

TEXT MINING

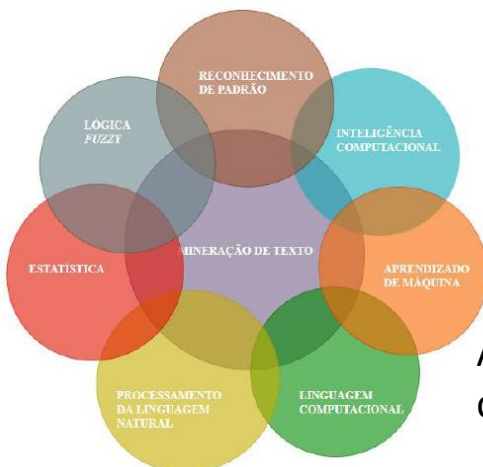
Por: Carla Cristina P. Cruz
Pesquisadora Voluntária - LSC/EGN
Rio, 30 de maio de 2020.



Fonte: MC.AI, 2020

Processo de Descoberta de Conhecimento pela análise de diferentes tipos de textos, que tem como objetivo encontrar termos relevantes em documentos de texto, estabelecendo padrões e relacionamentos entre eles com base na frequência e temática dos termos.

Imagem da Empresa Excelacom Inc. de 2019, que mostra a quantidade de dados gerados na Internet em apenas um minuto.



A autora, 2019



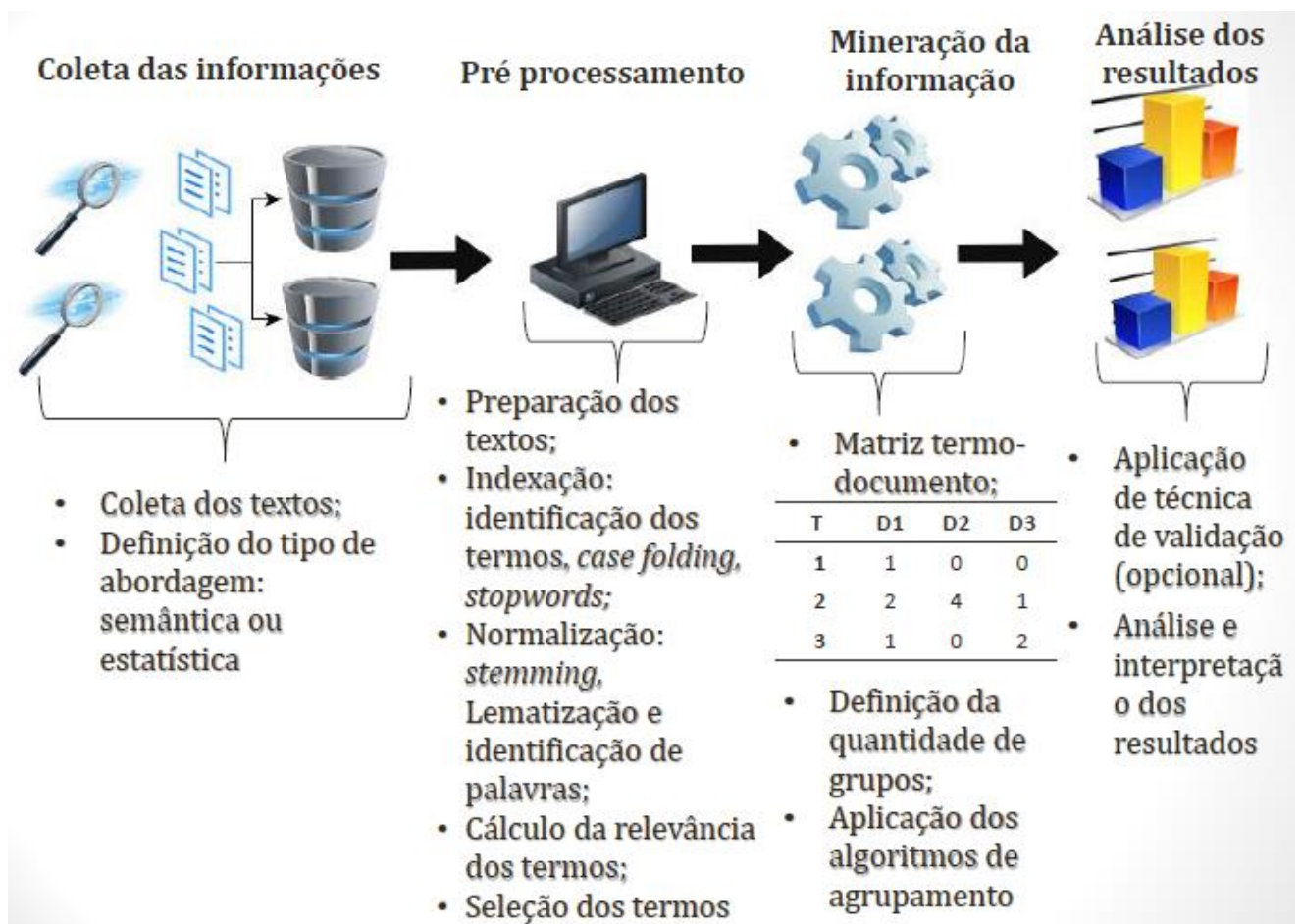
Fonte: Excelacom Inc. 2019

Algumas Áreas de Conhecimento contempladas pelo *Text Mining*

TEXT MINING

Para que os dados sejam obtidos, faz-se necessária a aplicação das técnicas do *KDT* - *Knowledge Discovery in Text* - conjunto de procedimentos para extrair e recuperar dados textuais.

