

## *Medical Mining*

**Felipe Vieira Nunes<sup>1</sup>, Vitor Silva de Freitas<sup>1</sup>, Vitor Angelo Silva<sup>1</sup>, Luan Soares<sup>1</sup>, Thiago Magela Rodrigues Dias<sup>1</sup>.**

<sup>1</sup>Centro Federal de Educacao Tecnologica – CEFET MG – Divinópolis, Minas Gerais  
– Brasil

felipe041198@hotmail.com, vitorsilvafreitas@hotmail.com, vitorangelo753@gmail.com,  
luan.soares.o@gmail.com , thiagomagela@gmail.com

**Abstract.** *This article describes the Medical Mining project, a portal that covers and delivers medical contents to the user, both basic and advanced. With the emergence of new diseases day after day and consequent medical advances, health professionals should be aware of all these new news. Often, these professionals end up looking for such information in other media outlets, including medical publication sites. The problem is that most of these sites that have functions similar to Medical Mining do not present a detailed analysis of scientific articles in the medical field, only offering summary information. With this, Medical Mining comes with the purpose of streamlining and facilitating the search of scientific articles in any area of health, providing the user with ease in finding and finding complete articles about what was researched, besides presenting tables and graphs adding important information Subject. Such functions that the portal promotes are made from the adoption of data mining techniques to improve the results presented, in addition to a small encyclopedia, through a website in the area of medicine.*

**Resumo.** *Este artigo descreve o projeto Medical Mining, um portal que abrange e oferece conteúdos médicos para o usuário, tanto básicos quanto avançados. Com o surgimento de novas doenças dia após dia e consequente avanços médicos, profissionais da área da saúde devem ficar atentos a todas essas novas notícias. Muitas vezes, estes profissionais acabam procurando tais informações em outros veículos de comunicação, inclusive sites de publicações médicas. O problema é que a maioria desses sites que tem funções semelhantes ao Medical Mining não apresentam uma análise detalhada dos artigos científicos da área da medicina, apenas oferecendo informações resumidas. Com isso, Medical Mining vem com o propósito de agilizar e facilitar a pesquisa de artigos científicos em qualquer área da saúde, disponibilizando ao usuário facilidade na busca e localização de artigos completos sobre o que foi pesquisado, além de apresentar tabelas e gráficos acrescentando informações importantes ao assunto pesquisado. Tais funções que o portal promove são feitas a partir da adoção de técnicas de mineração de dados para melhoria nos resultados apresentados, além de uma pequena enciclopédia, por meio de um site na área da medicina.*

## 1. INTRODUÇÃO

Em um mundo em que dia após dia surgem casos de novas doenças, como por exemplo, o *zika* vírus no Brasil [1], os avanços médicos acompanham esse ritmo e são maiores e cada vez mais rápidos. Sendo assim, médicos e outros profissionais de saúde devem estar sempre atualizados das últimas notícias e pesquisas na área da medicina. Para atingirem este objetivo, recorrem a sites de material científico de publicações médicas ou revistas científicas para se manterem atualizados, mas nem sempre são maneiras eficazes de pesquisa, já que as consultas geralmente apresentam resultados de forma resumida ou fracionada. Tendo em vista estes problemas, o site propõe analisar um conjunto composto por vários artigos de uma forma simples, rápida e completa, apresentando gráficos e tabelas para uma melhor visualização e um resumo de pontos importantes da pesquisa, sem apresentar longos textos de publicações, com a ajuda de técnicas de mineração de dados.

## 2. OBJETIVO

O objetivo do portal é facilitar a busca de artigos científicos da área da medicina em qualquer área da medicina, apresentando os artigos sobre os temas pesquisados, gráficos e tabelas com informações complementares para melhor visualização dos resultados. O portal tem como principais contribuições:

- 1- *Agilizar a pesquisa sobre qualquer doença e suas definições;*
- 2- *Demonstrar resultados de pesquisas sobre doenças e assuntos relacionados à elas.*

## 3. REFERENCIAL TEÓRICO

### 3.1 Trabalhos relacionados

Um primeiro aplicativo que tem um viés similar deste projeto é o *Read by QxMod*[2]. Caracteriza-se por apresentar os artigos em síntese, apenas listando-os, por exemplo, por popularidade ou data, sem oferecer nenhuma análise estatística. Mesmo assim, este aplicativo é muito popular, com avaliação de 4.6 na *Google Play* e mais de 50 mil downloads[3]. A Figura 1 apresenta a tela inicial do *Read by QxMod*.

