



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---

WALLACY SEBASTIAN APARECIDO JERONIMO DE ALMEIDA

ALGORITMO INTELIGENTE PARA OTIMIZAÇÃO DE  
BUSCAS EM FERRAMENTAS DE E-COMMERCE

---

LONDRINA

2023

WALLACY SEBASTIAN APARECIDO JERONIMO DE ALMEIDA

**ALGORITMO INTELIGENTE PARA OTIMIZAÇÃO DE  
BUSCAS EM FERRAMENTAS DE E-COMMERCE**

Dissertação apresentada ao Programa de Mestrado em Ciência da Computação da Universidade Estadual de Londrina para obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof(a). Dr(a). Helen Cristina de Mattos Senefonte

LONDRINA

2023

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

Sobrenome, Nome.

Título do Trabalho : Subtítulo do Trabalho / Nome Sobrenome. - Londrina, 2017.  
100 f. : il.

Orientador: Nome do Orientador Sobrenome do Orientador.

Coorientador: Nome Coorientador Sobrenome Coorientador.

Dissertação (Mestrado em Ciência da Computação) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Ciência da Computação, 2017.

Inclui bibliografia.

1. Assunto 1 - Tese. 2. Assunto 2 - Tese. 3. Assunto 3 - Tese. 4. Assunto 4 - Tese. I. Sobrenome do Orientador, Nome do Orientador. II. Sobrenome Coorientador, Nome Coorientador. III. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação. IV. Título.

WALLACY SEBASTIAN APARECIDO JERONIMO DE ALMEIDA

**ALGORITMO INTELIGENTE PARA OTIMIZAÇÃO DE  
BUSCAS EM FERRAMENTAS DE E-COMMERCE**

Dissertação apresentada ao Programa de Mestrado em Ciência da Computação da Universidade Estadual de Londrina para obtenção do título de Mestre em Ciência da Computação.

**BANCA EXAMINADORA**

---

Orientador: Prof(a). Dr(a). Helen Cristina  
de Mattos Senefonte  
Universidade Estadual de Londrina

---

Prof. Dr. Segundo Membro da Banca  
Universidade/Instituição do Segundo  
Membro da Banca – Sigla instituição

---

Prof. Dr. Terceiro Membro da Banca  
Universidade/Instituição do Terceiro  
Membro da Banca – Sigla instituição

---

Prof. Ms. Quarto Membro da Banca  
Universidade/Instituição do Quarto  
Membro da Banca – Sigla instituição

Londrina, 06 de maio de 2023.

*Este trabalho é dedicado às crianças adultas  
que, quando pequenas, sonharam em se  
tornar cientistas.*

## AGRADECIMENTOS

Os agradecimentos principais são direcionados à Gerald Weber, Miguel Frasson, Leslie H. Watter, Bruno Parente Lima, Flávio de Vasconcellos Corrêa, Otavio Real Salvador, Renato Machnievszc<sup>1</sup> e todos aqueles que contribuíram para que a produção de trabalhos acadêmicos conforme as normas ABNT com L<sup>A</sup>T<sub>E</sub>X fosse possível.

Agradecimentos especiais são direcionados ao Centro de Pesquisa em Arquitetura da Informação<sup>2</sup> da Universidade de Brasília (CPAI), ao grupo de usuários *latex-br*<sup>3</sup> e aos novos voluntários do grupo *abnT<sub>E</sub>X2*<sup>4</sup> que contribuíram e que ainda contribuirão para a evolução do abnT<sub>E</sub>X2.

