

Faculdade de Engenharia da Universidade do Porto



Text Mining Referentes Aplicações



Projeto FEUP 2020/2021 – Mestrado Integrado em Engenharia Informática e Computação :
(nome do coordenador geral) J. Magalhães Cruz

Equipa 1MIEIC01_1:

Supervisor: Jorge Barbosa Monitor: Rui Guedes

Estudantes & Autores:

Duarte Lopes up202006408@edu.fe.up.pt

Inês Cardoso up202005435@edu.fe.up.pt

Mariana Carvalho up202007620@edu.fe.up.pt

Milena Gouveia up202008862@edu.fe.up.pt

Nuno Neto up202005298@edu.fe.up.pt

Pedro Macedo up202007531@edu.fe.up.pt

Resumo

Hodiernamente, a quantidade de documentos armazenados em formato eletrónico tem crescido substancialmente. *Text mining* consiste no processo de obtenção de informação a partir de um texto escrito em linguagem natural. Este processo extrai dados ou semi-dados estruturados. Acredita-se que 80% da informação origina-se de uma forma não estruturada.

De facto, a busca de informação a partir de documentos muito extensos torna-se um processo muito complicado sem o auxílio de ferramentas computacionais. Portanto, torna-se muito importante o desenvolvimento de ferramentas que auxiliam neste processo, de forma automática e imediata.

Na verdade, a área do *text mining* tem vindo a captar o interesse das comunidades científicas e industriais. Este processo procura avaliar padrões, relações e regras a partir dos dados textuais que estão a ser analisados. Estes dados podem se apresentar em LN (linguagem natural), ou num formato semi-estruturado (por exemplo, documentos, documentos de email, e muitos outros).

Este relatório foi realizado com base em informações presentes em revistas científicas, livros técnicos, relatórios e teses, cujas referências se encontram na bibliografia.

Palavras-chave

Text mining; *Data mining*; Sistemas de informação inteligentes; Clusterização

Índice

Lista de figuras

Figura 1. Etapas do <i>text mining</i>	9
Figura 2. Clusterização de dados com apenas duas dimensões: dados e dados agrupados em clusters	11

Glossário

Palavras sentimentais -

Data mining - Processo que permite a análise e processo de dados online. O text mining é uma variante do data mining, mais orientado para a análise de textos.

Phishing - que tem como objetivo roubar informações e dados

Introdução

Todos os dias se disponibiliza uma grande quantidade de novas informações online. A maior parte destes dados não estão estruturados, o que dificulta a procura de informação e a posterior assimilação do que é realmente relevante. O facto do conteúdo não se encontrar organizado de uma maneira predefinida, nem associado a outros conceitos e contextos, torna-o simplesmente inútil. Os computadores acabam por tratá-lo como simples sequências de caracteres. O *text mining* nasce, assim, da necessidade de processar automaticamente estes novos dados, organizando-os e categorizando-os, para que se tornem em conhecimento e informação útil.

Deste modo, o objetivo deste relatório, realizado no âmbito da unidade curricular Projeto FEUP, é promover uma visão geral do conceito e importância do *text mining* na atualidade. A relevância da implementação de algoritmos com base nesta tecnologia será suportada pelas diferentes aplicações nas principais áreas interligadas ao mundo científico e técnico, como a saúde ou mesmo as redes sociais. Contudo, a mineração de textos ainda enfrenta alguns desafios relativos à ambiguidade intrínseca na informação, ou até mesmo nos próprios fornecedores da mesma. Estas desvantagens serão brevemente abordadas no final do trabalho com vista a questionar qual o futuro desta tecnologia.

1. Text Mining

2.1 O que é?

Procurando analisar o conceito do tema do trabalho, *text mining* ou mineração de texto, talvez uma abordagem lógica será começar por analisar os termos separadamente. De facto, assim, é possível ficar, rapidamente, com uma ideia geral do que se possa tratar.

mineração | s. f.

mi·ne·ra·ção

substantivo feminino

1. Trabalho ou exploração de minas.
2. Depuração do minério extraído.

"mineração", in Dicionário Priberam da Língua Portuguesa [em linha], 2008-2020, <https://dicionario.priberam.org/minera%C3%A7%C3%A3o> [consultado em 24-10-2020].

