
</tr>
▶ <tr>...</tr>
```

Árvore HTML " emprego.sapo.pt"

```
▼ <div class="row result clickcard" id="p_8e4d3ccd98a2cbdd" data-jk="8e4d3ccd98a2cbdd" data-tn-component="organicJob" data-tu>
  ▼ <h2 id="j1_8e4d3ccd98a2cbdd" class="jobtitle">
    ▼ <a href="/company/Triangulo/jobs/Nutricionistas-Canal-Farm%C3%A1cia-8e4d3ccd98a2cbdd?fccid=a3c35b192085c42e&vjs=3" target="_blank" rel="noopener nofollow" onmousedown="return rclk(this,jobmap[2],0);" onclick="setRefineByCookie([]); return rclk(this,jobmap[2],true,0);" title="NUTRICIONISTAS (m/f) Canal Farmácia - LISBOA" class="turnstileLink" data-tn-element="jobTitle">
      <b>NUTRICIONISTAS</b>
      " (m/f) Canal Farmácia - LISBOA"
    </a>
    "
    - "
    <span class="new">nova</span>
  </h2>
  <div style="display: inline">...</div>
  " - "
  <span class="location">Lisboa</span>
  ▼ <table width="100%" cellpadding="0" cellspacing="0" border="0">
    ▼ <tbody>
      ▶ <tr>...</tr>
    </tbody>
  </table>
</div>
```

Anexo B – Algoritmo de extração da plataforma

Repositório GitHub:

<https://github.com/Mr-Afonso/MasterThesis>

Anexo C – Dados focus group

Focus Group – Levantamento dos requisitos

Questão 1: Quais os aspetos mais importantes que uma plataforma web de agregação de ofertas de emprego deve ter?	
Nota 1	Um <i>website</i> de ofertas de emprego tem de ter filtros, tais como: área, descrição da oferta, data.
Nota 2	Fácil de usar. Não ter de fazer login, por exemplo.
Nota 3	Poder visualizar um grande número de ofertas em uma única página.
Nota 4	As ofertas de emprego estarem adequadas à pesquisa feita pelo utilizador, é um aspeto importante.
Nota 5	Mostrar ofertas atualizadas, e ainda válidas para submeter a candidatura.

Questão 2: Quais os aspetos mais fracos que considera existirem nas plataformas web de agregação de ofertas de emprego?	
Nota 1	Um dos aspetos mais fracos é o facto de não existir uma análise de tendências de ofertas de emprego.
Nota 2	Descrição das ofertas reduzida.
Nota 3	O facto de não mostrar se uma vaga já foi preenchida ou não, e o anúncio ainda continuar ativo é um ponto fraco.
Nota 4	Não cruzar o número de empregos com o número de licenciados em cada área. Por exemplo, existem 1000 vagas para enfermeiros, mas das faculdades saem todos os anos 3000 licenciados.
Nota 5	Ofertas de emprego repetidas.

Questão 3: Quais considera serem os benefícios do uso das plataformas web de agregação de ofertas de emprego no auxílio de novas oportunidades de emprego?	
Nota 1	Uma das maiores vantagens é a velocidade de pesquisa, e a facilidade de análise de cada uma das ofertas de emprego
Nota 2	Poder num curto período de tempo submeter várias candidaturas.
Nota 3	Cobrem um maior número de ofertas de emprego.
Nota 4	São convenientes, pois deixa de ser preciso pesquisar a mesma informação em diferentes <i>websites</i> .
Nota 5	Acesso a mais ofertas de emprego em um menor período de tempo.

Questão 4: Quais considera serem os inconvenientes do uso das plataformas web de agregação de ofertas de emprego no auxílio de novas oportunidades de emprego?	
Nota 1	Um dos inconvenientes é o fato de não sabermos se todas as ofertas de emprego apresentadas são mesmo todas as ofertas de emprego disponíveis.
Nota 2	Demasiada informação em uma única pesquisa.
Nota 3	Não vejo inconveniente.
Nota 4	Serem só um meio, isto é, tem de se aceder a outro site onde está a oferta de emprego propriamente dita.
Nota 5	Estes <i>websites</i> não são especializados numa área.

Focus Group – Avaliação da plataforma

Questão 1: Quais as mais-valias que considera existirem na plataforma web recém-criada em relação às suas congéneres?	
--	--

Nota 1	É interessante poder ver a percentagem de ofertas de emprego de uma área em relação ao número total de ofertas.
Nota 2	Não tem publicidade, e é fácil de usar.
Nota 3	A percentagem de ofertas de emprego relevantes ao utilizador é algo que nunca vi.
Nota 4	Poder ver qual a área que tem uma percentagem mais elevada de emprego.
Nota 5	Prático, e simples.

Questão 2: Quais os aspetos mais fracos que considera existirem na plataforma web recém-criada em relação às suas congéneres?

Nota 1	A velocidade de pesquisa é um dos aspetos a melhorar.
Nota 2	Não poder filtrar as ofertas por data.
Nota 3	Um pouco lento em algumas pesquisas.
Nota 4	Podia ter mais opções de pesquisa.
Nota 5	Velocidade ao realizar uma consulta.

Questão 3: Quais considera serem os benefícios do uso da plataforma web recém-criada no auxílio de novas oportunidades de emprego?

Nota 1	O fato de mostrar o peso atual de uma determinada área pode ser um benefício para quem está indeciso que área profissional seguir.
Nota 2	Centralizar a pesquisa de ofertas em um único <i>website</i> .
Nota 3	Mostra se uma área está com saída ou não.
Nota 4	Pode ajudar jovens a definir qual a melhor área a seguir segundo a percentagem de empregos em determinada área.
Nota 5	Poder consultar muitos <i>websites</i> em um único site.

Questão 4: Quais considera serem os inconvenientes do uso da plataforma web recém-criada no auxílio de novas oportunidades de emprego?

Nota 1	Um dos inconvenientes é ter de estar dependente de outros <i>websites</i> .
Nota 2	As ofertas de emprego podem variar de dia para dia, e induzir o utilizador em erro no cálculo da percentagem.
Nota 3	Ter de aceder a outros <i>websites</i> .
Nota 4	Algumas das ofertas são repetidas.
Nota 5	Não é o <i>website</i> final. Tem de se entrar em mais <i>websites</i> para se chegar à oferta final.