---

<sup>1</sup> Os nomes dos integrantes do primeiro projeto abnT<sub>E</sub>X foram extraídos de <<http://codigolivre.org.br/projects/abntex/>>

<sup>2</sup> <<http://www.cpai.unb.br/>>

<sup>3</sup> <<http://groups.google.com/group/latex-br>>

<sup>4</sup> <<http://groups.google.com/group/abntex2>> e <<http://abntex2.googlecode.com/>>

*“Não vos amoldeis às estruturas deste mundo, mas transformai-vos pela renovação da mente, a fim de distinguir qual é a vontade de Deus: o que é bom, o que Lhe é agradável, o que é perfeito.  
(Bíblia Sagrada, Romanos 12, 2))*

ALMEIDA, W. S. A. J.. **Algoritmo inteligente para otimização de buscas em ferramentas de e-commerce**. 2023. 23f. Dissertação (Mestrado em Ciência da Computação) – Universidade Estadual de Londrina, Londrina, 2023.

## RESUMO

**Palavras-chave:** Latex. Template ABNT-DC-UEL. Editoração de texto.



ALMEIDA, W. S. A. J.. **Title of the Work**. 2023. 23p. Master's Thesis (Master in Science in Computer Science) – State University of Londrina, Londrina, 2023.

## **ABSTRACT**

This is the english abstract. The Abstract in English should be faithful to the Resumo in Portuguese, but not a literal translation.

**Keywords:** Latex. ABNT-DC-UEL template. Text editoration.

## LISTA DE ILUSTRAÇÕES

Figura 1	– Imagem retirada do artigo A Review on Word Embedding Techniques for Text Classification, Birunda et. al: Tipos de Word Embedding [1]	. 14
Figura 2	– Imagem retirada do artigo Efficient Estimation of Word Representations in Vector Space, Mikolov et. al: Comparação entre CBOW e Skip-gram [2]	. . . . . 15
Figura 3	– Imagem retirada do artigo BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Devlin et. al: Procedimentos realizados no BERT para treiná-lo e ajustá-lo [3]	. . . . . 16

## LISTA DE TABELAS

## LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
BNDES	Banco Nacional de Desenvolvimento Econômico e Social
IBGE	Instituto Nacional de Geografia e Estatística
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
NBR	Norma Brasileira

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>13</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICO-METODOLÓGICA E ES-</b>	
	<b>TADO DA ARTE . . . . .</b>	<b>14</b>
<b>2.1</b>	<b>Processamento de Linguagem Natural . . . . .</b>	<b>14</b>
<b>2.2</b>	<b>Word Embedding . . . . .</b>	<b>14</b>
2.2.1	word2vec . . . . .	14
2.2.2	TF-IDF . . . . .	16
2.2.3	BERT . . . . .	16
<b>3</b>	<b>CONCLUSÃO . . . . .</b>	<b>17</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>18</b>
	<b>APÊNDICES</b>	<b>19</b>
	<b>APÊNDICE A – QUISQUE LIBERO JUSTO . . . . .</b>	<b>20</b>
	<b>ANEXOS</b>	<b>21</b>
	<b>ANEXO A – MORBI ULTRICES RUTRUM LOREM. . . . .</b>	<b>22</b>
	<b>Trabalhos Publicados pelo Autor . . . . .</b>	<b>23</b>

# 1 INTRODUÇÃO

A Inteligência Artificial (IA) ganhou importância neste último ano perante o público em geral, através de produtos como ChatGPT da OpenAI e Bard do Google. O surgimento de produtos como estes são resultados do incentivo dado pelas empresas das maiores áreas econômicas do mundo, como as áreas da saúde, educação e varejo, para que reduzissem custos e melhorassem o trabalho remoto durante a pandemia. Após este período, ainda há um crescente uso de IAs no dia-a-dia das pessoas, não só através dos produtos mencionados como em outros produtos do mercado, para reconhecimento de imagens e vídeos, processamento de textos, automatização de tarefas, análises no mercado financeiro, dentre muitas outras utilizações [4].

