Analisador Léxico-Morfológico Aplicado à Saúde

Um estudo léxico-morfológico aplicado à termos frequentemente utilizados na área de saúde

Gabriela T. Hirashima  
 Departamento de Ciência de Computação  
 Universidade Estadual de Londrina Londrina-PR, Brasil  
 gabriela.hirashima@uel.br

Cinthyan R. S. C. de Barbosa  
 Departamento de Ciência de Computação  
 Universidade Estadual de Londrina  
 Londrina-PR, Brasil  
 [cinthyan@uel.br](mailto:cinthyan@uel.br)

Pedro W. Hernandes  
 Departamento de Ciência de Computação  
 Universidade Estadual de Londrina  
 Londrina-PR, Brasil  
 deprao.pwhernndes@uel.br

ABSTRACT

In this study, we aimed to create a simple yet practical program to help with a problem that the Brazilian-Portuguese language has. Brazil is a vast and culturally rich country, which means that words could have different meanings or not even exist in different areas of the country. This clash of words often represents a barrier between doctors and patients, and may sometimes, lead to fatal mistakes. To help with that, a study of technical medical words, popular variants and definitions was used as a database and a Python code was developed with the goal of shortening the language barrier within the Brazilian-Portuguese language.

RESUMO

Nesse estudo, o objetivo foi de criar um simples, porém prático programa que auxilie em um grande problema que a língua Portuguesa do Brasil possui. Devido ao seu vasto tamanho e cultura, palavras podem possuir diferentes significados ou até mesmo não existir em diferentes regiões do país. Tal fato, dificulta a conversa entre médicos e pacientes, o que pode levar a erros fatais. Desta maneira, utilizando como base um estudo que aborda os termos técnicos utilizados com suas variantes populares e definições, um código escrito em Python foi desenvolvido para diminuir essas dificuldades encontradas na língua Portuguesa do Brasil.

Palavras-Chave ACM

Applied Computing; Natural Language Processing;

Palavras-Chave

Análise Léxico-Morfológica; Saúde e Tecnologia;

1 Introdução

Para abordar o assunto de maneira coerente, é preciso entender as variantes regionais que uma mesma palavra pode assumir. O problema de regiões diferentes do país possuírem diversos sotaques, variantes linguísticas e costumes, representam dificuldades para os nativos brasileiros e ainda mais, para os estrangeiros.

Professores e médicos brasileiros participantes do programa Língua Portuguesa para o Programa Mais Médicos para o Brasil, relataram a existência de duas linguagens no Sistema Único de Saúde(SUS): A linguagem técnica, falada por profissionais da área de saúde e a linguagem geral, utilizada por pacientes ou familiares de pacientes. (p.10)[3]

Este desentendimento entre os grupos não ocorre somente entre estrangeiros e nativos. Até mesmo dentro do país, os dialetos se expressam de maneiras extraordinariamente diferentes e que podem variar por diversos fatores como propriedade intelectual, classe social, contato cultural etc. Desta maneira, não é surpreendente notar a existência o distanciamento entre o profissional de saúde e a população geral, durante atendimentos médicos.

A partir deste ponto, o programa de analisador léxico-morfológico busca suprir a dificuldade em associação de palavras de cunho técnico e suas alternativas, oferecendo um método rápido e prático para o uso de profissionais e não profissionais.

2 Coleta de Dados

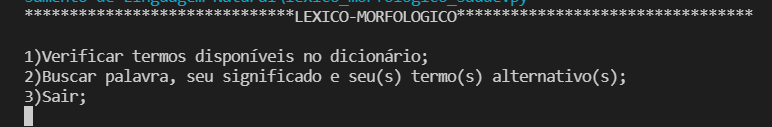
A principal fonte de dados utilizado para este trabalho é a de Silvestrow, em Contribuições dos Estudos Terminológicos Para os Profissionais da Saúde Básica do SUS[3]. Ao longo de 10 anos, os termos técnicos e seus respectivos coloquiais foram estudados e correlacionados, a fim de documentar as variações léxicas e morfológicas dos termos técnicos da área de saúde.

A partir das informações coletadas no documento citado[3], um banco de dados foi formado para a consulta, oferecendo alguns dos termos mais utilizados na área atendimento de saúde e armazenando os seguintes atributos citados: ‘Termo Pesquisado’, ‘Categoria’, ‘Variação(ões) Denominativa(s)’, ‘Variante(s) Popular(es)’ e ‘Definição’.