Através do significado literal de mineração, facilmente se percebe que, a mineração a partir de pedras ou areia, tem como grande objetivo, reduzir e particularizar a matéria original de modo a encontrar os verdadeiros tesouros, como por exemplo o ouro. Analogamente, a mineração de textos, parte de grandes quantidades de informação textual, muitas vezes não estruturada, ou seja, uma sequência de caracteres sem ligação a outros conceitos e contextos e por isso, inútil; para extrair verdadeiro conhecimento, por fim, útil. [3]

Estes dados não estruturados não são, necessariamente, apenas relativos a textos, pelo que podem estar presentes em imagens, vídeos, páginas da Internet, e-mails, entre outros [4]. No entanto, formatos que não o textual são abordados pelo *data mining*, outro tipo de ferramenta, que explora bases de dados com vista a descobrir padrões nessa mesma informação. Realmente, *text mining* é uma variante desta última tecnologia, da maneira em que, aquilo que é extraído, é encontrado, pois pertence a um padrão ou tendência. Porém, esta segunda ferramenta é reservada ao tratamento de dados em formato de texto, ainda que muitas vezes beneficie de algoritmos desenvolvidos para mineração de dados. [1]

2.2 Porquê *text mining*?

A informação que aparece na Internet é perto de infinita, visto que a cada momento qualquer pessoa é capaz de contribuir para a sua disponibilização. As fontes que partilham estes dados são intermináveis, logo é impossível acompanhar manualmente tudo o que se vai tornando acessível.

Assim, a palavra-chave aqui é automatizar, visto que não é viável os dados serem tratados um a um. Consequentemente, um indivíduo que pretenda pesquisar sobre um determinado assunto no *Google*, irá ver o processo de pesquisa favorecido se estiver implementada a tecnologia de *text mining* no motor de busca. Se tal não fosse o caso uma simples pesquisa tornar-se-ia insuportável, uma vez que a pessoa seria sobrecarregada por informação provavelmente inútil.

Supondo um trabalhador do apoio ao cliente de uma empresa que receba *feedback* dos seus compradores relativo aos problemas que encontram no seu produto e que tenha como objetivo solucioná-los. Este veria o seu trabalho facilitado se houvesse maneira de separar ou categorizar o *feedback* tendo em conta o tipo de problema. Este processo realizado manualmente, não só seria dispendioso, como extremamente monótono e demorado. Ainda por cima, a falha humana é um fator relevante e em contrapartida o trabalho feito por uma máquina será obviamente mais consistente, especialmente, se o trabalhador estiver a analisar informação subjetiva. Uma máquina que aprenda um algoritmo e treine a aplicação de um determinado critério irá garantir uma maior precisão e avaliação uniforme da informação. Neste caso, a aplicação da tecnologia de *text mining* otimizaria o setor de apoio ao cliente, o que consequentemente seria uma mais valia irrefutável para a empresa. [5]

2.3 Qual é o procedimento?

Para que a mineração de textos seja realizada com sucesso, é necessário ser seguida uma série de 5 etapas, que deverão ocorrer pela ordem que está representada no seguinte diagrama:

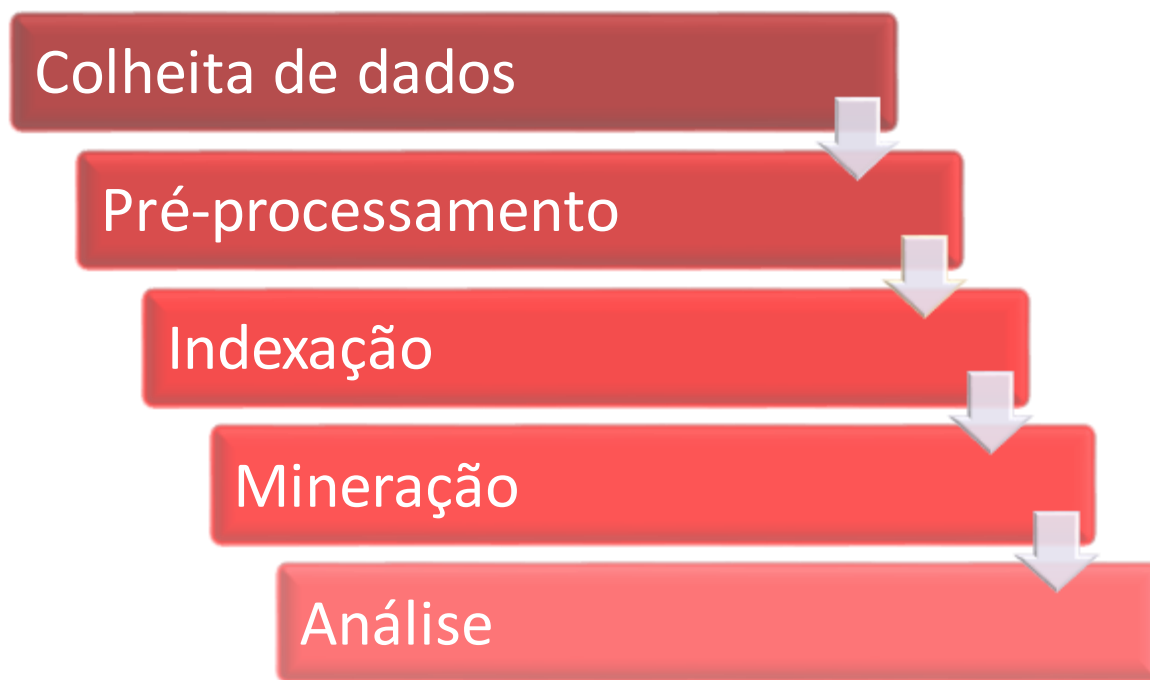


Figura 1. Etapas do *text-mining*

2.3.1. Colheita de dados

Em primeiro lugar, é necessário criar uma base de textos, pelo que ocorre a colheita de dados de diferentes recursos – desde sites, *e-mails*, redes sociais, notícias, ou arquivo de documentos. Tal processo pode ser totalmente automatizado, isto é, realizado unicamente pelo computador, ou orientado por um responsável pela sua execução (o *text miner*). [1]

2.3.2. Pré-processamento

Segundamente, ocorre o pré-processamento da informação. Neste processo acontece a identificação de conteúdos e a extração de recursos representativos. Isto é, dá-se:

- A **limpeza de texto**, na qual é removida a informação desnecessária ou indesejada, como anúncios nas páginas;
- A **segmentação de texto**, processo que divide o mesmo em objetos significativos, permitindo que o computador possa identificar frases, parágrafos, pontuações e espaços a partir dos próprios espaços em branco presentes. A segmentação é muito importante, pois sem ela os textos não poderiam vir a ser analisados, visto que estes seriam considerados pelo computador como uma simples sequência de caracteres.
- A **seleção de características**, na qual o texto é analisado de forma a ser obtido um conjunto de medidas quantitativas, como por exemplo a frequência de palavras num dado excerto. [11]

2.3.3. Indexação

Após o pré-processamento ocorre a indexação. Nesta etapa é formada uma estrutura semelhante a um índice de certos termos, bem como a sua localização – daí o nome desta etapa ser “indexação”. [1] [11]

A relevância desta etapa reside principalmente no facto de permitir, não só, que dados sejam recuperados e reestruturados, bem como que seja possível pesquisar rapidamente por documentos a partir de palavras-chaves. [11]