Quando observamos mais de perto a área do varejo, vemos o quanto a pandemia influenciou no crescimento dos *e-commerces* (termo usado para definir um comércio eletrônico). Durante este período, as ruas estavam com circulação restrita, os estabelecimentos possuíam limite de pessoas e muitas empresas do varejo necessitaram de uma mudança estratégica de venda, levando estas a vender pela Internet, seja através de *marketplaces* - modelo de negócio que possibilita a venda através de uma plataforma utilizada por vários lojistas -, de lojas virtuais, de redes sociais ou de aplicativos como WhatsApp [5]. E com o aumento do uso de e-commerces, naturalmente aumentaram também os problemas enfrentados pelos lojistas e pelos usuários que compram.

Desde que os sistemas ERP (Enterprise Resource Planning, ou Sistema de Gestão Integrado) surgiram na década de 90, as empresas têm recebido bem a solução para gerir os negócios [6], e com os e-commerces não foi diferente. Muitos e-commerces recorreram aos ERPs para gerenciar seus negócios, buscando não só ferramentas básicas (como controle de estoque, preços, entre outros) para os seus produtos, mas também uma forma de potencializar suas vendas através de integrações com outros sistemas [7]. Algumas das integrações de interesse foram os marketplaces: a empresa coloca seu produto à venda em seu e-commerce, e ao mesmo tempo em algum marketplace, sendo da responsabilidade do ERP gerenciar automaticamente os estoques das duas plataformas.

Junto às necessidades de automação devido ao grande volume de produtos de algumas empresas, surgiu também a necessidade de ferramentas que fizessem o seu produto ter mais visibilidade e praticidade ao integrá-los em marketplaces. Este estudo será focado na empresa MKX Lojas Virtuais - a qual desenvolve e-commerces para outras empresas [8] - e visa facilitar a correção e alteração de nomes de produtos, baseado em seus nomes e descrições atuais, para que a integração com o seu ERP e o marketplace Mercado Livre [9] obtenha aumento na precisão da busca de categorias de produtos do Mercado Livre.

## 2 FUNDAMENTAÇÃO TEÓRICO-METODOLÓGICA E ESTADO DA ARTE

### 2.1 Processamento de Linguagem Natural

Para que o estudo seja realizado, será necessário uma pesquisa aprofundada em alguns modelos para Processamento de Linguagem Natural (NLP - Natural Language Processing), sendo esta uma potencial solução para problemas que exigem a aquisição e transformação de conhecimento através de inteligência artificial [10].

### 2.2 Word Embedding

Uma maneira de representar textos é o método Word Embedding, ou Word Vector, que representa textos em vetores de palavras, atribuindo valores numéricos para cada palavra de acordo com o significado dela [11]. Como mostra a Figura 1, existem diversos algoritmos que se utilizam desta representação, e dentre eles foram escolhidos um de cada tipo para este estudo, descritos brevemente a seguir.

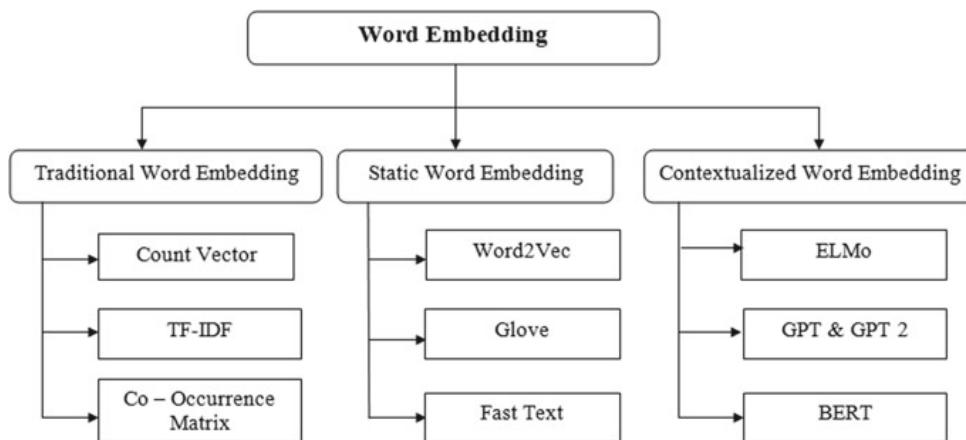


Figura 1 – Imagem retirada do artigo A Review on Word Embedding Techniques for Text Classification, Birunda et. al: Tipos de Word Embedding [1]