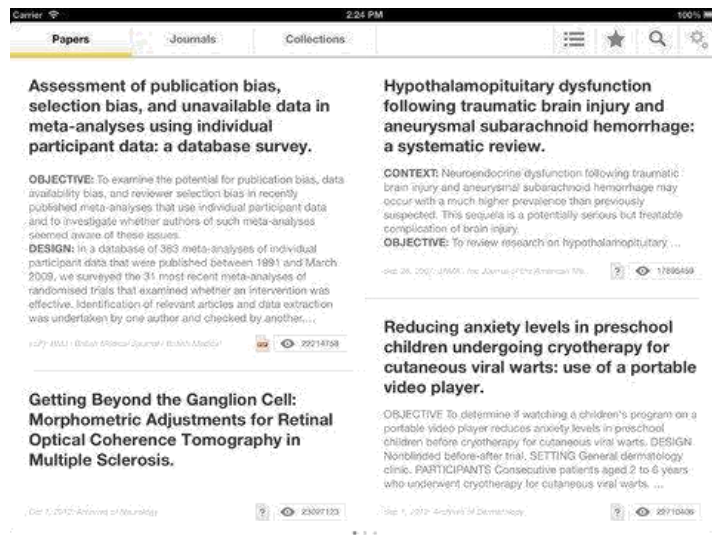


Figura 1: Read by QxMod.

Outra aplicação semelhante é o *PubMed PubReMiner*, uma plataforma de busca de livre acesso à base de dados *Medline* de citações e resumos de artigos científicos da área da medicina[4], com uma proposta similar ao portal *Medical Mining*. Este site apresenta as análises estatísticas por meio de gráficos em várias áreas da *Pubmed*; os autores mais envolvidos em uma pesquisa, jornais que mais publicam sobre determinada doença, entre outras análises. No entanto, a interface do site não é simples e intuitiva para usuário leigos, como pode ser visto na Figura 2.

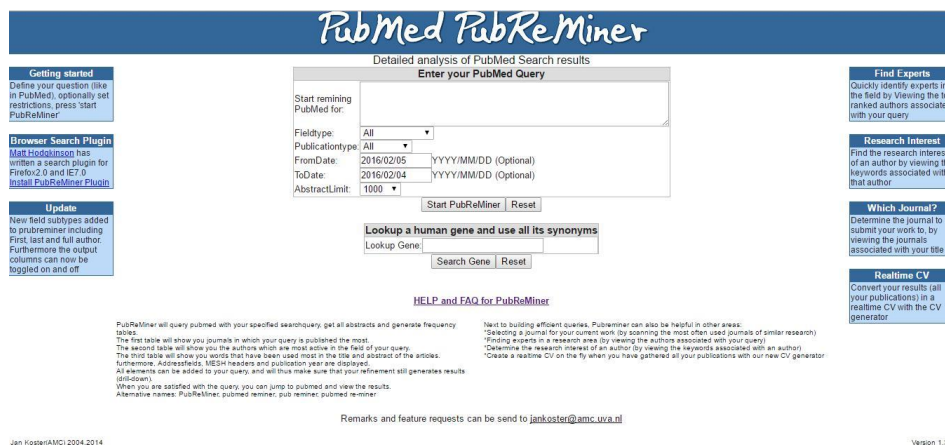


Figura 2: Pubmed PubReMiner.