No documento citado, utilizamos 42 palavras das 84 disponíveis [3] para o dicionário de termos técnicos e 70 palavras populares alternativas a termos técnicos [3] para o dicionário de palavras técnicas, totalizando 112 palavras disponíveis no trabalho.

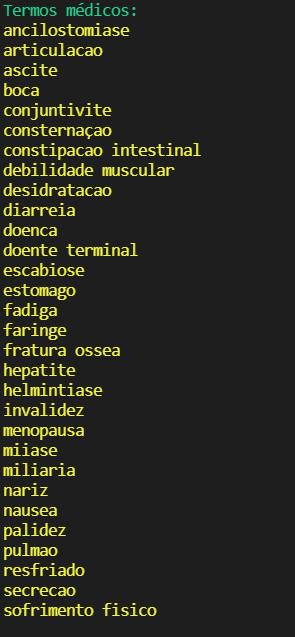
3 Funcionamento do Programa

O sistema possui o seguinte funcionamento: Ao inicializar o programa, um menu será mostrado ao usuário com as funções de ‘Verificar os termos disponíveis no dicionário’, ‘Buscar palavra, seu significado e seu(s) termo(s) alternativo(s)’ e uma opção para ‘Sair’, como mostrado na figura 1.

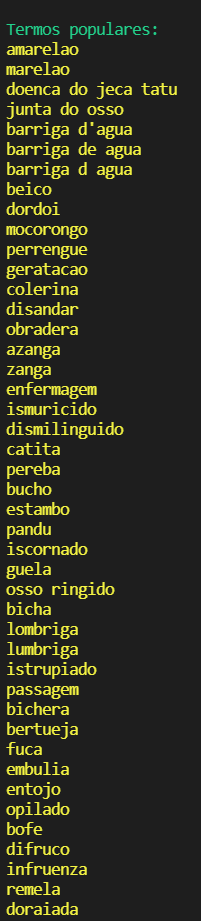


. **Figura 1:** Menu – Estrutura de Controle do Programa. [Fonte: Autores]

Ao selecionar a opção 1, o programa mostra listas de termos disponíveis para a busca em tons amarelados, considerando o grupo de termos médicos e de termos populares. Para colorir, foi utilizado a biblioteca Colorama disponível para Python. Na Figura 2 e 3, mostramos como estas listas são apresentadas.

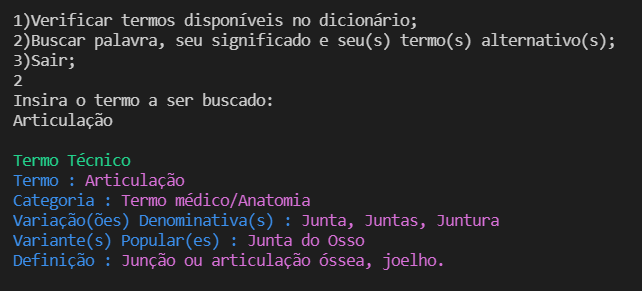


. **Figura 2:** Lista de palavras do dicionário de termos técnicos. [Fonte: Autores]



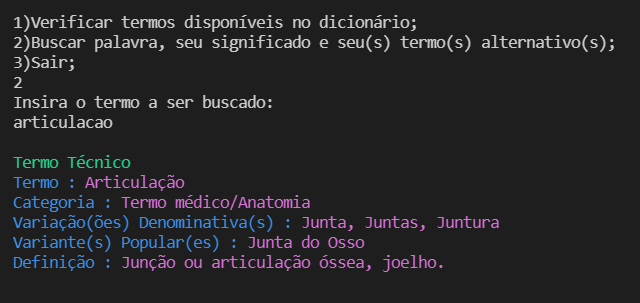
. **Figura 3:** Lista de palavras do dicionário de termos populares. [Fonte: Autores]

Ao selecionar a opção 2, o acesso às informações de cada palavra armazenada na estrutura de tabela de espalhamento (hashtable) é disponibilizada, porém, apenas quando esta é existente na base de dados. Na figura **4**, utilizamos um exemplo ao buscar a palavra “Articulação”.

