



Uso de text mining na determinação da relação entre instituições de investigação e empresas

SÉRGIO RICARDO MAGALHÃES MARQUES DA SILVA

Outubro de 2022

POLITÉCNICO DO PORTO
INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

Uso de *text mining* na determinação
da relação entre instituições de
investigação e empresas

Sérgio Ricardo Magalhães Marques da Silva

Mestrado em Engenharia Electrotécnica e de Computadores
Área de Especialização em Sistemas e Planeamento Industrial



DEPARTAMENTO DE ENGENHARIA ELETROTÉCNICA
Instituto Superior de Engenharia do Porto

Outubro, 2022

Esta dissertação satisfaz, parcialmente, os requisitos que constam da Ficha de Unidade Curricular de Tese/Dissertação, do 2º ano, do Mestrado em Engenharia Electrotécnica e de Computadores, Área de Especialização em Sistemas e Planeamento Industrial.

Candidato: Sérgio Ricardo Magalhães Marques da Silva, N.º 1060423,
1060423@isep.ipp.pt

Orientação Científica: Professor Doutor Filipe Miguel Tavares de Azevedo,
fta@isep.ipp.pt



DEPARTAMENTO DE ENGENHARIA ELETROTÉCNICA
Instituto Superior de Engenharia do Porto
Rua Dr. António Bernardino de Almeida, 431, 4200-072 Porto

Outubro, 2022

Agradecimentos

Gostaria de agradecer a um conjunto de pessoas que de diferentes formas me ajudaram:

Ao Professor Doutor Filipe Azevedo pela orientação, disponibilidade e partilha dos seus conhecimentos;

A todos os professores e colegas que me acompanharam durante o meu percurso académico;

Aos meus amigos e colegas de trabalho;

À Lara pelo incentivo e pelo apoio que sempre me deu.

Resumo

A excelência da investigação científica produzida é indissociável da inovação e criação de valor económico e social. Assim, a aproximação entre diferentes áreas científicas e de negócio, e a consolidação de formas de colaboração entre as instituições de investigação e o tecido económico e social, potencia a criação, transferência e valorização do conhecimento. Neste processo têm particular importância as unidades de investigação e as empresas, com papéis diferenciados, mas essenciais para completar o ciclo de inovação.

No mundo competitivo de hoje, a capacidade de extrair conhecimento útil de dados e tomar decisões de acordo com esse conhecimento é cada vez mais importante e essencial. O processo de aplicação de metodologias para extração de conhecimento a partir de dados textuais é denominado de *text mining*. A sua utilização tem como benefício a grande quantidade de informação importante latente neste formato e que não está disponível nos formatos clássicos de dados estruturados.

Neste trabalho foi desenvolvida uma ferramenta de apoio estratégico, que tem como objetivo potenciar a colaboração entre unidades de investigação e o tecido empresarial. Através de metodologias de *text mining*, nomeadamente sumário da informação e *topic modeling* com utilização do modelo *Latent Dirichlet Allocation*, é analisada a informação disponível nos *websites* das unidades de investigação e das empresas e são identificadas possíveis relações. É a primeira vez que é analisado este tipo de informação com recurso a metodologias de *text mining* e com o propósito de potenciar a relação entre estas estruturas.

A análise de resultados obtidos permitiu concluir que os mesmos estão dependentes da qualidade da informação disponível nos *websites* e da representatividade de todas as áreas. Caso estas condições sejam garantidas, esperam-se bons resultados relativos a possíveis relações, tendo sempre em consideração que estes resultados poderão não ser os mais óbvios tendo por base o conhecimento prévio das entidades em análise. Assim, a ferramenta deverá ser utilizada como apoio na tomada de decisão, devendo os resultados obtidos ser analisados de forma crítica e em complemento à experiência de especialistas.

Palavras-Chave: transferência e valorização do conhecimento, unidades de investigação, empresas, *text mining*, *topic modeling*, ferramenta *web*.

Abstract

The excellence of the scientific research produced is inseparable from innovation and the creation of economic and social value. Thus, the approximation between different scientific and business areas and the consolidation of forms of collaboration between research institutions and the economic and social fabric, boosts the creation, transfer and valorization of knowledge. In this process, research units and companies are particularly important, with different but essential roles to complete the innovation cycle.

In today's competitive world, the ability to extract useful knowledge from data and make decisions according to that knowledge is increasingly important and essential. The process of applying methodologies to extract knowledge from textual data is called text mining. Its use has the benefit of accessing to a large amount of important information latent in this format and which is not available in the classic structured data formats.

In this work a strategic support tool was developed, aiming to enhance collaboration between research units and the business fabric. Through the use of text mining methodologies, namely information summary and topic modeling using the Latent Dirichlet Allocation model, the information available on the websites of research units and companies is analyzed and possible connections are identified. It is the first time that this type of information is analyzed using text mining methodologies and with the purpose of enhancing the relationship between these structures.

The analysis of results led to the conclusion that these results depend on the quality of the information available on the websites and the representativeness of all areas. If these conditions are guaranteed, good results are expected regarding possible connections, always bearing in mind that these results may not be the most obvious, based on the prior knowledge of the entities under analysis. Therefore, the tool should be used as a support in decision making and the results obtained should be critically analyzed in addition to the experience of experts.

Keywords: knowledge transfer and valorization, research units, companies, text mining, topic modeling, web tool.

Índice

Lista de Figuras	vii
Lista de Tabelas	ix
Lista de Acrónimos	xi
1 Introdução	1
1.1 Contextualização e motivação	1
1.2 Objetivo e plano de trabalho	5
1.3 Organização do documento	6
2 Transferência e valorização do conhecimento	7
2.1 Introdução	8
2.2 Processo	9
2.3 Em Portugal	9
2.3.1 Unidades de Investigação	11
2.3.2 Empresas	13
3 <i>Text mining</i>	21
3.1 Introdução	22
3.2 Análise cientométrica com utilização de <i>text mining</i>	22
3.3 Bases metodológicas	27
3.4 <i>Topic modeling</i> com utilização do modelo <i>Latent Dirichlet Allocation</i>	33
4 Ferramenta desenvolvida	37
4.1 Objetivo	37
4.2 Desenvolvimento	38
4.2.1 Tecnologias	38
4.2.2 Aplicação de <i>text mining</i>	39
4.2.3 Ambiente gráfico	46
4.3 Resultados e discussão	47
4.3.1 Ambiente gráfico e funcionalidades	50
4.3.2 Nuvem de palavras	53
4.3.3 <i>Topic modeling</i>	61

5 Conclusões	79
5.1 Conclusões sobre o trabalho realizado	79
5.2 Sugestões de trabalho futuro	81
Referências	83
Anexo A Unidades de Investigação incluídas na ferramenta desenvolvida	93
Anexo B Empresas incluídas na ferramenta desenvolvida	105

Lista de Figuras

1.1	Objetivos para o desenvolvimento sustentável [3].	2
2.1	Transferência de conhecimento: da investigação ao impacto [31]. . .	9
2.2	O Sistema Nacional de Inovação (SNI) [5].	10
3.1	Evolução do número de documentos.	23
3.2	Documentos por país.	25
3.3	Mapeamento de tópicos e <i>clusters</i> com base em afinidades.	26
3.4	Mapeamento de tópicos e visualização sobreposta do ano médio de publicação.	27
3.5	Visão geral dos passos que compõem o processo de <i>text mining</i> . . .	28
3.6	Representação gráfica do modelo <i>Latent Dirichlet Allocation</i> (LDA) [72].	34
3.7	Aplicação da metodologia de <i>topic modeling</i> com utilização do algoritmo LDA.	35
4.1	Aplicação da metodologia de <i>topic modeling</i> com utilização do algoritmo LDA aplicado à ferramenta.	45
4.2	Estrutura da base de dados da ferramenta desenvolvida.	48
4.3	Estrutura de diretórios e ficheiros da ferramenta desenvolvida.	49
4.4	Ferramenta: <i>Home Page</i>	50
4.5	Ferramenta: Unidades de Investigação.	51
4.6	Ferramenta: Unidade de Investigação (descrição e nuvem de palavras). . .	51
4.7	Ferramenta: Unidade de Investigação (resultados de <i>topic modeling</i>). . .	52
4.8	Ferramenta: Unidade de Investigação (informação da empresa).	52
4.9	Ferramenta: Unidade de Investigação (informação da empresa - continuação).	53
4.10	Ferramenta: Empresas.	54
4.11	Ferramenta: empresa (descrição e nuvem de palavras).	54
4.12	Ferramenta: empresa (resultados de <i>topic modeling</i>).	55
4.13	Ferramenta: empresa (informação da Unidade de Investigação).	55
4.14	Ferramenta: empresa (informação da Unidade de Investigação - continuação).	56
4.15	Nuvem de palavras da informação de todas as Unidades de Investigação. . .	57

4.16	Nuvem de palavras da informação de todas as empresas.	57
4.17	Nuvem de palavras da Unidade de Investigação GECAD.	58
4.18	Nuvem de palavras da Unidade de Investigação INESC TEC.	59
4.19	Nuvem de palavras da Unidade de Investigação INET-md.	59
4.20	Nuvem de palavras da Unidade de Investigação ALGORITMI.	60
4.21	Nuvem de palavras da empresa EFACEC.	61
4.22	Nuvem de palavras da empresa FARFETCH.	62
4.23	Nuvem de palavras da empresa PORTUGALISOL.	62
4.24	Resultado da aplicação de <i>topic modeling</i> da Unidade de Investigação GECAD.	64
4.25	Relação com as empresas da Unidade de Investigação GECAD.	64
4.26	Resultado da aplicação de <i>topic modeling</i> da Unidade de Investigação INESC TEC.	65
4.27	Relação com as empresas da Unidade de Investigação INESC TEC.	66
4.28	Resultado da aplicação de <i>topic modeling</i> da Unidade de Investigação INET-md.	67
4.29	Relação com as empresas da Unidade de Investigação INET-md.	67
4.30	Resultado da aplicação de <i>topic modeling</i> da Unidade de Investigação ALGORITMI - tópico 2.	69
4.31	Resultado da aplicação de <i>topic modeling</i> da Unidade de Investigação ALGORITMI - tópico 3.	69
4.32	Relação com as empresas da Unidade de Investigação ALGORITMI - tópico 2.	70
4.33	Relação com as empresas da Unidade de Investigação ALGORITMI - tópico 3.	70
4.34	Resultado da aplicação de <i>topic modeling</i> da empresa EFACEC - tópico 2.	72
4.35	Relação com as Unidades de Investigação da empresa EFACEC - tópico 2.	72
4.36	Resultado da aplicação de <i>topic modeling</i> da empresa EFACEC - tópico 0.	73
4.37	Relação com as Unidades de Investigação da empresa EFACEC - tópico 0.	73
4.38	Resultado da aplicação de <i>topic modeling</i> da empresa FARFETCH.	74
4.39	Relação com as Unidades de Investigação da empresa FARFETCH.	75
4.40	Resultado da aplicação de <i>topic modeling</i> da empresa PORTUGALISOL.	76
4.41	Relação com as Unidades de Investigação da empresa PORTUGALISOL.	77

Lista de Tabelas

2.1	Classificação das Unidades de Investigação segundo a avaliação de 2017/2018 da Fundação para a Ciência e a Tecnologia (FCT).	12
2.2	Número e percentagem das Unidades de Investigação por domínio científico.	12
2.3	Número e percentagem dos investigadores por domínio científico. . .	13
2.4	Número e percentagem do financiamento por domínio científico. . . .	13
2.5	Número e percentagem das empresas com pessoas afetas à investigação e desenvolvimento por atividade económica.	14
2.6	Número, percentagem e média por empresa das “Pessoas ao serviço da empresa, remuneradas e não remuneradas: número médio de pessoas” das empresas com pessoas afetas à investigação e desenvolvimento, por atividade económica.	17
2.7	Número, percentagem e média por empresa das “Pessoas afetas à Investigação e Desenvolvimento” das empresas com pessoas afetas à investigação e desenvolvimento, por atividade económica.	18
2.8	Número, percentagem e média por empresa do “Operating revenue / turnover” das empresas com pessoas afetas à investigação e desenvolvimento, por atividade económica.	19
3.1	Documentos mais citados.	24
3.2	Áreas com maior número de documentos.	24
4.1	Número e percentagem das Unidades de Investigação por painel de avaliação FCT.	41
4.2	Número e percentagem das empresas por atividade económica. . . .	42
A.1	Unidades de Investigação incluídas na ferramenta desenvolvida. . . .	93
B.1	Empresas incluídas na ferramenta desenvolvida.	105

Lista de Acrónimos

AandHCI	<i>Arts and Humanities Citation Index</i>
ANI	Agência Nacional de Inovação
BKCI	<i>Book Citation Index</i>
CAE-Rev.3	Classificação Portuguesa de Atividades Económicas, Revisão 3
CPCI	<i>Conference Proceedings' Citation Index</i>
CSS	<i>Cascading Style Sheets</i>
ESCI	<i>Emerging Sources Citation Index</i>
FCT	Fundação para a Ciência e a Tecnologia
FOS	<i>Fields of Science and Technology</i>
GECAD	Grupo de Investigação em Engenharia e Computação Inteligente para a Inovação e o Desenvolvimento
HTML	<i>Hypertext Markup Language</i>
I&D	Investigação e Desenvolvimento
I&D+i	Investigação, Desenvolvimento e Inovação
I&I	Investigação e Inovação
IA	Inteligência Artificial
IES	Instituições de Ensino Superior
INESC TEC	INESC TEC - INESC Tecnologia e Ciência
INET-md	Instituto de Etnomusicologia - Centro de Estudos em Música e Dança
ISEP	Instituto Superior de Engenharia do Porto
LDA	<i>Latent Dirichlet Allocation</i>
NIF	Número de Identificação Fiscal

NLP	processamento avançado de linguagem natural
ODS	Objetivos de Desenvolvimento Sustentável
PIB	Produto Interno Bruto
SCIE	<i>Science Citation Index Expanded</i>
SCTN	Sistema Científico e Tecnológico Nacional
SNI	Sistema Nacional de Inovação
SQL	<i>Structured Query Language</i>
SSCI	<i>Social Science Citation Index</i>
UI	Unidade de Investigação
UM	Universidade do Minho
UP	Universidade do Porto
UTAD	Universidade de Trás-os-Montes e Alto Douro
WoS	<i>Web of Science</i>
WoS CC	<i>Web of Science Core Collection</i>
XML	<i>Extensible Markup Language</i>

Capítulo 1

Introdução

A excelência da investigação científica produzida é indissociável da inovação e criação de valor económico e social. Assim, a aproximação entre diferentes áreas científicas e de negócio, e a consolidação de formas de colaboração entre as instituições de investigação e o tecido económico e social, potencia a criação, transferência e valorização do conhecimento.

Neste capítulo será feita uma contextualização do problema abordado e temas relacionados, que levaram à motivação para desenvolver esta dissertação. De seguida será descrito o objetivo e o plano de trabalho para a sua concretização. Por fim, será descrita a organização do documento.

1.1 Contextualização e motivação

No contexto atual, o mundo enfrenta imensos desafios e encontrar um equilíbrio sustentável entre preocupações ecológicas, económicas e sociais, a transição digital e os principais desenvolvimentos políticos serão alguns dos principais motores de mudança para esta década [1].

Neste âmbito, importa fazer referência à Agenda 2030 para o desenvolvimento sustentável, adotada por todos os Estados Membros das Nações Unidas em 2015, que fornece um plano compartilhado para a paz e a prosperidade para as pessoas e o planeta, agora e no futuro. No seu cerne estão 17 Objetivos de Desenvolvimento Sustentável (ODS) (Figura 1.1), que são um apelo urgente à ação de todos os países

- desenvolvidos e em desenvolvimento - numa parceria global. Estes objetivos reconhecem que acabar com a pobreza e outras privações deve andar de mãos dadas com estratégias que melhorem a saúde e a educação, reduzam a desigualdade e estimulem o crescimento económico – ao mesmo tempo que se combatem as mudanças climáticas e se trabalha na preservação dos oceanos e florestas [2].



Figura 1.1: Objetivos para o desenvolvimento sustentável [3].

Para dar resposta a estes desafios, as Instituições de Ensino Superior (IES) e as empresas assumem um importante papel.

As IES têm que adequar as suas missões ao serviço da sustentabilidade, avaliando e responsabilizando-se pelo adequado ajuste dos princípios e valores, bem como de políticas e atividades. As IES têm como objetivo, entre outros, expandir as fronteiras do conhecimento, educar os cidadãos e impulsionar mudanças sociais. São lugares onde devem prosperar novas ideias e conseqüente criação de novos conhecimentos. A produção de conhecimento pode beneficiar do diálogo com a sociedade, envolvendo ativamente cidadãos e parceiros não académicos, como empresas, organizações não governamentais, autoridades públicas, entre outros [1].

Por sua vez, as empresas deverão assumir o compromisso de implementar iniciativas que abordem os ODS e suas metas, e contribuir com o seu poder de inovação para um futuro mais sustentável e inclusivo [3].

Assim, no contexto atual e no âmbito da evolução de uma sociedade baseada no conhecimento, a transferência de conhecimento e a colaboração entre IES e empresas assume um papel crucial na promoção do desenvolvimento [4]. A inovação a partir de atividades de Investigação e Desenvolvimento (I&D) é cada vez mais determinante para a criação de riqueza, a diferenciação de bens e serviços e a adaptação e transformação da estrutura produtiva com vista a aumentar o bem-estar social e económico da população [5].

A relevância do conhecimento na obtenção de resultados positivos e importantes no sistema empresarial tem sido analisada sob várias perspectivas na literatura internacional [6], nomeadamente e como exemplo, a criação e transferência de conhecimento dentro e entre organizações tem sido abordada em vários estudos [7, 8]. Embora os resultados positivos da transferência de conhecimento estejam no aumento das vantagens competitivas das empresas e na melhoria do desempenho organizacional e empresarial, a transferência de conhecimento nem sempre está isenta de análises críticas [9]. Considerando a economia baseada no conhecimento, é relevante perceber se o conhecimento é transferível, bem como quais são as variáveis que influenciam o desempenho organizacional em todos os tipos de empresas, incluindo *start-ups*, que requerem novas formas de financiamento [10].

Também a transferência de conhecimento e a relação entre o ensino superior e a investigação desenvolvida e a indústria, em concreto, têm sido objeto de estudo [11], sendo considerada um importante motor de inovação e crescimento económico, pois facilita a comercialização de novos conhecimentos científicos dentro das empresas [12]. Os investigadores também beneficiam da interação com a indústria, pois podem surgir novas linhas de investigação e financiamento adicional [13]. Nas últimas décadas, também aumentou o estudo relativo ao empreendedorismo académico com especial enfoque nas patentes, licenciamento, *spin-offs*, entre outros [14, 15]. No entanto, existem outras formas para os investigadores garantirem a aplicação dos conhecimentos, concentrando-se essas práticas predominantemente na troca de conhecimentos [16, 17]. De destacar também a identificação de práticas que podem facilitar a relação entre o ensino superior e as empresas, com a identificação de barreiras e facilitadores da transferência de conhecimento [11].

Em Portugal, segundo a Agência Nacional de Inovação (ANI) [5], têm sido dados passos significativos na criação de condições para a produção de conhecimento avançado, nomeadamente ao nível das IES. As IES lideram a produção de conhecimento em Portugal, mercê da capacidade de investigação dos seus ecossistemas científicos, onde as Unidades de Investigação representam um pilar fundamental. Temos hoje Universidades e Politécnicos que cobrem praticamente quase todas as áreas de conhecimento, dando resposta às necessidades da sociedade e do mercado. Nas últimas duas décadas, tem sido cada vez mais visível o desenvolvimento de iniciativas de aproximação e de ligação das IES às empresas, sendo cada vez mais frequente o desenvolvimento de projetos de investigação e de inovação colaborativa, resultando em novos produtos, processos ou soluções nos mercados. A capacidade de inovação, traduzida em produtos e serviços transacionáveis, resultante da transferência de conhecimento, é essencial para a competitividade das empresas, para a geração de empregos qualificados e para a melhoria do nível de vida das populações.

Um exemplo concreto é o projeto “UNorte Inova”, que pretende dar resposta à

necessidade de criação de um plano de ação que visa atuar no reforço de transferência de conhecimento científico e tecnológico para o sector empresarial. Este projeto foca-se no apoio às empresas nacionais e no desenvolvimento sistemático e sustentado das suas atividades de inovação e aceleração da criação de valor, estabelecendo uma ponte para as atividades de Investigação, Desenvolvimento e Inovação (I&D+i) que têm sido promovidas nas instituições de ensino superior – Universidade do Porto (UP), Universidade do Minho (UM) e Universidade de Trás-os-Montes e Alto Douro (UTAD). A promoção da competitividade da economia portuguesa, em particular da Região Norte, pela incorporação e valorização do conhecimento gerado por empresas, empreendedores e instituições do Sistema Científico e Tecnológico Nacional (SCTN), constitui o foco principal deste projeto [18]. As ações deste projeto assentam, resumidamente, em ações de sensibilização através de reuniões com empresas e investigadores; criação de oportunidades de *networking*, através de fóruns de *matchmaking* que envolverão investigadores, *start-ups* e empresas e encontros temáticos individuais; criação de um programa de provas de conceito; e criação de um balcão digital “Balcão UNorte”, que pretende apresentar às empresas e outras entidades a oferta de competências/ tecnologias/ serviços das universidades do norte do país.

As ações referidas no projeto “U.Norte Inova”, assim como outras ações que pretendam atuar no reforço de transferência de conhecimento científico e tecnológico para o sector empresarial, exigem, na sua maioria, um grande esforço e dispêndio de tempo, nomeadamente no mapeamento das estruturas envolvidas e suas competências, serviços e necessidades, por exemplo, com diversos mecanismos de interação e partilha de boas práticas. Todas estas ações necessitam de conhecimento prévio para uma melhor orientação de formação de grupos e parcerias para darem origem a resultados efetivos na sequência da transferência e valorização do conhecimento. Ou seja, é necessário saber, com base na experiência e em informação já aceite e difundida, quais as estruturas de investigação que poderão ser úteis e conseguirão dar resposta a problemas concretos e desafios das empresas. Esta tomada de decisão poderá ser difícil, pois exige, por um lado, especialistas das diferentes áreas, com conhecimentos das várias estruturas que poderão ser envolvidas no processo de transferência de conhecimento e ainda assim, por outro lado, esta experiência e conhecimentos, para além da necessidade de constante atualização, podem limitar a inclusão de outras estruturas, menos previsíveis e divulgadas.

Para dar resposta a esta e outras situações, numa era de informação digital, em que existe uma quantidade massiva de informação, é essencial a sua análise e processamento com o objetivo de a converter em informação útil, para que possa auxiliar na compreensão e na tomada de decisões.

A *Internet*, que possui uma ampla área de aplicação, é aceite como uma base de dados de informação relativa aos *websites* [19], que têm crescido de forma exponencial, crescimento acompanhado também por um forte crescimento dos seus

utilizadores [20]. As organizações têm investido quantidades significativas de recursos em aplicações *Web*, internas e externas [21], tendo-se tornado essencial para a comunicação organizacional a *Internet* e os *websites* corporativos [22].

Assim, no mundo competitivo de hoje, a capacidade de extrair conhecimento útil de dados e tomar decisões de acordo com esse conhecimento é cada vez mais importante e essencial. O processo de aplicação de uma metodologia para extração de conhecimento a partir de dados textuais, como, por exemplo, os textos presentes nos *websites*, é denominado de *text mining*.

Considerando, por um lado, o desafio de dinamizar a ligação entre a investigação, desenvolvida maioritariamente no âmbito das IES, e as empresas, dado que as ações usuais necessitam de um grande dispêndio de tempo e pessoas capacitadas, e por outro lado, a grande quantidade de informação disponível sobre estas estruturas e ainda o grande avanço na área computacional e metodologias de análise de dados, em particular dados textuais, torna-se pertinente a utilização destes avanços tecnológicos e disponibilidade de informação para ajudar a dar resposta a este desafio.

1.2 Objetivo e plano de trabalho

O objetivo desta dissertação é desenvolver uma ferramenta de apoio estratégico, que potencie a colaboração entre instituições de investigação e o tecido empresarial. Pretende-se através de metodologias de *text mining* analisar a informação disponível nos *websites* destas entidades e realizar a correspondência entre a investigação desenvolvida pelas instituições de investigação e os produtos e serviços disponibilizados pelas empresas.

Para alcançar este objetivo será necessário primeiramente perceber todo o processo e estruturas envolvidas na transferência e valorização do conhecimento. É necessário ter em consideração que as instituições de investigação e as empresas têm diferentes objetivos e beneficiam do conhecimento de maneira diferente, sendo necessário compartilhar metas para alcançar um entendimento comum do resultado esperado [11].

De seguida importa compreender o que envolve a área de *text mining*, o processo, as várias metodologias e aplicações. A aplicação de metodologias de *text mining* tem como objetivo obter informação nova, útil e não trivial a partir de um grande volume de dados e é um esforço cooperativo entre pessoas e computadores. Os melhores resultados são alcançados através do equilíbrio entre o conhecimento de especialistas (na descrição de problemas e objetivos e interpretação dos resultados) e recursos computacionais [23].

Definidas as instituições de investigação e empresas que serão analisadas, através da informação que disponibilizam nos seus *websites*, e os procedimentos de recolha e análise de dados, será desenvolvida uma ferramenta que possibilite de forma clara

e fácil o acesso aos resultados, resultados estes que deverão ajudar na tomada de decisão.

De referir que o objetivo desta dissertação não é desenvolver um novo processo de transferência de conhecimento, nem desenvolver uma nova metodologia de *text mining* ou estudar e definir um modelo ótimo que relacione a investigação desenvolvida e as necessidades das empresas. O que se pretende, que é novo e diferenciado, é criar uma ferramenta que potencie o processo de transferência de conhecimento; utilizar a informação disponível e acessível através de *websites* para este fim; e aplicar metodologias de *text mining* conhecidas para dar resposta a este problema em concreto.

Deste modo pretende-se também contribuir para a compreensão da dimensão das atividades inovadoras, das características das empresas inovadoras e dos fatores internos e sistémicos que influenciam a inovação, que é um pré-requisito para a investigação e análise de políticas direcionadas para a promoção da inovação [24].

1.3 Organização do documento

O relatório está organizado em cinco capítulos, que começa com a “Introdução”, o presente capítulo. No Capítulo 2 serão abordados os aspetos fundamentais que envolvem a transferência e valorização do conhecimento para uma melhor contextualização e perceção da pertinência da ferramenta desenvolvida, com a descrição da situação em Portugal. No Capítulo 3 será apresentado o tema *text mining*, com a abordagem dos conceitos e métodos necessários para uma melhor perceção e compreensão do trabalho realizado na construção da ferramenta e as suas funcionalidades. O Capítulo 4 será dedicado à explicação do processo de desenvolvimento da ferramenta, com demonstração das suas funcionalidades e resultados que se podem obter. As conclusões do trabalho desenvolvido serão descritas no Capítulo 5, assim como linhas de investigação para o trabalho futuro.

Importa referir que a presente dissertação não pretende providenciar uma análise exaustiva sobre os vários assuntos abordados, mas sim fornecer a informação necessária e esclarecedora, essencial para a compreensão da motivação e de todo o trabalho desenvolvido.

Capítulo 2

Transferência e valorização do conhecimento

O conhecimento é fundamental para o desenvolvimento das sociedades e um impulsionador da inovação. Num mundo globalizado e em constante mudança, tem aumentado a importância da produção de conhecimento com consequente criação de valor.

Neste processo têm particular importância as instituições de ensino e investigação e as empresas, com papéis diferenciados, mas essenciais para completar o ciclo de inovação. As IES, para além da educação e formação, assumem-se cada vez mais como instituições de investigação e abertas à sociedade, o que inclui, para além da criação e atualização do conhecimento, a sua transferência e consequente valorização socioeconómica. Por sua vez, o potenciar da inovação produz benefícios presentes e futuros para as empresas.

Neste capítulo são abordados os aspetos fundamentais que envolvem a transferência e valorização do conhecimento para uma melhor contextualização e perceção da pertinência da ferramenta desenvolvida. Será feita uma introdução do tema e a descrição do processo, que põe o conhecimento em ação, com ênfase, por fim, na transferência e valorização do conhecimento em Portugal e o importante papel das unidades de investigação e das empresas, estruturas que serão utilizadas na ferramenta desenvolvida.

2.1 Introdução

O conhecimento pode ser descrito como tendo duas dimensões – explícito e tácito. O conhecimento explícito é articulável, pode ser expresso na forma de palavras e números e é fácil de transmitir, enquanto que o conhecimento tácito, como está implicitamente na mente das pessoas, é difícil de articular e requer observação, demonstração e experiência para a sua transferência [25].

O conhecimento e a sua gestão têm recebido cada vez mais atenção nos últimos anos e existem dois processos que são considerados centrais para a gestão do mesmo: criação de conhecimento e transferência de conhecimento [26]. A criação de conhecimento é semelhante à exploração, na qual indivíduos e equipas criam novas ideias e conceitos, através da combinação do conhecimento existente [27]. A criação de conhecimento está estreitamente ligada à inovação de produtos e serviços [28]. A transferência de conhecimento, por outro lado, permite a exploração e aplicação do conhecimento existente para os propósitos de uma organização [25].

A transferência de conhecimento, que conseqüentemente levará à valorização do mesmo, pode ser definida como um processo de troca de conhecimento explícito ou tácito entre dois agentes, durante o qual um agente recebe e usa o conhecimento fornecido por outro. O processo de troca envolve dois atos complementares: o ato de dar ou entregar conhecimento por um agente (a fonte), complementado pelo ato de receber e usar o conhecimento por outro (o destinatário). Sem ambos, o processo de transferência é incompleto [25]. A esta definição deve ser acrescentado que as concepções tradicionais de um processo de inovação linear, são atualmente compreendidas como um processo multidirecional e interativo, envolvendo vários atores [29].

Surge assim o conceito de inovação, que pode ser definida como um novo ou melhorado produto ou processo (ou combinação dos dois), diferenciando-se significativamente dos seus anteriores e que tenha sido disponibilizado a potenciais utilizadores (produto) ou utilizado dentro de uma unidade (processo). As inovações derivam de atividades baseadas em conhecimento que envolvem a aplicação prática de informações e conhecimentos existentes ou recém-desenvolvidos [30].

O conceito de inovação é indissociável do conceito de competitividade. A competitividade de uma organização depende da sua capacidade de adquirir e organizar os seus recursos, de forma a tornar a sua atividade o mais eficiente possível, no sentido de potenciar um determinado retorno, financeiro ou não. Numa economia cada vez mais global e interdependente, a competitividade assenta, sobretudo, na diferenciação e criação de valor e, cada vez menos, em baixos custos de produção [5].

2.2 Processo

O processo de transferência de conhecimento refere-se ao fluxo no qual, o conhecimento é transmitido entre entidades, como já referido, sendo atualmente mais complexo, mais interrelacionado, mais abrangente e com diversos intervenientes (Figura 2.1).

Nos últimos vinte anos, verificou-se uma evolução do conceito de transferência de conhecimento, que passou do conceito mais tradicional de comercialização e monetização para uma abordagem mais abrangente, que apoia tanto a cocriação quanto a disseminação de resultados de investigação com e para vários *stakeholders* [31].