#### 2.2.1 word2vec

Foi desenvolvido pela Google em 2013, utiliza como métrica a similaridade de cossenos e possui duas variantes: CBOW (Continuous Bag of Words) e Skip-Gram. Na Figura 2, vemos que a variante CBOW tem como entrada as palavras ao redor da palavra que se quer prever, enquanto a variante Skip-gram recebe como entrada uma palavra e prevê as palavras ao redor desta.

Para treinar um modelo  $m$  da variante Skip-gram, temos:

$$m = \frac{1}{A} \sum_{a=1}^A v_a \quad (2.1)$$

onde

$A$  é a quantidade total de palavras em um documento  $d$ ,

$a = 1, 2, 3, \dots, A$ ,

$v_a$  é u

$$v_a = \sum_{-c \leq j \leq c, i \neq 0} \log p(w_{a+i} | w_a) \quad (2.2)$$

A Similaridade de Cosseno (Cosine Similarity) é uma métrica amplamente usada para identificar a similaridade entre dois textos, transformando estes textos em vetores e obtendo um valor que varia entre -1 e 1, onde -1 significa que um texto é totalmente oposto ao outro, e 1 significa que um texto é totalmente similar ao outro. Pode-se obter também o valor 0, onde não há similaridade [10]. Dado que os nomes dos produtos que serão utilizados neste estudo podem ser substituídos por nomes mais adequados para busca, a similaridade de cosseno pode ser útil não só como métrica mas também na verificação de palavras semelhantes ao substituir palavras escritas incorretamente por palavras gramaticalmente corretas.

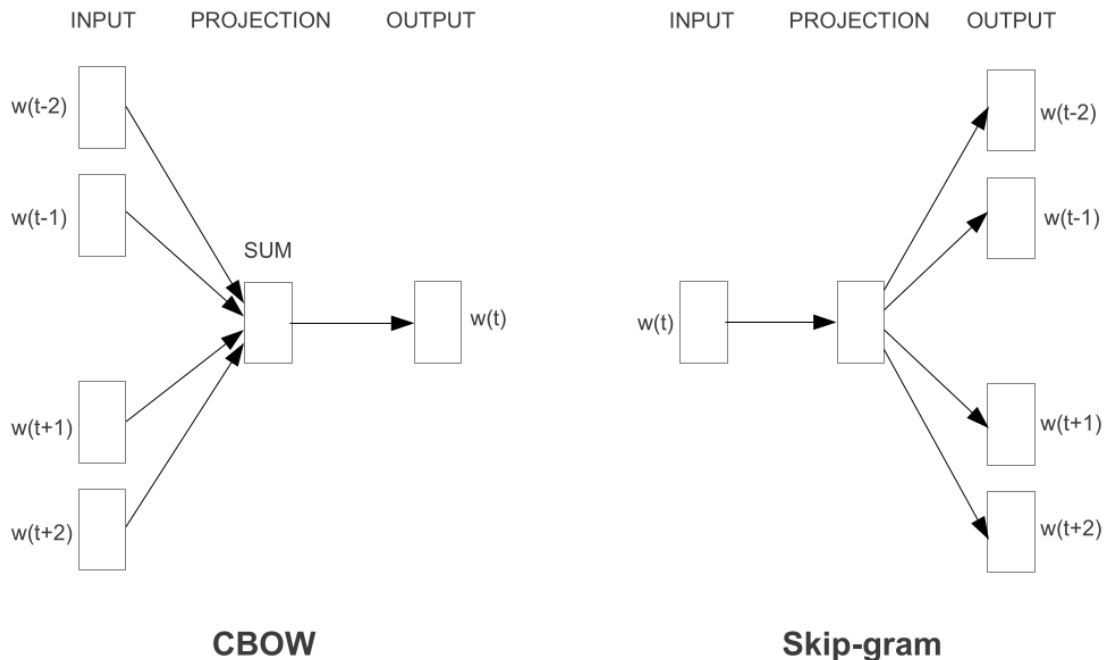


Figura 2 – Imagem retirada do artigo Efficient Estimation of Word Representations in Vector Space, Mikolov et. al: Comparação entre CBOW e Skip-gram [2]