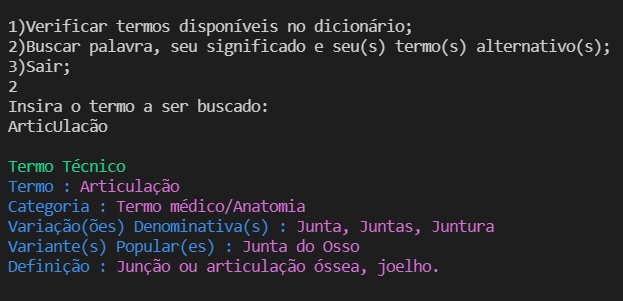


**. Figura 4**: Busca pela palavra ‘Articulação’. [Fonte: Autores]

É importante notar que na lista de palavras disponibilizadas pelo dicionário, como mostrado na Figura 2, ‘Articulação’ é listada como ‘articulacao’. Esta simplificação na definição da palavra, permite que durante a busca, a maioria das flexões de palavra (no que se diz a letras maiúsculas, minúsculas e acentos) sejam aceitas e processadas adequadamente. Na Figura **5** e **6**, temos exemplos de buscas realizadas de maneira alternativa. Podemos observar que a simplificação da palavra de busca, resulta em uma maior flexibilização e compreensão por parte do programa.

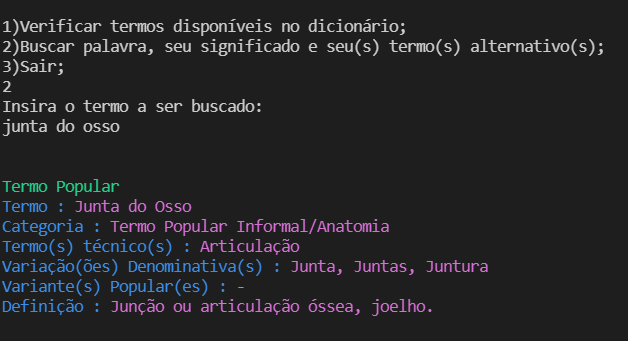


**Figura 5:** Busca pela palavra ‘articulacao’. [Fonte: Autores]



**Figura 6:** Busca pela palavra ‘ArticUlacão’. [Fonte: Autores]

Ainda seguindo nesta mesma linha, pode ser observar que a variantes popular da palavra ‘Articulação’ é Junta do Osso. Exemplificando uma pesquisa desta variante temos[Figura 7]:



**. Figura 7:** Busca pela palavra ‘Junta do Osso’. [Fonte: Autores]

Por fim, a opção de ‘Sair’ de número 3, apenas finaliza a execução do programa.

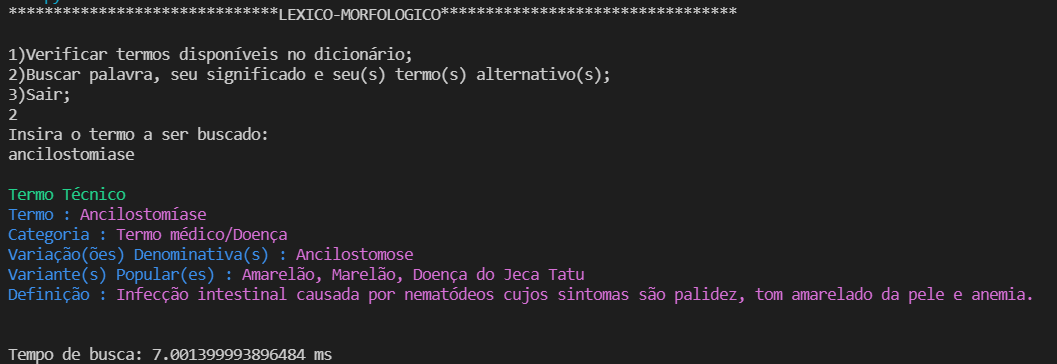
3 Código

A estrutura do código é separada entre definição dos elementos, dicionário com base em hashtable e a estrutura de controle de fluxo do programa. Também conta com uma função para manipular a entrada do usuário e converter para o formato aceito pelo dicionário criado.

O uso do dicionário neste caso, se justifica pelo volume de dados pequeno e também, pela praticidade e eficiência da função de hashing da própria linguagem Python.

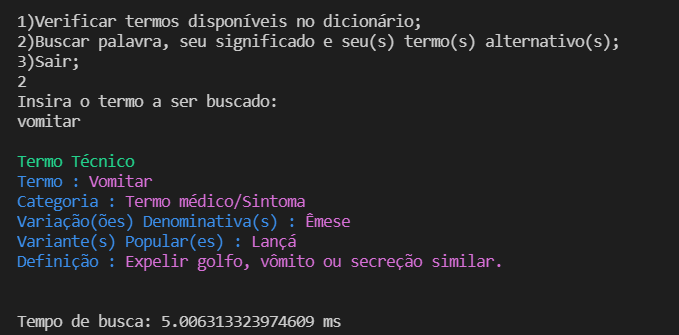
Na linguagem, a função de hashing recebe um elemento e retorna a chave como um valor inteiro. No caso de strings, os elementos sempre possuirão chaves diferentes devido ao valor aleatório que a função de hash cria.