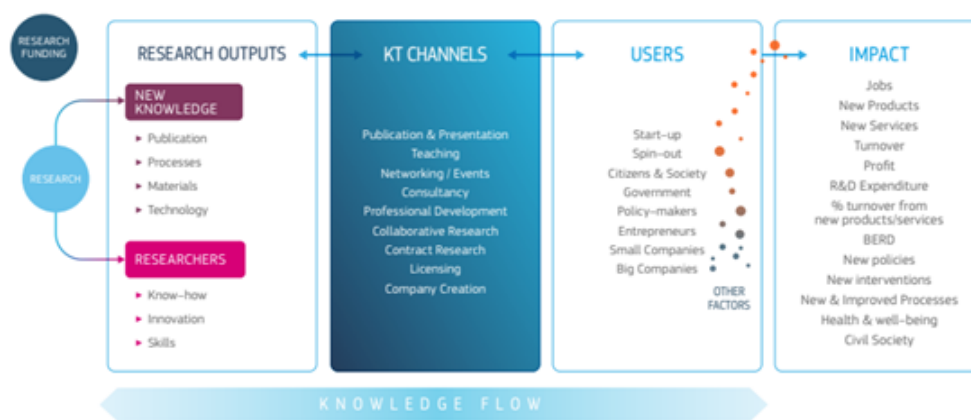


Figura 2.1: Transferência de conhecimento: da investigação ao impacto [31].

Em suma, a transferência de conhecimento cobre os processos de transferência de investigação, competências, experiência e ideias com o propósito de aumentar o retorno económico do seu investimento de forma a garantir benefícios culturais, educacionais e sociais para a sociedade. Consequentemente, é objetivada a valorização do mesmo conhecimento, como processo de criação de valor a partir do conhecimento, tornando-o adequado e/ou disponível para o uso económico e/ou social e traduzindo o mesmo conhecimento em produtos, processos e atividades empresariais competitivas [5].

2.3 Em Portugal

Segundo a Agência Nacional de Inovação [5], Portugal tem feito um caminho sólido no que respeita à sua capacidade de inovação e de criação de valor, contribuindo para que sejam garantidas novas ou significativamente melhoradas competências, essenciais para que organizações sediadas em Portugal consigam enfrentar uma realidade em constante evolução. Portugal tem acompanhado as melhores práticas

num conjunto de áreas tecnológicas e industriais, muito devido à articulação de esforços provenientes de diferentes tipos de atores, assegurado pela forte capacidade de desenvolver conhecimento e, também, pelo nível de qualificações dos recursos humanos em Portugal, nomeadamente dos investigadores. Segundo a FCT [32], as duas últimas décadas foram um período de notável desenvolvimento para o sistema português de investigação e inovação. O grande crescimento da produção científica foi sustentado pelo alargamento do número de investigadores, instituições de I&D e indicadores como a percentagem Produto Interno Bruto (PIB) dedicado à ciência, número de prémios a doutorados e número de publicações de referência internacional. Estes fatores, associados à existência de cada vez mais empresas com capacidade de inovação, justificam a existência de uma realidade nacional, onde a troca de experiências e trabalho em conjunto contribuem para o potencial de aprendizagem e de crescimento do país.

Existe uma multiplicidade de atores no sistema de inovação português com capacidade de desenvolver, transferir e valorizar conhecimento e gerar valor económico e social. Todas estas entidades desempenham o seu papel no sistema de inovação português e contribuem para a transferência e valorização do conhecimento. Importa assinalar que os tipos de entidades identificadas a azul escuro (Figura 2.2) têm uma participação mais ativa na transferência e valorização do conhecimento [5].



Figura 2.2: O SNI [5].

Assim, é possível agrupar as principais entidades do processo de transferência e valorização de conhecimento em três grupos: o “Ensino Superior e Investigação”, que inclui as Instituições de Ensino Superior, unidades de investigação, outros grupos de investigação e investigadores individuais; as “Empresas”, com diferentes dimensões e diferente utilização do conhecimento; e as “as estruturas intermediárias de apoio”.

2.3.1 Unidades de Investigação

As unidades de investigação representam um pilar fundamental no sistema científico e tecnológico português. São instituições de investigação públicas ou privadas, sem fins lucrativos, onde se desenvolve a maior parte da investigação e inovação não empresarial e devem reunir uma massa crítica adequada à sua missão e promover ambientes criativos, em que possam surgir novas ideias e onde os investigadores encontrem as condições adequadas à realização dos seus projetos científicos e ao desenvolvimento da sua carreira. Sempre que aplicável, devem reunir recursos interdisciplinares e multidisciplinares que potenciem a abordagem de problemas complexos e novos desafios sociais [33].

As unidades de investigação distribuem-se por diversas áreas de conhecimento e em todo o território nacional. Estas unidades podem ser entidades integradas em diferentes universidades ou politécnicos, ou constituírem parcerias entre universidades ou ainda possuírem um estatuto independente e funcionamento autónomo [33].

As unidades de investigação são avaliadas periodicamente, a cada 4-5 anos, por painéis de avaliadores internacionais, que atribuem uma classificação e um financiamento para o período plurianual seguinte. Os resultados do processo da última avaliação - 2017/2018 - determinaram o financiamento das Unidades de I&D com classificação igual a “Bom”, “Muito Bom” e “Excelente”, para o período 2020-2023 [33]. A avaliação é feita pela FCT, agência de financiamento da ciência nacional, que assume um papel central no Sistema de Investigação e Inovação nacional. É atualmente a principal financiadora não só do sector público de investigação, mas também das atividades de Investigação e Inovação (I&I) de natureza mais básica e estratégica do sector empresarial [34].

Os resultados da última avaliação estão disponíveis em [35] e uma caracterização das unidades com financiamento atribuído em [32].

Participaram na última avaliação 348 unidades de investigação, tendo sido atribuído financiamento a 312 (89,7%), que obtiveram a classificação igual a “Bom” (79 - 22,7%), “Muito Bom” (123 - 35,3%) e “Excelente” (110 - 31,6%). Das unidades que não obtiveram financiamento, 32 (9,2%) obtiveram a classificação “Fraco”, 3 (0,9%) “Insuficiente” e 1 (0,3%) unidade não classificada (Tabela 2.1).

Tabela 2.1: Classificação das Unidades de Investigação segundo a avaliação de 2017/2018 da FCT.

Classificação	N.º Unidades	Percentagem
Excelente	110	31,6%
Muito Bom	123	35,3%
Bom	79	22,7%
Fraco	32	9,2%
Insuficiente	3	0,9%
Não Classificada	1	0,3%

Considerando as 312 unidades que obtiveram financiamento, a distribuição por domínio científico (Tabela 2.2) mostra que o maior número de unidades pertence ao domínio “ciências sociais” (86 unidades, 27,6% do total) e o menor número ao domínio “ciências agrárias” (15 unidades, 4,8%). O domínio “ciências naturais”, que engloba as “ciências exatas”, inclui 63 unidades (20,2%), o domínio “humanidades” 62 (19,9%), o domínio “ciências da engenharia e tecnologias” 54 (17,3%) e o domínio “ciências médicas e da saúde” engloba 32 unidades (10,3%).

Tabela 2.2: Número e percentagem das Unidades de Investigação por domínio científico.

Domínio científico	N.º	%
Ciências naturais	63	20,2
Ciências da engenharia e tecnologias	54	17,3
Ciências médicas e da saúde	32	10,3
Ciências agrárias	15	4,8
Ciências sociais	86	27,6
Humanidades	62	19,9
Total	312	100

Os 18 523 investigadores reportados como integrados nas unidades de I&D (Tabela 2.3) encontram-se maioritariamente no domínio “ciências naturais” (4 612 investigadores, 24,9% do total), seguindo-se o domínio “ciências sociais” (3 990 investigadores, 21,5%), “humanidades” (3 313 investigadores, 17,9%), “Ciências da engenharia e tecnologias” (3 147 investigadores, 17,0%), “ciências médicas e da saúde” (2 509 investigadores, 13,5%) e “ciências agrárias” (952 investigadores, 5,1%). A média do número de investigadores por unidade varia de acordo com o domínio, verificando-se a média mais elevada no domínio “ciências médicas e da saúde” (78,4 investigadores por Unidade de Investigação (UI)) e mais baixa no domínio “ciências sociais” (46,4 investigadores por UI).

Tabela 2.3: Número e percentagem dos investigadores por domínio científico.

Domínio científico	N.º	%	Média por UI
Ciências naturais	4 612	24,9	73,2
Ciências da engenharia e tecnologias	3 147	17,0	58,3
Ciências médicas e da saúde	2 509	13,5	78,4
Ciências agrárias	952	5,1	63,5
Ciências sociais	3 990	21,5	46,4
Humanidades	3 313	17,9	53,4
Total	18 523	100	59,4

A maior percentagem do total de financiamento (Tabela 2.4) foi atribuído ao domínio “ciências naturais” (116 M€, 27,4%), seguindo-se “ciências sociais” (79 M€ 18,7%) e “ciências da engenharia e tecnologia” (79 M€, 18,6%), “humanidades” (67 M€, 15,8%), “ciências médicas e da saúde” (61 M€, 14,4%) e por último o domínio “ciências agrárias” (22 M€, 5,2%). Observa-se no domínio “ciências médicas e da saúde” o valor mais elevado da média de financiamento por unidade (1,9 M€ por UI) e no domínio “ciências sociais” o valor mais baixo (0,9 M€ por UI).

Tabela 2.4: Número e percentagem do financiamento por domínio científico.

Domínio científico	M€	%	Média por UI
Ciências naturais	116	27,4	1,8
Ciências da engenharia e tecnologias	79	18,6	1,5
Ciências médicas e da saúde	61	14,4	1,9
Ciências agrárias	22	5,2	1,5
Ciências sociais	79	18,7	0,9
Humanidades	67	15,8	1,1
Total	425	100	1,4

2.3.2 Empresas

Uma empresa pode ser definida como uma entidade jurídica (pessoa singular ou coletiva) correspondente a uma unidade organizacional de produção de bens e/ou serviços, usufruindo de uma certa autonomia de decisão, nomeadamente quanto à afetação dos seus recursos correntes. Uma empresa pode exercer uma ou várias atividades, em um ou em vários locais [36].

Para se obter informações sobre as empresas portuguesas foi utilizada a base de dados Sabi [37], que é uma base dados que contém informações abrangentes de

análise financeira das empresas portuguesas. Relativamente a 2020 (último ano com informação completa), a base de dados tem um total de 804 210 empresas, estando apenas 425 379 ativas.

Foi utilizado o campo “Pessoas ao serviço da empresa, afetas à Investigação e Desenvolvimento”, para selecionar, das 425 379 empresas ativas, as empresas com pelo menos uma pessoa afeta à investigação, tendo-se obtido 1 511 (0,4%) empresas.

Considerando as 1 511 empresas com pessoas afetas à investigação, a distribuição por atividade económica (Tabela 2.5), em que foi considerada a Classificação Portuguesa de Atividades Económicas, Revisão 3 (CAE-Rev.3) [38], mostra que o maior número de empresas, com uma grande diferença para as seguintes, pertence à secção “Indústrias transformadoras” (640 empresas, 42,4%), seguindo-se “Atividades de informação e de comunicação” (185 empresas, 12,2%) e “Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos” (178 empresas, 11,8%). As secções com o menor número de empresas são “Eletricidade, gás, vapor, água quente e fria e ar frio” (3 empresas, 0,2%), “Indústrias extrativas” (5 empresas, 0,3%) e “Atividades artísticas, de espetáculos, desportivas e recreativas” e “Outras atividades de serviços” com o mesmo número (9 empresas, 0,6%).

Tabela 2.5: Número e percentagem das empresas com pessoas afetas à investigação e desenvolvimento por atividade económica.

Atividade económica	N.º	%
Agricultura, produção animal, caça, floresta e pesca	28	1,9
Indústrias extrativas	5	0,3
Indústrias transformadoras	640	42,4
Eletricidade, gás, vapor, água quente e fria e ar frio	3	0,2
Captação, tratamento e distribuição de água; saneamento, gestão de resíduos e despoluição	14	0,9
Construção	85	5,6
Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos	178	11,8
Transportes e armazenagem	23	1,5
Alojamento, restauração e similares	39	2,6
Atividades de informação e de comunicação	185	12,2
Atividades financeiras e de seguros	11	0,7
Atividades imobiliárias	34	2,3
Atividades de consultoria, científicas, técnicas e similares	158	10,5
Atividades administrativas e dos serviços de apoio	32	2,1
Educação	14	0,9
Atividades de saúde humana e apoio social	44	2,9
Atividades artísticas, de espetáculos, desportivas e recreativas	9	0,6
Outras atividades de serviços	9	0,6
Total	1 511	100

No que respeita às “Pessoas ao serviço da empresa, remuneradas e não remuneradas: número médio de pessoas” (Tabela 2.6), do total das 188 023 pessoas, as percentagens mais elevadas encontram-se nas secções com maior número de empresas, nomeadamente “Indústrias transformadoras” (122 640 pessoas, 65,2%), “Atividades de informação e de comunicação” (21 372 pessoas, 11,4%) e “Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos” (15 760 pessoas, 8,4%). As secções com menos pessoas são “Atividades financeiras e de seguros” (24 pessoas, 0,01%), “Atividades artísticas, de espetáculos, desportivas e recreativas” (47 pessoas, 0,02%) e “Atividades imobiliárias” (53 pessoas, 0,03%). Relativamente à média de pessoas por empresa, a secção “Indústrias extrativas” apresenta o valor mais elevado (300,4 pessoas por empresa), seguindo-se “Captação, tratamento e distribuição de água; saneamento, gestão de resíduos e despoluição” (257,4 pessoas por empresa) e “Indústrias transformadoras” (191,6 pessoas por empresa). Os valores médios mais baixos observam-se nas secções “Atividades imobiliárias” (1,6 pessoas por empresa), “Atividades financeiras e de seguros” (2,2 pessoas por empresa) e “Atividades artísticas, de espetáculos, desportivas e recreativas” (5,2 pessoas por empresa).

Considerando agora as “Pessoas afetas à Investigação e Desenvolvimento” (Tabela 2.7), a secção “Indústrias transformadoras” apresenta o maior número (7 530 pessoas afetas à investigação, 50,2%), seguindo-se “Atividades de informação e de comunicação” (3 894 pessoas afetas à investigação, 26,0%) e “Atividades de consultoria, científicas, técnicas e similares” (1 424 pessoas afetas à investigação, 9,5%). Por sua vez, as secções com menos pessoas afetas à investigação são “Atividades artísticas, de espetáculos, desportivas e recreativas” (13 pessoas afetas à investigação, 0,1%), “Eletricidade, gás, vapor, água quente e fria e ar frio” (17 pessoas afetas à investigação, 0,1%) e “Outras atividades de serviços” (19 pessoas afetas à investigação, 0,1%). Quanto ao valor médio por empresa, a secção “Atividades de informação e de comunicação” apresenta o valor mais elevado (21,0 pessoas afetas à investigação por empresa), seguindo-se “Indústrias transformadoras” (11,8 pessoas afetas à investigação por empresa) e “Indústrias extrativas” (10,2 pessoas afetas à investigação por empresa). As secções com menor valor médio são “Atividades imobiliárias” (1,4 pessoas afetas à investigação por empresa), “Atividades artísticas, de espetáculos, desportivas e recreativas” (1,4 pessoas afetas à investigação por empresa) e “Atividades financeiras e de seguros” (2,1 pessoas afetas à investigação por empresa).

A maior percentagem relativa ao volume de negócios (“Operating revenue / turnover”) (Tabela 2.8) verifica-se na secção “Indústrias transformadoras” (23 860,0 M€, 67,2%), seguindo-se “Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos” (4 892,5 M€, 13,8%) e “Atividades de informação e de comunicação” (4 011,7 M€, 11,3%). As secções com menor percentagem são

“Atividades artísticas, de espetáculos, desportivas e recreativas” (2,3 M€, 0,006%), “Atividades financeiras e de seguros” (2,9 M€, 0,008%) e “Atividades imobiliárias” (3,8 M€, 0,01%). A secção “Indústrias extrativas” apresenta o valor médio mais elevado (61,4 M€por empresa), seguindo-se “Eletricidade, gás, vapor, água quente e fria e ar frio” (38,3 M€por empresa) e “Indústrias transformadoras” (37,3 M€por empresa) e as secções “Atividades imobiliárias” (0,1 M€por empresa), “Atividades artísticas, de espetáculos, desportivas e recreativas” (0,3 M€por empresa) e “Atividades financeiras e de seguros” (0,3 M€por empresa) apresentam os valores médios mais baixos.

Tabela 2.6: Número, percentagem e média por empresa das “Pessoas ao serviço da empresa, remuneradas e não remuneradas: número médio de pessoas” das empresas com pessoas afetas à investigação e desenvolvimento, por atividade económica.

Atividade económica	N.º	%	Média por empresa
Agricultura, produção animal, caça, floresta e pesca	837	0,4	29,9
Indústrias extrativas	1 502	0,8	300,4
Indústrias transformadoras	122 640	65,2	191,6
Eletricidade, gás, vapor, água quente e fria e ar frio	102	0,1	34,0
Captação, tratamento e distribuição de água; saneamento, gestão de resíduos e despoluição	3 604	1,9	257,4
Construção	2 517	1,3	29,6
Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos	15 760	8,4	88,5
Transportes e armazenagem	3 166	1,7	137,7
Alojamento, restauração e similares	778	0,4	19,9
Atividades de informação e de comunicação	21 372	11,4	115,5
Atividades financeiras e de seguros	24	< 0,1	2,2
Atividades imobiliárias	53	< 0,1	1,6
Atividades de consultoria, científicas, técnicas e similares	4 684	2,5	29,6
Atividades administrativas e dos serviços de apoio	3 124	1,7	97,6
Educação	267	0,1	19,1
Atividades de saúde humana e apoio social	7 329	3,9	166,6
Atividades artísticas, de espetáculos, desportivas e recreativas	47	< 0,1	5,2
Outras atividades de serviços	217	0,1	24,1
Total	188 023	100	124,4

Tabela 2.7: Número, percentagem e média por empresa das “Pessoas afetas à Investigação e Desenvolvimento” das empresas com pessoas afetas à investigação e desenvolvimento, por atividade económica.

Atividade económica	N.º	%	Média por empresa
Agricultura, produção animal, caça, floresta e pesca	77	0,5	2,8
Indústrias extrativas	51	0,3	10,2
Indústrias transformadoras	7 530	50,2	11,8
Eletricidade, gás, vapor, água quente e fria e ar frio	17	0,1	5,7
Captação, tratamento e distribuição de água; saneamento, gestão de resíduos e despoluição	68	0,5	4,9
Construção	353	2,4	4,2
Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos	794	5,3	4,5
Transportes e armazenagem	51	0,3	2,2
Alojamento, restauração e similares	211	1,4	5,4
Atividades de informação e de comunicação	3 894	26,0	21,0
Atividades financeiras e de seguros	23	0,2	2,1
Atividades imobiliárias	47	0,3	1,4
Atividades de consultoria, científicas, técnicas e similares	1 424	9,5	9,0
Atividades administrativas e dos serviços de apoio	250	1,7	7,8
Educação	52	0,3	3,7
Atividades de saúde humana e apoio social	127	0,8	2,9
Atividades artísticas, de espetáculos, desportivas e recreativas	13	0,1	1,4
Outras atividades de serviços	19	0,1	2,1
Total	15 001	100	9,9

Tabela 2.8: Número, percentagem e média por empresa do “Operating revenue / turnover” das empresas com pessoas afetas à investigação e desenvolvimento, por atividade económica.

Atividade económica	M€	%	Média por empresa
Agricultura, produção animal, caça, floresta e pesca	96,3	0,3	3,4
Indústrias extrativas	306,8	0,9	61,4
Indústrias transformadoras	23 860,0	67,2	37,3
Eletricidade, gás, vapor, água quente e fria e ar frio	115,0	0,3	38,3
Captação, tratamento e distribuição de água; saneamento, gestão de resíduos e despoluição	498,6	1,4	35,6
Construção	396,6	1,1	4,7
Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos	4 892,5	13,8	27,5
Transportes e armazenagem	248,3	0,7	10,8
Alojamento, restauração e similares	20,1	0,1	0,5
Atividades de informação e de comunicação	4 011,7	11,3	21,7
Atividades financeiras e de seguros	2,9	< 0,1	0,3
Atividades imobiliárias	3,8	< 0,1	0,1
Atividades de consultoria, científicas, técnicas e similares	488,0	1,4	3,1
Atividades administrativas e dos serviços de apoio	107,9	0,3	3,4
Educação	11,0	< 0,1	0,8
Atividades de saúde humana e apoio social	448,2	1,3	10,2
Atividades artísticas, de espetáculos, desportivas e recreativas	2,3	< 0,1	0,3
Outras atividades de serviços	18,0	0,1	2,0
Total	35 528,1	100	23,5

Capítulo 3

Text mining

Os dados qualitativos são atrativos por várias razões, são dados ricos, completos, holísticos, “reais”. Contudo, recolher e analisar este tipo de dados exige muito trabalho e tempo, nomeadamente por existir uma grande variedade e quantidade de informação [39]. A utilização de *text mining* abrange um vasto conjunto de métodos com uma coisa em comum: texto como informação de entrada. A sua utilização tem como benefício a grande quantidade de informação importante latente neste formato e que não está disponível nos formatos clássicos de dados estruturados [40].

Neste capítulo são apresentados os aspetos fundamentais que envolvem o tema *text mining*, com a abordagem dos conceitos e métodos necessários para uma melhor perceção e compreensão do trabalho realizado na construção da ferramenta e as suas funcionalidades. De notar que existe uma grande variedade de metodologias, conceitos e estudos relacionados com o tema, mas que não foi considerada essencial a sua abordagem ou aprofundamento para se cumprir o objetivo proposto neste trabalho.

O capítulo começa com uma introdução do tema *text mining*, seguindo-se uma breve análise cientométrica com utilização de *text mining*. Após esta introdução e contextualização, será feita uma abordagem das bases metodológicas e em específico da técnica utilizada na ferramenta desenvolvida - *topic modeling* com utilização do modelo LDA.

3.1 Introdução

Text mining (em português “prospecção/mineração de dados textuais” [41]) é uma área interdisciplinar, que tem como objetivo extrair conhecimento (ou informações, padrões) a partir de dados textuais, que são não estruturados ou semiestruturados. *Text mining* está assim relacionada com várias áreas científicas, que incluem, por exemplo, *machine learning*, recuperação de informação, processamento de linguagem natural, estatística, reconhecimento de padrões e inteligência artificial [42]. Devido à sua transversalidade, tem vindo a ganhar popularidade, conseguindo reunir as denominadas ciências “leves” e “duras”. Tem vindo também a acumular várias metodologias, que a tornam flexível para vários propósitos e difusa o suficiente para aproximar a compreensão de texto semelhante à humana e a extração de conhecimento [43].

Historicamente, as raízes do que hoje se chama de *text mining* estão na área de recuperação de informação [44]. A classificação de documentos é semelhante em muitos aspetos à indexação de documentos, que foi extensivamente estudada no final dos anos 1950 e 1960 [45, 46]. O agrupamento de documentos e medição de similaridade de documentos também são temas antigos [47]. A representação de um documento como *bag of words*, com uma medida de frequência associada a cada palavra, tornou-se popular na década de 1970 [48]. Com o desenvolvimento na década de 1980 de métodos de Inteligência Artificial (IA), os mesmos também foram aplicados a problemas que hoje se classificam como *text mining*, sendo particularmente verificado na categorização de texto [49].

O verdadeiro impulso para o desenvolvimento das técnicas usadas atualmente deve-se à disponibilidade de computação rápida e barata e também à grande quantidade de dados textuais disponíveis em formato digital [44].

3.2 Análise cientométrica com utilização de *text mining*

A cientometria analisa os aspetos quantitativos da produção, divulgação e uso da informação científica com o objetivo de alcançar uma melhor compreensão dos mecanismos da investigação científica [50]. A aplicação desta técnica pode fornecer uma visão macroscópica de uma grande quantidade de literatura científica através de uma análise quantitativa com utilização de técnicas de *text mining*. Assim, é possível mapear o desenvolvimento científico de uma determinada área, identificar padrões relacionados com, por exemplo, os autores, revistas e países [51].

Para a análise cientométrica, com o objetivo de estudar a produção científica relativa ao tópico *text mining*, foi utilizada a base de dados *Web of Science* (WoS) [52] e o *software* VOSviewer. A WoS da Clarivate Analytics é uma das principais

bases de dados de documentos científicos, pesquisa de citações e plataforma de informação analítica. A WoS é utilizada tanto como ferramenta de pesquisa, devido à sua vasta biblioteca acadêmica, bem como uma fonte de informação para estudos de larga escala em inúmeras áreas científicas [53, 54]. A principal coleção da plataforma é a *Web of Science Core Collection* (WoS CC), que compreende seis índices principais de citações: *Science Citation Index Expanded* (SCIE); *Social Science Citation Index* (SSCI); *Arts and Humanities Citation Index* (AandHCI); *Conference Proceedings' Citation Index* (CPCI); *Book Citation Index* (BKCI); *Emerging Sources Citation Index* (ESCI) [51]. O VOSviewer é um *software*, disponível gratuitamente, desenvolvido na Universidade de Leiden, para construção e visualização de mapas bibliométricos [55].

A pesquisa na base de dados WoS foi realizada no dia 26 de agosto de 2022, tendo sido utilizado o termo de pesquisa “*text mining*” no campo “*topic*” (este campo procura o termo inserido no título do artigo, no *abstract* e nas palavras-chave do autor e do artigo [52]). Desta pesquisa resultaram 14 918 documentos indexados.

O primeiro artigo indexado referente ao tema é datado de 1997, com o título “*Mining in the phrasal Frontier*”, foi escrito por Ahonen, H; Heinson, O; Klemetinen, M; Verkamo, AI e foi publicado na revista “*Principles of data mining and knowledge discovery*” [56]. O artigo descreve a aplicação de métodos usuais de *data mining* a dados textuais, com a descrição do processo de *text mining* desenvolvido.

O número de documentos tem aumentado desde então, verificando-se uma tendência de crescimento mais acentuada a partir de 2012 (Figura 3.1). De ressaltar que a indexação de documentos poderá não estar completa nos últimos anos, nomeadamente 2020 e 2021.

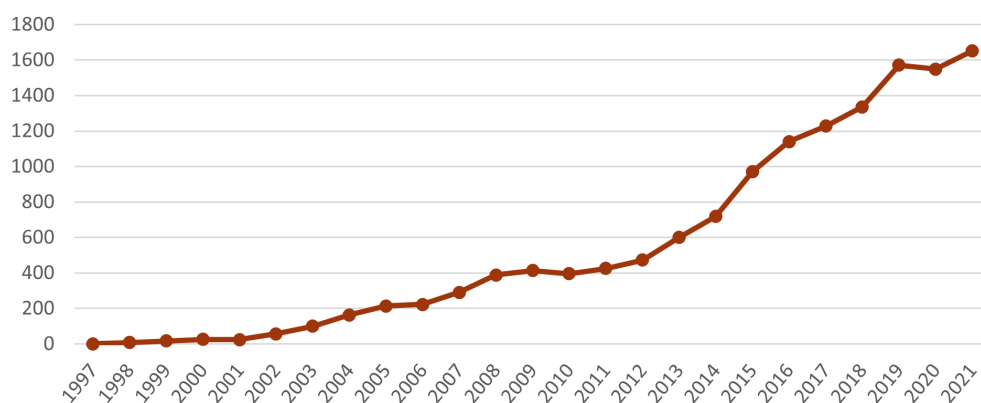


Figura 3.1: Evolução do número de documentos.

O documento com mais citações (6 256) é o artigo “*STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets*” [57]. O artigo, da área de bioquímica e biologia molecular, aborda o tópico *text mining* como uma das funcionalidades da base

de dados STRING, que, de uma maneira simplista, tem como objetivo reunir, pontuar e integrar todas as fontes publicamente disponíveis de informações de interação proteína-proteína e complementá-las com previsões computacionais. Na Tabela 3.1 são apresentados os artigos mais citados, com mais de 1 000 citações. Com exceção do quinto artigo, que pertence às áreas de ciências da computação e informática médica, todos os outros são das áreas de bioquímica e biologia molecular.

Tabela 3.1: Documentos mais citados.

Título	Citações
STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets [57]	6 256
The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible [58]	3 981
Large-scale automated synthesis of human functional neuroimaging data [59]	1 729
miRBase: from microRNA sequences to function [60]	1 450
miRWalk - Database: Prediction of possible miRNA binding sites by "walking" the genes of three genomes [61]	1 302
CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database [62]	1 155
DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants [63]	1 021

Apesar do maior número de documentos ser publicado nas áreas de ciências da computação e engenharia (Tabela 3.2), como esperado, devido às técnicas que vêm sendo desenvolvidas de *text mining*, um grande número de documentos foi publicado em áreas distintas, em que as metodologias de *text mining* são aplicadas, tendo estes documentos, muitas vezes, mais citações, como visto anteriormente.

Tabela 3.2: Áreas com maior número de documentos.

Área	Documentos
Computer Science	8 183
Engineering	2 975
Business & Economics	1 129
Mathematical & Computational Biology	963
Biochemistry & Molecular Biology	848
Information Science & Library Science	829
Medical Informatics	689
Science & Technology - Other Topics	685
Telecommunications	655
Operations Research & Management Science	517

Relativamente aos países com mais documentos (Figura 3.2), destacam-se os Estados Unidos da América e a China, com 3 337 e 2 489 documentos, respetivamente.

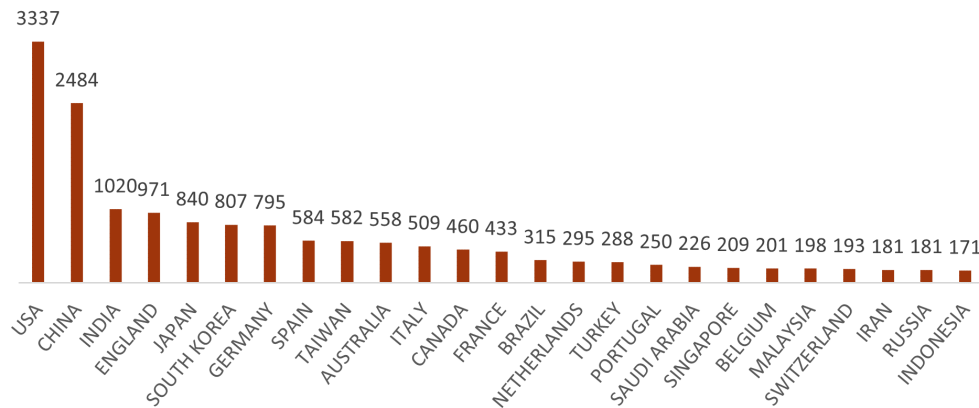


Figura 3.2: Documentos por país.

Foi ainda realizada uma análise de tópicos através da utilização do VOSviewer, para criar um mapa com a representação das palavras-chave principais e as suas relações, utilizando a análise semântica. Através dos algoritmos de *text mining* foram identificados cinco grupos principais de acordo com as suas semelhanças (Figura 3.3). Os grupos foram criados com base na frequência de termos que ocorrem em simultâneo e o tamanho de cada círculo representa o número de vezes que um termo ocorreu [51].

O grupo de cor azul, onde aparece a palavra-chave “*text mining*”, representa as metodologias mais relacionadas com as áreas da computação e inteligência artificial, como “*machine learning*”, “*classification*”, “*web mining*” e “*opinion mining*”. No grupo verde estão as palavras relacionadas com temas mais recentes, como “*twitter*”, “*social media*” e “*big data*” e a metodologia “*sentiment analysis*”, que se utiliza muitas vezes para analisar dados textuais das redes sociais. No grupo amarelo estão palavras relacionadas com a avaliação e gestão da informação científica, como “*science*”, “*bibliometric analysis*”, “*manegement*”, “*performence*” e “*innovation*”. No grupo vermelho estão áreas em que as metodologias de *text mining* são muito aplicadas, como “*biomedical tex mining*”, “*gene ontology*” e “*protein*”, já referido anteriormente na análise dos documentos mais citados, em conjunto com metodologias “*named entity recognition*” e “*retrieval*” e palavras relacionadas com extração de conhecimento, como “*information*”, “*extraction*”, “*knowledge*” e “*database*”. Por fim, o grupo roxo, o mais pequeno, é relativo a “*natural language processing*”, estando este termo relacionado com um grande conjunto de técnicas que são aplicadas ao longo de todo o processo de *text mining*.