### 3.2 Materiais e métodos

Para realizar a extração dos dados da internet utilizados neste trabalho de forma automática utiliza-se a linguagem de programação Python. Segundo Luiz Eduardo Borges[5], *Python* é uma linguagem de alto nível, de

tipagem forte, dinâmica e interpretada sendo esta a forma mais simples de extrair dados da internet, incluindo redes sociais [6]. Já a coleta automática de conteúdo da web é parte de um processo maior conhecido como extração de dados. Para Marcelino Pereira dos Santos Silva[7], a mineração de dados, é a análise estatística de um DCBD (descoberta de conhecimento em bancos de dados) denominada como uma análise automática ou semi-automática de grandes quantidades de dados em um banco de dados). Ainda para este autor, três áreas estão envolvidas na mineração de dados, a inteligência artificial (aprendizagem da máquina), banco de dados (recursos para manipular uma base de dados) e estatística (avaliação e validação de resultados) conforme a Figura 3.

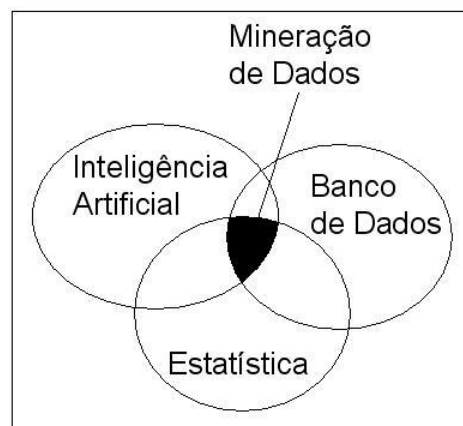


Figura 3 : Três áreas, que segundo Marcelino Pereira, caracterizam a mineração de dados.

Assim, em uma análise mais complexa sobre a mineração de dados, surge o conceito de mineração de texto. Esta é a forma com que o portal *Medical Mining* trabalha com as informações. Mineração de texto, segundo Eliana e Delcio Lago, se define pela extração de informação útil em documentos de textos não-estruturados[8]. Já que mais de 80% das informações que são oferecidas por plataformas na internet estão no formato texto, a mineração por técnica de mineração de textos possui um grande valor comercial [8]. Alguns passos são determinantes para a denominação de mineração de texto e KDD, listados a seguir :

- 1 – Dados: Identificar um armazenamento de dados externo;
- 2 – Seleção de dados: Selecionar os dados dessa fonte por meio da extração efetuada por um algoritmo;
- 3 – Processamento: Eliminar dados incompatíveis com a análise a ser feita, seja por estarem incompletos, repetidos ou danificados;

4 – Transformação : Dados sem impurezas devem ser armazenados para efetuar análise;

5 – Mineração de dados : Métodos para efetuar aprendizado de máquina; identificando padrões, estatísticas, classificação, clusterização e/ou modelos gráficos;

6 – Interpretação/ Avaliação : Os resultados devem ser apresentados de alguma forma que algum usuário leigo possa compreender.

## 4. DESENVOLVIMENTO

### 4.1 A base de dados

A base das informações é o banco de dados, do qual se extrai as informações para a realização das análises estatísticas, a mineração de dados. No caso do *Medical Mining*, a base de dados escolhida foi a *Pubmed* por apresentar mais de 26 milhões de artigos científicos da área da medicina, tornando-a a plataforma mais completa em artigos relacionados à área da medicina [4].

A base de dados é formada por dados brutos a serem analisados, compostos por artigos científicos. Esta página da internet é um buscador da base de dados da *Medline*, que consiste um repositório de artigos da literatura internacional da área médica e biomédica. Atualmente, reúne artigos de 4.800 revistas médicas e de 70 revistas ao redor do mundo[9]. Em vista disso, apresenta uma grande quantidade de artigos em que pontos chave possam ser analisados. A Figura 4 mostra um exemplo de busca pelo termo *zika* realizada no dia 31/05/2016. Como resultado, apareceram a quantidade de artigos, 851, na parte superior, e o número de artigos mostrados por página, de 1 a 20, além dos links para os resumos dos artigos, sublinhados em azul, como o link “*Zika Virus : What Clinicians need to know*”.