### 2.2.2 TF-IDF

Denominado TF-IDF por ser a multiplicação entre a frequência de uma palavra em um documento (Term Frequency - TF) e a frequência inversa de aparições dessa palavra em outros documentos (Inverse Document Frequency - IDF). É muito utilizado e simples de ser implementado. Enquanto o TF foca na quantidade de vezes que uma palavra aparece em um documento, o IDF tem como foco a raridade da palavra em outros documentos, e dessa forma é possível obter a relevância da palavra em um determinado documento [12].

Apesar de existirem algumas variantes, geralmente utiliza-se a seguinte equação matemática:

$$p_d = \text{tf-idf}_{p,d} = \text{tf}_{p,d} \cdot \text{idf}_p \quad (2.3)$$

sendo

$p$  a palavra buscada,

$d$  o documento utilizado,

$$\text{tf}_{p,d} = \frac{\text{Quantidade de vezes que aparece } p \text{ no documento } d}{\text{Quantidade total de palavras no documento } d},$$

$$\text{idf}_p = \log_2\left(\frac{\text{Quantidade total de documentos}}{\text{Quantidade de documentos que aparece } p}\right).$$

### 2.2.3 BERT

Apresentado pela Google em 2019, BERT (Bidirectional Encoder Representations from Transformers) foca em seu pré-treinamento, onde recebe uma entrada, esconde palavras aleatórias e o objetivo é tentar prever a palavra original da palavra escondida baseando-se somente em seu contexto [3].

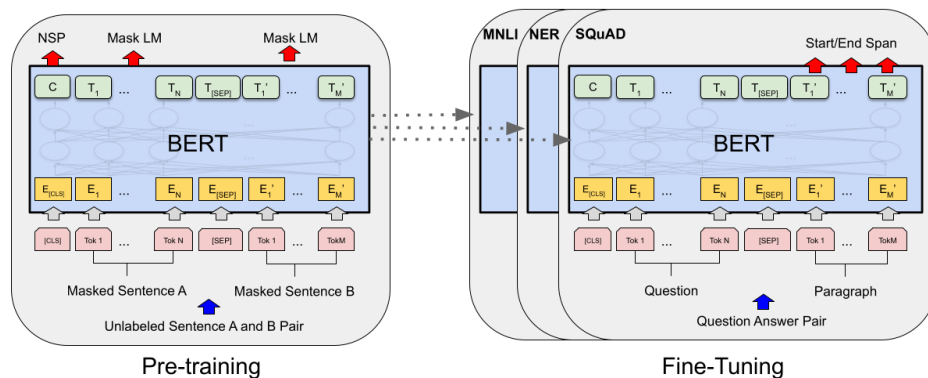


Figura 3 – Imagem retirada do artigo BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Devlin et. al: Procedimentos realizados no BERT para treiná-lo e ajustá-lo [3]