Na Figura 3.4 pode-se observar que o grupo verde (da Figura 3.3, agora representado nos tons mais próximos do amarelo) apresenta as palavras utilizadas em

documentos mais recentes, em média, o que corrobora o descrito na explicação do grupo. Por sua vez, as palavras relacionadas com as áreas em que se aplica *text mining* (pertencentes ao grupo vermelho e agora representadas num tom roxo) aparecem nos documentos mais antigos, em média. De notar que palavras-chave como “*text mining*”, “*natural language processing*”, “*classification*”, “*knowledge*”, “*science*” e “*innovation*”, aparecem em documentos que em média estão no centro da escala, ou seja, são palavras que têm vindo sempre a ser utilizadas.

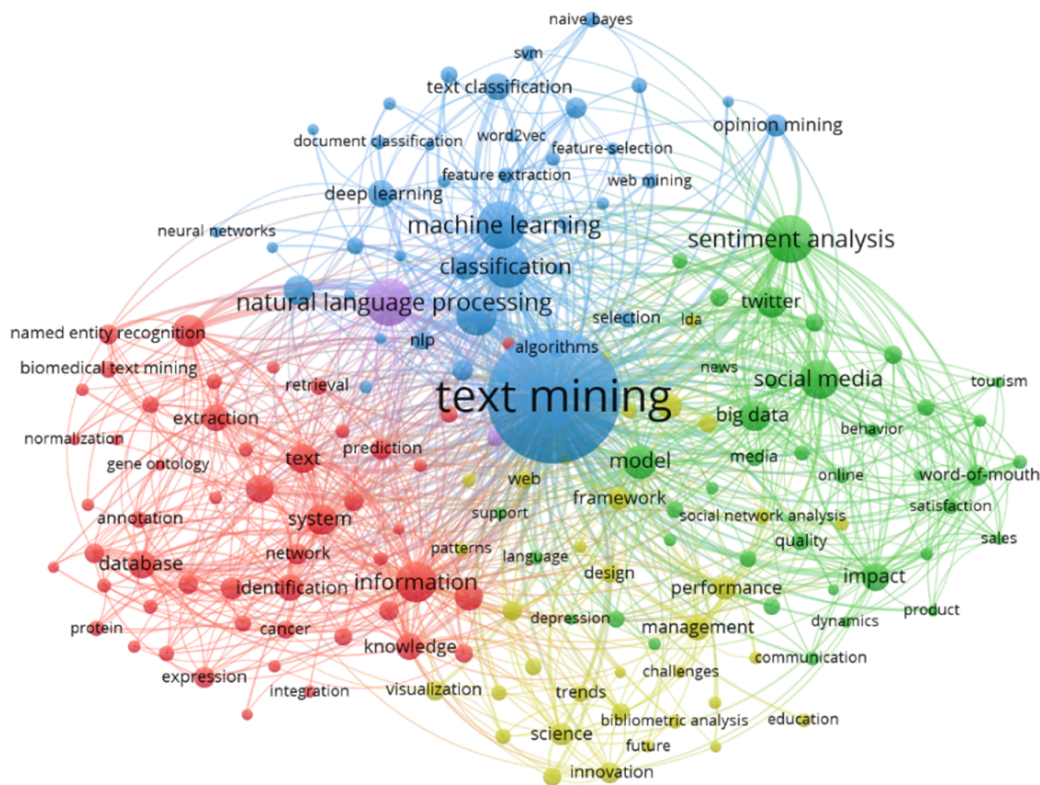


Figura 3.3: Mapeamento de tópicos e *clusters* com base em afinidades.

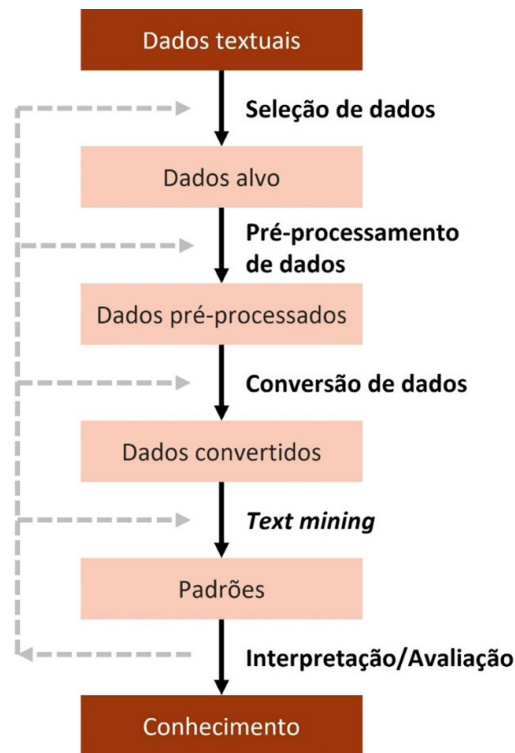


Figura 3.5: Visão geral dos passos que compõem o processo de *text mining*.

blogs, notícias, questionários, pesquisas ou mensagens, e poderão ser obtidos manualmente e automaticamente [43].

Dada a grande variedade de dados textuais e objetivos de estudo, existem diferentes processos de recolha e seleção de dados. Se em alguns casos a informação relevante está definida e acessível, noutros casos pode ser necessário realizar um processo de recolha de dados. Por exemplo, para uma aplicação *Web* que inclua vários *websites* autónomos, pode-se implementar uma ferramenta de *software*, como um *web crawler*, que recolhe os documentos. Noutros casos, pode-se ter um processo de registo anexado a um fluxo de dados de entrada por um período de tempo. De referir ainda que no caso do conjunto de dados textuais ser extremamente grande, podem ser utilizadas técnicas de amostragem para seleccionar um subconjunto de documentos relevantes [64].

- **Pré-processamento de dados**

O passo de pré-processamento de dados é um dos mais desafiadores, principalmente influenciado pelo facto dos textos, do ponto de vista do computador, serem coleções de palavras bastante desestruturadas. Uma análise de *text mining* normalmente começa com um conjunto dados textuais de entrada

muito heterogêneo [40]. O pré-processamento, ou seja, a aplicação de métodos para limpar e estruturar os dados textuais de entrada para uma análise posterior é uma componente central, que afeta diretamente os resultados numa aplicação prática de *text mining* [65, 66, 67]. Existem vários métodos de pré-processamento, dependendo a sua utilização do estudo e seu objetivo. Seguidamente são abordados os métodos mais comumente usados e referidos na literatura.

Tokenization é o procedimento de dividir um texto em palavras, frases ou outras partes significativas denominadas *tokens*, ou por outras palavras, é uma forma de segmentação de texto [66] e é um primeiro passo importante que apesar de parecer fácil, está repleto de pequenas decisões, como, por exemplo, lidar com apóstrofes e hífen, capitalização, pontuação, números, *strings* alfanuméricas, se a quantidade de espaço em branco é significativa, se se deve impor um valor máximo nos *tokens* ou o que fazer com caracteres não imprimíveis [68].

Stemming e *lemmatization* são métodos para normalizar as palavras nas suas formas básicas, para reduzir o tamanho do vocabulário de um conjunto de dados textuais [67]. Enquanto que *stemming* é um método para reduzir uma palavra à sua raiz ou radical, *lemmatization*, em linguística, é o processo de agrupar as diferentes formas flexionadas de uma palavra para que possam ser analisadas como um único elemento [69].

Outra técnica de pré-processamento é a remoção de palavras irrelevantes. *Stopwords* são palavras tão comuns num idioma (como, por exemplo na língua portuguesa, “e”, “da” ou “em”) que não contêm muita informação. Assim, é comum remover estas palavras antes de uma análise mais aprofundada [40]. Para além da remoção de *stopwords*, por vezes também se removem termos que aparecem com pouca frequência, pois, teoricamente, se o objetivo é encontrar padrões de utilização de termos nos documentos, os termos usados com pouca frequência não contribuirão com muita informação sobre similaridade de documentos. Por outro lado, esta remoção pode reduzir bastante o tamanho do vocabulário, o que pode tornar mais rápido o processo de análise seguinte [65].

Em muitos casos pode ser vantajoso conhecer os sinónimos para um determinado termo, pois podem-se identificar palavras distintas com o mesmo significado. Este processo pode ser considerado como um tipo de análise semântica de nível baixo [40].

De referir que os estudos recentes sugerem que não só os procedimentos de pré-processamento que se aplicam, mas também a sua ordem, influenciam significativamente os resultados das técnicas de análise de texto subsequentes [65].

- **Conversão de dados**

Na conversão de dados, os dados textuais são convertidos para um formato que facilite o seu processamento por algoritmos de *text mining* como, por exemplo, o formato de vetores, sequências ou tabelas [42].

- ***Text mining***

De modo a dar resposta ao problema definido inicialmente, neste passo é definida e aplicada uma metodologia, ou mais do que uma, de *text mining* com o objetivo de encontrar padrões de dados, que serão avaliados no passo seguinte [64].

Existem várias técnicas de *text mining* e com o aumento dos dados textuais disponíveis e avanço computacional, mais técnicas ou variações das já existentes vão sendo desenvolvidas. De seguida serão abordadas algumas das técnicas mais utilizadas e referidas na literatura, sendo na secção seguinte abordada com mais detalhe a metodologia utilizada na ferramenta desenvolvida - *topic modeling* com utilização do modelo *Latent Dirichlet Allocation*.

Um dos métodos de análise mais simples de *text mining* é a avaliação baseada na contagem. Os termos com as frequências de ocorrência mais altas num texto são classificados como importantes. Apesar de sua simplicidade, esta abordagem é amplamente utilizada em *text mining* [70], pois é de fácil interpretação e é pouco dispendiosa computacionalmente [40].

Text classification (em português “classificação de texto” [41]) é a atribuição de documentos de texto a categorias predefinidas de acordo com o seu conteúdo. Ao classificar um documento a única informação utilizada é o próprio conteúdo do documento. Um grande número de técnicas de seleção de características e *machine learning* têm sido aplicadas à classificação de texto. As abordagens típicas extraem características de cada documento e usam os vetores de características como entrada para um esquema que aprende a classificar os documentos. Para cada categoria é construído um modelo que utiliza palavras como características e frequências de ocorrência de palavras como valores das características. Os documentos de cada categoria são “exemplos positivos” e os demais documentos são “exemplos negativos”. O modelo prevê se uma categoria é ou não atribuída a um novo documento com base nas palavras contidas no documento e suas frequências [42]. A classificação automática de texto tem muitas aplicações práticas, que incluem a indexação para recuperação de documentos, extração automática de metadados, desambiguação de sentido de palavras pela deteção dos tópicos que um documento abrange, organização e manutenção de grandes catálogos de recursos da *web*, encaminhamento automático de *e-mails* para os departamentos apropriados ou deteção de *spam*, ou

ainda previsão de movimentos de ações com base em experiências anteriores através de recolha de notícias que apareceram antes de um aumento ou queda nos preços das ações [42, 44].

Text clustering (em português “agrupamento/ aglomeração/ classificação de texto” [41]) é um método não supervisionado, no qual não existem categorias ou classes predefinidas, procurando-se definir grupos de documentos semelhantes. Os métodos de agrupamento não exigem dados pré-classificados e os próprios algoritmos geralmente são muito mais intensivos em computação do que os métodos supervisionados. Normalmente, a qualidade do agrupamento é considerada melhor se os conteúdos dos documentos dentro de um grupo forem mais semelhantes e entre os grupos mais diferentes [42]. Como exemplo, considerando o suporte técnico de uma empresa de computadores, pode-se descobrir que o maior grupo de reclamações é devido a problemas de rede, através de documentos que indicam muitos pedidos de ajuda sobre esse problema [44].

Information Extraction (em português “extração de informação”) é o processo de preenchimento de *templates* a partir de linguagem natural como dados de entrada, uma das principais subáreas de *text mining*. Um *template* é uma estrutura composta por campos, que são preenchidos por partes individuais de informação estruturada. *Machine learning* tem sido aplicada na extração de informação com a procura de regras de correspondência de padrões que extraem informação para os campos do *template* [42]. Um exemplo de aplicação é a análise de documentos relativos a empresas com texto não estruturado e extrair, por exemplo, o volume de vendas e os códigos do setor. De notar que o atributo medido não terá uma posição fixa no texto e poderá não ser descrito da mesma forma em diferentes documentos [44].

Information retrieval (em português “recuperação de informação” [41]) é o processo de extração de padrões relevantes de acordo com um determinado conjunto de palavras ou frases. Por outras palavras, considerando um conjunto de documentos, através de “pistas” sobre os documentos que se pretendem recuperar do conjunto, obtêm-se os documentos correspondentes como respostas à procura. A recuperação de informações é o tópico mais comumente associado a documentos *online*, na medida em que o motor de procura é uma das funcionalidades fundamentais de utilização de Internet [71, 44].

Text summarization (em português “resumo de texto”) é um método para produzir uma representação condensada do texto de entrada. Podem ser feitas distinções entre diferentes tipos de resumos: um extrato onde se selecionam frases-chave dispersas por todo o documento ou um resumo que contém material que não está presente no texto de entrada, ou pelo menos é expresso de uma maneira diferente. O objetivo principal de um resumo indicativo é sugerir

o conteúdo do documento sem revelar detalhes do mesmo, podendo servir para atrair a leitura do documento completo. Exemplos deste tipo de resumo são as capas de livros e *trailers* de filmes. Um resumo informativo tem como objetivo representar o documento original. Assim, deve conter todas as informações pertinentes e necessárias para transmitir as informações principais e omitir as informações auxiliares. Outra distinção é entre um resumo genérico, destinado a um público amplo, e um focado no tópico, adaptado às necessidades de um determinado grupo. Os resumos também podem ser produzidos a partir de um único documento ou a partir de vários documentos, envolvendo múltiplas fontes de informação que se sobrepõem e se complementam, sendo por vezes contraditórias. Logo, para além de identificação e trabalho de redundância entre documentos, também é necessário identificar o que é novo e garantir que o resumo final seja coerente e completo [42].

Opinion mining, também conhecida como análise de sentimentos, refere-se ao processo de análise de dados de opinião. Os problemas típicos de análise incluem a classificação de subjetividade/objetividade com a identificação se os dados textuais contêm a opinião de utilizadores; a classificação de sentimentos com a previsão da polaridade do sentimento de utilizadores (por exemplo, positivo ou negativo); a sumarização de opinião com o fornecimento de uma representação condensada de conjunto de dados textuais de opinião de utilizadores; e *anti-spam* de opinião com a deteção se os dados de opinião são escritos por remetentes de *spam* [42].

- **Interpretação/Avaliação**

No passo de interpretação/avaliação, os padrões extraídos através de metodologias de *text mining* são interpretados e avaliados e os considerados relevantes são emitidos como conhecimento final [42]. A contribuição de especialistas das áreas em estudo para a interpretação, desempenha um papel importante no processo [69].

No final de todos os passos, o conhecimento obtido poderá ser utilizado diretamente, incorporado noutra análise para ações adicionais, ou simplesmente documentado e apresentado às partes interessadas. Este processo também inclui a verificação e resolução de potenciais conflitos com conhecimento previamente reconhecido (ou extraído) [64].

De referir que apesar do passo relativo à definição e implementação de *text mining* ser o mais abordado na literatura, os restantes passos são tão importantes (e por vezes mais) para o sucesso da aplicação de todo o processo na prática [64].

3.4 *Topic modeling com utilização do modelo Latent Dirichlet Allocation*

A ausência de uma metodologia genérica para realizar uma análise de dados textuais tornou-se num grande desafio na investigação na área de *text mining*, na medida em que os modelos de *text mining* diferem para cada caso, pois cada área possui um conjunto de palavras específicas com significados semânticos diferentes. Uma das técnicas mais preponderante e amplamente utilizada para *text mining* é *topic modeling* [69].

Topic modeling é uma abordagem estatística para agrupar documentos de texto, assumindo que cada documento é uma função de variáveis latentes intituladas tópicos. Na modelação dos tópicos, um tópico é composto por uma lista de palavras geradas por meio de métodos estatísticos apropriados [72]. Um texto pode ser um capítulo de livro, uma publicação de um *blog*, um *e-mail*, respostas a perguntas abertas de um questionário e qualquer tipo de texto não estruturado. De notar que a utilização desta abordagem não tem como objetivo perceber a semântica das palavras dos documentos de texto. *Topic modeling* geralmente é considerado como um método "não supervisionado" porque envolve um processo de inferência em vez de assumir o conteúdo dos tópicos em consideração e tem sido utilizado numa grande variedade de áreas [69].

Topic modeling com utilização do modelo Latent Dirichlet Allocation (LDA), introduzido pela primeira vez por Blei, Ng e Jordan em 2003 [72], é uma técnica de análise computacional que pode ser usada para investigar a estrutura temática latente de uma coleção de dados textuais, dentro do paradigma estatístico Bayesiano. O algoritmo LDA é o principal e mais utilizado na modelação de tópicos [69, 73] e combina uma abordagem indutiva com medidas estatísticas, o que o torna particularmente adequado para análises exploratórias e descritivas [74].

Tendo por base [72], a palavra (w) é a unidade básica de dados discretos, definida como um item de um vocabulário; o documento é uma sequência de N palavras denotadas por $\mathbf{w} = (w_1, w_2, \dots, w_N)$, onde w_n é a palavra na posição n da sequência; o corpus é a coleção de M documentos denotados por $D = \mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M$. O algoritmo LDA é um modelo generativo probabilístico de tópicos, cuja ideia básica é que um documento é composto por uma mistura aleatória de tópicos latentes e cada tópico é caracterizado por uma distribuição que define a probabilidade de cada palavra aparecer num determinado tópico.

Na Figura 3.6 é apresentada uma representação gráfica do modelo LDA, onde se pode observar que é um modelo hierárquico de três níveis (*corpus*-documento-palavra). LDA assume que cada documento (M), que é composto por um número de palavras (N), pode ser representado como uma distribuição de probabilidade de

Dirichlet de tópicos latentes. β e α são os parâmetros *a priori* de *Dirichlet*, w representa a palavra observada, θ é a distribuição de tópicos para cada documento e z representa a atribuição de uma palavra a um determinado tópico. *Topic modeling* com LDA é utilizada para identificar os tópicos dos documentos. No processo generativo, o algoritmo assume que uma palavra pertence a um tópico e que um documento pertence a pelo menos um tópico [69].

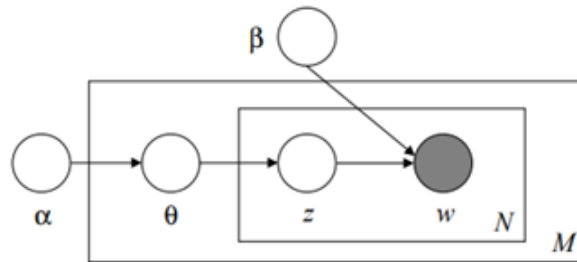


Figura 3.6: Representação gráfica do modelo LDA [72].

Para definir o modelo é necessário determinar o número de tópicos e os parâmetros *a priori*. O objetivo deste processo é encontrar a configuração de parâmetros que conduza ao modelo mais apropriado, considerando os dados e objetivo da investigação. A definição do número de tópicos pode ajustar o nível de pormenor do modelo. Geralmente, quanto maior o número de tópicos, mais específicos e restritos são os tópicos resultantes. No entanto, considerar um número elevado de tópicos pode resultar em entidades semelhantes que não são distinguidas de forma significativa. Ao mesmo tempo, um número reduzido de tópicos pode levar a entidades muito amplas, combinando diferentes aspetos que devem ser separados. Para determinar um número adequado de tópicos, usualmente executam-se vários modelos candidatos com números variados de tópicos e posteriormente os modelos resultantes são comparados tendo em consideração diferenças significativas e interpretabilidade. A escolha dos parâmetros *a priori*, α e β , não é muito abordada, sendo geralmente aplicados valores padrão ou procedimentos de otimização definidos nos *softwares* utilizados para análise de dados [74].

Relativamente à interpretação e validade dos tópicos resultantes, a abordagem mais direta é a análise das palavras com as maiores probabilidades para cada tópico, tentando-se encontrar um rótulo que descreva o conteúdo do mesmo. Muitas vezes, também é utilizada a leitura de uma amostra de documentos com elevadas proporções do respetivo tópico. Estas estratégias são aplicadas para garantir a validade semântica dos tópicos. Também são utilizadas métricas de diagnóstico quantitativas, como por exemplo *topic coherence* ou medidas de informação mútua. Enquanto que a coerência do tópico mede a frequência com que as principais palavras de um tópico ocorrem simultaneamente, a informação mútua pretende identificar qual das principais palavras contribui com a informação mais significativa para um determinado

tópico. Para verificar se os tópicos são suficientemente distintos entre si ou para encontrar padrões de semântica entre os tópicos pode ser utilizado o agrupamento hierárquico. De referir também estratégias de validação externa, que podem incluir avaliações de especialistas [74].

Na Figura 3.7 é apresentada uma representação da aplicação da metodologia de *topic modeling* com utilização do algoritmo *Latent Dirichlet Allocation* (adaptado de [74]).

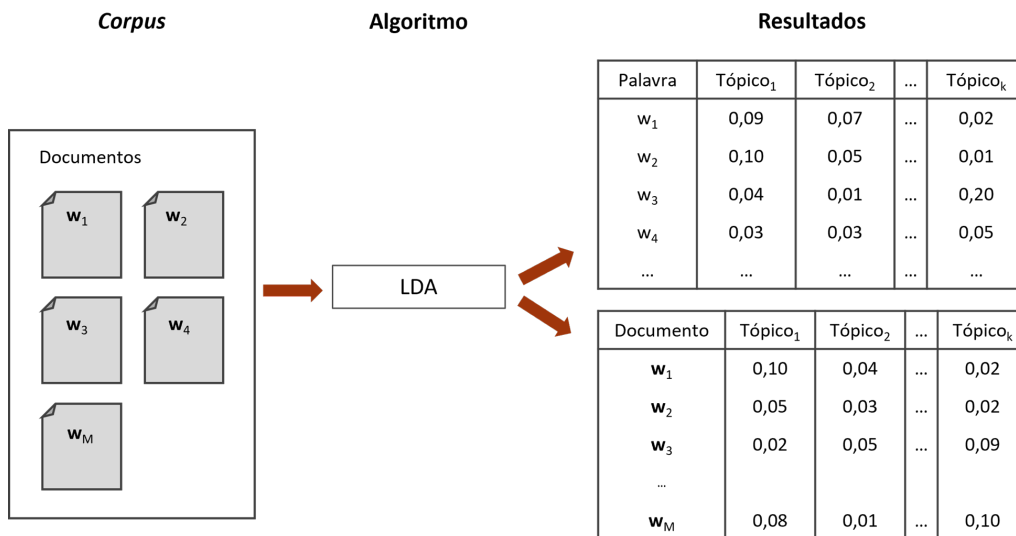


Figura 3.7: Aplicação da metodologia de *topic modeling* com utilização do algoritmo LDA.

Capítulo 4

Ferramenta desenvolvida

No contexto da transferência e valorização do conhecimento e assumindo que nem toda a investigação resulta em inovação, as unidades de investigação e as empresas precisam de alinhar os seus objetivos para que a sua relação seja profícua.

Neste capítulo são abordados todos os procedimentos realizados no desenvolvimento da ferramenta de apoio ao processo de transferência e valorização do conhecimento, proposta nesta dissertação, através da utilização de informação que as unidades de investigação e as empresas disponibilizam nos seus *websites* e com a aplicação de metodologias de *text mining*. De notar que neste processo, o contributo de especialistas desempenha um papel importante e assim, as interfaces de visualização, são fundamentais para uma análise eficiente.

Em primeiro lugar será descrito o objetivo geral da ferramenta desenvolvida. Em seguida será descrito e explicado os procedimentos mais relevantes para o desenvolvimento da ferramenta, nomeadamente as tecnologias utilizadas, os passos da análise de *text mining* e o ambiente gráfico desenvolvido. Para uma melhor perceção do resultado obtido, será apresentada a ferramenta e explicada a sua utilização, assim como os resultados obtidos e discussão.

4.1 Objetivo

O objetivo geral da ferramenta é dar apoio aos utilizadores na interface entre as unidades de investigação e as empresas. Os utilizadores poderão ser pessoas envolvidas em estruturas de apoio à transferência de conhecimento, responsáveis de unidades de

investigação ou empresas, ou de uma forma geral qualquer pessoa que pretenda perceber com que empresas determinada unidade de investigação apresenta uma maior relação, de acordo com a informação disponibilizada nos *websites*, e vice-versa, ou seja, com que unidades de investigação determinada empresa apresenta uma maior relação. Assim, a ferramenta a desenvolver pretende apoiar a tomada de decisão, tendo em consideração os seguintes pontos:

- Recolha da informação dos *websites* das unidades de investigação e das empresas;
- Através de metodologia de *text mining*, sumário da informação (termos com as frequências de ocorrência mais altas);
- Através de metodologia de *text mining*, identificação de relações (*topic modeling*);
- Interface *user friendly*;
- Plataforma dinâmica, com possíveis atualizações.

4.2 Desenvolvimento

O desenvolvimento da ferramenta englobou vários procedimentos, nomeadamente, a configuração da estrutura, a construção da base de dados, a definição da estrutura de arquivos e o *workflow*. Foi implementada a aplicação das metodologias de *text mining* e foram também construídas as necessárias páginas *web* para a interface com o utilizador, sendo para isso necessária a utilização de diferentes linguagens, *frameworks* e bibliotecas.

4.2.1 Tecnologias

A linguagem de programação base foi Python [75], utilizada nos vários passos da análise de *text mining* e na construção da aplicação *web*. Foram utilizadas várias bibliotecas, sendo as mais específicas e utilizadas para a recolha e análise de dados textuais: (i) “Beautiful Soup”, uma biblioteca baseada no mecanismo de análise HTML/XML e utilizada para extrair, analisar e editar informações de *websites* [76, 77]; (ii) “spaCy”, uma biblioteca gratuita e de código aberto para processamento avançado de linguagem natural (NLP) em Python, que foi desenvolvida para criar aplicações que processem e “compreendam” grandes volumes de texto e pode ser utilizada para construir sistemas de extração de informações ou de compreensão de linguagem natural [78]. Esta biblioteca vem com uma série de segmentação de instruções pré-treinadas e suporta *tokenization* e treino para mais de 60 idiomas. Possui ainda modelos de rede neuronal para marcação, análise, reconhecimento de

entidade nomeada, classificação de texto, entre outros [79]; (iii) “wordcloud”, utilizada para aplicar a técnica de visualização de dados, que representa dados textuais em que o tamanho de cada palavra indica a sua frequência ou importância [80]; (iv) “pandas”, uma biblioteca que fornece estruturas de dados rápidas, flexíveis e expressivas, projetadas para tornar o trabalho com dados “relacionais” ou “rotulados” fácil e intuitivo [81]; e (v) “Scikit-learn”, que fornece implementações de última geração de muitos algoritmos de *machine learning*, mantendo um interface fácil de usar e totalmente integrado com a linguagem Python [82].

Na criação da aplicação *web* foram ainda utilizadas diferentes linguagens e *frameworks*: (i) *Hypertext Markup Language* (HTML), uma linguagem de marcação *standard* utilizada para descrever a estrutura de páginas *web*. Esta linguagem permite publicar documentos *on-line* com, por exemplo, títulos, texto, tabelas ou imagens [83]; (ii) “JavaScript”, uma linguagem de programação orientada para páginas *web* no lado do cliente, que proporciona uma melhor apresentação e interatividade das páginas. Trata-se de uma linguagem orientada a objetos e como não tem que utilizar os recursos de um servidor, aumenta a rapidez do processo [84]; (iii) *Cascading Style Sheets* (CSS), uma linguagem para descrever a apresentação de páginas *web*, incluindo cores, *layouts* e fontes. Permite adaptar a apresentação a diferentes tipos de dispositivo de visualização. O CSS é independente do HTML e pode ser usado com qualquer linguagem de marcação baseada em *Extensible Markup Language* (XML). A separação de HTML e CSS facilita a manutenção de *websites*, a partilha de *style sheets* entre páginas e a adaptação de páginas para diferentes ambientes [83]; (iv) “Bootstrap”, uma *framework front-end open source* que visa facilitar e agilizar o processo de desenvolvimento *web*. Esta *framework* inclui modelos de design baseados em HTML e CSS, além de *plug-ins* JavaScript opcionais. O Bootstrap oferece também a capacidade de criar facilmente *designs* responsivos, sendo um dos mais utilizados nesta área [85]; (v) *Structured Query Language* (SQL), uma linguagem padrão para aceder e gerir bases de dados relacionais [86]; e (vi) “flask”, uma *framework* que foi projetada para facilitar a criação de aplicações *web*. As aplicações desenvolvidas com recurso à biblioteca “flask” têm a capacidade de escalar para versões mais complexas. A “flask” oferece sugestões, mas não impõe nenhuma dependência ou *layout* podendo ser integrada com outras bibliotecas [87].

4.2.2 Aplicação de *text mining*

Como descrito no Capítulo 2, o processo de *text mining* envolve um conjunto de passos, que foram seguidos e para os quais são apresentadas e explicadas as opções consideradas.

- **Seleção de dados**

O primeiro passo é a seleção de dados textuais apropriados, que serão analisados nos passos seguintes. Considerando o objetivo do estudo, pretende-se analisar a informação disponível nos *websites* das unidades de investigação e das empresas.

A opção de considerar as unidades de investigação avaliadas pela FCT e as empresas com pessoas afetas à investigação (informação já analisada no Capítulo 2), para representar as estruturas envolvidas no processo de transferência e valorização do conhecimento, foi a primeira decisão. Na impossibilidade de considerar mais estruturas, devido a limitações de informação disponível, recursos e tempo, optou-se por considerar as unidades de investigação avaliadas pela FCT pois são as estruturas onde se desenvolve a maior parte da investigação em Portugal e existe informação disponível sobre as mesmas. Apesar de apenas a 312 unidades ter sido atribuído financiamento, foram consideradas as 348 que participaram no processo de avaliação, pois pretende-se que a ferramenta seja dinâmica e que possa incorporar mais estruturas, estando deste modo já incluídas as unidades que neste processo não obtiveram financiamento, mas no próximo processo poderão vir a ter, e para as quais já existe informação disponível. Relativamente às empresas, utilizou-se a base de dados Sabi, por disponibilizar várias informações, nomeadamente o *website*, e dado o elevado número de empresas incluídas na base de dados, optou-se por considerar as 1 511 empresas com pessoas afetas à investigação, por se entender que estariam mais orientadas para a inovação, sendo mais provável identificar relações com as unidades de investigação.

A opção por utilizar a informação disponível nos *websites* deve-se essencialmente: (i) à facilidade de acesso e disponibilidade da informação, não sendo necessário a recolha através de contactos, que pode envolver uma grande logística, como se se considerar o exemplo em estudo onde estão envolvidas centenas de entidades, acrescentando ainda a dificuldade de harmonização da informação devido às várias fontes; (ii) à disponibilização de informação atual, pois os *websites* são atualizados ao longo tempo, o que torna a ferramenta dinâmica, com resultados que vão refletindo a evolução das unidades e empresas; (iii) e é a primeira vez que é analisada este tipo de informação com o propósito deste estudo.