**Search results**  
Items: 1 to 20 of 851

<< First < Prev Page 1 of 43 Next > Last >>

[Zika Virus: What Clinicians Need to Know](#)  
1. Dennehy PH.  
R I Med J (2013). 2016 Feb;99(2):13-4. No abstract available.  
PMID: 27236867  
[Similar articles](#)

[\[Relevance of laboratory diagnosis of Zika during pregnancy and its possible association with Guillain-Barré syndrome\]](#)  
2. Alvarado-Socarras JL, Sepúlveda-Arias JC, Zambrano LI, Rodriguez-Morales AJ.  
Med Clin (Barc). 2016 May 25. pii: S0025-7753(16)30090-2. doi: 10.1016/j.medcli.2016.03.044. [Epub ahead of print]  
Spanish. No abstract available.  
PMID: 27236614  
[Similar articles](#)

[Zika Virus Infection: Current Concerns and Perspectives](#)  
3. Maharajan MK, Ranjan A, Chu JF, Foo WL, Chai ZX, Lau EY, Ye HM, Theam XJ, Lok YL.  
Clin Rev Allergy Immunol. 2016 May 28. [Epub ahead of print] Review.  
PMID: 27236440  
[Similar articles](#)

Figura 4 : Site da *Pubmed* com o termo pesquisado *zika* dia 31/05/2016.

Na Figura 5 é mostrado o resultado ao clicar em algum link da Figura 4, como o título do artigo, os autores, as informações sobre os autores, o *abstract* (resumo) e as *keywords* (palavras chave).

[Methods Mol Biol.](#) 2016;1426:297-309. doi: 10.1007/978-1-4939-3618-2\_27.

### **Production of Chikungunya Virus-Like Particles and Subunit Vaccines in Insect Cells.**

Metz SW<sup>1,2</sup>, Pijlman GP<sup>3</sup>.

#### **Author information**

<sup>1</sup>Department of Microbiology and Immunology, University of North Carolina, Chapel Hill, NC, USA.

<sup>2</sup>Laboratory of Virology, Wageningen University, Droevendaalsesteeg 1, 6708 NW, Wageningen, The Netherlands.

<sup>3</sup>Laboratory of Virology, Wageningen University, Droevendaalsesteeg 1, 6708 NW, Wageningen, The Netherlands. [gorben.pijlman@wur.nl](mailto:gorben.pijlman@wur.nl).

#### **Abstract**

Chikungunya virus is a reemerging human pathogen that causes debilitating arthritic disease in humans. Like dengue and Zika virus, CHIKV is transmitted by *Aedes* mosquitoes in an epidemic urban cycle, and is now rapidly spreading through the Americas since its introduction in the Caribbean in late 2013. There are no licensed vaccines or antiviral drugs available, and only a few vaccine candidates have passed Phase I human clinical trials. Using recombinant baculovirus expression technology, we have generated CHIKV glycoprotein subunit and virus-like particle (VLP) vaccines that are amenable to large scale production in insect cells. These vaccines, in particular the VLPs, have shown high immunogenicity and protection against CHIKV infection in different animal models of CHIKV-induced disease. Here, we describe the production, purification, and characterization of these potent CHIKV vaccine candidates.

**KEYWORDS:** Baculovirus; Chikungunya virus; Insect cells; Production and purification; Secreted E1 and E2; Virus-like particles

PMID: 27233282 [PubMed - in process]



Figura 5 : Resumo de um artigo com Autor, *Abstract* e *Keywords*.

## **4.2 Dados da área da medicina e funcionalidades**

Os dados da área da medicina consistem em dados relacionados a tópicos de pesquisa em medicina e biomedicina e são relacionados entre si, por exemplo, os medicamentos e vacinas, afim de apresentar dados estatísticos. Portanto, os dados devem ser relacionados com a base de dados da *Pubmed*[4] descrita no tópico 4.1. Por enquanto, podem ser analisados cinco itens (medicamentos, vacinas, outras enfermidades, tendência de pesquisa, localidades mais afetadas) mas a quantidade de combinações é diversa, sendo apenas necessário saber os detalhes e termos chaves a serem pesquisados. Por exemplo, caso o usuário deseje descobrir qual parte do corpo humano é mais afetada pela “malária”, o site fará essa combinação, mostrando como resultado o órgão fígado.