Os dicionários são efetivos neste caso, já que possuem velocidade além da efetividade, ambas características importantes para o desempenho do programa. Para a visualização da rapidez introduzida ao funcionamento, vamos exemplificar a busca pelo primeiro elemento inserido no dicionário técnico, “ancilostomiase”, mostrando o tempo necessário para completar a operação [Figura 8].



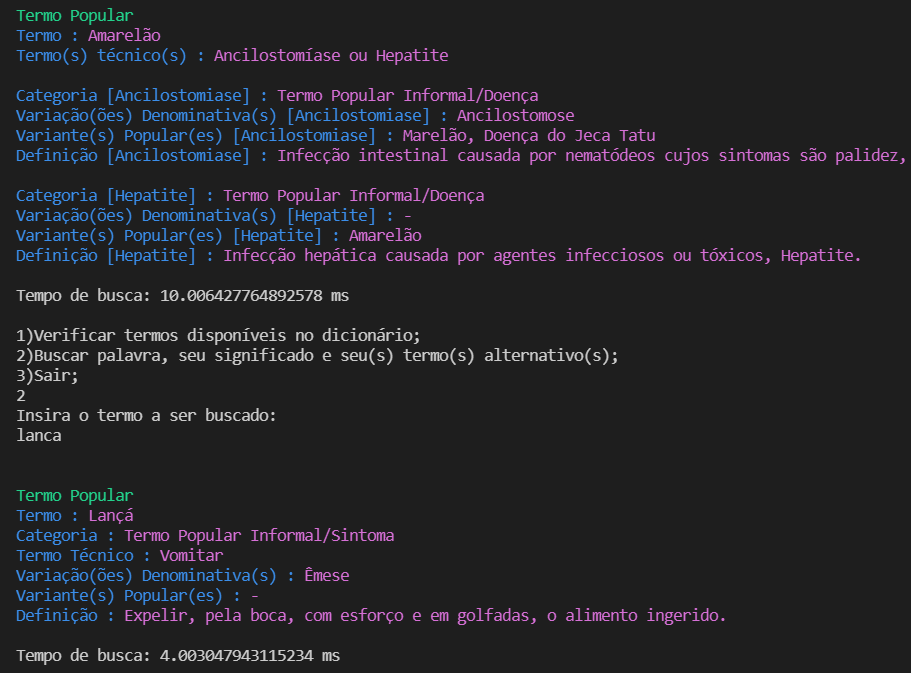
**Figura 8:** Busca por palavra ‘Ancilostomíase’. [Fonte: Autores]

Podemos observar na Figura 8 que o tempo total para buscar e retornar o primeiro elemento inserido é de 7.001ms. Agora o último elemento inserido, o verbo “vomitar”.



**Figura 9:** Busca por palavra ‘Vomitar’. [Fonte: Autores]

É possível notar que o tempo levado para encontrar tal termo é de 5.006ms. Com o dicionário popular, o tempo para encontrar o primeiro e o segundo termos inseridos (Amarelão e Lançá) é ilustrado na Figura 10.

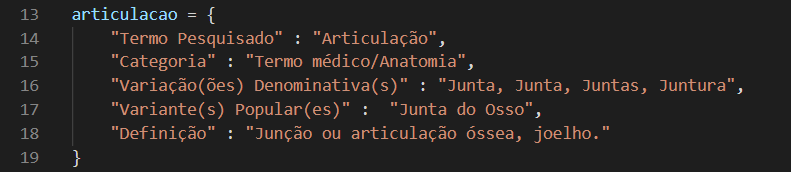


**Figura 10:** Busca por palavras ‘amarelao’ e ‘lanca’. [Fonte: Autores]

A partir desses testes, é perceptível que não há variações de tempo grandes o suficiente para que haja problema. Com exceção do termo ‘Amarelão’ que requer mais dados exibidos, a variação de tempo é pequena e não causa incômodo ao usuário, um dos objetivos do programa .