Assim, tendo por base as 348 unidades e 1 511 empresas, foram retiradas as que não tinham um *website* disponível ou que não era possível obter a informação e ainda as que não tinham o *website* em língua inglesa. Dado que o objetivo era aplicar uma análise de dados textuais, era necessária a uniformização da língua. A opção pela língua inglesa deveu-se, por um lado,

ao facto da maioria da investigação ser feita em inglês, nomeadamente quando considerada a produção e divulgação científica e por outro, estas entidades apresentam na sua maioria o *website* em inglês ou em inglês e português (sendo recolhida a informação em inglês). De notar que a investigação realizada nas áreas de ciências sociais e humanidades, principalmente, é feita muitas vezes em português, devido às especificidades destas áreas, o que neste caso pode prejudicar os resultados obtidos pela ferramenta nestas áreas. Uma opção seria traduzir a informação dos *websites*, mas as soluções disponíveis, como por exemplo o “Google Translate API”, têm um custo associado.

Depois de eliminadas as unidades e empresas que não satisfazem estas condições, na ferramenta estão disponíveis 284 unidades de investigação (listadas no Anexo A) e 601 empresas (listadas no Anexo B).

Considerando as 284 unidades de investigação e o respetivo painel de avaliação da FCT (Tabela 4.1), a maior percentagem de unidades pertence às “ciências sociais” (81 unidades, 28,5%), seguindo-se as “artes e humanidades” (46 unidades, 16,2%), as “ciências da engenharia e tecnologias” (45 unidades, 15,8%), “ciências exatas” (35 unidades, 12,3%), “ciências da saúde” (32 unidades, 11,3%), “áreas temáticas” (24 unidades, 8,5%) e, por fim, “ciências naturais” (21 unidades, 7,4%).

Tabela 4.1: Número e percentagem das Unidades de Investigação por painel de avaliação FCT.

Painel de avaliação	N.º	%
Ciências naturais	21	7,4
Ciências Exatas	35	12,3
Ciências da engenharia e tecnologias	45	15,8
Ciências da saúde	32	11,3
Ciências sociais	81	28,5
Artes e Humanidades	46	16,2
Áreas Temáticas	24	8,5
Total	284	100

Relativamente às 601 empresas (Tabela 4.2), mais de metade das empresas pertence à secção “Indústrias transformadoras” (395 empresas, 65,7%), seguindo-se “Atividades de informação e de comunicação” (83 empresas, 13,8%), “Atividades de consultoria, científicas, técnicas e similares” (48 empresas, 8,0%) e “Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos” (38 empresas, 6,3%), tendo as restantes secções pouca representatividade.

Tabela 4.2: Número e percentagem das empresas por atividade económica.

Atividade económica	N.º	%
Agricultura, produção animal, caça, floresta e pesca	4	0,7
Indústrias extrativas	2	0,3
Indústrias transformadoras	395	65,7
Captação, tratamento e distribuição de água; saneamento, gestão de resíduos e despoluição	3	0,5
Construção	8	1,3
Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos	38	6,3
Transportes e armazenagem	5	0,8
Alojamento, restauração e similares	3	0,5
Atividades de informação e de comunicação	83	13,8
Atividades imobiliárias	2	0,3
Atividades de consultoria, científicas, técnicas e similares	48	8,0
Atividades administrativas e dos serviços de apoio	4	0,7
Atividades de saúde humana e apoio social	4	0,7
Outras atividades de serviços	2	0,3
Total	601	100

A perceção das áreas de atuação das unidades e das empresas é importante para a análise das relações, pois unidades e empresas de áreas semelhantes, poderão ser identificadas com maior facilidade. Considerando todas as secções da CAE-Rev.3, verifica-se que não existem empresas com atividade económica nas secções “Eletricidade, gás, vapor, água quente e fria e ar frio”, “Atividades financeiras e de seguros”, “Administração pública e defesa; segurança social obrigatória”, “Educação”, “Atividades artísticas, de espetáculos, desportivas e recreativas”, “Atividades das famílias empregadoras de pessoal doméstico e atividades de produção das famílias para uso próprio” e “Atividades dos organismos internacionais e outras instituições extraterritoriais”, podendo limitar as relações de unidades de investigação destas áreas.

O processo de recolha dos dados foi feito através da utilização de um *web crawler*, para recolher os *hiperlinks* dos *websites* das unidades de investigação e das empresas e posteriormente a informação constante nas respetivas páginas *web*. Devido, novamente, a limitações de recursos e tempo, optou-se por recolher o *hiperlink* da página inicial e todos os *hiperlinks* referenciados na mesma. Ao todo, foram identificadas e recolhida informação de 39 476 páginas *web* (15 520 das unidades e 23 956 das empresas), contemplando 184 292 934 palavras (77 922 138 das unidades e 106 370 796 das empresas). Aplicou-se ainda o limite de 3 000 000 de caracteres por unidade ou empresa.

- **Pré-processamento de dados**

O passo de pré-processamento de dados é um dos mais desafiadores e importantes para garantir a qualidade dos dados, tendo sido realizadas algumas etapas.

Como foi utilizado um procedimento de *web crawling* para recolher informação de páginas *web*, foi necessário remover o código de programação *web*, através da utilização da biblioteca “Beautiful Soup”.

Através da biblioteca “spaCy”, foi feita a verificação das frases que não estavam em inglês, sendo as mesmas removidas. Apesar de se ter optado por *websites* em inglês, alguns conteúdos não estavam nesta língua.

Seguidamente foi feita a *tokenization* e foram removidos os caracteres especiais, todas as letras foram convertidas para minúsculas e foram removidos também os espaços em branco extra, pontuação e números.

Procedeu-se ainda à remoção das *stopwords*, através da biblioteca “spaCy”.

Por fim, foi realizado um processo de lematização, para a normalização de palavras, ou seja, foi realizado o processo de agrupar as diferentes formas flexionadas de uma palavra para que possam ser analisadas como um único elemento, através da biblioteca “spaCy”.

No final do pré-processamento, obteve-se, para posterior análise, um total de 11 374 050 palavras (5 508 265 das unidades de investigação e 5 865 785 das empresas).

- **Conversão de dados**

De modo a facilitar o processamento dos dados nas análises seguintes, os mesmos foram guardados num formato csv e foram convertidos em *data frame*, com a utilização da biblioteca “pandas”.

- ***Text mining***

Como referido no Capítulo 3, existem várias metodologias de *text mining*, tendo sido aplicadas no desenvolvimento da ferramenta duas, sumário da informação baseado na contagem de palavras e *topic modeling* com utilização do modelo LDA.

Com o objetivo de ter uma perceção do que cada *website* das unidades de investigação e das empresas aborda, foi utilizada uma avaliação baseada na contagem, onde os termos com as frequências de ocorrência mais altas num texto são classificados como importantes, obtendo-se assim um sumário da informação. Para representar esta informação foram utilizadas as “nuvens de palavras”, com a utilização da biblioteca “wordcloud”.

As nuvens de palavras surgiram como um método de visualização de texto simples e visualmente atrativo. São utilizadas em vários contextos como um meio de fornecer uma visão geral, destilando o texto até às palavras que aparecem com maior frequência. Uma área de grande aplicação das nuvens de palavras é o sumário de texto. As nuvens de palavras geradas para um corpo de texto podem servir como ponto de partida para uma análise mais profunda [88].

Com o objetivo de analisar possíveis relações entre as unidades de investigação e as empresas foi utilizada a metodologia *topic modeling* com utilização do modelo LDA. Através desta abordagem, pretende-se agrupar documentos de texto, neste caso cada *website* corresponde a um documento, e assume-se que um documento é composto por uma mistura aleatória de tópicos latentes e cada tópico é caracterizado por uma distribuição que define a probabilidade de cada palavra aparecer num determinado tópico.

Uma grande vantagem deste método é que os tópicos são inferidos a partir dos dados em análise, sem a necessidade de qualquer conhecimento prévio. Como os tópicos não estão definidos inicialmente, nenhuma informação sobre os mesmos é diretamente observável nos dados. O algoritmo LDA resolve o problema, inferindo tópicos através de padrões recorrentes de ocorrência de palavras nos documentos [74].

Para definir o modelo é necessário determinar o número de tópicos e os parâmetros *a priori*. Para o número de tópicos foi considerado o número de áreas científicas utilizadas pela FCT, que é também o número de grandes áreas consideradas na revisão da classificação “*Fields of Science and Technology (FOS)*” no Manual de Frascati [89], ou seja, foi considerada a existência de 6 tópicos. A definição dos parâmetros *a priori* foi feita através dos procedimentos de otimização definidos na biblioteca “Scikit-learn”, que fornece a implementação de muitos algoritmos de *machine learning* de última geração, como referido anteriormente.

De modo a considerar apenas informação distintiva para a definição dos tópicos, foram definidas duas condições quanto à ocorrência das palavras: (i) as palavras não podem aparecer em mais de 90% dos documentos (neste caso, dos *websites*); e (ii) as palavras têm de aparecer em mais de 3 documentos. Deste modo pretende-se garantir, por um lado, que as palavras não são muito comuns, não diferenciando assim o conteúdo dos documentos e, por outro lado, que as palavras não são muito específicas, dificultando a relação com outros documentos.

Para uma melhor perceção da metodologia implementada e dos resultados esperados, na Figura 4.1 encontra-se a adaptação da Figura 3.7 para esta

análise, considerando como exemplo analisar uma determinada Unidade de Investigação.

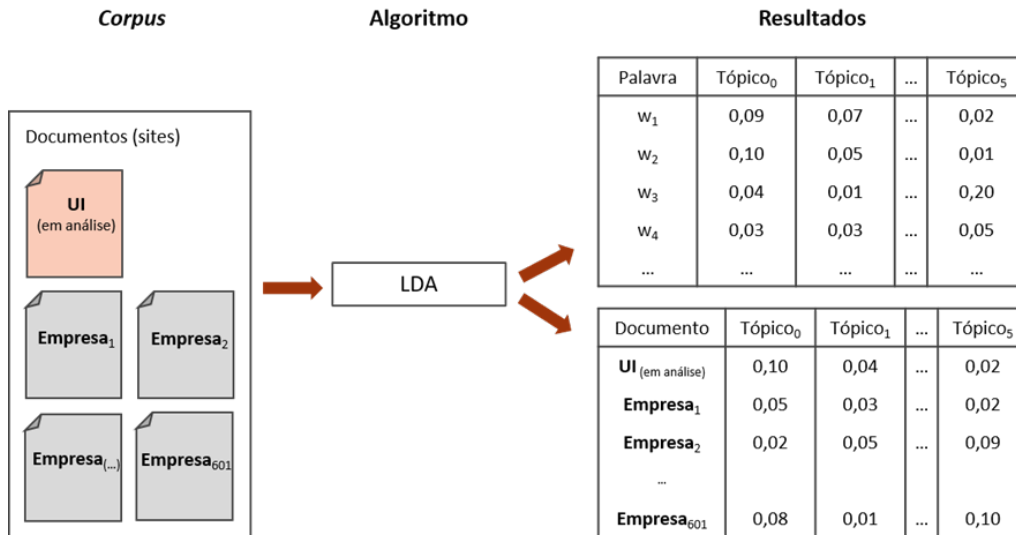


Figura 4.1: Aplicação da metodologia de *topic modeling* com utilização do algoritmo LDA aplicado à ferramenta.

Considerando uma unidade que se pretende analisar e perceber com que empresas poderá ter uma maior relação, na definição do modelo será considerada a informação do *website* dessa unidade e dos *websites* de todas as empresas, ou seja, o modelo definido terá por base 602 documentos (ou *websites*), 1 documento da unidade e 601 documentos das empresas.

Caso se considere uma empresa para analisar, pretende-se perceber com que unidades de investigação pode ter uma maior relação, na definição do modelo será considerada a informação do *website* dessa empresa e dos *websites* de todas as unidades, ou seja, o modelo definido terá por base 285 documentos (ou *websites*), 1 documento da empresa e 284 documentos das unidades.

No desenvolvimento da ferramenta, para o conjunto das 284 unidades de investigação e 601 empresas, foram definidos 885 modelos, um para cada unidade e cada empresa. Para cada modelo é identificada a probabilidade da Unidade de Investigação em análise e das empresas (ou da empresa em análise e das unidades) pertencer a cada um dos 6 tópicos e as palavras com maior probabilidade de pertencer a cada tópico.

- **Interpretação/Avaliação**

No passo de interpretação/avaliação, os padrões extraídos através de metodologias de *text mining* são interpretados e avaliados e como já referido, a contribuição de especialistas das áreas em estudo para a interpretação, desempenha um papel importante no processo. No caso da metodologia aplicada,

esta avaliação passa pela interpretação e validação dos tópicos resultantes, sendo a abordagem mais direta a análise das palavras com as maiores probabilidades para cada tópico, tentando-se encontrar um rótulo que descreva o conteúdo do mesmo.

No presente estudo, não se pretende rotular os tópicos para que representem, de certa forma, subconjuntos de documentos. O que se pretende é utilizar os mesmos para identificar grupos de empresas relacionados com uma determinada Unidade de Investigação (ou grupos de unidades relacionados com uma determinada empresa), através dos referidos padrões recorrentes, considerando a ocorrência de palavras nos documentos. Ou seja, as probabilidades de pertencer aos diferentes tópicos, darão indicação de uma possível maior ou menor relação entre as unidades de investigação e as empresas, não sendo para isso necessário rotular e interpretar de forma clara os tópicos obtidos. Até porque neste estudo são definidos 885 modelos com os respetivos 6 tópicos, sendo todos eles diferentes.

De qualquer forma, a contribuição de especialistas será sempre um benefício para interpretar os resultados obtidos. Aliás, a ferramenta pretende ser de apoio e não de disponibilização de resultados finais.

4.2.3 Ambiente gráfico

A disponibilização da ferramenta foi feita através de uma aplicação *web*, pois pretende-se que a ferramenta seja de fácil acesso, *user friendly* e com informação e resultados atuais. O ambiente gráfico foi desenvolvido com recurso à *framework* “Flask” e a uma base de dados relacional conectada a uma estrutura de ficheiros robusta.

A estrutura da base de dados (Figura 4.2) desenvolvida para suportar a ferramenta encontra-se dividida em seis tabelas: “empresas”, “ui”, “empconnui1”, “empconnui2”, “uiconnemp1” e “uiconnemp2”. Foram utilizadas relações do tipo “1 para 1” e “1 para n” [90]. Em ligações “1 para 1”, a uma entrada na tabela 1 corresponde uma entrada na tabela 2, verificando-se o mesmo no sentido contrário. Em ligações “1 para n”, a uma entrada na tabela 1 podem corresponder várias entradas na tabela 2, enquanto que a uma entrada na tabela 2 corresponde apenas uma entrada na tabela 1. A tabela “ui” contém informação base acerca das unidades de investigação em análise, entre as informações disponibilizadas encontra-se o código FCT, o nome, o acrónimo, a instituição de gestão, a avaliação e financiamento da FCT, o painel de avaliação, *website*, entre outras. A tabela “empresas”, por sua vez, contém informação base acerca das empresas e entre os dados disponibilizados encontra-se o nome, Número de Identificação Fiscal (NIF), localização, *website*, CAE-Rev.3, volume de negócios, número de pessoas alocados à investigação, entre outras. As tabelas “empconnui1” e “uiconnemp1” disponibilizam a identificação do tópico com

maior probabilidade para cada empresa ou unidade, consoante a tabela em análise, a distribuição de probabilidades de pertencer a cada tópico e as palavras com maior probabilidade de pertencer a cada tópico. A tabela “uiconnemp2” disponibiliza para cada unidade em análise, a distribuição de probabilidades das empresas pertencerem a cada tópico. A tabela “empconnui2” disponibiliza para cada empresa em análise, a distribuição de probabilidades das unidades pertencerem a cada tópico.

Relativamente à estrutura de ficheiros (Figura 4.3), considerando a estrutura da *framework* “flask”, foram criados os diretórios “static” e “templates” e os ficheiros “app.py”, “dboperations.py” e “requirements.txt”. Dentro do diretório “static” existe o diretório “image”, para guardar as imagens, que são separadas ainda em dois diretórios (“general” e “wordcloud”) e o diretório “styles”, onde estão definidos os estilos do *website* em CSS. No diretório “templates” estão definidos todos os modelos de páginas utilizadas na aplicação. O ficheiro “app.py” é responsável por todo o roteamento da aplicação, o ficheiro “dboperations.py” contém todas as ligações à base de dados e o ficheiro “requirements.txt” guarda os requisitos para a criação do ambiente.

4.3 Resultados e discussão

A análise dos resultados obtidos deve ter em consideração que a informação utilizada na aplicação de metodologias de *text mining* foi extraída dos *websites* das unidades de investigação e das empresas, tendo a mesma sido reduzida no passo da seleção de dados. Assim e dada a complexidade de todo o trabalho realizado pelas unidades e empresas e ainda os diferentes objetivos que pretendem atingir com a informação que disponibilizam através dos *websites*, os resultados obtidos na utilização da ferramenta poderão não refletir as dimensões que de certa forma os utilizadores podem esperar, pela experiência e conhecimento prévio das entidades, por exemplo. Por outro lado, esta divergência poderá resultar em colaborações menos esperadas e que de outra forma não seriam analisadas.

Assim, a ferramenta deverá ser utilizada como apoio na tomada de decisão, devendo os resultados obtidos ser analisados de forma crítica e em complemento à experiência de especialistas.

Dada a complexidade dos resultados, tanto em termos de quantidade (é apresentada informação para 284 unidades de investigação e 601 empresas, com a definição de 885 modelos) como de diversidade (as unidades de investigação e empresas são entidades distintas e abrangem diversas áreas), não seria viável, em termos de tempo e conhecimentos, uma análise exaustiva de todos os resultados. Neste subcapítulo pretende-se apresentar a ferramenta desenvolvida e suas funcionalidades e exemplos concretos de unidades de investigação e empresas, que apresentam resultados diferentes.

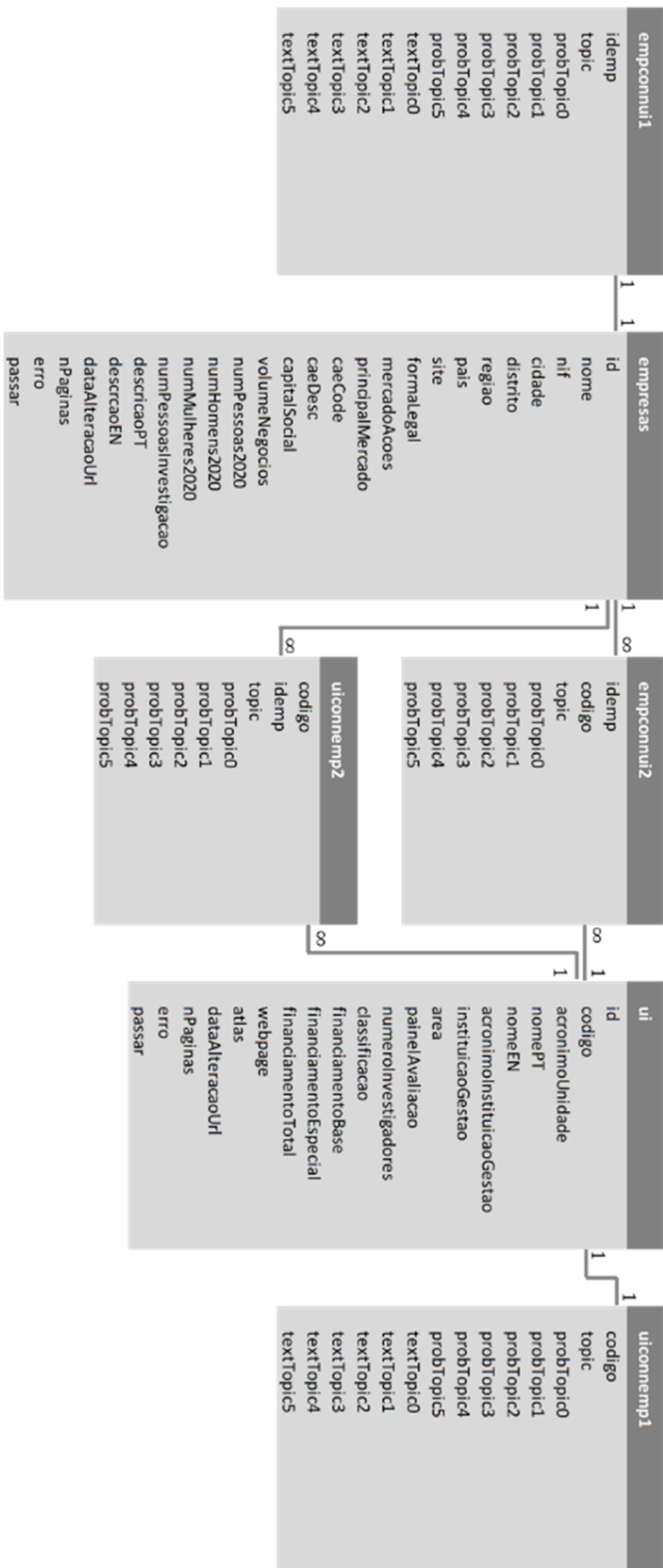


Figura 4.2: Estrutura da base de dados da ferramenta desenvolvida.

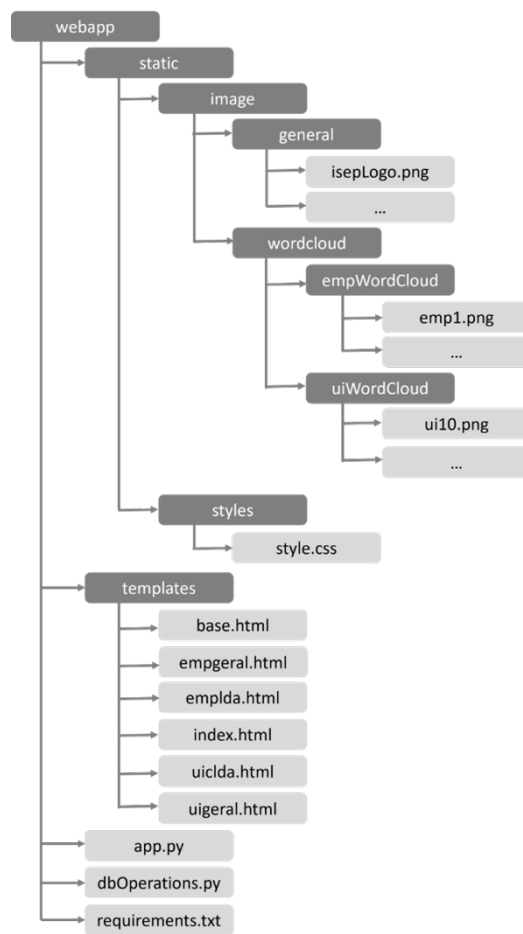


Figura 4.3: Estrutura de diretórios e arquivos da ferramenta desenvolvida.

4.3.1 Ambiente gráfico e funcionalidades

A ferramenta desenvolvida resultou numa página *web* com um menu inicial, as páginas “Home”, “Unidades” e “Empresas”.

Na página “Home” é apresentada uma descrição geral da ferramenta e indicações sobre a sua utilização (Figura 4.4).



Figura 4.4: Ferramenta: *Home Page*.

No separador “Unidades” encontram-se listadas as unidades de investigação consideradas no desenvolvimento da ferramenta, sendo possível filtrar pelo nome, e deverá ser selecionada a Unidade de Investigação que se pretende analisar (Figura 4.5).

Para cada Unidade de Investigação são apresentadas informações descritivas (“Acrónimo”, “Código FCT”, “Nome PT”, “Nome EN”, “Acrónimo da Instituição de Gestão Principal” e “Nome da Instituição de Gestão Principal”), informações relativas ao processo de avaliação FCT (“Área do Atlas”, “Painel de Avaliação”, “Avaliação FCT”, “Financiamento” e “N.º de Investigadores”) e ainda o “Website analisado” e a “Data da informação” (Figura 4.6). É também apresentada uma nuvem de palavras da informação presente no *website* da unidade, onde as palavras maiores aparecem com maior frequência.

Os resultados da aplicação de *topic modeling* são apresentados através de um gráfico com a distribuição da probabilidade da unidade pertencer a cada um dos seis tópicos e as vinte palavras com maior probabilidade de pertencer ao tópico selecionado (Figura 4.7). Por defeito, aparece o tópico com a maior probabilidade, sendo possível selecionar cada um dos restantes. Aparecem também as 20 empresas com maior probabilidade de pertencer ao tópico, por ordem decrescente.

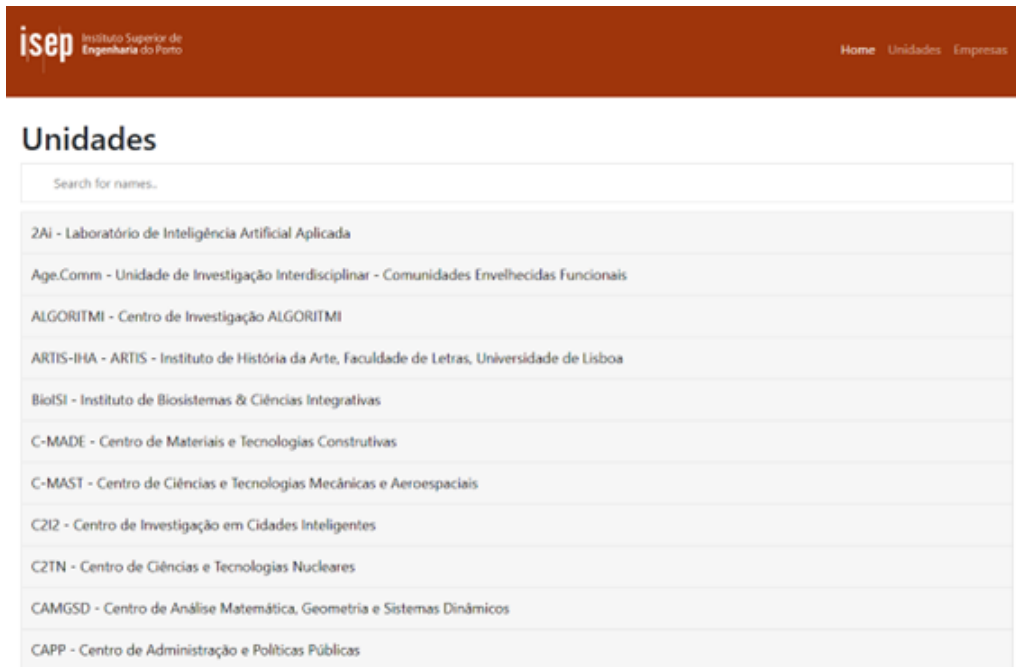


Figura 4.5: Ferramenta: Unidades de Investigação.



Figura 4.6: Ferramenta: Unidade de Investigação (descrição e nuvem de palavras).

De referir que o número do tópico é meramente indicativo, não tendo qualquer relação a ordem em que os mesmos aparecem, nem são comparáveis com os resultados das outras unidades ou empresas.

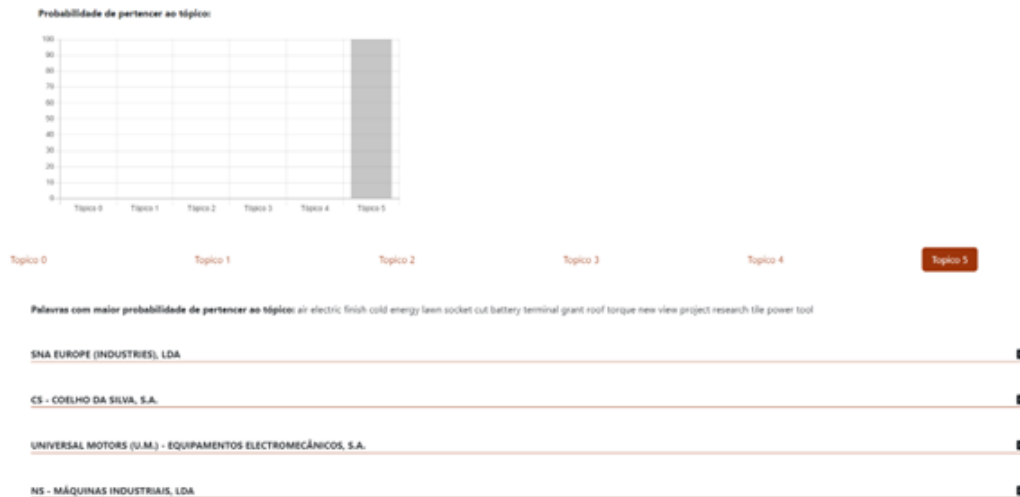


Figura 4.7: Ferramenta: Unidade de Investigação (resultados de *topic modeling*).

É ainda possível expandir a informação das empresas, através do clique no sinal “+” (Figura 4.7) e são apresentadas informações descritivas, nuvem de palavras e a probabilidade de pertencer aos diferentes tópicos (Figura 4.8 e Figura 4.9).



Figura 4.8: Ferramenta: Unidade de Investigação (informação da empresa).

O separador “Empresas” é semelhante ao separador “Unidades” no que diz respeito à informação disponibilizada e procedimentos. Neste caso, encontram-se listadas as empresas consideradas no desenvolvimento da ferramenta, sendo possível



Figura 4.9: Ferramenta: Unidade de Investigação (informação da empresa - continuação).

filtrar pelo nome, e deverá ser selecionada a empresa que se pretende analisar (Figura 4.10).

Para cada empresa são apresentadas informações descritivas (“Nome”, “NIF”, “Cidade”, “Distrito” e “Região”), informações relativas à atividade económica (“CAE”, “Descrição do CAE”, “Capital Social”, “Volume de Negócios”, “Número de Pessoas” e “N.º de Pessoas em Investigação”), descrição em português e em inglês e ainda o “Website analisado” e a “Data da informação” (Figura 4.11 e Figura 4.12). É também apresentada uma nuvem de palavras da informação presente no *website* da empresa.

Da mesma forma, os resultados da aplicação de *topic modeling* são apresentados através de um gráfico com a distribuição da probabilidade da empresa pertencer a cada um dos seis tópicos e as vinte palavras com maior probabilidade de pertencer ao tópico selecionado. Por defeito, aparece o tópico com a maior probabilidade, sendo possível selecionar cada um dos restantes. Aparecem também as 20 unidades com maior probabilidade de pertencer ao tópico, por ordem decrescente.

Por fim, é possível expandir a informação das unidades, através do clique no sinal “+” (Figura 4.12), e são apresentadas também informações descritivas e a probabilidade de pertencer aos diferentes tópicos (Figura 4.13 e Figura 4.14).

4.3.2 Nuvem de palavras

As nuvens de palavras apresentadas para cada Unidade de Investigação e empresa têm como objetivo fornecer uma visão geral da informação presente nos *websites*. As palavras com tamanho de letra maior representam as palavras utilizadas com mais frequência no texto presentes nos *websites*.



Figura 4.12: Ferramenta: empresa (resultados de *topic modeling*).