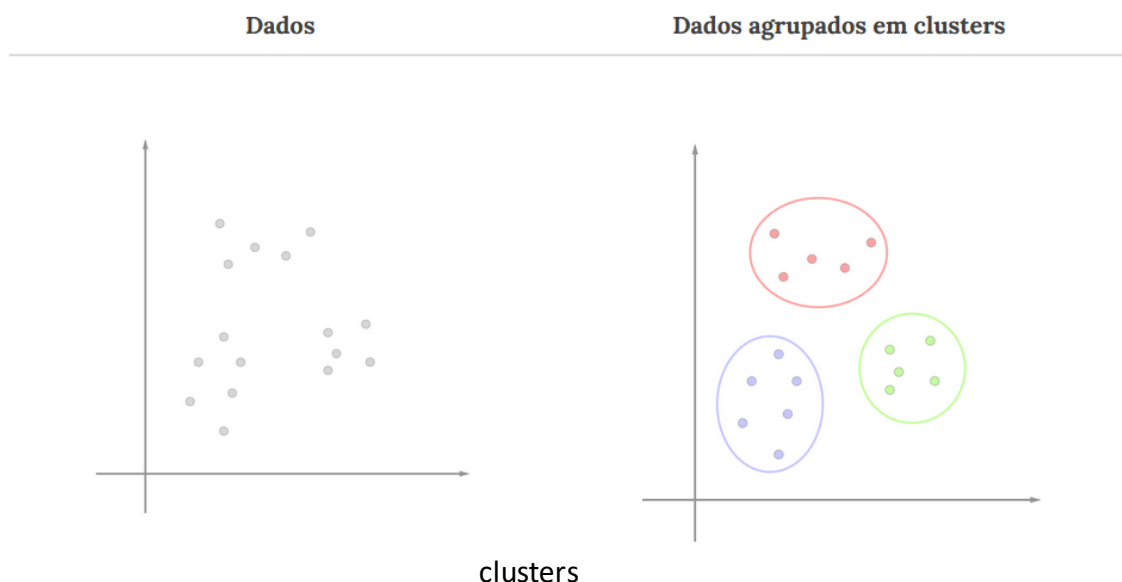
2.3.4. Mineração

A mineração é a etapa em que diferentes técnicas de *data mining* são aplicadas aos dados já sujeitos aos dois processos anteriormente descritos, de forma a que sejam revelados novos conhecimentos. As técnicas mais utilizadas são:

- **Seleção de informação**, na qual são extraídos os documentos mais importantes com base no que se procura; [9]
- **Extração de informação**, na qual são extraídos elementos relevantes com base no que se procura, de determinados documentos; [9]

- **Sumarização**, em que o texto é reduzido mantendo-se apenas os pontos mais importantes, com base na informação desejada; [10]
- **Classificação**, também conhecida como processamento de Linguagem natural, no qual os textos em linguagem normal são atribuídos a um conjunto predefinido de tópicos, dependendo de seu conteúdo; [9]
- **Clusterização**, que é um processo da mineração também conhecido como “agrupamento”. Isto, pois, procura identificar estruturas intrínsecas nas informações textuais e organizar os seus dados em subgrupos ou “clusters”. A imagem a seguir é uma boa representação desta técnica de Clusterização: [10]

Figura 2. Clusterização de dados com apenas duas dimensões: dados e dados agrupados em



Em suma, é na etapa de mineração em que ocorre a identificação de termos específicos e das relações existentes entre estes.

2.3.5. Análise

Os dados produzidos na etapa de “Mineração” não estão estruturados. Assim, é nesta última etapa denominada de “Análise” que os dados são estruturados e avaliados, para que possam ser interpretados com relação às questões que o minerador de texto deseja investigar. [11]

3. Aplicações

Como já foi referido, a mineração de textos tem a sua importância mesmo nas atividades mais simples que se possam realizar online. Assim, seguir-se-á a referência a algumas das aplicações desta ciência na área comercial, nas redes sociais, na análise de sentimentos, na medicina e no correio eletrónico.

3.1 Área comercial

Uma das aplicações mais comuns de *text mining* é na área comercial, que usufrui desta tecnologia implementada em diversos fatores. Exemplos destes são:

- uma procura facilitada de novidades, tendências e informações na *Web*, permitindo o mercado manter-se atualizado;
- o atendimento ao cliente, este que beneficia da ferramenta da maneira em que o tratamento da informação obtida do *feedback* dos clientes se torna mais fácil;
- evolução das empresas e organizações, para o qual o *feedback* é essencial, uma vez que é partindo dos comentários e recomendações dos seus clientes, que a concretização destes através dos produtos ou serviços oferecidos, melhora.