Etapas do KDT



Fonte: A autora, 2019.

TEXT MINING

Existem várias técnicas utilizadas para obtenção de dados textuais, cujas as mais usuais são:

Principais Técnicas de *Text Mining*

Sumarização

- Síntese dos pontos de maior significado

Classificação

- Identificação por características



Categorização

- Identificação dos temas principais

Agrupamento

- Técnica de agregamento por grupos de textos similares em função da estrutura natural. Avalia a dimensionalidade, com identificação de *outliers* inferindo hipóteses pela associação

Fonte: A autora, 2019.

TEXT MINING

Alguns Tipos de Visualização dos Resultados

Wordcloud

Em português, Nuvem de palavras, exibe as palavras mais citadas e a frequência com que as mesmas são citadas (palavras mais citadas aparecem em um tamanho de destaque).



Figura. Palavras mais citadas no Twitter do Ministério da Defesa em 10/04/2020
Fonte: A autora, 2020.

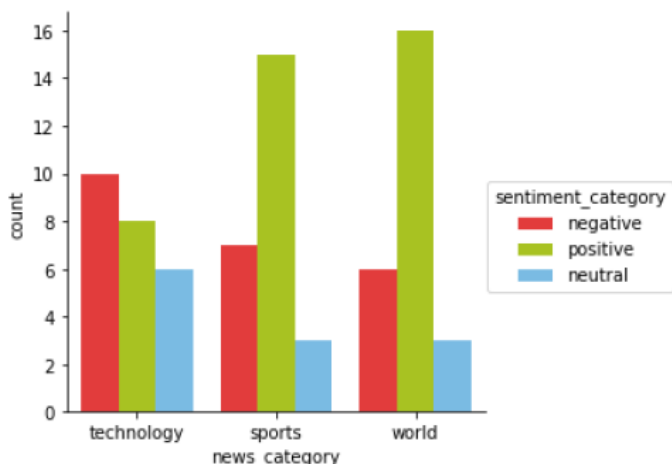


Figura. Análise de sentimentos de três categorias

Fonte: https://medium.com/@himanshu_23732/sentiment-analysis-with-textblob-6bc2eb9ec4ab.

Sentimental Analysis

Em português, Análise de Sentimentos, “capta” os sentimentos positivos, negativos e neutros de acordo com as palavras encontradas em um determinado texto.

TEXT MINING

Alguns Tipos de Visualização dos Resultados

Web/Semantic Network

Em português, Web/Rede Semântica exibe os *clusters* (grupos) e as palavras que pertencem a ele. Além disso, mostra a ligação (relação) entre os grupos e palavras, podendo exibir também a frequência das palavras.