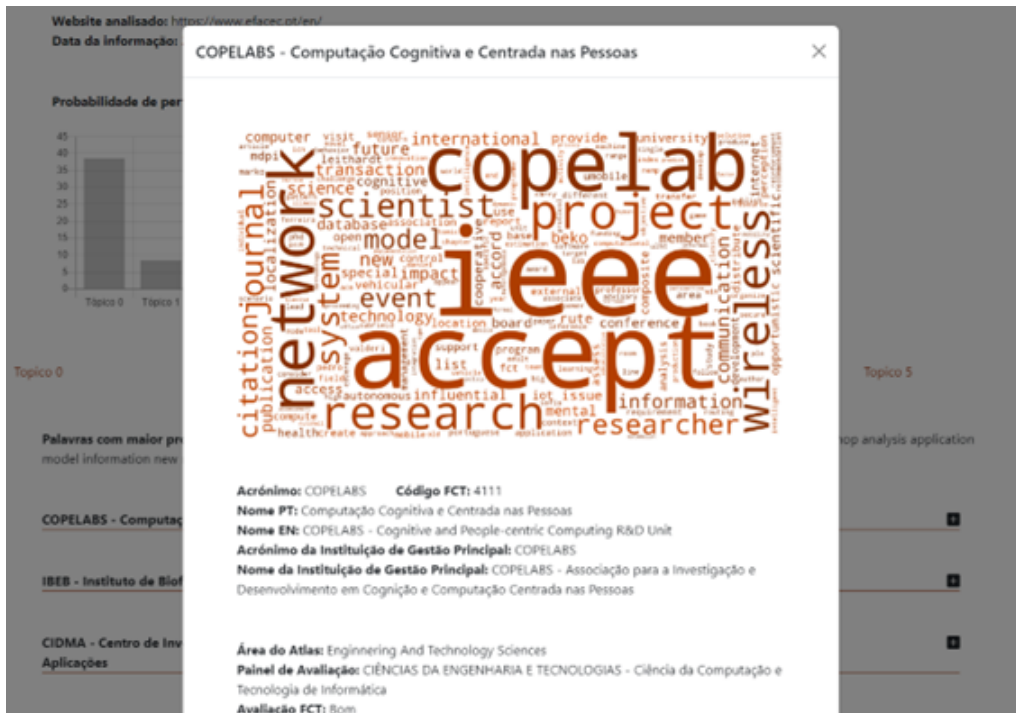


Figura 4.13: Ferramenta: empresa (informação da Unidade de Investigação).



Figura 4.14: Ferramenta: empresa (informação da Unidade de Investigação - continuação).

A análise das nuvens de palavras permite, por um lado, ter uma melhor perceção das atividades realizadas, áreas de atuação, serviços e produtos disponibilizados, por exemplo, caso a informação disponibilizada nos *websites* o consiga refletir. Por outro lado, esta análise é importante para uma melhor compreensão dos resultados obtidos pela análise de *topic modeling*. Da mesma forma que é difícil ter uma boa perceção da dimensão e características das unidades e empresas, caso a informação dos *websites* não o reflita, também a atribuição a um tópico com base nas palavras do *website* poderá não ser a mais adequada.

Para uma melhor perceção de toda a informação contida nos *websites* das unidades de investigação e que foi analisada, utilizou-se também uma nuvem de palavras (Figura 4.15). As palavras que mais se destacam são palavras relacionadas com a atividade de investigação, tais como “research”, “project”, “international”, “university” e “publication”, como era esperado. Também são visíveis, com menor destaque, palavras como “management”, “art”, “economic”, “social”, “technology”, “language” e “health”, por exemplo, que poderão ser mais importantes na diferenciação dos tópicos.

Por sua vez, na nuvem de palavras de toda a informação considerada das empresas (Figura 4.16), o destaque vai para palavras relacionadas com os serviços e produtos prestados e características que importa destacar, como “new”, “use”, “solution”, “product”, “service”, “privacy” e “information”, por exemplo. Com menos destaque, mas que poderão ser importantes para a identificação de tópicos, surgem

Um exemplo de Unidade de Investigação em que as palavras em destaque da nuvem de palavras não são tão diferenciadoras é a unidade ALGORITMI - Centro de Investigação ALGORITMI da Universidade do Minho, da área “Ciências da Engenharia e Tecnologias – Ciência da Computação e Tecnologia de Informática”. A nuvem de palavras (Figura 4.20) destaca, por exemplo, “project”, “research”, “member”, “publication”, “program”, “phd” e “leadership”.



Figura 4.20: Nuvem de palavras da Unidade de Investigação ALGORITMI.

Relativamente às empresas e utilizando o exemplo demonstrativo, a nuvem de palavras da empresa EFACEC ENERGIA - MÁQUINAS E EQUIPAMENTOS ELÉCTRICOS, S.A., que se dedica ao “exercício da indústria, comércio, instalação, reparação e manutenção de material eléctrico e mecânico, nomeadamente o destinado à produção e distribuição de energia; estudo, desenvolvimento, conceção, projeto, produção, comercialização, instalação e manutenção de sistemas, equipamentos e instalações electrónicos, eléctricos e eletromecânicos, de telecomunicações e de software, para os diversos setores industriais e de serviços; fornecimento de sistemas de gestão, comando e controlo de redes de energia eléctrica e instalações industriais; desenvolvimento, produção e comercialização de placas e circuitos electrónicos; e a sociedade dedica-se também à prestação de serviços de formação profissional”, da área “Indústrias transformadoras”, destaca as áreas de atuação “energy”, “power”, “system”, “technology” e “digital” e ainda características dos seus serviços como “solution”, “new”, “support” e “strong”(Figura 4.21).

No *website* da empresa FARFETCH PORTUGAL, UNIPESSOAL, LDA, da área “Atividades de informação e de comunicação”, que é uma empresa de “conceção e

Quando se está a analisar uma Unidade de Investigação, os dados utilizados para a definição do modelo e respetivos tópicos são a informação da unidade em análise e a informação de todas as empresas. Assim, a informação das empresas, nomeadamente as áreas de atuação e tipo de serviços que prestam, terão uma maior influência na definição dos tópicos. Como já referido anteriormente, o conjunto de empresas que constam na ferramenta não contempla todas as áreas.

Por outro lado, quando se está a analisar uma empresa, os dados utilizados para a definição do modelo e respetivos tópicos são a informação da empresa em análise e de todas as unidades de investigação, tendo estas uma maior influência. Neste caso, as unidades de investigação contemplam todas as áreas, mas o número de unidades por área difere.

De notar que, na apresentação das empresas ou unidades de investigação com maior relação com a unidade ou empresa em análise, estas aparecem listadas por ordem decrescente de pertencer ao tópico em análise e por vezes, a probabilidade mais elevada de pertencer ao tópico em análise pode não indicar uma maior relação na prática, mas dever-se à constituição dos restantes tópicos, que não estão mesmo relacionados, aumentando assim a probabilidade do único tópico em que existe alguma relação.

Utilizando os mesmos exemplos e começando pela Unidade de Investigação GECAD - Grupo de Investigação em Engenharia e Computação Inteligente para a Inovação e o Desenvolvimento, o resultado da aplicação de *topic modeling* indica uma probabilidade de 100% de pertencer ao tópico 5 (Figura 4.24), em que as 20 palavras com maior probabilidade de pertencer ao tópico incluem “energy”, “power”, “project” e “research”, também destacadas na nuvem de palavras na unidade, incluindo também palavras relacionadas com a área de atuação. As empresas que apresentam maior probabilidade de pertencer a este tópico (Figura 4.25) são empresas da área das “Indústrias transformadoras” (como a SNA EUROPE (INDUSTRIES), LDA – “indústria, comercialização e distribuição, por grosso ou a retalho, de todo o género de limas, ferramentas, manuais e industriais e máquinas industriais”; CS - COELHO DA SILVA, S.A. – “indústria de materiais de barro para a construção civil”; UNIVERSAL MOTORS (U.M.) - EQUIPAMENTOS ELECTROMECHANICOS, S.A. – “indústria, comércio e reparação de equipamento eletromecânico”; NS - MÁQUINAS INDUSTRIAIS, LDA – “produção de máquinas lixadoras para a indústria metalomecânica - acabamento e desgaste de superfícies”).

Considerando a Unidade de Investigação INESC TEC - INESC Tecnologia e Ciência (INESC TEC), o resultado indica uma probabilidade próxima de 100% de pertencer ao tópico 4 (Figura 4.26), que neste caso também contém nas 20 palavras com maior probabilidade de pertencer ao tópico, palavras identificadas na nuvem de palavras do INESC TEC, como “innovation” e “information” e ainda palavras relacionadas.

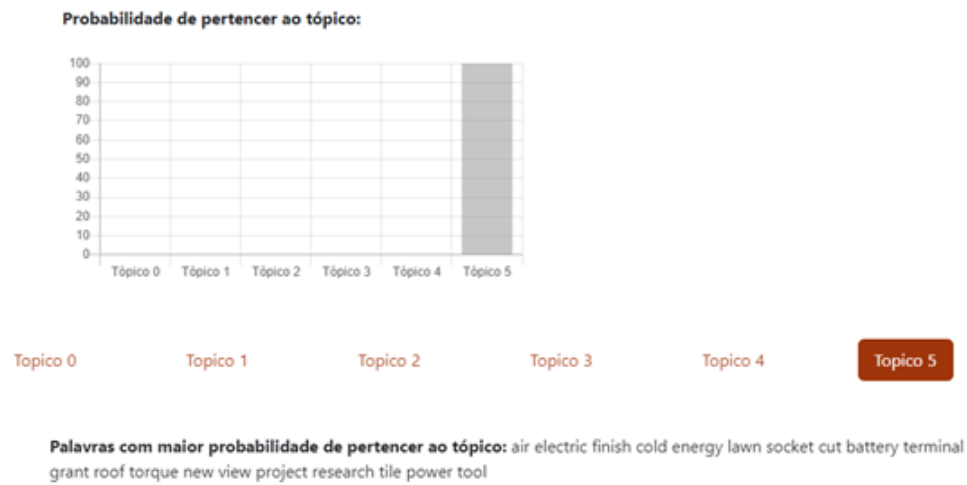


Figura 4.24: Resultado da aplicação de *topic modeling* da Unidade de Investigação GECAD.

SNA EUROPE (INDUSTRIES), LDA	+
CS - COELHO DA SILVA, S.A.	+
UNIVERSAL MOTORS (U.M.) - EQUIPAMENTOS ELECTROMECÂNICOS, S.A.	+
NS - MÁQUINAS INDUSTRIAIS, LDA	+
HUSQVARNA PORTUGAL, S.A.	+
CALIBRAFRUTA - EQUIPAMENTOS HORTOFRUTÍCOLAS, LDA	+
YILPORT IBERIA, S.A.	+
CAUDALIE PORTUGAL, UNIPessoal, LDA	+
DOMINÓ - INDÚSTRIAS CERÂMICAS, S.A.	+
VIEIRA DE SOUSA - VINES & WINES, LDA	+

Figura 4.25: Relação com as empresas da Unidade de Investigação GECAD.

As empresas que apresentam maior probabilidade de pertencer ao tópico 4 (Figura 4.27), fazem parte da área “Indústrias transformadoras” (como a FASTER - PRODUTOS ALIMENTARES, LDA – “produção e comércio de produtos alimentares; FLUPOL - SURFACE ENGINEERING, S.A. – “desenvolvimento e comercialização de soluções de engenharia de superfície pela aplicação de revestimentos funcionais/ fabrico e comercialização de peças com revestimentos funcionais/ comercialização de peças e semiprodutos”) e também da área “Atividades de informação e de comunicação” (como a WIT-SOFTWARE - CONSULTORIA E SOFTWARE PARA A INTERNET MÓVEL, S.A. – “serviços de consultoria de *software* para a *net* móvel”; MOONGY, S.A. – “desenvolvimento de projetos, prestação de serviços de consultoria em particular nas áreas de engenharia, com especial enfoque para as áreas de tecnologias e das infraestruturas, bem como comercialização de *software* e *hardware* e demais equipamentos conexos com as atividades desenvolvidas”).

As unidades de investigação GECAD e INESC TEC pertencem à mesma área, mas como se verificou, obtiveram resultados diferentes relativamente às palavras pertencentes ao tópico com maior probabilidade e às empresas identificadas. Algumas das empresas, pela descrição que apresentam (retirada da base de dados Sabi) poderão não apresentar uma relação muito óbvia com as unidades consideradas, mas o âmbito da atuação não é de todo desapropriado.



Figura 4.26: Resultado da aplicação de *topic modeling* da Unidade de Investigação INESC TEC.

Os resultados da unidade INET-md - Instituto de Etnomusicologia - Centro de Estudos em Música e Dança indicam uma probabilidade quase de 100% de pertencer ao tópico 1 (Figura 4.28), mas as palavras relativas ao mesmo não se enquadram com os resultados da nuvem de palavras, que de certa forma conseguia transmitir as dimensões da unidade.

As empresas com maior probabilidade de pertencer ao tópico (Figura 4.29) são

FASTER - PRODUTOS ALIMENTARES, LDA	■
FLUPOL - SURFACE ENGINEERING, S.A.	■
WIT-SOFTWARE - CONSULTORIA E SOFTWARE PARA A INTERNET MÓVEL, S.A.	■
CTR - CONSULTORIA, TÉCNICA E REPRESENTAÇÕES, LDA	■
MOONGY, S.A.	■
RESPOL - RESINAS, S.A.	■
INTRAPLÁS - INDÚSTRIA TRANSFORMADORA DE PLÁSTICOS, S.A.	■
GLN - ENGINEERING, MOLDING AND PLASTICS, S.A.	■
OLI - SISTEMAS SANITÁRIOS, S.A.	■
SIA - SOCIEDADE DE INOVAÇÃO AMBIENTAL, LDA	■

Figura 4.27: Relação com as empresas da Unidade de Investigação INESC TEC.

da área “Indústrias transformadoras” (como a WINPROVIT - SOLUÇÕES INTELIGENTES, S.A. – “manutenção e reparação de máquinas de escritório, de contabilidade e de material informático e atividades conexas à informática. *Outsourcing* de recursos humanos, tecnologias de informação e comunicações, bem como todos os serviços conexas”; FASTER - PRODUTOS ALIMENTARES, LDA – “produção e comércio de produtos alimentares”; CTR - CONSULTORIA, TÉCNICA E REPRESENTAÇÕES, LDA – “fabricação e comercialização de produtos de higiene, limpeza e cuidados do lar, fabricação e comercialização de peças plásticas/consultoria técnica, estudos e projetos serviços prestados às empresas, comissões e representações”).

A unidade INET-md pertence à área “Artes e Humanidades - Artes e Design, Desenvolvimento Artístico e Musical” e como visto anteriormente é uma área não contemplada pelas empresas em análise. Esta unidade é um exemplo de que uma elevada probabilidade de pertencer a um tópico não significa que a unidade (o mesmo se aplica às empresas) esteja muito relacionada com o mesmo (considerando as palavras com maior probabilidade de pertencer ao tópico e também as empresas identificadas), mas significa neste caso que ainda está menos relacionada com os restantes tópicos. Verifica-se assim a influência que as empresas têm na constituição dos tópicos, quando está em análise uma unidade (o mesmo se verifica quando se analisa uma empresa, tendo as unidades uma maior influência), não permitindo de certa forma que a informação da unidade em análise tenha influência na determinação

dos tópicos. Também através da análise das empresas identificadas, em que algumas coincidem com as identificadas no caso da unidade INESC TEC, se verifica a influência das empresas, ou seja, apesar de se alterar a unidade, os grupos de empresas identificados (correspondentes aos tópicos) apresentam semelhanças.



Figura 4.28: Resultado da aplicação de *topic modeling* da Unidade de Investigação INET-md.

WINPROVIT - SOLUÇÕES INTELIGENTES, S.A.	■
FASTER - PRODUTOS ALIMENTARES, LDA	■
WIT-SOFTWARE - CONSULTORIA E SOFTWARE PARA A INTERNET MÓVEL, S.A.	■
CTR - CONSULTORIA, TÉCNICA E REPRESENTAÇÕES, LDA	■
WAVECOM - SOLUÇÕES RÁDIO, S.A.	■
GLN - ENGINEERING, MOLDING AND PLASTICS, S.A.	■
OLI - SISTEMAS SANITÁRIOS, S.A.	■
SIMOLDES - PLÁSTICOS, S.A.	■
GLNMOLDS, S.A.	■
FAMOLDE - FABRICAÇÃO E COMERCIALIZAÇÃO DE MOLDES, S.A.	■

Figura 4.29: Relação com as empresas da Unidade de Investigação INET-md.

Considerando a unidade ALGORITMI - Centro de Investigação ALGORITMI da Universidade do Minho, que apresenta uma nuvem de palavras sem palavras com grande destaque relativamente à investigação desenvolvida, nos resultados da aplicação de *topic modeling* (Figura 4.30) é possível identificar dois tópicos, o tópico 2 e o tópico 3, com probabilidades próximas (52% e 47%, respetivamente). Tendo

em conta a área a que pertence a unidade, “Ciências das Engenharias e Tecnologias - Ciência da Computação e Tecnologia de Informática”, os dois conjuntos de palavras com a maior probabilidade de pertencer ao tópico 2 (Figura 4.30) e tópico 3 (Figura 4.31), apesar de não representarem de forma clara a unidade, poderão estar um pouco relacionados com a mesma.

Relativamente às empresas com maior probabilidade de pertencer ao tópico 2 (Figura 4.32), pertencem na maioria à área das “Indústrias transformadoras” (como a CS - COELHO DA SILVA, S.A. – “indústria de materiais de barro para a construção civil”; FACAL - ENGENHARIA DE FACHADAS, LDA – “indústria de revestimentos de edifícios em aço, alumínio e vidro. Empreiteiros de estruturas em aço e vidro”; RACLAC, S.A. – “fabricação de outras preparações e de artigos farmacêuticos”; LUSOFORMA – “indústria e comércio de embalagens de alumínio”), aparecendo também da área das “Atividades de consultoria, científicas, técnicas e similares” (como a SOPSEC - SOCIEDADE DE PRESTAÇÃO DE SERVIÇOS DE ENGENHARIA CIVIL, S.A. – “prestação de serviços no domínio da consultoria, elaboração e gestão de estudos e projetos de engenharia civil, de arquitetura e de urbanismo, gestão da qualidade em empreendimentos da construção e produção de cartografia”). Verifica-se neste caso uma variedade de empresas relativamente à área de atuação.

Caso se considere o tópico 3 (Figura 4.33), também existe uma grande variedade relativamente às empresas com maior probabilidade de pertencer ao tópico, abrangendo áreas como o “Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos” (como a ABBVIE, LDA – “fabrico, comércio e distribuição de produtos farmacêuticos e de saúde”), “Atividades de informação e de comunicação (como a MOONGY, S.A.- “desenvolvimento de projetos, prestação de serviços de consultoria em particular nas áreas de engenharia, com especial enfoque para as áreas de tecnologias e das infraestruturas, bem como comercialização de *software* e *hardware* e demais equipamentos conexos com as atividades desenvolvidas”) e “Indústrias transformadoras” (como a OLI - SISTEMAS SANITÁRIOS, S.A. – “conceção, produção e comercialização de autoclismos e componentes em matéria plástica (...)).

Considerando agora as empresas para análise, as unidades de investigação terão uma maior influência na definição dos tópicos.

Os resultados da empresa EFACEC ENERGIA - MÁQUINAS E EQUIPAMENTOS ELÉCTRICOS, S.A., indicam que o tópico 2 (Figura 4.34) apresenta a probabilidade mais elevada (45%) e as palavras com maior probabilidade de pertencer ao tópico incluem “energy” e “technology”, presentes na nuvem de palavras da empresa. As restantes palavras estão mais relacionadas com a atividade de investigação. A unidade com maior probabilidade de pertencer ao tópico 2 (Figura 4.35) é da área “Ciências da Engenharia e Tecnologias - Ciência da Computação e Tecnologia de



Figura 4.30: Resultado da aplicação de *topic modeling* da Unidade de Investigação ALGORITMI - tópico 2.



Figura 4.31: Resultado da aplicação de *topic modeling* da Unidade de Investigação ALGORITMI - tópico 3.

CS - COELHO DA SILVA, S.A.	+
SOPSEC - SOCIEDADE DE PRESTAÇÃO DE SERVIÇOS DE ENGENHARIA CIVIL, S.A.	+
FACAL - ENGENHARIA DE FACHADAS, LDA	+
RACLAC, S.A.	+
LUSOFORMA - INDÚSTRIA E COMÉRCIO DE EMBALAGENS, S.A.	+
FASSALUSA - PRODUÇÃO E COMERCIALIZAÇÃO DE MATERIAIS DE CONSTRUÇÃO, LDA	+
ALVES RIBEIRO, S.A.	+
DOMINÓ - INDÚSTRIAS CERÂMICAS, S.A.	+
TECNIARTE - PROJECTOS E CONSTRUÇÕES S.A.	+
RENOVA - FÁBRICA DE PAPEL DO ALMONDA, S.A.	+

Figura 4.32: Relação com as empresas da Unidade de Investigação ALGORITMI - tópico 2.

ABBVIE, LDA	+
MOONGY, S.A.	+
OLI - SISTEMAS SANITÁRIOS, S.A.	+
FUJITSU TECHNOLOGY SOLUTIONS, LDA	+
BRISTOL - MYERS SQUIBB FARMACÊUTICA PORTUGUESA, S.A.	+
BIOGEN PORTUGAL - SOCIEDADE FARMACÊUTICA, UNIPESSOAL, LDA	+
UNIPARTNER IT SERVICES, S.A.	+
XZ - CONSULTORES, S.A.	+
TECHNOPHAGE - INVESTIGAÇÃO E DESENVOLVIMENTO EM BIOTECNOLOGIA, S.A.	+
THALES PORTUGAL, S.A.	+

Figura 4.33: Relação com as empresas da Unidade de Investigação ALGORITMI - tópico 3.

Informática” (COPELABS - Computação Cognitiva e Centrada nas Pessoas), sendo as seguintes das “Ciências Exatas - Física e Matemática” (como o IBEB - Instituto de Biofísica e Engenharia Biomédica; CIDMA - Centro de Investigação e Desenvolvimento em Matemática e Aplicações; CMAFCIO - Centro de Matemática, Aplicações Fundamentais e Investigação Operacional; CFTP - Centro de Física Teórica de Partículas; GFMUL - Grupo de Física-Matemática da Universidade de Lisboa).

A empresa apresenta também uma probabilidade de pertencer ao tópico 0 (39%) próxima do tópico 2, contendo o grupo de palavras com maior probabilidade de pertencer ao tópico 0, palavras com relevância na nuvem de palavras da empresa, como “power”, “technology”, “new” e “energy” (Figura 4.36). Considerando este tópico, as unidades com maior probabilidade de pertencer ao mesmo distribuem-se por várias áreas (Figura 4.32), nomeadamente “Ciências da Engenharia e Tecnologias – Engenharia Eletrotécnico e de Computadores” (como o Grupo de Investigação em Engenharia e Computação Inteligente para a Inovação e o Desenvolvimento (GECAD) - Grupo de Investigação em Engenharia e Computação Inteligente para a Inovação e o Desenvolvimento), “Artes e Humanidades - Artes e Design, Desenvolvimento Artístico e Musical” (como o CESEM - Centro de Estudos de Sociologia e Estética Musical; INET-md - Instituto de Etnomusicologia - Centro de Estudos em Música e Dança), “Ciências Exatas – Matemática” (como o CMA/FCT/UNL - Centro de Matemática e Aplicações) e “Ciências Sociais - Ciências da Linguagem” (como o CEL - Centro de Estudos em Letras).

Neste exemplo, quando considerado o tópico 2, que apresenta maior probabilidade, as empresas com maior probabilidade de pertencerem ao tópico são de áreas semelhantes, fazendo sentido para a empresa em análise.

Por outro lado, quando considerado o tópico 0 (Figura 4.37), é identificada a Unidade de Investigação GECAD em primeiro lugar, o que também faz sentido visto pertencer à área “Ciências da Engenharia e Tecnologias – Engenharia Eletrotécnico e de Computadores”. A análise dos restantes tópicos, quando apresentam probabilidades próximas da probabilidade do tópico mais elevado pode ser útil para identificar outras relações.

Este exemplo serve também para perceber que os resultados relativamente às melhores relações podem diferir, caso se analise a unidade ou a empresa. Neste exemplo, ao analisar a unidade GECAD, a empresa EFACEC ENERGIA - MÁQUINAS E EQUIPAMENTOS ELÉCTRICOS, S.A. não foi identificada (tendo a unidade probabilidade de 100% de pertencer ao tópico identificado, não fazendo sentido analisar os restantes tópicos), mas na análise da empresa, a unidade já foi identificada (quando analisado o segundo tópico com maior probabilidade da empresa, mas próximo do primeiro).

Relativamente à FARFETCH PORTUGAL, UNIPESSOAL, LDA, nos resultados obtidos verifica-se uma elevada probabilidade no tópico 4 (Figura 4.38) e as

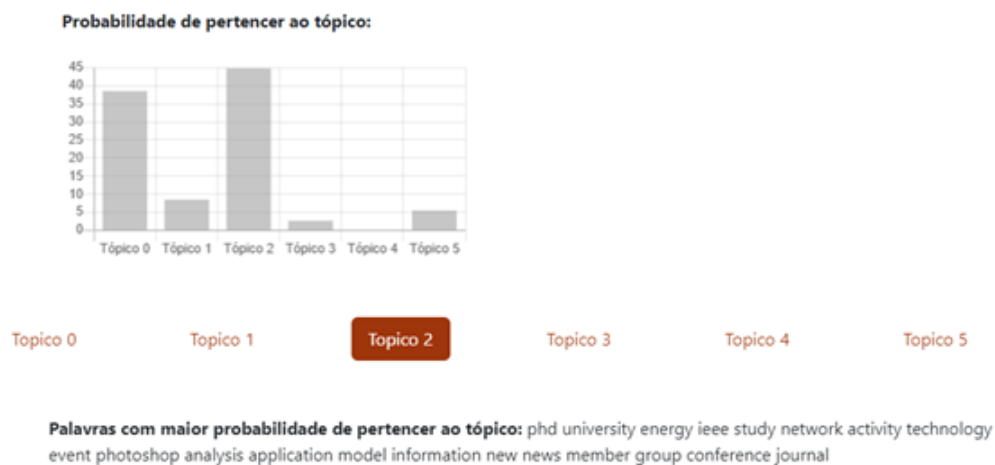


Figura 4.34: Resultado da aplicação de *topic modeling* da empresa EFACEC - tópico 2.

COPELABS - Computação Cognitiva e Centrada nas Pessoas	+
IBEB - Instituto de Biofísica e Engenharia Biomédica	+
CIDMA - Centro de Investigação e Desenvolvimento em Matemática e Aplicações	+
CMAFCIO - Centro de Matemática, Aplicações Fundamentais e Investigação Operacional	+
CFTP - Centro de Física Teórica de Partículas	+
GMUL - Grupo de Física-Matemática da Universidade de Lisboa	+
IEETA - Instituto de Engenharia Eletrónica e Informática de Aveiro	+
PRAXIS - PRAXIS - Centro de Filosofia, Política e Cultura	+
UACC - Laboratório de Inteligência Artificial e Ciência de Computadores	+
CEAFEL - Centro de Análise Funcional, Estruturas Lineares e Aplicações	+

Figura 4.35: Relação com as Unidades de Investigação da empresa EFACEC - tópico 2.

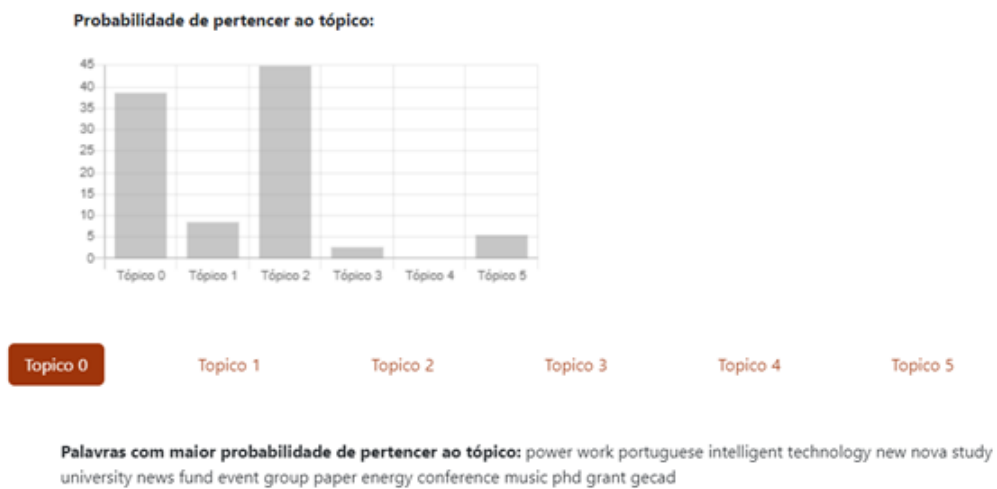


Figura 4.36: Resultado da aplicação de *topic modeling* da empresa EFACEC - tópico 0.

GECAD - Grupo de Investigação em Engenharia e Computação Inteligente para a Inovação e o Desenvolvimento	■
CESEM - Centro de Estudos de Sociologia e Estética Musical	■
INET-md - Instituto de Etnomusicologia - Centro de Estudos em Música e Dança	■
CMA/FCT/UNL - Centro de Matemática e Aplicações	■
CEL - Centro de Estudos em Letras	■
CEDIS - CEDIS - Centro de Investigação & Desenvolvimento sobre Direito e Sociedade	■
ALGORITMI - Centro de Investigação ALGORITMI	■
CLUNL - Centro de Linguística da Universidade Nova de Lisboa	■
CTS - Centro de Tecnologias e Sistemas	■

Figura 4.37: Relação com as Unidades de Investigação da empresa EFACEC - tópico 0.

palavras com maior probabilidade de pertencer ao tópico, apesar de não incluírem as destacadas na análise da nuvem de palavras da empresa, estão relacionadas com a área empresarial, fazendo sentido estes resultados para a empresa em análise. As unidades com maior probabilidade de pertencer ao mesmo tópico (Figura 4.39) acompanham a interpretação das palavras e são maioritariamente da área “Ciências Sociais - Economia e Gestão” (como o NIPE - Núcleo de Investigação em Políticas Económicas e Empresariais; Nova SBE - Nova School of Business and Economics; cef.up - Centro de Economia e Finanças da Universidade do Porto; CeBER - Centro de Investigação em Economia e Gestão da Universidade de Coimbra).

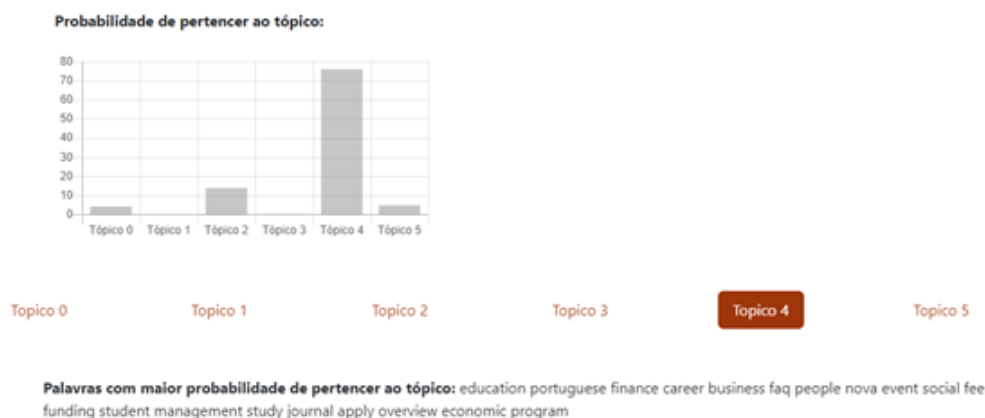


Figura 4.38: Resultado da aplicação de *topic modeling* da empresa FARFETCH.

Por fim, os resultados da análise de *topic modeling* para a empresa PORTUGALISOL NORTE - ISOLAMENTOS INDUSTRIAIS, LDA apresentam probabilidades dispersas pelos tópicos, não se verificando um tópico com probabilidade muito elevada (Figura 4.40). Como já referido, quando na nuvem de palavras não aparecem palavras com muito destaque, os resultados de *topic modeling* tendem a não destacar um só tópico e as unidades com maior probabilidade de pertencer ao tópico mais elevado, neste caso é o tópico 3, distribuem-se por várias áreas, como neste exemplo (Figura 4.41), em que pertencem à área “Ciências da Engenharia e Tecnologias - Engenharia Química e Biológica, e Química Ambientalmente Sustentável” (como o LSRE-LCM - Laboratório de Processos de Separação e Reacção - Laboratório de Catálise e Materiais), “Ciências da Saúde - Investigação Clínica e de Translação” (como o MedInUP - Centro de Investigação Farmacológica e Inovação Medicamentosa), “Ciências da Saúde - Saúde Pública, Enfermagem, Tecnologias da Saúde e do Desporto, Reabilitação e Bem-estar” (como a FP-ENAS - Unidade de Investigação UFP em Energia, Ambiente e Saúde), “Ciências Exatas - Química” (como o REQUIMTE - Laboratório Associado para a Química Verde - Tecnologias e Processos Limpos) e “Ciências Sociais – Psicologia” (como o CINEICC - Centro de Investigação em Neuropsicologia e Intervenção Cognitivo Comportamental).