Considerando técnicas de *text mining* aplicadas em informação online relativa às empresas, quer seja a sua divulgação nas redes sociais, referência em *blogs* ou revistas, rapidamente se identifica aspetos que podem ser modificados de forma a evoluir e moldar a imagem da marca com base nas preferências dos seus clientes.

Passando a um exemplo mais concreto; supondo alguém que usufrui de um determinado serviço ou produto necessita de entrar em contacto com um especialista da empresa para solucionar um problema encontrado. Decerto, a utilização de um algoritmo de mineração de texto permite que a solicitação do cliente seja analisada minuciosamente, classificando a urgência do pedido, e reencaminhando automaticamente esse indivíduo para as pessoas indicadas. Isto, sem que a empresa tenha de passar por estas etapas manualmente. [6]

De facto, o processo de atendimento ao cliente torna-se superior, do ponto de vista económico, uma vez que elimina a necessidade de contratar mão de obra, da precisão, porque diminui a quantidade de erros que poderiam ser causados pela intervenção humana ligada à subjetividade, e da rapidez da capacidade de resposta ao cliente. [5]

3.2 Redes sociais

Uma das aplicações mais recentes do *text mining* é nas plataformas de *streaming* e de transmissão de vídeos, como o “*Youtube*” ou a “*Twitch*”. Estas contêm ferramentas capazes de melhorar a segurança da comunidade e do próprio criador de conteúdo, recorrendo ao uso de algoritmos de mineração de texto.

De facto, estas inovações são bastante valorizadas nestes tempos em que as redes sociais têm um impacto tão grande na sociedade. No caso do “*Youtube*” que tem um total de 1,325,000,000 de utilizadores [7], necessita deste género de tecnologia para que não se torne num espaço tóxico onde, uma percentagem do bilião de pessoas que utilizam o site, partilham as suas opiniões negativas. Assim, parte-se para um exemplo das ferramentas de mineração de texto aplicadas nestas plataformas: a opção do dono do vídeo ou da *stream* poder seleccionar palavras que não ache benéficas para a sua comunidade e ter automaticamente apagados todos os comentários que apresentem tais palavras.

3.3 Análise Sentimental

Atualmente, são realizadas várias pesquisas de opinião pública, sejam estas feitas para a obtenção de *feedback*, por parte de uma empresa ou serviço, quer como parte de um estudo social ou económico, por exemplo.

Tais pesquisas apresentam-se muitas vezes como questionários de resposta fechada e pré-determinada. No entanto, quando é necessária uma justificação mais profunda ou elaborada, os organizadores da pesquisa tendem a apresentar perguntas de resposta aberta. Os resultados dessa questão seriam subjetivos, visto que apresentariam uma opinião pessoal. Nestes casos, é necessária a intervenção do *text mining* para categorizar as respostas e analisar o sentimento global expresso em cada resposta, dado que realizá-lo manualmente tomaria mais tempo e seria mais dispendioso a nível financeiro. [6]

3.4 Área da saúde

A medicina é uma área onde a produção de informação de texto é vasta. Para que a comunidade científica usufrua de uma base de estudo mais variada e completa; tome decisões baseadas e suportadas por informação e estudos atualizados, os profissionais de saúde acabam por beneficiar da aplicação da mineração de textos. De facto, recorrendo a softwares que têm como base o algoritmo de *text mining*, torna-se mais simples a realização do diagnóstico médico do paciente, a busca de tratamentos e medicamentos, ou até o conhecimento e investigação de novas doenças. [6]

“Medline” é o exemplo de um destes softwares. Este serviço permite aceder aos dados bibliográficos da Biblioteca Nacional de Medicina dos Estados Unidos da América, auxiliando então as pessoas relacionadas com a comunidade médica no seu trabalho.