### **4.2.1 Medicamentos**

No local de pesquisar medicamentos no site, o usuário pode descobrir se foi criado algum medicamento para determinada doença, se algum outro medicamento mais antigo esteja sendo pouco citado em artigos médicos, seguindo a base de dados do site *ATC Code* [10]. Exemplos de medicamentos são: *OlaFlur* (prevenção de cárie dental), e *Sodium Floride* (tratamento de osteoporose).

#### 4.2.2 Vacinas

Essa opção possibilita ao usuário pesquisar e descobrir, por exemplo, se foi criada alguma vacina para determinada doença ou se esta esteja sendo pouco citada em artigos médicos, segundo a base de dados do site CDC[11] . Alguns exemplos de vacinas são: BCG (*Bacilo Calmette-Guérin*), que é aplicada em recém nascidos para a prevenção da tuberculose[12], e a vacina contra o HPV (aplicada em meninas e recentemente também em meninos a partir dos 10 anos de idade para a prevenção do vírus HPV)[13][14].

#### 4.2.3 Outras enfermidades

Essa opção disponibiliza ao usuário descobrir se alguma doença pode estar relacionada com a pesquisa, ou possa estar relacionada também á novos sintomas, seguindo a base de dados do site ICD (Centro Internacional de Doenças)[15]. Alguns exemplos de outras enfermidades relacionadas são : pneumonia e tuberculose.

#### 4.2.4 Tendência de Pesquisa

Com essa opção, o usuário pode descobrir se a doença está em tendência de pesquisa pelos cientistas. A quantidade de artigos em cada ano e, portanto, o ano que determinada doença foi mais citada.

#### 4.2.5 Localidades Mais Afetadas

Essa opção disponibiliza o usuário descobrir em que país uma determinada doença está sendo mais citada. Logo, é possível revelar em quais países uma doença está agindo com maior intensidade, de forma que seja possível revelar possíveis epidemias.

#### 4.3 Mineração e coleta de dados

A mineração de dados está relacionado a coleta de dados dos sites descritos no seção 3, como *Pubmed* e *ATC Code*, para a montagem das bases e preenchimento de todas tabelas do banco de dados.

De uma maneira simplificada, Figura 6, o primeiro passo é utilizar um “coletor de dados” para extrair informações da web e preencher o banco de dados.



Figura 6: Simples esquema do início do processo de mineração de dados.



Em um momento inicial, as informações estão simplesmente expostas para a extração em uma entidade chamada “web”. Elas estão acessíveis em linhas de códigos HTMLs, a estrutura fundamental de uma página web. Na Figura 7, é descrito como opera a coleta de dados de uma forma mais detalhada. O site, representado em 1, é o layout tradicional que um usuário enxerga. No item 2 é como o código pra gerar esse site está disposto, em HTML. No item 3 é como o nosso programa, que realmente é um bot (aplicação de software que simula ações humanas repetidas vezes de maneira padrão) extrai os dados desse HTML de forma automática [16]. No item 4, o programa faz a filtragem dos dados, para que não haja sujeira e, por fim, no item 5 armazena no banco de dados. Esses itens se repetem em um processo recursivo, a menos que se programe um comando para encerrar a ação ou a conexão caia: seja por que o servidor bloqueou o acesso ou a conexão na máquina que esteja executando o código seja interrompido.