NIPE - Núcleo de Investigação em Políticas Económicas e Empresariais	1
Nova SBE - Nova School of Business and Economics	1
cef.up - Centro de Economia e Finanças da Universidade do Porto	1
CINEICC - Centro de Investigação em Neuropsicologia e Intervenção Cognitivo Comportamental	1
CeBER - Centro de Investigação em Economia e Gestão da Universidade de Coimbra	1
INET-md - Instituto de Etnomusicologia - Centro de Estudos em Música e Dança	1
NECE - Núcleo de Estudos em Ciências Empresariais	1
REM - Research in Economics and Mathematics	1
CICS.NOVA - Centro Interdisciplinar de Ciências Sociais	1
CIPsi - Centro de Investigação em Psicologia	1

Figura 4.39: Relação com as Unidades de Investigação da empresa FARFETCH.

Nos exemplos em que não existe um tópico com uma probabilidade muito elevada, mas sim probabilidades idênticas para dois ou mais tópicos, como neste exemplo da empresa PORTUGALISOL NORTE - ISOLAMENTOS INDUSTRIAIS, LDA e no exemplo da unidade ALGORITMI - Centro de Investigação ALGORITMI da Universidade do Minho, os grupos de unidades ou empresas com maior probabilidade tendem a ser mais diversificados relativamente às áreas a que pertencem. Ou seja, quando não existe uma preponderância de um tópico, os resultados tendem a ser mais diversificados, não se observando grupos de unidades de investigação ou empresas que desenvolvam atividades semelhantes entre si e também em relação à unidade ou empresa em análise.

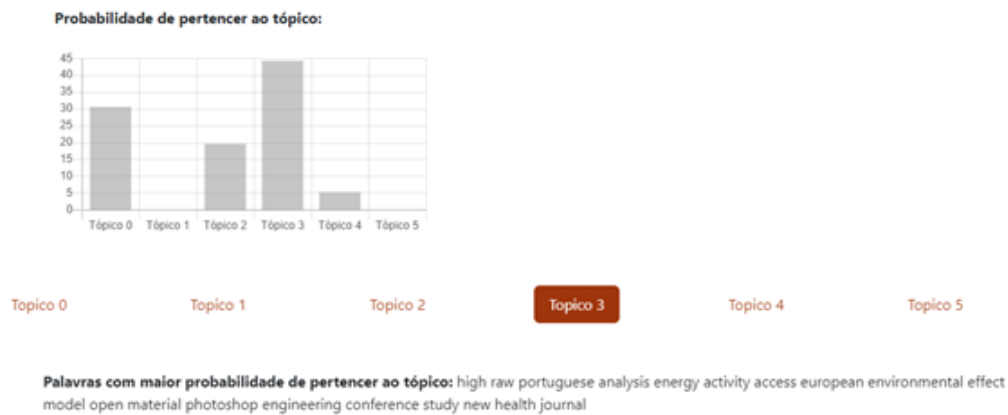


Figura 4.40: Resultado da aplicação de *topic modeling* da empresa PORTUGALISOL.

LSRE-LCM - Laboratório de Processos de Separação e Reacção - Laboratório de Catálise e Materiais	2
MedInUP - Centro de Investigação Farmacológica e Inovação Medicamentosa	2
FP-ENAS - Unidade de Investigação UFP em Energia, Ambiente e Saúde	2
REQUIMTE - Laboratório Associado para a Química Verde - Tecnologias e Processos Limpos	2
CINEICC - Centro de Investigação em Neuropsicologia e Intervenção Cognitivo Comportamental	2
CEFT - Centro de Estudos de Fenómenos de Transporte	2
CQVR - Centro de Química - Vila Real	2
CIMO - Centro de Investigação de Montanha	2
iNOVA4Health - iNOVA4Health - Programa de Medicina Translacional	2
IIFACTS - Instituto de Investigação e Formação Avançada em Ciências e Tecnologias da Saúde	2

Figura 4.41: Relação com as Unidades de Investigação da empresa PORTUGALISOL.

Capítulo 5

Conclusões

Neste capítulo serão apresentadas as conclusões sobre o trabalho desenvolvido e ainda sugestões de trabalho futuro, tendo em conta aspetos que podem ser melhorados.

5.1 Conclusões sobre o trabalho realizado

O trabalho desenvolvido resultou na criação de uma ferramenta *web* que tem como objetivo dar apoio na interface entre as unidades de investigação e as empresas, potenciando assim a transferência e valorização do conhecimento.

Para alcançar este objetivo foi necessário perceber o processo de transferência e valorização do conhecimento, que é cada vez mais complexo, interrelacionado e abrangente e envolve diversos intervenientes. Em Portugal assumem particular importância neste processo as unidades de investigação e as empresas, que têm objetivos diferentes e beneficiam do conhecimento de maneira diferente, sendo importante a consolidação e densificação da sua relação de forma a potenciar um entendimento comum do resultado esperado.

Considerando o desafio de dinamizar a relação entre as unidades de investigação e as empresas e dado o aumento da informação disponível através dos *sites* das mesmas, tornou-se pertinente a utilização de metodologias de *text mining* para extrair informação a partir destes dados textuais. Verificou-se que existem várias técnicas de *text mining* e o aumento dos dados textuais disponíveis e avanço computacional potenciam a sua evolução. No desenvolvimento da ferramenta optou-se por utilizar

duas metodologias, o sumário da informação baseado na contagem de palavras e *topic modeling* com utilização do modelo LDA.

O sumário da informação baseado na frequência das palavras foi representado através de uma nuvem de palavras. Na ferramenta, para cada unidade de investigação e cada empresa é representada uma nuvem de palavras da informação que disponibilizam nos respectivos *websites*. Pela análise de alguns exemplos, percebeu-se por um lado que, para algumas unidades e empresas, as nuvens de palavras refletem e permitem ter uma percepção das atividades realizadas, áreas de atuação, serviços e produtos disponibilizados. Por outro lado, existem exemplos em que a nuvem de palavras reflete apenas informações gerais do âmbito da investigação e do âmbito empresarial, não sendo possível diferenciar as áreas de atuação. Esta análise é também importante para uma melhor compreensão dos resultados obtidos pela análise de *topic modeling*.

A aplicação de *topic modeling* teve como objetivo identificar relações entre unidades de investigação e empresas, através da informação presente nos respectivos *websites*. Para cada unidade de investigação é possível identificar quais as empresas que apresentam uma maior relação, ou seja, que apresentam maior probabilidade de pertencer ao mesmo tópico, sendo possível identificar também as palavras com maior probabilidade de pertencer a cada tópico. Da mesma forma é possível identificar as unidades de investigação mais relacionadas com uma empresa.

Verificou-se que, da mesma forma que é difícil ter uma boa percepção através da nuvem de palavras da dimensão e características das unidades e empresas, caso a informação dos *websites* não o reflita, também a atribuição a um tópico com base nesta informação poderá não ser a mais adequada.

Quando se está a analisar uma unidade de investigação e considerando os dados utilizados, a informação das empresas terá uma maior influência para a definição do modelo e respetivos tópicos. Por sua vez, quando se está a analisar uma empresa, será a informação das unidades de investigação a ter maior influência. De notar que as unidades de investigação presentes na ferramenta contemplam todas as áreas, mas as empresas não, podendo comprometer assim os resultados das unidades das áreas não contempladas.

Dada a diferença da informação considerada na definição do modelo quando se analisa uma unidade ou uma empresa, os resultados relativamente às melhores relações podem diferir, ou seja, na análise de uma unidade pode ser identificada uma empresa com elevada probabilidade de pertencer ao mesmo tópico, mas na análise desta empresa, como o modelo definido é outro e diferente, a unidade de investigação pode não apresentar uma elevada probabilidade de pertencer ao mesmo tópico.

A análise dos tópicos terá de ser cautelosa pois, caso a informação presente no *website* consiga transmitir a dimensão da unidade ou da empresa e a respetiva área de atuação esteja contemplada no conjunto da informação, esperam-se bons

resultados relativos a possíveis relações, tendo sempre em consideração que estes resultados poderão não ser os mais óbvios tendo por base o conhecimento prévio das entidades em análise. Por outro lado, caso estas condições não se verifiquem, os resultados poderão não ser os melhores. A probabilidade elevada de pertencer a um tópico pode não indicar uma maior relação na prática, mas dever-se à constituição dos restantes tópicos, que não apresentam qualquer relação.

Assim, a ferramenta deverá ser utilizada como apoio na tomada de decisão, devendo os resultados obtidos ser analisados de forma crítica e em complemento à experiência de especialistas.

Por fim, referir que a ferramenta é dinâmica, com possibilidade de atualização ao longo do tempo, o que permite incorporar as alterações da informação dos *websites*, conseguindo assim acompanhar o contínuo e crescente desenvolvimento científico e tecnológico.

5.2 Sugestões de trabalho futuro

O trabalho desenvolvido cumpre com o objetivo proposto de criação de uma ferramenta de apoio estratégico, que potencie a colaboração entre instituições de investigação e o tecido empresarial. Ainda assim, considera-se que a ferramenta desenvolvida poderá ser um ponto de partida para uma ferramenta mais abrangente, eficaz e eficiente.

Ao longo do processo foram tomadas várias decisões, tendo em consideração limitações de recursos e tempo. Seguindo os vários passos da análise de *text mining*, na seleção de dados, consideraram-se para integrar a ferramenta, as unidades de investigação e empresas com pessoas afetas à investigação, dada a disponibilidade e quantidade de informação, restringindo desde já o âmbito da ferramenta. Assim, poderão ser acrescentadas outras estruturas de investigação e empresas, proporcionando outras opções de análise e uma maior representatividade de todas as áreas. Quanto à eliminação das unidades e empresas em que não foi possível recolher a informação dos *websites*, poderão ser desenvolvidos e aplicados outros métodos de recolha. A exclusão de *websites* que não estejam em inglês, poderá ser resolvida com recurso a ferramentas de tradução. Por fim, optou-se por recolher o *hiperlink* da página inicial e os *hiperlinks* referenciados na mesma, podendo esta recolha ser alargada a todos os *hiperlinks* disponíveis.

O passo de pré-processamento poderá ser melhorado, com a aplicação de outras técnicas e até análise da ordem de aplicação das mesmas.

Relativamente à metodologia aplicada - *topic modeling* com utilização do modelo LDA - poderão ser feitos vários estudos, nomeadamente de otimização do número de tópicos e dos parâmetros *a priori*, análise de diferentes condições para utilizar informação distintiva e análise exaustiva dos resultados, com estudo da influência

das áreas de atuação das entidades nos mesmos, por exemplo. Poderá ser também estudada a aplicação de outros modelos na análise de *topic modeling*. De referir que esta metodologia não é isenta de limitações, não tendo, por exemplo, como objetivo perceber a semântica das palavras e tende a focar-se apenas na frequência das mesmas. Poderão também ser utilizadas, com posterior comparação de resultados, outras metodologias de *text mining*.

Por fim, referir a importância que poderá ter a contribuição de especialistas de várias áreas no passo de interpretação e avaliação.

Referências

- [1] European University Association, *Universities without walls: A vision for 2030*, 2021. [Citado nas páginas 1 e 2]
- [2] United Nations, “The 17 goals: Sustainable development.” Available at <https://sdgs.un.org/goals>. (Last accessed in 01/09/2022). [Citado na página 2]
- [3] Conselho empresarial para o Desenvolvimento Sustentável, “Objetivos de desenvolvimento sustentável.” Available at <https://www.ods.pt/>. (Last accessed in 02/09/2022). [Citado nas páginas vii e 2]
- [4] I. Estrada, D. Faems, N. M. Cruz, and P. P. Santana, “The role of interpartner dissimilarities in industry-university alliances: Insights from a comparative case study,” *Research Policy*, vol. 45, no. 10, pp. 2008–2022, 2016. [Citado na página 2]
- [5] ANI - Agência Nacional de Inovação, *Rede de Transferência e Valorização do Conhecimento no âmbito do Ensino Superior*, 2021. [Citado nas páginas vii, 2, 3, 8, 9 e 10]
- [6] R. Lombardi, “Knowledge transfer and organizational performance and business process: past, present and future researches,” *Business Process Management Journal*, vol. 25, no. 1, pp. 2–9, 2019. [Citado na página 3]
- [7] L. Argote and J. M. Guo, “Routines and transactive memory systems: Creating, coordinating, retaining, and transferring knowledge in organizations,” *Research in Organizational Behavior*, vol. 36, pp. 65–84, 2016. [Citado na página 3]
- [8] A. J. Gil and F. J. Carrillo, “Knowledge transfer and the learning process in spanish wineries,” *KNOWLEDGE MANAGEMENT RESEARCH & PRACTICE*, vol. 14, no. 1, pp. 60–68, 2016. [Citado na página 3]
- [9] L. Argote and P. Ingram, “Knowledge transfer: A basis for competitive advantage in firms,” *ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES*, vol. 82, no. 1, pp. 150–169, 2000. [Citado na página 3]
- [10] R. Lombardi, R. Trequattrini, and G. Russo, “Innovative start-ups and equity crowdfunding,” *Proceedings of 2nd Business Systems Laboratory - International Symposium "Systems thinking for a sustainable economy. Advancements in*

- economic and managerial theory and practice*, vol. 19, 01 2014. [Citado na página 3]
- [11] E. de Wit-de Vries, W. A. Dolfsma, H. J. van der Windt, and M. P. Gerkema, “Knowledge transfer in university-industry research partnerships: a review,” *Journal of Technology Transfer*, vol. 44, no. 4, pp. 1236–1255, 2019. [Citado nas páginas 3 e 5]
- [12] T. Baycan and R. R. Stough, “Bridging knowledge to commercialization: the good, the bad, and the challenging,” *ANNALS OF REGIONAL SCIENCE*, vol. 50, no. 2, SI, pp. 367–405, 2013. [Citado na página 3]
- [13] P. D’Este and M. Perkmann, “Why do academics engage with industry? the entrepreneurial university and individual motivations,” *JOURNAL OF TECHNOLOGY TRANSFER*, vol. 36, no. 3, pp. 316–339, 2011. [Citado na página 3]
- [14] A. Agrawal, “University-to-industry knowledge transfer: literature review and unanswered questions,” *INTERNATIONAL JOURNAL OF MANAGEMENT REVIEWS*, vol. 3, no. 4, pp. 285–302, 2001. [Citado na página 3]
- [15] S. Shane, “Economic development through entrepreneurship: Government, university and business linkages,” *Economic Development through Entrepreneurship: Government, University and Business Linkages*, 01 2005. [Citado na página 3]
- [16] A. Salter and B. Martin, “The economic benefits of publicly funded basic research: a critical review,” *RESEARCH POLICY*, vol. 30, no. 3, pp. 509–532, 2001. [Citado na página 3]
- [17] A. T. Alexander and S. J. Childe, “Innovation: a knowledge transfer perspective,” *PRODUCTION PLANNING & CONTROL*, vol. 24, no. 2-3, SI, pp. 208–225, 2013. [Citado na página 3]
- [18] U.Norte Inova, “U.norte inova.” Available at <http://www.unorteinova.pt/>. (Last accessed in 01/09/2022). [Citado na página 4]
- [19] S. Cebi, “Determining importance degrees of website design parameters based on interactions and types of websites,” *Decision Support Systems*, vol. 54, no. 2, pp. 1030–1043, 2013. [Citado na página 4]
- [20] C. Y. Lee and S. Kim, “Stochastic dynamics for the popularity of growing websites,” *Journal of the Korean Physical Society*, vol. 44, no. 2, pp. 442–445, 2004. [Citado na página 5]
- [21] J. M. Pearson and A. M. Pearson, “An exploratory study into determining the relative importance of key criteria in web usability: A multi-criteria approach,”

- Journal of Computer Information Systems*, vol. 48, no. 4, pp. 115–127, 2008. [Citado na página 5]
- [22] P. Capriotti and A. Moreno, “Corporate citizenship and public relations: The importance and interactivity of social responsibility issues on corporate websites,” *Public Relations Review*, vol. 33, no. 1, pp. 84–91, 2007. [Citado na página 5]
- [23] M. Kantardzic, *Data mining : concepts, models, methods, and algorithms*. New Jersey: John Wiley IEEE Xplore, 2nd ed., 2011. [Citado na página 5]
- [24] A. N. de Inovação, “Ani notícias: Manual de oslo.” Available at <https://www.ani.pt/noticias/not%C3%ADcias-ani/manual-de-oslo-2018/>. (Last accessed in 25/08/2022). [Citado na página 6]
- [25] J. A. Kumar and L. S. Ganesh, “Research on knowledge transfer in organizations: a morphology,” *Journal of Knowledge Management*, vol. 13, no. 4, pp. 161–174, 2009. [Citado na página 8]
- [26] G. von Krogh, I. Nonaka, and M. Aben, “Making the most of your company’s knowledge: A strategic framework,” *Long Range Planning*, vol. 34, no. 4, pp. 421–439, 2001. [Citado na página 8]
- [27] J. Nahapiet and S. Ghoshal, “Social capital, intellectual capital, and the organizational advantage,” *Academy of Management Review*, vol. 23, no. 2, pp. 242–266, 1998. [Citado na página 8]
- [28] M. Subramaniam and M. A. Youndt, “The influence of intellectual capital on the types of innovative capabilities,” *Academy of Management Journal*, vol. 48, no. 3, pp. 450–463, 2005. [Citado na página 8]
- [29] N. Hewitt-Dundas, “Research intensity and knowledge transfer activity in uk universities,” *Research Policy*, vol. 41, no. 2, pp. 262–275, 2012. [Citado na página 8]
- [30] European Commission. Statistical Office of the European Union and Organisation for Economic Co-operation and Development, *Oslo manual 2018 : guidelines for collecting, reporting and using data on innovation*. The Measurement of Scientific, Technological and Innovation Activities, Paris: OECD, 4th edition. ed., 2018. [Citado na página 8]
- [31] European Commission. Directorate General for Research, *Metrics for knowledge transfer from public research organisations in Europe : report from the European Commission’s Expert Group on Knowledge Transfer Metrics : expert group report*. EUR 23894, Luxembourg: Office for Official Publications of the European Communities, 2009. [Citado nas páginas vii e 9]

- [32] Fundação para a Ciência e a Tecnologia, *Portuguese Atlas of Research Units*, 2022. [Citado nas páginas 10 e 11]
- [33] Fundação para a Ciência e a Tecnologia, “Fct - fundação para a ciência e a tecnologia.” Available at <https://www.fct.pt/index.phtml.pt>. (Last accessed in 05/05/2022). [Citado na página 11]
- [34] Fundação para a Ciência e a Tecnologia, *Diagnóstico do Sistema de Investigação e Inovação: desafios, forças e fraquezas rumo a 2020*, 2013. [Citado na página 11]
- [35] Fundação para a Ciência e a Tecnologia, “Resultados finais após audiência prévia do processo de avaliação de unidades de i&d.” Available at https://www.fct.pt/apoios/unidades/avaliacoes/2017/docs/Avaliacao_UnidadesI&D2017_2018_Resultados_finais.xlsx, 2019. (Last accessed in 09/05/2022). [Citado na página 11]
- [36] Instituto Nacional de Estatística, “Detalhe do conceito empresa.” Available at <https://smi.ine.pt/Conceito/Detalhes/4456>, 1994. (Last accessed in 09/09/2022). [Citado na página 13]
- [37] Bureau van Dijk Electronic Publishing Ltd, “Base de dados sabi.” Available at <https://sabi.bvdinfo.com>. (Last accessed in 09/09/2022). [Citado na página 13]
- [38] Instituto Nacional de Estatística, *Classificação Portuguesa das Actividades Económicas Rev.3*, 2007. [Citado na página 14]
- [39] M. B. Miles, “Qualitative data as an attractive nuisance - problem of analysis,” *Administrative Science Quarterly*, vol. 24, no. 4, pp. 590–601, 1979. [Citado na página 21]
- [40] I. Feinerer, K. Hornik, and D. Meyer, “Text mining infrastructure in r,” *Journal of Statistical Software*, vol. 25, no. 5, pp. 1–54, 2008. [Citado nas páginas 21, 29 e 30]
- [41] Sociedade Portuguesa de Estatística and Associação Brasileira de Estatística, “Glossário de estatística.” Available at <https://www.spestatistica.pt/glossario>. (Last accessed in 01/09/2022). [Citado nas páginas 22, 30 e 31]
- [42] Y. Cai and J. Sun, *Text Mining*, pp. 3061–3065. Boston, MA: Springer US, 2009. [Citado nas páginas 22, 27, 30, 31 e 32]
- [43] E. Isaeva and D. Aldarova, “Text-mining in terms of methodology and development,” *Proceedings of the 2021 Ieee Conference of Russian Young Researchers in Electrical and Electronic Engineering (Elconrus)*, pp. 413–416, 2021. [Citado nas páginas 22 e 28]

- [44] S. M. Weiss, N. Indurkha, T. Zhang, and F. Damerau, eds., *Text Mining : Predictive Methods for Analyzing Unstructured Information*. New York, NY: Springer New York : Imprint: Springer, 1st 2005. ed., 2005. [Citado nas páginas 22 e 31]
- [45] H. P. Luhn, “The automatic creation of literature abstracts,” *Ibm Journal of Research and Development*, vol. 2, no. 2, pp. 159–165, 1958. [Citado na página 22]
- [46] M. E. Maron and J. L. Kuhns, “On relevance, probabilistic indexing and information retrieval,” *Journal of the Acm*, vol. 7, no. 3, pp. 216–244, 1960. [Citado na página 22]
- [47] N. Jardine and C. J. van Rijsbergen, “The use of hierarchic clustering in information retrieval,” *Information Storage and Retrieval*, vol. 7, no. 5, pp. 217–240, 1971. [Citado na página 22]
- [48] G. Salton, A. Wong, and C. S. Yang, “Vector-space model for automatic indexing,” *Communications of the Acm*, vol. 18, no. 11, pp. 613–620, 1975. [Citado na página 22]
- [49] P. J. Hayes and S. P. Weinstein, “Construe/tis: A system for content-based indexing of a database of news stories,” 1991. [Citado na página 22]
- [50] P. Chellappandi and S. C. Vijayakumar, “Bibliometrics, scientometrics, webometrics/ cybermetrics, informetrics and altmetrics - an emerging field in library and information science research,” *Shanlax International Journal of Education*, vol. 7, pp. 5–8, 2018. [Citado na página 22]
- [51] I. Rodriguez-Rodriguez, J. V. Rodriguez, N. Shirvanizadeh, A. Ortiz, and D. J. Pardo-Quiles, “Applications of artificial intelligence, machine learning, big data and the internet of things to the covid-19 pandemic: A scientometric review using text mining,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 16, 2021. [Citado nas páginas 22, 23 e 25]
- [52] Clarivate, “Web of science.” Available at <https://www.webofscience.com/wos/woscc/basic-search>, 2008. (Last accessed in 26/08/2022). [Citado nas páginas 22 e 23]
- [53] J. W. Zhu and W. S. Liu, “A tale of two databases: the use of web of science and scopus in academic papers,” *Scientometrics*, vol. 123, no. 1, pp. 321–335, 2020. [Citado na página 23]
- [54] K. Li, J. Rollins, and E. Yan, “Web of science use in published research and review papers 1997-2017: a selective, dynamic, cross-domain, content-based analysis,” *Scientometrics*, vol. 115, no. 1, pp. 1–20, 2018. [Citado na página 23]

- [55] N. J. van Eck and L. Waltman, “Software survey: Vosviewer, a computer program for bibliometric mapping,” *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010. [Citado na página 23]
- [56] H. Ahonen, O. Heinonen, M. Klemettinen, and A. I. Verkamo, “Mining in the phrasal frontier,” *Principles of Data Mining and Knowledge Discovery*, vol. 1263, pp. 343–350, 1997. [Citado na página 23]
- [57] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, and C. Mering, “String v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets,” *Nucleic Acids Research*, vol. 47, no. D1, pp. D607–D613, 2019. [Citado nas páginas 23 e 24]
- [58] D. Szklarczyk, J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, L. J. Jensen, and C. von Mering, “The string database in 2017: quality-controlled protein-protein association networks, made broadly accessible,” *Nucleic Acids Research*, vol. 45, no. D1, pp. D362–D368, 2017. [Citado na página 24]
- [59] T. Yarkoni, R. A. Poldrack, T. E. Nichols, D. C. V. Essen, and T. D. Wager, “Large-scale automated synthesis of human functional neuroimaging data,” *Nature Methods*, vol. 8, no. 8, pp. 665–U95, 2011. [Citado na página 24]
- [60] A. Kozomara, M. Birgaoanu, and S. Griffiths-Jones, “mirbase: from microrna sequences to function,” *Nucleic Acids Research*, vol. 47, no. D1, pp. D155–D162, 2019. [Citado na página 24]
- [61] H. Dweep, C. Sticht, P. Pandey, and N. Gretz, “mirwalk - database: Prediction of possible mirna binding sites by "walking" the genes of three genomes,” *Journal of Biomedical Informatics*, vol. 44, no. 5, pp. 839–847, 2011. [Citado na página 24]
- [62] B. F. Jia, A. R. Raphenya, B. Alcock, N. Waglechner, P. Y. Guo, K. k Tsang, B. A. Lago, B. M. Dave, S. Pereira, A. N. Sharma, S. Doshi, M. Courtot, R. Lo, L. E. Williams, J. G. Frye, T. Elsayegh, D. Sardar, E. L. Westman, A. C. Pawlowski, T. A. Johnson, F. S. L. Brinkman, G. D. Wright, and A. G. McArthur, “Card 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database,” *Nucleic Acids Research*, vol. 45, no. D1, pp. D566–D573, 2017. [Citado na página 24]
- [63] J. Pinero, A. Bravo, N. Queralt-Rosinach, A. Gutierrez-Sacristan, J. Deu-Pons, E. Centeno, J. Garcia-Garcia, F. Sanz, and L. I. Furlong, “Disgenet: a comprehensive platform integrating information on human disease-associated genes

- and variants,” *Nucleic Acids Research*, vol. 45, no. D1, pp. D833–D839, 2017. [Citado na página 24]
- [64] U. Fayyad, G. PiatetskyShapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” *Ai Magazine*, vol. 17, no. 3, pp. 37–54, 1996. [Citado nas páginas 27, 28, 30 e 32]
- [65] M. J. Denny and A. Spirling, “Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it,” *Political Analysis*, vol. 26, no. 2, pp. 168–189, 2018. [Citado na página 29]
- [66] A. K. Uysal and S. Gunal, “The impact of preprocessing on text classification,” *Information Processing & Management*, vol. 50, no. 1, pp. 104–112, 2014. [Citado na página 29]
- [67] C. P. Chai, “Comparison of text preprocessing methods,” *Natural Language Engineering*, 2022. [Citado na página 29]
- [68] M. P. Singh, ed., *The Practical Handbook of Internet Computing*. New York: Chapman and Hall/CRC., 2004. [Citado na página 29]
- [69] D. Buenano-Fernandez, M. Gonzalez, D. Gil, and S. Lujan-Mora, “Text mining of open-ended questions in self-assessment of university teachers: An lda topic modeling approach,” *Ieee Access*, vol. 8, pp. 35318–35330, 2020. [Citado nas páginas 29, 32, 33 e 34]
- [70] A. Davi, D. Haughton, N. Nasr, G. Shah, M. Skaletsky, and R. Spack, “A review of two text-mining packages: Sas textmining and wordstat,” *American Statistician*, vol. 59, no. 1, pp. 89–103, 2005. [Citado na página 30]
- [71] R. Talib, M. K. Hanif, S. Ayesha, and F. Fatima, “Text mining: Techniques, applications and issues,” *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 11, pp. 414–418, 2016. [Citado na página 31]
- [72] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *Journal of Machine Learning Research*, vol. 3, no. 4-5, pp. 993–1022, 2003. [Citado nas páginas vii, 33 e 34]
- [73] M. M. Wang, S. B. Gao, W. L. Gui, J. Q. Ye, and S. S. Mi, “Investigation of pre-service teachers’ conceptions of the nature of science based on the lda model,” *Science & Education*, 2022. [Citado na página 33]
- [74] D. Maier, A. Waldherr, P. Miltner, G. Wiedemann, A. Niekler, A. Keinert, B. Pfetsch, G. Heyer, U. Reber, T. Haussler, H. Schmid-Petri, and S. Adam, “Applying lda topic modeling in communication research: Toward a valid and

- reliable methodology,” *Communication Methods and Measures*, vol. 12, no. 2-3, pp. 93–118, 2018. [Citado nas páginas 33, 34, 35 e 44]
- [75] Python Software Foundation, “Python.” Available at <https://www.python.org/>. (Last accessed in 20/09/2022). [Citado na página 38]
- [76] C. Zheng, G. He, and Z. Peng, “A study of web information extraction technology based on beautiful soup,” *JOURNAL OF COMPUTERS*, vol. 10, no. 6, pp. 381–387, 2015. [Citado na página 38]
- [77] L. Richardson, “Beautiful soup documentation.” Available at <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>. (Last accessed in 16/09/2022). [Citado na página 38]
- [78] ExplosionAI GmbH, “spacy: Facts & figures.” Available at <https://spacy.io/usage/facts-figures>. (Last accessed in 10/09/2022). [Citado na página 38]
- [79] Python Software Foundation, “Pypi: spacy.” Available at <https://pypi.org/project/spacy/>. (Last accessed in 18/08/2022). [Citado na página 39]
- [80] Python Software Foundation, “Pypi: wordcloud.” Available at <https://pypi.org/project/wordcloud/>. (Last accessed in 10/08/2022). [Citado na página 39]
- [81] Python Software Foundation, “Pypi: Pandas.” Available at <https://pypi.org/project/pandas/>. (Last accessed in 10/08/2022). [Citado na página 39]
- [82] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in python,” *JOURNAL OF MACHINE LEARNING RESEARCH*, vol. 12, pp. 2825–2830, 2011. [Citado na página 39]
- [83] MIT, ERCIM, Keio, and Beihang, “W3c: Html e css.” Available at <https://www.w3.org/standards/webdesign/htmlcss>. (Last accessed in 14/08/2022). [Citado na página 39]
- [84] Mozilla Corporation’s, “Mdn web docs: Javascript..” Available at <https://developer.mozilla.org/en-US/docs/Web/JavaScript>. (Last accessed in 14/08/2022). [Citado na página 39]
- [85] MIT, “Bootstrap: Site oficial.” Available at <https://getbootstrap.com/>. (Last accessed in 14/08/2022). [Citado na página 39]
- [86] ANSI, “Ansi: Sql..” Available at <https://blog.ansi.org/2018/10/sql-standard-iso-iec-9075-2016-ansi-x3-135/#gref>. (Last accessed in 11/08/2022). [Citado na página 39]

-
- [87] Netlify, “The pallets projects: Flask..” Available at <https://palletsprojects.com/p/flask/>. (Last accessed in 14/08/2022). [Citado na página 39]
- [88] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl, “Word cloud explorer: Text analytics based on word clouds,” in *2014 47TH HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES (HICSS)* (R. Sprague, ed.), Proceedings of the Annual Hawaii International Conference on System Sciences, pp. 1833–1842, Univ Hawaii, Shidler Coll Business; IEEE Comp Soc, 2014. [Citado na página 44]
- [89] DGECC - Geração-Geral de Estatísticas da Educação e Ciência, “Classificação de domínios científicos e tecnológicos, 2007 (fos)..” Available at https://estatistica.dgeec.mec.pt/docs/docs_cdr/Classificacao_FOS_VersaoPortuguesa.pdf. (Last accessed in 02/10/2022). [Citado na página 44]
- [90] M. Kleppmann, ed., *ODesigning Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. O’Reilly Media, 2017. [Citado na página 46]

Anexo A

Unidades de Investigação incluídas na ferramenta desen- volvida

Tabela A.1: Unidades de Investigação incluídas na ferramenta desen-
volvida.