3.5 Correio eletrónico

E-mails são um dos métodos mais populares e adotados de comunicação. No entanto, uma das desvantagens dos *e-mails* é a quantidade indesejada de anúncios não solicitados que se recebe na “caixa de entrada”. Geralmente estes *e-mails* são considerados perigosos pois são utilizados “para *phishing* de dados sensíveis”, ou então a propagação de vírus. Pelo que é necessário categorizá-los como *spam*, atenuando ou até mesmo evitando os possíveis danos causados por eles. [2] [8]

De acordo com o *Nucleus Research Inc.* (2007), “a força de trabalho dos Estados Unidos da América perde mais de 71 bilhões de dólares por ano na produtividade gasta ao controlar o *spam*. Detetar efetivamente *e-mails* de *spam*, não só reduz perdas financeiras, como também melhora as satisfações dos utilizadores de *e-mails*” [2]. A mineração de textos é a via para que os *e-mails* de *spam* sejam detetados corretamente, mesmo que não seja atingida uma precisão de 100%, devido a alguns dos desafios, ainda presentes.

Contudo, de uma maneira bastante simplificada, é possível distinguir dois modos de detetar *spam* realizados através de *text mining*: deteção baseada no remetente do *e-mail* e a baseada no conteúdo do mesmo. Enquanto que o primeiro se centra na informação do emissor, como o seu nome de utilizador e o estilo de escrita, para a categorização, o segundo baseia-se nas próprias palavras e termos extraídos do texto. [3]

De facto, a mineração de textos pode ser a chave para perceber o porquê de alguns dos *e-mails* enviados inofensivamente serem etiquetados como *spam*. Tal ocorre porque, quando o *e-mail* é sujeito ao algoritmo de *text mining*, algumas palavras são consideradas como “*spam trigger words*”. Isto é, ao serem detetadas algumas palavras normalmente usadas por “*spammers*”, como “gratuito” ou “grátis”, “parabéns”, “isso não é *spam*” e “grande oferta”, os *e-mails* são direcionados para a pasta de *spam*. [12]

Na verdade, é graças à aplicação de técnicas de mineração de textos nos correios eletrónicos que têm vindo a ser detetadas muitas fraudes, constituindo por isso, um modo de segurança na internet.

4. Principais desafios de *text mining*

Ao longo deste relatório já foram apresentados vários aspetos que revelam o quão benéfico *text mining* pode ser, já que permite que o trabalho em dadas áreas se torne mais simples e eficaz. Contudo, este também apresenta algumas desvantagens e problemas a serem corrigidos, os quais serão referidos de seguida.

4.1. Erros de escrita nos textos

Alguns dos desafios do *text mining* surgem principalmente pelo facto de, por vezes, a informação na Internet se encontrar com erros ortográficos ou de escrita, que dificultam o processo de mineração de textos, como por exemplo na etapa de classificação de informação. Tal ocorre principalmente em redes sociais, nas quais o uso de abreviaturas, *tags* e a ausência de acentos ou espaços, – derivados da informalidade associada a este tipo de ambiente - dificultam o processo de análise de texto. [1]

4.2. Ambiguidade na linguagem natural

A ambiguidade inerente à linguagem humana leva a que sejam cometidos erros ao longo do processo de mineração, dado que uma palavra pode ter inúmeros significados. [1]

Afinal, a existência de palavras que tanto podem ter uma conotação positiva como negativa, dependendo do contexto no qual estão inseridas, influencia o sucesso do procedimento de análise de textos. O mesmo problema ocorre:

- Em frases sarcásticas, com ou sem palavras sentimentais, como se pode observar em: “Ótimo! O meu carro deixou de funcionar”, na qual o sistema de *text mining* pode interpretar a frase como uma expressão positiva, apesar desta ter uma conotação negativa;
- Em frases que contenham palavras sentimentais, mas que não expressem nenhum sentimento ou opinião, como: “Pode dizer-me qual destas câmaras é a melhor?”.