### 3 CONCLUSÃO

## REFERÊNCIAS

- [1] BIRUNDA, S. S.; DEVI, R. K. A review on word embedding techniques for text classification. In: RAJ, J. S. et al. (Ed.). *Innovative Data Communication Technologies and Application*. Singapore: Springer Singapore, 2021. p. 267–281. ISBN 978-981-15-9651-3.
- [2] MIKOLOV, T. et al. *Efficient Estimation of Word Representations in Vector Space*. 2013.
- [3] DEVLIN, J. et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019.
- [4] FORUM, W. E. *COVID-19 increased the use of AI. Here's why it's here to stay*. 2021. <<https://www.weforum.org/agenda/2021/02/covid-19-increased-use-of-ai-here-s-why-its-here-to-stay/>>. [Online; acessado em 07-Setembro-2023].
- [5] E-commerce Brasil. *Evolução do e-commerce: cinco fatos para entender melhor o mercado*. 2022. <<https://www.ecommercebrasil.com.br/artigos/evolucao-do-e-commerce>>. [Online; acessado em 07-Setembro-2023].
- [6] AL-MASHARI, M. Enterprise resource planning (erp) systems: a research agenda. *Industrial Management & Data Systems*, MCB UP Ltd, v. 102, n. 3, p. 165–170, Jan 2002. ISSN 0263-5577. Disponível em: <<https://doi.org/10.1108/02635570210421354>>.
- [7] ORACLE. *Your Complete Guide to Modern ERP*. 2017. <https://www.oracle.com/webfolder/s/assets/ebook/modern-erp/index.html>. [Online; acessado em 12-Setembro-2023].
- [8] VIRTUAIS, M. L. *Sobre a MKX e-commerce - Desenvolvimento de Lojas Virtuais e Sites*. 2019. <https://www.mkx.com.br/empresa/>. [Online; acessado em 15-Setembro-2023].
- [9] LIBRE, M. *Documentação do Mercado Livre*. 2020. [https://developers.mercadolivre.com.br/pt\\_br/guia-para-produtos](https://developers.mercadolivre.com.br/pt_br/guia-para-produtos). [Online; acessado em 15-Setembro-2023].
- [10] EISENSTEIN, J. *Natural Language Processing*. [S.l.]: The MIT Press, 2019.
- [11] TURING. *A Guide on Word Embeddings in NLP*. 2022. <<https://www.turing.com/kb/guide-on-word-embeddings-in-nlp>>. [Online; acessado em 08-Setembro-2023].
- [12] RAMOS, J. *Using TF-IDF to determine word relevance in document queries*. 2003.

## Apêndices

## APÊNDICE A – QUISQUE LIBERO JUSTO

Quisque facilisis auctor sapien. Pellentesque gravida hendrerit lectus. Mauris rutrum sodales sapien. Fusce hendrerit sem vel lorem. Integer pellentesque massa vel augue. Integer elit tortor, feugiat quis, sagittis et, ornare non, lacus. Vestibulum posuere pellentesque eros. Quisque venenatis ipsum dictum nulla. Aliquam quis quam non metus eleifend interdum. Nam eget sapien ac mauris malesuada adipiscing. Etiam eleifend neque sed quam. Nulla facilisi. Proin a ligula. Sed id dui eu nibh egestas tincidunt. Suspendisse arcu.

## **Anexos**

## ANEXO A – MORBI ULTRICES RUTRUM LOREM.

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.

## TRABALHOS PUBLICADOS PELO AUTOR

Trabalhos publicados pelo autor durante o programa.

Publicações principais do trabalho.

1. Jose da silva, autor2 da silva, orientador da silva, **Título do artigo**, local onde foi publicado, mês/ano, editora, número de página, isbn, etc. (Qualis CC 2017, xx)
2. Jose da silva, autor2 da silva, orientador da silva, **Título do artigo**, local onde foi publicado, mês/ano, editora, número de página, isbn, etc. (Qualis CC 2017, xx)
3. Jose da silva, autor2 da silva, orientador da silva, **Título do artigo**, local onde foi publicado, mês/ano, editora, número de página, isbn, etc. (Qualis CC 2017, xx)

Publicações complementares.

1. Jose da silva, autor2 da silva, orientador da silva, **Título do artigo**, local onde foi publicado, mês/ano, editora, número de página, isbn, etc. (Qualis CC 2017, xx)
2. Jose da silva, autor2 da silva, orientador da silva, etc. **Título do artigo**, local onde foi publicado, mês/ano, editora, número de página, isbn, (Qualis CC 2017, xx)