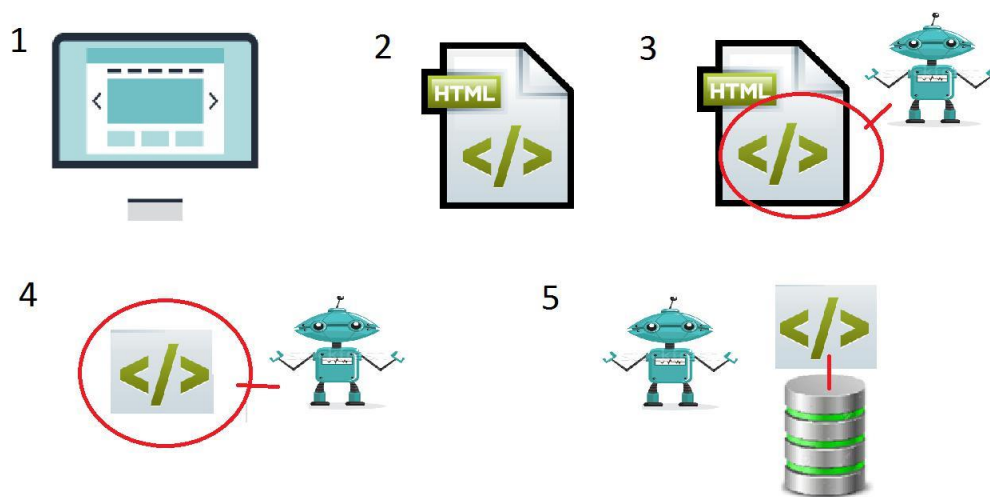


Figura 7: Funcionamento da extração de dados do site pelo programa.

Em uma visão ainda mais detalhada, é uma mostra das ferramentas para executar a extração dos dados da internet. São utilizados a linguagem *Python* [4] e a *IDE PyCharm* [17], juntamente com algumas bibliotecas que possibilitam o que o *Python* naturalmente não faz. As bibliotecas são o *BeautifulSoup* [18], para a extração das tags de HTML de um site e *Selenium* [19], para executar tarefas ainda mais complexas, como apertar um botão de cadastro, rolar a barra de pesquisa, atualizar a página, entre outras funções.

Essas bibliotecas por si só já aceleram o processo de extração, todavia o carregamento dos layouts, como o item 1 da Figura 7, lentificam a extração. Para contornar esse problema é utilizado um navegador sem interface, sem carregamentos de CSS ou JavaScript, apenas com PHP e HTML básicos. Seu nome é *PhantomJS* [20] e é um dos maiores pilares para a otimização da parte de mineração.



O processo de analisar as linhas do código se dá com a manipulação de tags (palavra-chave ou termo associado com uma informação que o descreve e permite uma classificação da informação) do HTML. Por exemplo, extrair o conteúdo que está em uma tag como “href” ou extrair atributos na própria tag, como uma classe de um “a”.

Uma forma de aperfeiçoar esse método é utilizar a filtragem das informações. Palavras com acento, com codificações diferentes, por exemplo, dão um problema quando se realiza as comparações para estabelecer dados estatísticos. Então, é ideal que esse processo seja feito antes do preenchimento no Banco de Dados.

#### 4.4 Armazenamento

O cadastro no Banco de Dados se constitui das bases extraídas no seção 4.2, além dos artigos da base de dados do tópico 2 e das relações que são estabelecidas entre elas. Desse modo, as relações são executadas sem necessidade de baixar nenhum arquivo na hora da consulta, já que elas já estão previamente armazenadas no servidor. Assim, a relação se dará puramente pelo Banco de Dados localmente. Segue o conteúdo do modelo de banco de dados feito no *MySQL Workbench* na Figura 8.