Para além disto, também são enfrentados problemas quando diversos tipos de menções ou pronomes se referem a uma mesma entidade ou objeto, devendo estes ser identificados corretamente. A título de exemplo:

- As expressões como: “Marcelo”, “Presidente de Portugal” e “Presidente Marcelo Rebelo de Sousa”, devem referir-se todas à mesma pessoa;
- Na frase: “O Porto é uma cidade extraordinária. É um lugar que toda a gente deveria visitar. Os seus monumentos são muito reconhecidos”, o algoritmo deve reconhecer o pronome “seus” e “um lugar” como objetos referentes à cidade do Porto.

É então essencial que o programa considere o contexto da frase no seu todo, de forma a interpretar melhor o seu sentido.

4.3. Necessidade de atualização do algoritmo

Algo muito importante para os programadores é manterem os sistemas de mineração de textos atualizados. Isto, pois, ao longo do tempo vão sendo estabelecidas novas regras ortográficas para uma dada língua, que devem ser reconhecidas pelo algoritmo de modo a que a extração de informação seja bem-sucedida. [1]

4.4. Existência de grandes volumes de dados

Por último, outro desafio do *text mining* é o facto de existir uma quantidade imensurável de dados, que podem ocupar o espaço de até 100GB. Além disso, os algoritmos manipulam uma diversificada panóplia de palavras, contidas em toda a coleção, e levam a que surja a necessidade de computadores com maior capacidade de armazenamento. [1]

Conclusão

A constante inovação da tecnologia provocou um aumento de possíveis vias e modos de comunicação, que consequentemente se tornaram indispensáveis à sociedade. Os softwares e redes sociais, que estão em permanente evolução, têm sido os propulsores da partilha de informação, uma vez que são plataformas que promovem a distribuição, a interação e a publicação de conhecimento.

De facto, a velocidade a que tanta informação é disponibilizada e partilhada implica um constante acompanhamento deste novo conteúdo. O text mining vantajosamente substitui trabalho, que realizado manualmente, seria impossível de terminar. Para além do mais, uma pessoa que tentasse analisar todos os textos, de modo a categorizá-los e separá-los por conteúdo, com a finalidade de estruturar e posteriormente utilizar a informação intrínseca, iria certamente cair em erro em algum momento. A falha humana é facilmente atingida aquando desta atividade devido à subjetividade dos textos disponibilizados na Internet.

Porém, a automatização da análise e da extração de informação realmente importante classificam o text mining como uma ferramenta indispensável para o sucesso de quase todas as atividades realizadas online. A sua aplicação em motores de busca, correio eletrónico, softwares de partilha de informação científica, táticas empresariais, entre outros, promove a otimização tanto dos próprios serviços como também da experiência dos seus utilizadores.

Em suma, a realização deste relatório veio provar a utilidade e a característica indispensável da mineração de textos na evolução da tecnologia, visto que esta última promove o exponencial aumento de informação e a primeira aproveita-se desses dados como matéria-prima, procurando minerar verdadeiros tesouros, que neste contexto, representam informação útil.

Este relatório tem como objetivo transmitir informação sobre text mining e suas aplicações, visto que nos dias de hoje há muita informação e há a necessidade dessa informação ser devidamente organizada para facilitar a pesquisa e a assimilação da informação, sendo assim, um grande apoio para a sociedade. Ao longo do relatório, são abordados vários temas como: o conceito de text mining, descrição do procedimento de colheita de dados, identificação das principais aplicações, descrição das mesmas e os principais desafios.

Em suma, há duas formas de abordagem, utilizando este método, que são: uma abordagem onde as informações são retiradas de textos e são transformadas em um conteúdo mais organizado e outra abordagem, onde as informações são retiradas de documentos textuais e são transformadas em dados estruturados.

De facto, é uma ferramenta muito útil pois, pode ser usada para a resolução de um problema em concreto ou para a pesquisa de alguma informação, sendo muito usada no ramo do negócio, da medicina, redes sociais, entre outras.