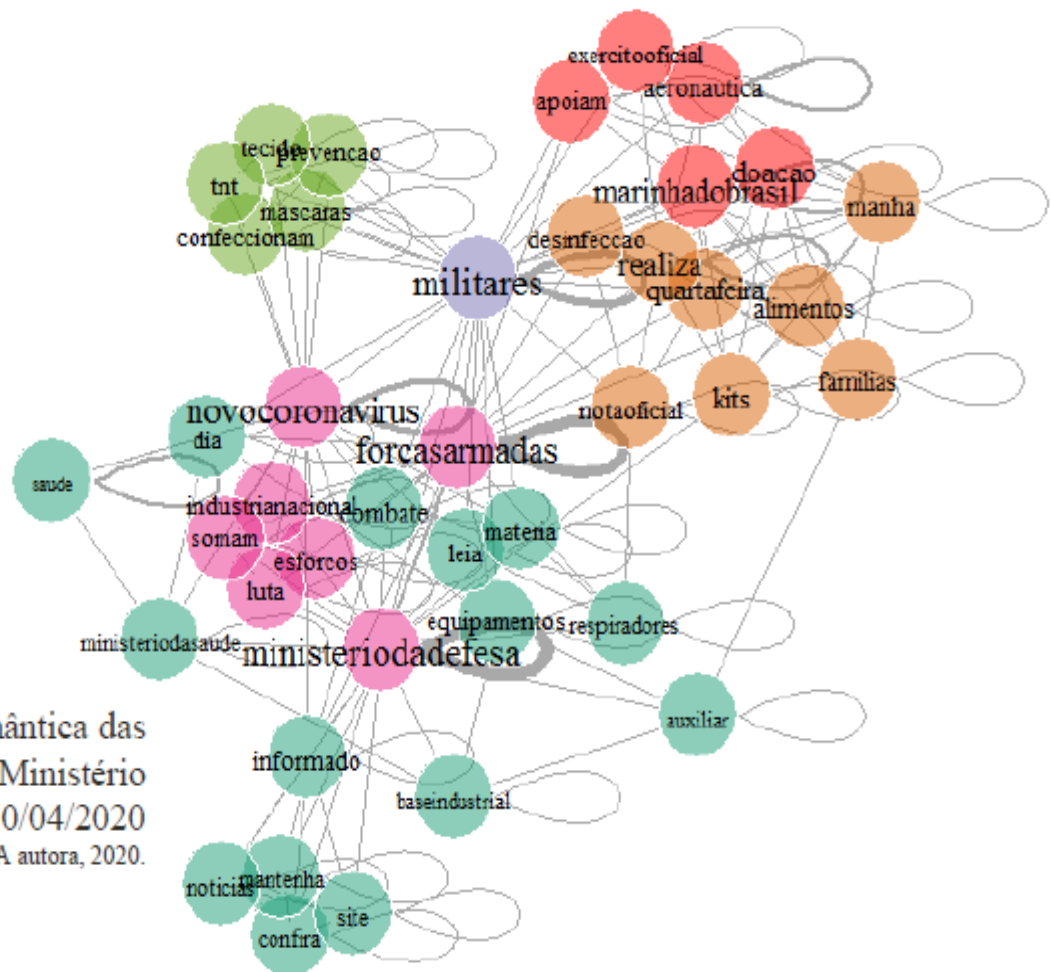


Figura. Web Semântica das palavras do Twitter do Ministério da Defesa em 10/04/2020
Fonte: A autora, 2020.

TEXT MINING

Vantagem x Desvantagem do *Text Mining*

VANTAGEM



Uso de um material bruto pouco explorado, como fonte de dados.

DESVANTAGEM



Limitações sobre a quantidade de texto pode ser analisado conjuntamente com o método escolhido inicialmente pelo pesquisador, podendo levar à observação de uma parcela pequena do problema e influenciar no resultado final.

Fonte: A autora, 2020.

TEXT MINING

***Text Mining* e Cenários Prospectivos**



Fonte: Emerging Payments Africa, 2017

Em cenários prospectivos, o *Text Mining* seria uma ferramenta quantitativa, mais precisamente da Bibliometria, usada em análise prospectiva para prover uma base de evidências, ou mesmo como ferramenta de previsão e extrapolação de tendências;

Além disso, requer a construção de algoritmos e métodos estatísticos para extração de frequências de palavras similares de qualquer tipo de base de dados textual e o uso da interpretação de analistas humanos.

TEXT MINING

Referências

DIXON, M. **An Overview of Document Mining Technology**, 1997. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=6B148E1215673D6A30D0535B7D70EA71?doi=10.1.1.56.5351&rep=rep1&type=pdf>. Acesso em: 20 mai. 2020.

EPA, Africa. **Project Futures is looking for prospective team members**, 2017. Disponível em: www.elpafrica.com/2017/09/25/project-futures-looking-prospective-team-members/. Acesso em: 20 mai. 2020.

EXCELACOM Inc. **One Minute in 2019**. Infográfico. Adaptado. Acesso em: 22 mai. 2020.

FREITAS, J. E. de F. **O Sistema de Inovação no Setor de Defesa no Brasil: Proposta de uma Metodologia de Análise Prospectiva e seus possíveis Cenários**. Tese (Doutorado em Administração) – Universidade de Brasília, 2013. Disponível em: <https://repositorio.unb.br/handle/10482/13395>. Acesso em: 22 mai. 2020.

LOHIYA, H. **Sentiment Analysis with TextBlob**. Medium, 4 jul. 2018. Disponível em: https://medium.com/@himanshu_23732/sentiment-analysis-with-textblob-6bc2eb9ec4ab. Acesso em: 21 mai. 2020.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. **Mineração de Textos**. Goiânia: Universidade Federal de Goiás, 2007. Disponível em: http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_005-07.pdf. Acesso em: 20 mai. 2020.

TEXT Mining for Dummies: Text Classification with Python. **MC.AI**, 8 mar. 2020. Disponível em: <https://mc.ai/text-mining-for-dummies-text-classification-with-python/>. Acesso em: 20 mai. 2020.

WIVES, L. K. **Tecnologias de Descoberta de Conhecimento em Textos Aplicadas à Inteligência Competitiva**. Tese (Doutorado em Computação) - Universidade Federal do Rio Grande do Sul, 2002. Disponível em: <https://seer.ufrgs.br/cadernosdeinformatica/article/view/v1n1p25-28>. Acesso em: 20 mai. 2020.