Início da tabela	
Acrónimo	Nome
2Ai	Laboratório de Inteligência Artificial Aplicada
Age.Comm	Unidade de Investigação Interdisciplinar - Comunidades Envelhecidas Funcionais
ALGORITMI	Centro de Investigação ALGORITMI
ARTIS-IHA	ARTIS - Instituto de História da Arte, Faculdade de Letras, Universidade de Lisboa
BioISI	Instituto de Biosistemas & Ciências Integrativas
C-MADE	Centro de Materiais e Tecnologias Construtivas
C-MAST	Centro de Ciências e Tecnologias Mecânicas e Aeroespaciais
C2I2	Centro de Investigação em Cidades Inteligentes
C2TN	Centro de Ciências e Tecnologias Nucleares

Continuação da tabela A.1	
Acrónimo	Nome
CAMGSD	Centro de Análise Matemática, Geometria e Sistemas Dinâmicos
CAPP	Centro de Administração e Políticas Públicas
CARME	Centro de Investigação Aplicada em Gestão e Economia
CBA	Centro de Biotecnologia dos Açores
CBIOS	Centro de Investigação em Biociências e Tecnologias da Saúde
CBMA	Centro de Biologia Molecular e Ambiental
CBMR	Centro de Investigação em Biomedicina
CCMAR	Centro de Ciências do Mar do Algarve
CCUL	Centro Cardiovascular da Universidade de Lisboa
CDRSP	Centro para o Desenvolvimento Rápido e Sustentado de Produto
cE3c	Centro de Ecologia, Evolução e Alterações Ambientais
CEAD	Centro de Investigação em Educação de Adultos e Intervenção Comunitária
CEAFEL	Centro de Análise Funcional, Estruturas Lineares e Aplicações
CEAU	Centro de Estudos de Arquitectura e Urbanismo
CEAUL/ULICES	Centro de Estudos Anglísticos da Universidade de Lisboa
CEAUP	Centro de Estudos Africanos da Universidade do Porto
CeBER	Centro de Investigação em Economia e Gestão da Universidade de Coimbra
CEC	Centro de Estudos Comparatistas
CEC-FLUL	Centro de Estudos Clássicos
CECH/FL/UC	Centro de Estudos Clássicos e Humanísticos da Faculdade de Letras da Universidade de Coimbra
CECS	Centro de Estudos de Comunicação e Sociedade
CEDH	Centro de Investigação para o Desenvolvimento Humano
CEDIS	CEDIS - Centro de Investigação & Desenvolvimento sobre Direito e Sociedade
CeDRI	Centro de Investigação em Digitalização e Robótica Inteligente
CEEApIA	Centro de Estudos em Economia Aplicada do Atlântico

Continuação da tabela A.1	
Acrónimo	Nome
CEF	Centro de Estudos Florestais
cef.up	Centro de Economia e Finanças da Universidade do Porto
CEFAGE	Centro de Estudos e Formação Avançada em Gestão e Economia da Universidade de Évora
CeFEMA	Centro de Física e Engenharia de Materiais Avançados
CEFH	Centro de Estudos Filosóficos e Humanísticos
CEFITEC	Centro de Física e Investigação Tecnológica
CEFT	Centro de Estudos de Fenómenos de Transporte
CEG	Centro de Estudos Geográficos - Universidade de Lisboa
CEGE	Centro de Estudos de Gestão e Economia
CEHR-UCP	Centro de Estudos de História Religiosa
CEHUM	Centro de Estudos Humanísticos
CEI-IUL	Centro de Estudos Internacionais
CEID-CRCFL	Centro de Estudos e Investigação em Direito
CeiED	Centro de Estudos Interdisciplinares em Educação e Desenvolvimento
CEJEA	Centro de Estudos Jurídicos, Económicos e Ambientais
CEL	Centro de Estudos em Letras
CEMAT	Centro de Matemática Computacional e Estocástica
CEMMPRE	Centro de Engenharia Mecânica, Materiais e Processos
CENSE	Centro de Investigação em Ambiente e Sustentabilidade
CENTEC	Centro de Engenharia e Tecnologia Naval e Oceânica
CENTRA	Centro de Astrofísica e Gravitação
CEOS.PP	Centro de Estudos Organizacionais e Sociais do Politécnico do Porto
CEOT	Centro de Electrónica, Optoelectrónica e Telecomunicações
CEPESE	Centro de Estudos da População, Economia e Sociedade
CEPS	Centro de Ética, Política e Sociedade
CERENA	Centro de Recursos Naturais e Ambiente
CERIS	Instituto de Investigação e Inovação em Engenharia Civil para a Sustentabilidade

Continuação da tabela A.1	
Acrónimo	Nome
CERNAS	Centro de Estudos de Recursos Naturais, Ambiente e Sociedade
CES	Centro de Estudos Sociais
CESAM	Centro de Estudos do Ambiente e do Mar
CESEM	Centro de Estudos de Sociologia e Estética Musical
CETAPS	Centro de Estudos Ingleses, de Tradução e Anglo-Portugueses
CETRAD	Centro de Estudos Transdisciplinares para o Desenvolvimento
CF-UM-UP	Centro de Física das Universidades do Minho e do Porto
CFisUC	Centro de Física da Universidade de Coimbra
CFTC	Centro de Física Teórica e Computacional da Universidade de Lisboa
CFTP	Centro de Física Teórica de Partículas
CFUL	Centro de Filosofia da Universidade de Lisboa
CGEO	Centro de Geociências
CHAIA	Centro de História da Arte e Investigação Artística
CHRC	Centro de investigação Integrada em Saúde - Investigação, Educação e Inovação em Investigação Clínica e Saúde Pública
CHSC	Centro de História da Sociedade e da Cultura
CI-IPOP	Centro de Investigação do Instituto Português de Oncologia do Porto
CI&DEI	Centro de Estudos em Educação e Inovação
CIAC	Centro de Investigação em Artes e Comunicação - CIAC
CIAFEL	Centro de Investigação em Actividade Física, Saúde e Lazer
CIAS	Centro de Investigação em Antropologia e Saúde
CIAUD	Centro de Investigação em Arquitectura Urbanismo e Design
CIBB	Centro de Inovação em Biomedicina e Biotecnologia
CIBIT	Centro de Imagem Biomédica e Investigação Translacional
CICANT	Centro de Investigação em Comunicação Aplicada, Cultura e Novas Tecnologias
CICECO	CICECO-Instituto de Materiais de Aveiro

Continuação da tabela A.1	
Acrónimo	Nome
CICGE	Centro de Investigação em Ciências Geo-Espaciais
CICP	Centro de Investigação em Ciência Política
CICPSI	Centro de Investigação em Ciência Psicológica
CICS-UBI	Centro de Investigação em Ciências da Saúde
CICS.NOVA	Centro Interdisciplinar de Ciências Sociais
CIDAF	Centro de Investigação do Desporto e Actividade Física/Universidade de Coimbra
CIDEHUS.UÉ	Centro Interdisciplinar de História, Culturas e Sociedades da Universidade de Évora
CIDEM	Centro de Investigação e Desenvolvimento em Engenharia Mecânica
CIDMA	Centro de Investigação e Desenvolvimento em Matemática e Aplicações
CIDP	Centro de Investigação de Direito Público
CIDP	Centro de Investigação de Direito Privado
CIDPCC	Centro de Investigação em Direito Penal e Ciências Criminais
CIE	Centro de Investigação em Educação
CIE-ISPA	Centro de Investigação em Educação
CIEBA	Centro de Investigação e de Estudos em Belas-Artes
CIEC-UM	Centro de Investigação em Estudos da Criança
CIEd	Centro de Investigação em Educação
CIEG	Centro Interdisciplinar de Estudos de Género
CIEPQPF	Centro de Investigação em Engenharia dos Processos Químicos e dos Produtos da Floresta
CIEQV	Centro de Investigação em Qualidade de Vida
CIES-IUL	Centro de Investigação e Estudos de Sociologia
CIETI	Centro de Inovação em Engenharia e Tecnologia Industrial
CIIC	Centro de Investigação em Informática e Comunicações
CIICESI	Centro de Inovação e Investigação em Ciências Empresariais e Sistemas de Informação
CIIE - U.Porto	Centro de Investigação e Intervenção Educativas
CiiEM	Centro de Investigação Interdisciplinar Egas Moniz
CIIMAR	Centro Interdisciplinar de Investigação Marinha e Ambiental

Continuação da tabela A.1

Acrónimo	Nome
CIISA	Centro de Investigação Interdisciplinar em Sanidade Animal
CIJE	Centro de Investigação Jurídico-Económica
CIMA UALG	Centro de Investigação Marinha e Ambiental
CIMO	Centro de Investigação de Montanha
CINEA-IPS	Centro de Investigação em Energia e Ambiente
CINEICC	Centro de Investigação em Neuropsicologia e Intervenção Cognitivo Comportamental
CINTESIS	Centro de Investigação em Tecnologias e Serviços de Saúde
CinTurs	Centro de Investigação em Turismo, Sustentabilidade e Bem-estar
CIP	Centro de Investigação em Psicologia
CIPD	Centro de Investigação em Psicologia para o Desenvolvimento
CIPER	Centro Interdisciplinar de Estudo da Performance Humana
CIPES	Centro de Investigação de Políticas do Ensino Superior
CIPES	Centro de Investigação em Política, Economia e Sociedade
CIPsi	Centro de Investigação em Psicologia
CIQUP	Centro de Investigação em Química da Universidade do Porto
CIR	Centro de Investigação em Reabilitação
CIS-IUL	Centro de Investigação e Intervenção Social
CISA	Centro de Investigação em Saúde e Ambiente
CISE	Centro de Investigação em Sistemas Electromecatrónicos
CISTER	Centro de Investigação em Sistemas Computacionais Embebidos e de Tempo-Real
CISUC	Centro de Informática e Sistemas da Universidade de Coimbra
CITAB	Centro de Investigação e de Tecnologias Agro-Ambientais e Biológicas
CITAD	Centro de Investigação em Território, Arquitectura e Design
CITAR	Centro de Investigação em Ciência e Tecnologia das Artes

Continuação da tabela A.1	
Acrónimo	Nome
CITCEM	Centro de Investigação Transdisciplinar Cultura, Espaço e Memória
ciTechCare	Centro de Inovação em Tecnologias e Cuidados de Saúde
CITEUC	Centro de Investigação da Terra e do Espaço da Universidade de Coimbra
CITTA	Centro de Investigação do Território, Transportes e Ambiente
CI TUA	Centro para a Inovação em Território, Urbanismo e Arquitetura
CI TUR	Centro de Investigação, Desenvolvimento e Inovação em Turismo
CIUHCT	Centro Interuniversitário de Historia das Ciencias e da Tecnologia
CLIPIS	Centro Lusíada de Investigação em Política Internacional e Segurança
CLISSIS	Centro Lusíada de investigação em Serviço Social e Intervenção Social
CLUL	Centro de Linguística da Universidade de Lisboa
CLUNL	Centro de Linguística da Universidade Nova de Lisboa
CLUP	Centro de Linguística da Universidade do Porto
CMA-UBI	Centro de Matemática e Aplicações da Universidade da Beira Interior
CMA/FCT/UNL	Centro de Matemática e Aplicações
CMAFCIO	Centro de Matemática, Aplicações Fundamentais e Investigação Operacional
CMAT	Centro de Matemática da Universidade do Minho
CMEMS-UMinho	Unidade de Investigação em Microssistemas Eletromecânicos
CMUC	Centro de Matemática da Universidade de Coimbra
CMUP	Centro de Matemática da Universidade do Porto
COMEGI	Centro de Investigação em Organizações, Mercados e Gestão Industrial
CONSTRUCT	Instituto de I&D em Estruturas e Construções
COPELABS	Computação Cognitiva e Centrada nas Pessoas
CPUP	Centro de Psicologia da Universidade do Porto
CQ-UM	Centro de Química da Universidade do Minho
CQC	Centro de Química de Coimbra

Continuação da tabela A.1	
Acrónimo	Nome
CQE	Centro de Química Estrutural
CQM	Centro de Química da Madeira
CQVR	Centro de Química - Vila Real
CR	Programa Champalimaud de Investigação
CRC-W	Centro de Estudos de Bem-Estar Psicológico, Familiar e Social
CRIA	Centro em Rede de Investigação em Antropologia
CSG	CSG - Investigação em Ciências Sociais e Gestão
CTS	Centro de Tecnologias e Sistemas
DigiMedia	Centro de Investigação em Média Digitais e Interação
DINÂMIA'CET-IUL	DINÂMIA'CET-IUL, Centro de Estudos Sobre a Mudança Socioeconómica e o Território
DiSAC	Unidade de I&D em Serviços, Aplicações e Conteúdos Digitais
DREAMS	Centro Interdisciplinar de Desenvolvimento e Investigação em Ambiente, Gestão Aplicada e Espaço
EPIUnit	Unidade de Investigação em Epidemiologia - Instituto de Saúde Pública da Universidade do Porto
esad-idea	esad idea, investigação em design e arte
FibEnTech	Materiais Fibrosos e Tecnologias Ambientais
FP-ENAS	Unidade de Investigação UFP em Energia, Ambiente e Saúde
GECAD	Grupo de Investigação em Engenharia e Computação Inteligente para a Inovação e o Desenvolvimento
GFMUL	Grupo de Física-Matemática da Universidade de Lisboa
GILT	Games, Interaction & Learning Technologies
GREEN-IT	Biorecursos para a Sustentabilidade
GreenUPorto	Centro de investigação em Produção Agroalimentar Sustentável
H&TRC	Centro de Investigação em Saúde e Tecnologia
HEI-Lab	Laboratórios Digitais de Ambientes e Interações Humanas
HERCULES	Laboratório HERCULES - Herança Cultural, Estudos e Salvaguarda
i2ADS	Instituto de Investigação em Arte, Design e Sociedade
i3N	Instituto de Nanoestruturas, Nanomodelação e Nanofabricação

Continuação da tabela A.1	
Acrónimo	Nome
i3S	Instituto de Investigação e Inovação em Saúde
IA	Instituto de Astrofísica e Ciências do Espaço
iBB	Instituto de Bioengenharia e Biociências
IBEB	Instituto de Biofísica e Engenharia Biomédica
ICArEHB	Centro Interdisciplinar de Arqueologia e Evolução do Comportamento Humano
ICNOVA	Instituto de Comunicação da NOVA
ICS-ULisboa	Instituto de Ciências Sociais da Universidade de Lisboa
ICT	Instituto de Ciências da Terra
ID+	Instituto de Investigação em Design, Media e Cultura
IDL	Instituto Dom Luiz
IEETA	Instituto de Engenharia Eletrónica e Informática de Aveiro
IEF	Instituto de Estudos Filosóficos
IELT	Instituto de Estudos de Literatura e Tradição - Patrimónios, Artes e Culturas
IEM	Instituto de Estudos Medievais
IF	Instituto de Filosofia
IFILNOVA	Instituto de Filosofia da Nova
IFIMUP	Instituto de Física de Materiais Avançados, Nanotecnologia e Fotónica - Universidade do Porto
IHA	Instituto de História da Arte
IHC	Instituto de História Contemporânea
IINFACTS	Instituto de Investigação e Formação Avançada em Ciências e Tecnologias da Saúde
IITAA (CITA-A)	Instituto de Investigação e Tecnologia Agrária e do Ambiente dos Açores
IJ	Instituto Jurídico da Faculdade de Direito da Universidade de Coimbra
IJP	IJP - Instituto Jurídico Portucalense
ILC	Instituto de Literatura Comparada
iMed.ULisboa	Instituto de Investigação do Medicamento
iMM	Instituto de Medicina Molecular
InBIO	Rede de Investigação em Biodiversidade e Biologia Evolutiva
inED	Centro de Investigação e Inovação em Educação

Continuação da tabela A.1	
Acrónimo	Nome
INESC Coimbra	Instituto de Engenharia de Sistemas e Computadores de Coimbra
INESC MN	INESC Microsistemas e Nanotecnologias - Instituto de Engenharia de Sistemas e Computadores para os Microsistemas e as Nanotecnologias
INESC TEC	INESC TEC - INESC Tecnologia e Ciência
INESC-ID	Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa
INET-md	Instituto de Etnomusicologia - Centro de Estudos em Música e Dança
iNOVA4Health	iNOVA4Health - Programa de Medicina Translacional
INPP	Instituto de Desenvolvimento Humano Portucalense
IO	Instituto do Oriente
IPC	Instituto de Polímeros e Compósitos
IPFN	Instituto de Plasmas e Fusão Nuclear
IPRI-NOVA	Instituto Português de Relações Internacionais - Universidade NOVA de Lisboa
IS-UP	Instituto de Sociologia da Universidade do Porto
ISAMB	Instituto de Saúde Ambiental
ISISE	Instituto para a Sustentabilidade e Inovação em Estruturas de Engenharia
ISR-UC	Instituto de Sistemas e Robótica - ISR – COIMBRA
ISRC	Centro de Investigação em Estudos Interdisciplinares
ISTAR-IUL	Centro de Investigação em Ciências da Informação, Tecnologias e Arquitetura
IT	Instituto de Telecomunicações
JusGov	Centro de Investigação em Justiça e Governação
Lab2PT	Laboratório de Paisagens, Património e Território
LabCom	LabCom - Comunicação e Artes
LAETA	Laboratório Associado de Energia, Transportes e Aeronáutica
LARSyS	Laboratório de Robótica e Sistemas de Engenharia
LASIGE	LASIGE - Extreme Computing
LEAF	Centro de Investigação em Agronomia, Alimentos, Ambiente e Paisagem
LEAU	Laboratório Experimental de Arquitectura e Urbanismo

Continuação da tabela A.1	
Acrónimo	Nome
LEPABE	Laboratório de Engenharia de Processos, Ambiente, Biotecnologia e Energia
LIBPhys	Laboratório de Instrumentação, Engenharia Biomédica e Física da Radiação
LIDA	Laboratório de Investigação em Design e Artes
LIP	Laboratório de Instrumentação e Física Experimental de Partículas
LSRE-LCM	Laboratório de Processos de Separação e Reacção - Laboratório de Catálise e Materiais
MARE	Centro de Ciências do Mar e do Ambiente
MedInUP	Centro de Investigação Farmacológica e Inovação Medicamentosa
MEtRICs	Centro de Engenharia Mecânica e Sustentabilidade de Recursos
MOSTMICRO-ITQB	Microbiologia Molecular, Estrutural e Celular - Instituto de Tecnologia Química e Biológica António Xavier – MOSTMICRO
NECE	Núcleo de Estudos em Ciências Empresariais
NIPE	Núcleo de Investigação em Políticas Económicas e Empresariais
Nova SBE	Nova School of Business and Economics
OBSERVARE	Observatório de Relações Exteriores
Okeanos - UAc	Centro I&D Okeanos - Universidade dos Açores
PRAXIS	PRAXIS - Centro de Filosofia, Política e Cultura
QFM-UC	Unidade de I&D Química-Física Molecular
RATIO LEGIS	RATIO LEGIS - Centro de Investigação e desenvolvimento em ciências Jurídicas
RECI	Unidade de Investigação em Educação e Intervenção Comunitária
REM	Research in Economics and Mathematics
REMIT	REMIT - Investigação em Economia, Gestão e Tecnologias da Informação
REQUIMTE	Laboratório Associado para a Química Verde - Tecnologias e Processos Limpos
SYSTEC	Centro de Sistemas e Tecnologias
TECHN&ART	Centro de Tecnologia, Restauro e Valorização das Artes

Continuação da tabela A.1

Acrónimo	Nome
UACC	Laboratório de Inteligência Artificial e Ciência de Computadores
UCIBIO	Unidade de Ciências Biomoleculares Aplicadas
UCISA: E	Unidade de Investigação em Ciências da Saúde: Enfermagem
UIDEF	Unidade de Investigação e Desenvolvimento em Educação e Formação
UMIB	Unidade Multidisciplinar de Investigação Biomédica
UnIC	Unidade de Investigação e Desenvolvimento Cardiovascular
UNIDCOM/IADE	Unidade de Investigação em Design e Comunicação - UNIDCOM/IADE
VALORIZA	Centro de Investigação para a Valorização de Recursos Endógenos
VICARTE	Vidro e Cerâmica para as Artes
WJCR	Centro de Investigação William James

Fim da tabela

Anexo B

Empresas incluídas na ferramenta desenvolvida

Tabela B.1: Empresas incluídas na ferramenta desenvolvida.

Início da tabela

Nome
5 ESTAÇÕES - INFORMAÇÃO E SERVIÇOS, LDA
A.FERREIRA & FILHOS, S.A.
A.SAMPAIO & FILHOS - TÊXTEIS, S.A.
A.T.B. - ACABAMENTOS TÊXTEIS DE BARCELOS, LDA
AAPICO ÁGUEDA, S.A.
AAPICO MAIA, S.A.
ABBVIE, LDA
ABER HYDRAULICS, S.A.
ACI - AUTOMOTIVE COMPOUNDING INDUSTRY, LDA
ACLIMPEX, LDA
ACTUASYS, LDA
ADICO-ADELINO DIAS COSTA - MOBILIÁRIO METÁLICO, LDA
ADP FERTILIZANTES, S.A.
AGROLEX II - RAÇÕES, LDA
ALELUIA - CERÂMICAS, S.A.
ALGARDATA - SISTEMAS INFORMÁTICOS, S.A.

Continuação da tabela B.1

Nome

ALGORITMIA ÁGIL - SISTEMAS DE INFORMAÇÃO, LDA
ALPHABET DIMENSION, LDA
ALS LIFE SCIENCES PORTUGAL, S.A.
ALVES RIBEIRO, S.A.
AMAZON & FORMOSA, LDA
AMBAR - PASSION, S.A.
AMCOR FLEXIBLES PORTUGAL, LDA
AMORIM CORK COMPOSITES, S.A.
AMT - CONSULTING, S.A.
AMTROL-ALFA - METALOMECÂNICA, S.A.
ÂNGELO COIMBRA, S.A.
ANKIX SYSTEMS, LDA
ANSELL PORTUGAL - INDUSTRIAL GLOVES, SOCIEDADE UNIPES-
SOAL, LDA
APPGENERATION - SOFTWARE TECHNOLOGIES, LDA
APPS4MOBILITY INTERNATIONAL, LDA
AQUAPOR - SERVIÇOS, S.A.
ARAMAGUE - FÁBRICA DE ARTIGOS DE ARAME, S.A.
ARCABOA - INDÚSTRIA DE FRIO, S.A.
ARPITEX - COMÉRCIO E INDÚSTRIA DE CONFECÇÕES, LDA
ARSOPI - INDÚSTRIAS METALÚRGICAS ARLINDO S.PINHO,
LDA
ARTINVOGUE, S.A.
ARVÓLEA, SOCIEDADE UNIPES-
SOAL, LDA
ASM TAPS, S.A.
ASPÖCK PORTUGAL, S.A.
ASSECO PST PORTUGAL - BUSINESS & SOFTWARE SOLUTI-
ONS, S.A.
AUCHAN RETAIL PORTUGAL, S.A.
AUDAXYS - SOFTWARE E SISTEMAS, S.A.
AUTO RIBEIRO, LDA
AVE - GESTÃO AMBIENTAL E VALORIZAÇÃO ENERGÉTICA,
S.A.
AVELEDA, S.A.
AXIANSEU - DIGITAL SOLUTIONS, S.A.
BAÇA VIAGENS E TURISMO, LDA
BAQUELITE LIZ, S.A.
BARATA & RAMILO, S.A.

Continuação da tabela B.1

Nome

BARBOT - INDÚSTRIA DE TINTAS, S.A.
BARROS & MOREIRA, S.A.
BASTOS VIEGAS, S.A.
BBG, S.A.
BEAMIAN, S.A.
BEBÉ VIDA - CIÊNCIAS PARA A VIDA, S.A.
BEE2FIRE, S.A.
BEYONDEVICES, S.A.
BI-SILQUE - PRODUTOS DE COMUNICAÇÃO VISUAL, S.A.
BICAFÉ - TORREFACÇÃO E COMÉRCIO DE CAFÉ, LDA
BIDIRCO - ENGINEERING SYSTEMS, LDA
BIMMS, LDA
BIOGEN PORTUGAL - SOCIEDADE FARMACÊUTICA, UNIPES-
SOAL, LDA
BIOJAM, S.A.
BIQ - HEALTH SOLUTIONS, LDA
BISELARTE - SOCIEDADE DE VIDROS, S.A.
BLACHERE PORTUGAL - COMÉRCIO DE ILUMINAÇÃO, LDA
BLOCOTELHA - STEEL CONSTRUCTIONS, S.A.
BLUEPHARMA - INDÚSTRIA FARMACÊUTICA, S.A.
BÖLLINGHAUS STEEL, LDA
BONDUELLE (PORTUGAL) - AGROINDÚSTRIA, S.A.
BOSCH CAR MULTIMÉDIA PORTUGAL, S.A.
BOSCH SECURITY SYSTEMS - SISTEMAS DE SEGURANÇA, S.A.
BRAIN MATTERS - INOVAÇÃO E TECNOLOGIA PORTUGAL,
UNIPES-
SOAL, LDA
BRESFOR - INDÚSTRIA DO FORMOL, S.A.
BRISTOL - MYERS SQUIBB FARMACÊUTICA PORTUGUESA,
S.A.
BSOLUS - BUSINESS SOLUTIONS, LDA
BY FOODS, S.A.
BYSIDE - LEAD ACTIVATION, S.A.
CABOPOL - POLYMER COMPOUNDS, S.A.
CALIBRAFRUTA - EQUIPAMENTOS HORTOFRUTÍCOLAS, LDA
CAMBRA-SISTEM - ENGENHARIA DE SISTEMAS, S.A.
CARD4B - SYSTEMS, S.A.
CAREWILL, LDA
CARLOS ALBERTO & FILHOS, S.A

 Continuação da tabela B.1

Nome

CASAS EM MOVIMENTO, LDA
 CATARI INDÚSTRIA, S.A.
 CAUDALIE PORTUGAL, UNIPessoal, LDA
 CELFINET - CONSULTORIA EM TELECOMUNICAÇÕES, S.A.
 CELFOCUS, S.A.
 CELOPLÁS - PLÁSTICOS PARA A INDÚSTRIA, S.A.
 CENTROCAR - CENTRO DE EQUIPAMENTOS MECÂNICOS, S.A.
 CEREALIS - PRODUTOS ALIMENTARES, S.A.
 CEREALTO SINTRA FOODS, S.A.
 CERIART - CERÂMICA ARTÍSTICA, S.A.
 CERISOL - ISOLADORES CERÂMICOS, S.A.
 CERTECA - INDÚSTRIAS CERÂMICAS, S.A.
 CEV, S.A.
 CEVARGADO - ALIMENTOS COMPOSTOS, UNIPessoal, LDA
 CHEMITEK - QUÍMICA AVANÇADA, S.A.
 CHIPEMULATOR MPWE, LDA
 CICLO-FAPRIL - INDÚSTRIAS METALÚRGICAS, S.A.
 CIN - CORPORAÇÃO INDUSTRIAL DO NORTE, S.A.
 CINCA - COMPANHIA INDUSTRIAL DE CERÂMICA, S.A.
 CIPAN - COMPANHIA INDUSTRIAL PRODUTORA DE ANTIBIÓ-
 TICOS, S.A.
 CLARANET PORTUGAL, S.A.
 CLDN AGENCIES, UNIPessoal, LDA
 CLDN CARGO PORTUGAL, UNIPessoal, LDA
 COFACO AÇORES - INDÚSTRIA DE CONSERVAS, S.A.
 COLEP CONSUMER PRODUCTS PORTUGAL, S.A.
 COLQUÍMICA - INDÚSTRIA NACIONAL DE COLAS, S.A.
 COMPONIT, LDA
 COMPTA - EMERGING BUSINESS, S.A.
 CONFEITARIA MONTEVERDE, LDA
 CONFETIL, S.A.
 CONSULAI - CONSULTORIA AGRO-INDUSTRIAL, LDA
 CONTINENTAL MABOR - INDÚSTRIA DE PNEUS, S.A.
 COPO TÊXTIL PORTUGAL, S.A.
 CORAL - VISION EUROPA, S.A.
 CORK SUPPLY PORTUGAL 4 - ROLHAS TÉCNICAS, UNIPES-
 SOAL, LDA
 CORK SUPPLY PORTUGAL, S.A.