Referências bibliográficas

- [1] Aranha, Christian e Passos, Emmanuel. 2006. *A Tecnologia de Mineração de Textos*. RESI-Revista Eletrônica de Sistemas de Informação, Nº2. Atualizado em 2006, acessado a 23 outubro de 2020. <https://core.ac.uk/download/pdf/237009834.pdf>
- [2] Tang, Guanting, Pei, Jian e Luk Wo-Shun. October 2013. *Email Mining: Tasks, Common Techniques, and Tools*. School of Computing Science, Simon Fraser University, Burnaby BC, Canadá. Atualizado em outubro de 2013, acessado a 23 de outubro de 2020. https://www.researchgate.net/publication/257482113_Email_mining_Tasks_common_techniques_and_tools
- [3] Han, Jiawei, Kamber, Micheline e Pei, Jian. 2012. *Data Mining: Concepts and Techniques*. 3 ed, *Data Mining*. Massachusetts: Elsevier, Inc.
- [4] Hernandez, Thiago. 2019. Atualizado em 2019, acessado a 24 de outubro de 2020. *O que são os dados estruturados e para que servem?* <https://ibsdocs.com.br/o-que-sao-os-dados-estruturados-e-para-que-servem/>
- [5] Monkeylearn, Inc. *Text Mining: The Beginner's Guide*. Atualizado em 2020, acessado a 24 de outubro de 2020. <https://monkeylearn.com/text-mining/>
- [6] Carrilho Junior, João Ribeiro. 2007. *Desenvolvimento de uma Metodologia para Mineração de Textos*. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro. https://www.maxwell.vrac.puc-rio.br/11675/11675_1.PDF
- [7] Borges, Luís Tomé, Oliveira, Pedro e Pedroso, Eduardo. 2016. *Estatísticas*. Atualizado em 2016, acessado a 26 de outubro. <http://web.tecnico.ulisboa.pt/~ist178552/wordpress/estatisticas/>
- [8] Khan, Zahra e Qamar, Usman. 2016. *Text Mining Approach to Detect Spam*. Department of Computer Engineering, NUST, College of EME. Rawalpindi, Pakistan. Atualizado em

fevereiro de 2016, acessado em 25 de outubro de 2020. [https://www.researchgate.net/publication/295737772 Text Mining Approach to Detect Spam in Emails](https://www.researchgate.net/publication/295737772_Text_Mining_Approach_to_Detect_Spam_in_Emails)

[9] Forte, Ana Catarina Barbosa. 2015. *Análise de comentários de clientes com o auxílio a técnicas de Text Mining para determinar o nível de (in)satisfação*. Tese de Mestrado em Modelação, Análise de Dados e Sistemas de Apoio à Decisão. Faculdade de Economia da Universidade do Porto. Porto.
file:///C:/Users/INS~1/AppData/Local/Temp/dissertacao_Ana_Forte.pdf

[10] Rai, Abhinav. 2019. *What is Text Mining: Techniques and Applications*. Atualizado em 2019, acessado a 24 de outubro de 2020. <https://www.upgrad.com/blog/what-is-text-mining-techniques-and-applications/>.

[11] OpenMinTeD. 2018. *Text Mining 101*. Atualizado em 2018, acessado a 25 de outubro de 2020. <http://openminded.eu/text-mining-101/>

[12] jivochat. 2016. *12 razões pelas quais seus emails acabam indo para a caixa de SPAM*. Atualizado em 2016, acessado a 19 de outubro de 2020. <https://www.jivochat.pt/blog/comunicacao/porque-emails-vao-para-spam.html>

[13] Honda, Hugo. 5 de outubro de 2017. *Introdução Básica à Clusterização*. Laboratório de Aprendizado de Máquina em Finanças e Organizações (LAMFO). Atualizado em 2017, acessado em 26 de outubro de 2020. https://lamfo-unb.github.io/2017/10/05/Introducao_basica_a_clusterizacao/

Ideia de uma aplicação do text mining

O aparecimento da pandemia do coronavírus abalou o mundo e provocou um impacto inquestionável na sociedade. Foi necessário adaptar os nossos estilos de vida de um momento para outro, de modo a prevenir o alastramento do vírus que até ao momento causou tanto sofrimento e transtorno em todos os países.

“A app *STAY AWAY COVID* pretende identificar potenciais exposições a pessoas infetadas com COVID-19.” <https://www.sns.gov.pt/apps/stayaway-covid/>