Cada elemento do dicionário técnico(hashtable) possui os seguintes atributos: Sua chave-geral (o termo técnico simplificado), responsável por identificar o elemento, e os valores, ‘Termo’, ‘Categoria’, ‘Variação(ões) Denominativa(s)’, ‘Variante(s) Popular(es)’ e ‘Definição’. Já elementos do dicionário de termos populares possuem: ‘Termo’, ‘Categoria’, ‘Termo Pesquisado’, ‘Variação(ões) Denominativa(s)’, ‘Variante(s) Popular(es)’ e ‘Definição’.

Para demonstrar visualmente, novamente utilizaremos o termo “Articulação” para exemplificar sua definição na Figura 11.



**Figura 11:** Definição da palavra ‘Articulação’. [Fonte: Autores]

Desta maneira, os elementos são criados já dentro do programa. Para a criação dos dicionários, os elementos e as chaves são associados da seguinte maneira: a chave entre aspas, separada por dois pontos e o elemento definido no escopo. No código criado, temos duas estruturas, para os termos técnicos e populares. Um é chamado de ‘dicionario\_tecnico’ e possui o formato apresentado na Figura 12, e assim como o nome supõe, relaciona os termos utilizados pelos profissionais de saúde ao outros alternativos.



**Figura 12:** Dicionário Técnico. [Fonte: Autores]

O outro dicionário criado é de palavras populares [Figura 13] (as mesmas relacionadas as suas alternativas). Desta maneira, o vocabulário é expandido com o objetivo de atender a maioria das variações dos termos inseridos no léxico.



**Figura 13:** Dicionário Popular. [Fonte: Autores]

Para o controle de fluxo de programa, uma condição de while associada a condições de if são aninhadas para garantir o funcionamento do código, o qual o encerramento é obtido apenas quando a condição de saída é passada no console. (Condição de número 3 da Figura 1)

4 Conclusão

A língua portuguesa é rica ao ser vista por vários ângulos, e mais ainda, quando analisada morfologicamente. De maneira orgânica, as palavras possuem diversas flexões e variantes, o que por si, já a torna mais complexa do que muitas outras línguas. Considerando essa dificuldade natural que a própria linguagem traz, a variação regional de sotaques e costumes, elevam o nível de acessibilidade de forma exponencial.

Esta complexidade se importa para a comunicação na área da saúde, o qual já é observada há muito tempo. O desentendimento causado por profissionais da área de saúde por utilizarem somente termos técnicos, e o dos pacientes por utilizarem termos populares ou regionais desconhecidos pelos profissionais, se torna uma barreira perigosa durante diagnósticos, consultas ou procedimentos corriqueiros.

A partir desta dor observada no ambiente de saúde, um programa como o citado neste trabalho, pode ser de extremo auxílio, se desenvolvido com mais profundeza.

Embora instrumentos de pesquisa possam auxiliar na busca de tais termos desconhecidos, uma ferramenta especializada tendo como base um trabalho extenso e completo sobre o léxico aplicado a saúde, se sobressairia de maneira positiva e confiável. Um possível uso, por exemplo, seria um analisador de sentenças completas. Aliando os significados dos termos citados nos dicionários com a análise de sentença natural, seria possível realizar uma “tradução” de uma frase técnica para uma de mais fácil compreensão.

Desta forma, é possível concluir que a continuação e o refinamento do estudo poderiam criar um objeto auxiliar de extrema importância da saúde pública brasileira, aproximando os profissionais de saúde e os pacientes, e consequentemente, melhorando a qualidade do sistema de saúde.

REFERENCES

[1] VERANZA, Solange Coelho. *Contextualizando o léxico como objeto de estudo: considerações sobre sinonímia e referência*. Scielo Brasil. Acesso em Novembro de 2021, Disponível em: https://www.scielo.br/j/delta/a/x3ZVpH4tTgWCpSRfbnWkTHG/?lang=pt

[2] REY-DEBOVE, Josette. 1984. *Léxico e Dicionário*. Tradução de Clóvis Barleta de Morais. Alfa, São Paulo.

[3] SILVESTROW, Tatiana Martins Mendes. 2017. *Contribuições dos Estudos Terminológicos para os Profissionais da Saúde Básica do SUS.* Belo Horizonte, Faculdade de Letras da UFMG.

[4] MATROMATTED, Davide. Python Hashtable: Understanding dictionaries. The Python Corrner.Disponível em <https://thepythoncorner.com/posts/2020-08-21-hash-tables-understanding-dictionaries//>. Acesso em Novemrbo de 2021.