Continuação da tabela B.1

Nome

CORTADORIA NACIONAL DE PÊLO, S.A.
COSME TEK, LDA
COSTA & DIAS, LDA
COTESI - COMPANHIA DE TÊXTEIS SINTÉTICOS, S.A.
COURO AZUL - INDÚSTRIA E COMÉRCIO DE COUROS, S.A.
CREDIN PORTUGAL - PRODUTOS ALIMENTARES, S.A.
CRITICAL TECHWORKS, S.A.
CRSA - CONNECT ROBOTICS SISTEMAS AUTÓNOMOS, LDA
CRUZ MARTINS & WAHL, LDA
CS - COELHO DA SILVA, S.A.
CTR - CONSULTORIA, TÉCNICA E REPRESENTAÇÕES, LDA
DAMACENO & ANTUNES - TECIDOS DE DECORAÇÃO, LDA
DAMEL - CONFECÇÃO DE VESTUÁRIO, LDA
DANIPACK - INDÚSTRIA DE PLÁSTICOS, S.A.
DECUNIFY - SOLUÇÕES DE COMUNICAÇÕES, S.A.
DEIMOS - ENGENHARIA, S.A.
DERMA-LEATHER - COMÉRCIO E INDÚSTRIA DE PELES, S.A.
DERMOTECA - PRODUTOS QUÍMICOS E DERMATOLÓGICOS,
S.A.
DESTINOS - ARTE CERÂMICA, S.A.
DF - ELASTOMER SOLUTIONS, LDA
DIERA - FÁBRICA DE REVESTIMENTOS, COLAS E TINTAS, LDA
DIFERENCIAL - ELECTROTÉCNICA GERAL, LDA
DIGIWEST - WIRELESS AND EMBEDDED SOLUTIONS, LDA
DOCUMÁTICA - SISTEMAS DE PROCESSAMENTO DE MEIOS DE
PAGAMENTO E IMAGEM, LDA
DOMINGOS DE SOUSA & FILHOS, S.A.
DOMINÓ - INDÚSTRIAS CERÂMICAS, S.A.
DR.JOAOQUIM CHAVES - LABORATÓRIO DE ANÁLISES CLÍ-
CAS, S.A.
DRT PLASTIC INJECTION, LDA
DUAL BORGSTENA TEXTILE PORTUGAL, UNIPessoal, LDA
EBANKIT - OMNICHANNEL INNOVATION, S.A.
ECCO'LET (PORTUGAL) - FÁBRICA DE SAPATOS, LDA
ECOSTEEL, S.A.
EDISOFT - EMPRESA DE SERVIÇOS E DESENVOLVIMENTO DE
SOFTWARE, S.A.
EDITORA HOGREFE, LDA

 Continuação da tabela B.1

Nome

EFACEC ENERGIA - MÁQUINAS E EQUIPAMENTOS ELÉCTRICOS, S.A.
 EFAFLU - BOMBAS E VENTILADORES, S.A.
 EFAPEL - EMPRESA FABRIL DE PRODUTOS ELÉCTRICOS, S.A.
 EGITRON - ENGENHARIA E AUTOMAÇÃO INDUSTRIAL, LDA
 EI-TAG, LDA
 EMBALCER - EQUIPAMENTOS E MATERIAIS DE EMBALAGEM, LDA
 ENAME, S.A.
 ENDIPREV, S.A.
 ENDUTEX - REVESTIMENTOS TÊXTEIS, S.A.
 ENEIDA, WIRELESS & SENSORS, S.A.
 ENERGY PULSE SYSTEMS, LDA
 ENKROTT - GESTÃO E TRATAMENTO DE ÁGUAS, S.A.
 ENZIFARMA - DIAGNÓSTICA E FARMACÊUTICA, S.A.
 EPOLI - ESPUMAS DE POLIETILENO, S.A.
 ERNESTO MORGADO, S.A.
 ESMALGLASS (PORTUGAL) - PRODUTOS CERÂMICOS, S.A.
 ESMALTICER - ESMALTES CERÂMICOS, LDA
 ESMOMOLDE - FABRICO DE MOLDES, LDA
 ETICADATA - SOFTWARE, LDA
 EUMEL - EMPRESA DE UTILIDADES METÁLICAS, LDA
 EUROFROZEN - INDÚSTRIA E COMÉRCIO DE PRODUTOS ALIMENTARES, S.A.
 EUROPASTRY PORTUGAL, S.A.
 EURORALEX - CONFECÇÕES, S.A.
 EUROSHELVES - ESTANTES METÁLICAS, LDA
 EUROSURFACES PORTUGAL, S.A.
 EVIOSYS PACKAGING PORTUGAL, S.A.
 EXATRONIC, LDA
 EXIDE TECHNOLOGIES, LDA
 EXPORPLÁS - INDÚSTRIA DE EXPORTAÇÃO DE PLÁSTICOS, S.A.
 FABOR - FÁBRICA DE ARTEFACTOS DE BORRACHA, S.A.
 FÁBRICA DE CONSERVAS A POVEIRA, S.A.
 FACAL - ENGENHARIA DE FACHADAS, LDA
 FACOTIL - FÁBRICA DE COLAS E TINTAS, S.A.
 FAFEDRY - FAST FASHION FINISHING, UNIPessoal, LDA

Continuação da tabela B.1

Nome

FAIRJOURNEY BIOLOGICS, S.A.
FAMO - INDÚSTRIA DE MOBILIÁRIO DE ESCRITÓRIO, LDA
FAMOLDE - FABRICAÇÃO E COMERCIALIZAÇÃO DE MOLDES,
S.A.
FANCYSTAGE, UNIPessoal, LDA
FARFETCH PORTUGAL, UNIPessoal, LDA
FASSALUSA - PRODUÇÃO E COMERCIALIZAÇÃO DE MATERI-
AIS DE CONSTRUÇÃO, LDA
FASTER - PRODUTOS ALIMENTARES, LDA
FAURECIA - ASSENTOS DE AUTOMÓVEL, LDA
FCC AQUALIA, S.A. - SUCURSAL EM PORTUGAL
FERESPE - FUNDIÇÃO DE FERRO E AÇO, LDA
FERPINTA - INDÚSTRIAS DE TUBOS DE AÇO DE FERNANDO
PINHO TEIXEIRA, S.A.
FERREIRA MARTINS & FILHOS - MADEIRAS E DERIVADOS, S.A.
FERRO - INDÚSTRIAS QUÍMICAS (PORTUGAL), LDA
FERTIPRADO - SEMENTES E NUTRIENTES, LDA
FIBOPE PORTUGUESA - FILMES BIORIENTADOS, S.A.
FICO CABLES - FÁBRICA DE ACESSÓRIOS E EQUIPAMENTOS
INDUSTRIAIS, LDA
FILKEMP - INDÚSTRIA DE FILAMENTOS, S.A.
FIRST SOLUTIONS - SISTEMAS DE INFORMAÇÃO, S.A.
FISOLA - FÁBRICA DE ISOLADORES ELÉTRICOS, LDA
FLEXACO - CONCENTRADOS E ADITIVOS PLÁSTICOS, S.A.
FLEXITEX - FÁBRICA DE TECIDOS, S.A.
FLEXPUR - POLÍMEROS DE POLIURETANO, S.A.
FLUPOL - SURFACE ENGINEERING, S.A.
FONT SALEM PORTUGAL, S.A.
FOOT BY FOOT - REVESTIMENTOS TÊXTEIS, LDA
FOURSTEEL - PRODUTOS EM INOX, LDA
FRAVIZEL - EQUIPAMENTOS METALOMECÂNICOS, S.A.
FREITAS & SILVA, S.A.
FRESENIUS KABI PHARMA PORTUGAL, LDA
FREZITE - FERRAMENTAS DE CORTE, S.A.
FUJITSU TECHNOLOGY SOLUTIONS, LDA
FUNDILUSA - FUNDIÇÕES PORTUGUESAS, LDA
FWD, S.A.
G9TELECOM, S.A.

Continuação da tabela B.1

Nome

GARDENGATE, S.A.
GENE EXPRESS - SERVIÇOS GENÓMICOS PARA DIAGNÓSTICO E INVESTIGAÇÃO, LDA
GEOINVESTE - COMÉRCIO E INDÚSTRIA DE ROCHAS, S.A.
GEORGINA NETO DOS SANTOS, LDA
GEPACK - EMPRESA TRANSFORMADORA DE PLÁSTICOS, S.A.
GEPROCEA I - SERVIÇOS DE ENGENHARIA, LDA
GL INTERNATIONAL FOOD, S.A.
GLN - ENGINEERING, MOLDING AND PLASTICS, S.A.
GLNMOLDS, S.A.
GLNPLAST, S.A.
GLOBALSOFT - CLOUD BUSINESS AND SOFTWARE CONSULTING, LDA
GLORY GLOBAL SOLUTIONS (PORTUGAL), S.A.
GONFRIO - INDÚSTRIA DE FRIO, S.A.
GOTELECOM, S.A.
GRAPHENEST, S.A.
GRAPHICSLEADERPACKAGING - ARTES GRÁFICAS, S.A.
GRÉS PANARIA PORTUGAL, S.A.
GRETEL - PRODUTOS CERÂMICOS, S.A.
GROHE PORTUGAL - COMPONENTES SANITÁRIOS, LDA
GRUPO ANTOLIN LUSITÂNIA - COMPONENTES AUTOMÓVEL, UNIPESSOAL, LDA
GUM CHEMICAL SOLUTIONS, S.A.
H.B.FULLER, ISAR - RAKOLL, S.A.
HABIDECOR - INDÚSTRIA TÊXTIL PARA HABITAÇÃO, S.A.
HENRIQUE VIEIRA & FILHOS, S.A.
HEXASTEP, S.A.
HFA - HENRIQUE, FERNANDO & ALVES S.A.
HIDRO-LINHA - SOCIEDADE COMERCIAL DE COMPONENTES HIDRAULICOS, LDA
HIH, S.A.
HIKMA FARMACÊUTICA (PORTUGAL), S.A.
HITACHI ASTEMO ABRANTES, S.A.
HOVIONE FARMACIÊNCIA, S.A.
HPRD - HEALTH PRODUCTS RESEARCH AND DEVELOPMENT, LDA
HRV - EQUIPAMENTOS DE PROCESSO, S.A.

 Continuação da tabela B.1

Nome

HUSQVARNA PORTUGAL, S.A.
 HUTCHINSON (PORTO) - TUBOS FLEXÍVEIS, SOCIEDADE UNIPES-
 SOAL, LDA
 HYDRACOOING, LDA
 HYPELABS, UNIPES-
 SOAL, LDA
 I-SENSIS PERFUME DESIGN, LDA
 I.C.C. - INDÚSTRIAS E COMÉRCIO DE CALÇADO, S.A.
 IBER-OLEFF - COMPONENTES TÉCNICOS EM PLÁSTICO, S.A.
 IBERFAR - INDÚSTRIA FARMACÊUTICA, S.A.
 ICNAS - PRODUÇÃO, UNIPES-
 SOAL, LDA
 IDMIND - ENGENHARIA DE SISTEMAS, LDA
 IMMUNETHEP, S.A.
 IMOTOUR - SOCIEDADE DE TURISMO, S.A.
 IMPERIAL - PRODUTOS ALIMENTARES, S.A.
 IMPETUS PORTUGAL - TÊXTEIS, S.A.
 IMPRENSA NACIONAL - CASA DA MOEDA, S.A.
 INCOMEUF - ENGINEERING, LDA
 INDASA - INDÚSTRIA DE ABRASIVOS, S.A.
 INDELAGUE - INDÚSTRIA ELÉTRICA DE ÁGUEDA, S.A.
 INDIE CAMPERS CENTRAL SERVICES, UNIPES-
 SOAL, LDA
 INDINOR - INDÚSTRIAS QUÍMICAS, S.A.
 INDÚSTRIAS METÁLICAS VENEPORTE, S.A.
 INDUTI TECNOLOGIA, UNIPES-
 SOAL, LDA
 INFOS - INFORMÁTICA E SERVIÇOS, S.A.
 INICIATIVAS DE MEIOS - ACTIVIDADES PUBLICITÁRIAS, LDA
 INJEX - PINHEIRO DE LACERDA, LDA
 INOVA+ - INNOVATION SERVICES, S.A.
 INOVAFIL - FIAÇÃO, S.A.
 INOVAPRIME - SERVIÇOS EM TECNOLOGIAS DE INFORMA-
 ÇÃO, LDA
 INOVRETAIL, S.A.
 INOXANTUA - INSTALAÇÕES EM INOX, LDA
 INSTITUTO NCFORENSES - CIÊNCIAS FORENSES, UNIPES-
 SOAL, LDA
 INTERECYCLING - SOCIEDADE DE RECICLAGEM, S.A.
 INTERMOLDE - MOLDES VIDREIROS INTERNACIONAIS, LDA
 INTRAPLÁS - INDÚSTRIA TRANSFORMADORA DE PLÁSTICOS,
 S.A.

 Continuação da tabela B.1

Nome

INTRUM PORTUGAL, UNIPESSOAL, LDA
 INVEPE - INDÚSTRIA DE VEÍCULOS PESADOS, S.A.
 IRMÃOS VILA NOVA, S.A.
 ISOLAGO - INDÚSTRIA DE PLÁSTICOS, S.A.
 ISRETAIL - SOLUÇÕES INFORMÁTICAS DE RETALHO, LDA
 ISSHO TECHNOLOGY, LDA
 ITALAGRO - INDÚSTRIA DE TRANSFORMAÇÃO DE PRODUTOS ALIMENTARES, S.A.
 ITELMATIS, LDA
 ITSCRED - INFORMATION TECHNOLOGY, S.A.
 ITSECTOR - SISTEMAS DE INFORMAÇÃO, S.A.
 J.CANÃO, LDA
 J.J.M.ESPERANÇA, LDA
 JACINTO MARQUES DE OLIVEIRA - SUCESSORES, LDA
 JACL PRODUCTS - PRODUÇÃO DE COMPONENTES AUTOMÓVEIS, UNIPESSOAL, LDA
 JANELA DIGITAL - INFORMÁTICA E TELECOMUNICAÇÕES, S.A.
 JBMC - INDÚSTRIA DE PLÁSTICOS, S.A.
 JOALPE - INDÚSTRIA DE EXPOSITORES, S.A.
 JOFEBAR, S.A.
 JOPER - INDÚSTRIA DE EQUIPAMENTOS AGRÍCOLAS, S.A.
 JOSÉ ANTÓNIO SAMPAIO TEIXEIRA & COMPANHIA, LDA
 JOSÉ JÚLIO JORDÃO, LDA
 JPM - AUTOMAÇÃO E EQUIPAMENTOS INDUSTRIAIS, S.A.
 JSL - MATERIAL ELÉCTRICO, S.A.
 JULAR - MADEIRAS, S.A.
 KCSIT, S.A.
 KEMET ELECTRONICS PORTUGAL, S.A.
 KEMI - PINE ROSINS PORTUGAL, S.A.
 KENETTO, LDA
 KNOWLEDGEWORKS - CONSULTORIA EM SISTEMAS DE INFORMAÇÃO, LDA
 KÖRBER SUPPLY CHAIN PT, S.A.
 LABORATÓRIO EDOL - PRODUTOS FARMACÊUTICOS, S.A.
 LABORATÓRIOS ATRAL, S.A.
 LABORATÓRIOS BASI - INDÚSTRIA FARMACÊUTICA, S.A.
 LACTICÍNIOS DO PAIVA, S.A.

Continuação da tabela B.1

Nome

LANKHORST EURONETE PORTUGAL, S.A.
LEICA - APARELHOS ÓPTICOS DE PRECISÃO, S.A.
LIGHTENJIN II - INDÚSTRIA DE ILUMINAÇÃO, LDA
LIVTC PORTUGAL, LDA
LOAD INTERACTIVE, LDA
LS - GESTÃO EMPRESARIAL E IMOBILIÁRIA, S.A.
LUCEMPLAST, LDA
LUSO FINSA - INDÚSTRIA E COMÉRCIO DE MADEIRAS, S.A.
LUSODATA - SERVIÇOS INFORMÁTICOS, S.A.
LUSOFORMA - INDÚSTRIA E COMÉRCIO DE EMBALAGENS,
S.A.
LUSOREDE - REDES, ENERGIA E SISTEMAS DE INFORMAÇÃO,
LDA
M.A.R.- KAYAKS, LDA
M.A.SILVA - CORTIÇAS, S.A.
M.FERREIRA & FILHAS, LDA
M.RODRIGUES, S.A.
MAKE IT SPECIAL, LDA
MALÓ CLINIC, S.A.
MANULENA, LDA
MARCOVIL - METALOMECÂNICA DE VISEU, S.A.
MARQUES, S.A.
MATCH PROFILER - CONSULTADORIA E DESENVOLVIMENTO
DE SISTEMAS DE GESTÃO, LDA
MATECHNICS - MASTER TECHNOLOGY SOLUTIONS, LDA
MATIAS & ARAÚJO, S.A.
MAZONI - INDÚSTRIA DE CALÇADO, S.A.
MCGA PORTUGAL, UNIPESSOAL, LDA
MECALBI - ENGINEERING SOLUTIONS, LDA
MEHLER PORTUGAL, LDA
MEKKIN, S.G.P.S., S.A.
MENDES GONÇALVES, S.A.
MEO - SERVIÇOS DE COMUNICAÇÕES E MULTIMÉDIA, S.A.
MESKLIFLOWER, UNIPESSOAL, LDA
MESOSYSTEM, S.A.
METALOGALVA - IRMÃOS SILVAS, S.A.
METALSOLVUS, UNIPESSOAL, LDA
METALÚRGICA PROGRESSO DE VALE DE CAMBRA, S.A.

 Continuação da tabela B.1

Nome

METALUSA INDUSTRIAL, S.A.
 MI CASA ES TU CASA, LDA
 MICAU - INDÚSTRIAS ALIMENTARES E COMÉRCIO GERAL, S.A.
 MICROPLÁSTICOS, S.A.
 MIND SOLUTIONS - MANAGEMENT AND INDUSTRIAL SOLUTIONS, LDA
 MOCOFFEE EUROPE, UNIPESSOAL, LDA
 MOLDIT - INDÚSTRIA DE MOLDES, S.A.
 MOONGY, S.A.
 MOTOFIL, S.A.
 MOVECHO, S.A.
 MOVELAR - MÁRIO F. RINO, LDA
 MTEX NEW SOLUTION, S.A.
 MULTIVISION - CONSULTORIA EM TELECOMUNICAÇÕES E INFORMÁTICA, LDA
 MWTECHNOLOGIES, LDA
 NAVARRA - EXTRUSÃO DE ALUMÍNIO, S.A.
 NCP - FABRICO DE PRODUTOS METÁLICOS, S.A.
 NDRIVE, NAVIGATION SYSTEMS, S.A.
 NEADVANCE - MACHINE VISION, S.A.
 NEUTROPLAST - INDÚSTRIA DE EMBALAGENS PLÁSTICAS, S.A.
 NEVES & COMPANHIA, LDA
 NEWSTAMP - ESTAMPAGEM DE COMPONENTES METÁLICOS, LDA
 NEXXPRO - FÁBRICA DE CAPACETES, S.A.
 NIGEL - CONGELADORA JOSÉ NICOLAU, LDA
 NIMCO PORTUGAL, LDA
 NLS, LDA
 NOMAD TECH, LDA
 NOS COMUNICAÇÕES, S.A.
 NOT GUILTY - THE RIGHT WAY, LDA
 NOVABASE BUSINESS SOLUTIONS - SOLUÇÕES DE CONSULTORIA, DESENVOLVIMENTO, INTEGRAÇÃO, OUTSOURCING, MANUTENÇÃO E OPERAÇÃO DE SISTEMAS DE INFORMAÇÃO, S.A.
 NOVARES PORTUGAL, S.A.
 NOVARROZ - PRODUTOS ALIMENTARES, S.A.
 NS - MÁQUINAS INDUSTRIAIS, LDA

Continuação da tabela B.1

Nome

NUTPOR - PRODUTOS ALIMENTARES, UNIPESOAL, LDA
NUTRY4ALL, UNIPESOAL, LDA
O FELIZ - METALOMECÂNICA, S.A.
OGMA - INDÚSTRIA AERONÁUTICA DE PORTUGAL, S.A.
OLESA - INDÚSTRIA DE MOLDES, S.A.
OLI - SISTEMAS SANITÁRIOS, S.A.
OLICARGO, S.A.
OLIVEIRA SÁ, S.A.
OMNOVA SOLUTIONS PORTUGAL, S.A.
OMYA, S.A.
ORICA MINING SERVICES PORTUGAL, S.A.
ORO AGRI EUROPE, S.A.
ORTHOS XXI, UNIPESOAL, LDA
OSVALDO MATOS, S.A.
P & R - TÊXTEIS, S.A.
P.V. - PORTO VINTAGE - GUESTHOUSE, LDA
PALBIT, S.A.
PALMETAL - ARMAZENAGEM E SERVIÇOS, S.A.
PANICONGELADOS - MASSAS CONGELADAS, S.A.
PANIKE - INDÚSTRIA DE PRODUTOS ALIMENTARES CONGELADOS, S.A.
PAPELEIRA COREBOARD, S.A.
PDM E FC - PROJECTO, DESENVOLVIMENTO, MANUTENÇÃO, FORMAÇÃO E CONSULTADORIA, S.A.
PEAR PANEL, SOCIEDADE UNIPESOAL, LDA
PECOL - SISTEMAS DE FIXAÇÃO, S.A.
PECOL AUTOMOTIVE, S.A.
PENTAPLAST, UNIPESOAL, LDA
PERFINOX - INDÚSTRIA METALÚRGICA, S.A.
PETRATEX - CONFECÇÕES, S.A.
PETROTEC - INOVAÇÃO E INDÚSTRIA, S.A.
PINOPINE - PRODUTOS QUÍMICOS, S.A.
PIZARRO, S.A.
PLANETA - PLÁSTICOS, S.A.
PLASMITEC - PLÁSTICOS TÉCNICOS DO MINHO, LDA
PLAYVEST, S.A.
PLM PLURAL, S.A.
PLURIFIL - PRODUTOS QUÍMICOS, LDA

Continuação da tabela B.1

Nome

POLISPORT PLÁSTICOS, S.A.
POLIVERSAL - PLÁSTICOS E TECNOLOGIA, S.A.
POLIVOUGA - INDÚSTRIA DE PLÁSTICOS, S.A.
POLOPIQUE II - TECIDOS, S.A.
PORTUGALISOL NORTE - ISOLAMENTOS INDUSTRIAIS, LDA
PPRR - PLÁSTICOS, S.A.
PRADECON - CONSTRUÇÕES METÁLICAS, S.A.
PREVINIL - EMPRESA PREPARADORA DE COMPOSTOS VINÍ-
LICOS, S.A.
PRIBERAM - INFORMÁTICA, S.A.
PRIMAVERA - BUSINESS SOFTWARE SOLUTIONS, S.A.
PRIMUS VITÓRIA - AZULEJOS, S.A.
PRIO ENERGY, S.A.
PRK ATLANTIC HOTEL PORTO, S.A.
PROEF EURICO FERREIRA PORTUGAL, S.A.
PROFIMETRICS - SOFTWARE SOLUTIONS, S.A.
PROLACTO - LACTÍCIOS DE SÃO MIGUEL, S.A.
PULVERIZADORES ROCHA, S.A.
PURATOS - PRODUTOS E ARTIGOS PARA A INDÚSTRIA ALI-
MENTAR, S.A.
PUREM TONDELA, UNIPessoal, LDA
PUREVER FRIEMO, S.A.
PUREVER INDUSTRIAL SOLUTIONS, S.A.
QUALISEG - ENGENHARIA E GESTÃO, LDA
QUIDGEST - CONSULTORES DE GESTÃO, S.A.
QUIMISERVE - QUÍMICOS E SERVIÇOS, LDA
QUITÉRIOS - FÁBRICA DE QUADROS ELÉCTRICOS, LDA
RACLAC, S.A.
RAMALHOS, S.A.
RAPORAL, S.A.
READINESS IT - SYSTEMS INTEGRATION, S.A.
REAL GAMES, UNIPessoal, LDA
RECER - INDÚSTRIA DE REVESTIMENTOS CERÂMICOS, S.A.
RELVAS II - ROLHAS DE CHAMPANHE, S.A.
RENOVA - FÁBRICA DE PAPEL DO ALMONDA, S.A.
RESIBRAS - COMPANHIA PORTUGUESA DE RESINAS PARA
ABRASIVOS, S.A.
RESPOL - RESINAS, S.A.

Continuação da tabela B.1

Nome

RIOPELE - TÊXTEIS, S.A.
RIVAZ - QUÍMICA, S.A.
RNM - PRODUTOS QUÍMICOS, S.A.
S.ROQUE - MÁQUINAS E TECNOLOGIAS LASER, S.A.
SAFA - CONSTRUÇÕES ELECTROMECÂNICAS, LDA
SANINDUSA - INDÚSTRIA DE SANITÁRIOS, S.A.
SANITANA - FÁBRICA DE SANITÁRIOS DE ANADIA, S.A.
SCHAEFFLER PORTUGAL, UNIPessoal, LDA
SCHINDLER - ASCENSORES E ESCADAS ROLANTES, S.A.
SCHRÉDER ILUMINAÇÃO, S.A.
SECIL - COMPANHIA GERAL DE CAL E CIMENTO, S.A.
SEND IT - SOFTWARE E SERVIÇOS PARA TELECOMUNICAÇÕES, S.A.
SENSE TEST - SOCIEDADE DE ESTUDOS DE ANÁLISE SENSORIAL A PRODUTOS ALIMENTARES, LDA
SEW - EURODRIVE PORTUGAL, LDA
SF MOLDES, S.A.
SGL COMPOSITES, S.A.
SHAMIR OPTICAL, LDA
SHIJI PORTUGAL, S.A.
SIA - SOCIEDADE DE INOVAÇÃO AMBIENTAL, LDA
SICOESTE - SOCIEDADE INDUSTRIAL DE CUTELARIAS, LDA
SIEMENS GAMESA RENEWABLE ENERGY BLADES, S.A.
SIEMENS, S.A.
SIGIF IMO, LDA
SIKA PORTUGAL - PRODUTOS CONSTRUÇÃO E INDÚSTRIA, S.A.
SIMOLDES - PLÁSTICOS, S.A.
SINUTA FCE - FERRAMENTAS DE CORTE E ESTAMPAGEM, LDA
SINUTA, S.A.
SISTRADÉ - SOFTWARE CONSULTING, S.A.
SMART INOVATION, LDA
SNA EUROPE (INDUSTRIES), LDA
SNPS PORTUGAL, UNIPessoal, LDA
SOCIEDADE AGRÍCOLA DE NOGUEIRA, LDA
SOCIEDADE DE PERFUMARIAS NALLY, LDA
SOCIEDADE DE TRANSPORTES COLECTIVOS DO PORTO, E.I.M., S.A.

 Continuação da tabela B.1

Nome

SOCIEDADE VINÍCOLA TERRAS DE VALDIGEM, S.A.
 SODECIA SAFETY & MOBILITY GUARDA, S.A.
 SOFTINSA - ENGENHARIA DE SOFTWARE AVANÇADO, LDA
 SOFTMORPHIS, LDA
 SOGRAPE VINHOS, S.A.
 SOLANCIS - SOCIEDADE EXPLORADORA DE PEDREIRAS, S.A.
 SOLIDAL - CONDUTORES ELÉCTRICOS, S.A.
 SOLTRÁFEGO - SOLUÇÕES DE TRÂNSITO, ESTACIONAMENTO
 E COMUNICAÇÕES, S.A.
 SOLZAIMA - EQUIPAMENTOS PARA ENERGIAS RENOVÁVEIS,
 S.A.
 SOMENGIL - SOLUÇÕES INTEGRADAS DE ENGENHARIA, S.A.
 SOMINCOR - SOCIEDADE MINEIRA DE NEVES-CORVO, S.A.
 SOPAC - SOCIEDADE PRODUTORA DE ADUBOS COMPOSTOS,
 S.A.
 SOPREFA - COMPONENTES INDUSTRIAIS, S.A.
 SOPSEC - SOCIEDADE DE PRESTAÇÃO DE SERVIÇOS DE EN-
 GENHARIA CIVIL, S.A.
 SORGILA - SOCIEDADE DE ARGILAS, S.A.
 SOUL2SOLE, LDA
 SÓVIRAS - COMPONENTES PARA CALÇADO, LDA
 SPIRALPACK - MANIPULADOS DE PAPEL, S.A.
 ST IBÉRICA, SOCIEDADE UNIPESSOAL, LDA
 ST+I - SERVIÇOS TÉCNICOS DE INFORMÁTICA, UNIPESSOAL,
 LDA
 STEELPRO - ENGENHARIA INDUSTRIAL, LDA
 SUCCESS GADGET - NANOTECNOLOGIA E NOVOS MATERIAIS,
 LDA
 SUMA - SERVIÇOS URBANOS E MEIO AMBIENTE, S.A.
 SUMOL+COMPAL - MARCAS, S.A.
 SUNBD - SUN BUSINESS DEVELOPMENT, LDA
 SUPER BOCK BEBIDAS, S.A.
 SYSADVANCE - SISTEMAS DE ENGENHARIA, S.A.
 TAPEÇARIAS FERREIRA DE SÁ, S.A.
 TEARFIL - INDÚSTRIA TÊXTIL, S.A.
 TECHNOEDIF - ENGENHARIA, S.A.
 TECHNOPHAGE - INVESTIGAÇÃO E DESENVOLVIMENTO EM
 BIOTECNOLOGIA, S.A.

Continuação da tabela B.1

Nome

TECNIARTE - PROJECTOS E CONSTRUÇÕES S.A.
 TECNIFREZA - INDÚSTRIA DE MOLDES, S.A.
 TECNILAB-PORTUGAL - SOCIEDADE DE PLANEAMENTO TÉCNICO E CIENTÍFICO, S.A.
 TECNIMEDE - SOCIEDADE TÉCNICO-MEDICINAL, S.A.
 TEKA PORTUGAL, S.A.
 TEKEVER UAS, S.A.
 TENENTE & ELOSEGUI - DESENVOLVIMENTO E PRODUÇÃO, LDA
 TENSAI INDÚSTRIA, S.A.
 TETRAMOLD - INDÚSTRIA DE MOLDES, LDA
 TEXCON-IMPORTAÇÃO E EXPORTAÇÃO UNIPessoal,, LDA
 TEXLAAUTOMOTIVE - TÊXTEIS, LDA
 TÊXTEIS J.F.ALMEIDA, S.A.
 THALES PORTUGAL, S.A.
 THE BRICK WALL AND PROPERTIES - MEDIAÇÃO IMOBILIÁRIA, LDA
 TIMESTAMP - BUSINESS INTELLIGENCE & WAREHOUSING, LDA
 TIMESTAMP - SISTEMAS DE INFORMAÇÃO, S.A.
 TMG - TECIDOS PLASTIFICADOS E OUTROS REVESTIMENTOS PARA A INDÚSTRIA AUTOMÓVEL, S.A.
 TRIANGLE'S - CYCLING EQUIPMENTS, S.A.
 TUPAI - FÁBRICA DE ACESSÓRIOS INDUSTRIAIS, S.A.
 U.E.M. - UNIDADE DE ESTRUTURAS METÁLICAS, S.A.
 UARTRÓNICA - ELECTRÓNICA, LDA
 UBIWHERE, LDA
 UNICOR 2 - PRODUTOS DE CORTIÇA, LDA
 UNIPARTNER IT SERVICES, S.A.
 UNIVERSAL MCCANN CONNECTIONS - PUBLICIDADE, UNIPessoal, LDA
 UNIVERSAL MOTORS (U.M.) - EQUIPAMENTOS ELECTROMECÂNICOS, S.A.
 UNYKVIS, LDA
 VAISA - AGRICULTURA INTENSIVA, S.A.
 VALBOPAN - FIBRAS DE MADEIRA, S.A.
 VALENTE & LOPES, LDA
 VAPESOL - FÁBRICA DE COMPONENTES PARA CALÇADO, LDA

 Continuação da tabela B.1

Nome

VERDASCA & VERDASCA, S.A.
 VESAM - ENGENHARIA, S.A.
 VICAIMA - INDÚSTRIA DE MADEIRAS E DERIVADOS, S.A.
 VIDRARIA MORTÁGUA - VIDROS E ESPELHOS, S.A.
 VIDRIMOLDE - INDÚSTRIA INTERNACIONAL DE MOLDES, LDA
 VIEIRA ARAÚJO, S.A.
 VIEIRA DE CASTRO - PRODUTOS ALIMENTARES, S.A.
 VIEIRA DE SOUSA - VINES & WINES, LDA
 VIGOBLOCO - PRÉ-FABRICADOS, S.A.
 VILARTEX - EMPRESA DE MALHAS VILARINHO, LDA
 VILT - PORTUGAL, S.A.
 VIMÉTRICA - SOLUÇÕES DE VISÃO ARTIFICIAL, UNIPESSOAL,
 LDA
 VINCO - VÁLVULAS, S.A.
 VIROC PORTUGAL - INDÚSTRIAS DE MADEIRA E CIMENTO,
 S.A.
 VIRTUAL CAMPUS, LDA
 VIRTUALCARE, LDA
 VISHAY - ELECTRÓNICA, PORTUGAL, LDA
 VIÚVA LAMEGO - CERÂMICAS DE SINTRA, LDA
 VLB TEC, UNIPESSOAL, LDA
 VMUSE, LDA
 VOLKSWAGEN AUTOEUROPA, LDA
 WAKARU - CONSULTING, LDA
 WAVECOM - SOLUÇÕES RÁDIO, S.A.
 WEADD, LDA
 WINPROVIT - SOLUÇÕES INTELIGENTES, S.A.
 WISE4AUTOMATION, LDA
 WIT-SOFTWARE - CONSULTORIA E SOFTWARE PARA A IN-
 TERNET MÓVEL, S.A.
 WONDERCOM, GESTÃO DE PROJECTOS E EQUIPAMENTOS DE
 INFORMAÇÃO E TELECOMUNICAÇÕES, LDA
 WONDERLIGHT, UNIPESSOAL, LDA
 XAREVISION, S.A.
 XZ - CONSULTORES, S.A.
 YD YNVISIBLE, S.A.
 YILPORT IBERIA, S.A.
 ZAGOPE - CONSTRUÇÕES E ENGENHARIA, S.A.

Continuação da tabela B.1

Nome

ZENITHWINGS, LDA

ZENTIVA PORTUGAL, LDA

ZOLLERN & COMANDITA

Fim da tabela