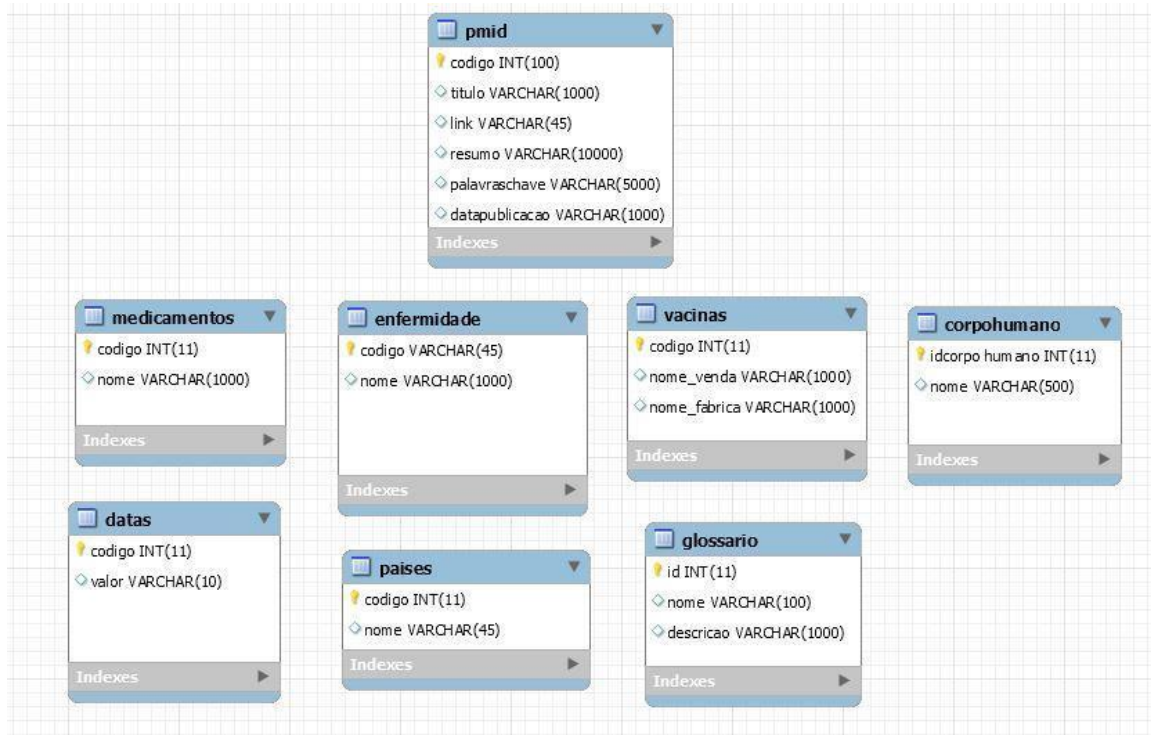


Figura 8: Modelo do Banco de Dados do site *Medical Mining*.

A partir da Figura 8, pode-se inferir a seguinte relação: São 8 tabelas sem chaves estrangeiras, com os títulos “pmid”, “medicamentos”, “enfermidade”, “vacinas”, “corpohumano”, “datas”, “paises”, “glossario”. A tabela “pmid” se refere aos artigos da Pubmed, com os atributos código, titulo, link, resumo, datapublicacao e palavraschave. A tabela Medicamentos com os atributos código e nome. A datas, por sua vez, com código e valor. A enfermidade com código e nome. A de países com código e nome. A

de vacinas com código, nome\_venda e nome\_fabrica. Glossario com id, nome e descrição.

A comparação entre tabelas se dá sempre entre 2, com a “pmid” fixa. Por conseguinte, quando uma consulta é realizada no site, por exemplo, escolhendo as “localidades mais afetadas” e o termo “bruxismo” (tópico 4.5) as tabelas selecionadas para a comparação são “países” e “pmid” utilizando o comando sql *like* [21].

Nas Figuras 9 e Figura 10 são exemplos de como está o layout das tabelas do banco de dados. A Figura 10 referente à tabela “pmid”, aos artigos e a Figura 11 à tabela “enfermidades”,

codigo	titulo	link	resumo	palavraschave	datapublicacao
26921283	Parents Who Supply ...	http://www.ncbi.nlm....	Parents who perceiv...	Inexistente nesse artigo	Pediatrics. 2016 Mar;137(3):e2015261
26921284	Social Transition: Sup...	http://www.ncbi.nlm....	Inexistente nesse a...	Inexistente nesse artigo	Pediatrics. 2016 Mar;137(3):e2015435
26921285	Mental Health of Tran...	http://www.ncbi.nlm....	Socially transitioned ...	Inexistente nesse artigo	Pediatrics. 2016 Mar;137(3):e2015322
26921286	Parental Choice of Re...	http://www.ncbi.nlm....	Preference-based r...	Inexistente nesse artigo	Pediatrics. 2016 Mar;137(3):e2015285
26921287	Fibrin deposition follo...	http://www.ncbi.nlm....	Coagulation cascad...	Inexistente nesse artigo	Blood. 2016 Jun 2;127(22):2751-62. d
26921288	The Changing Landsc...	http://www.ncbi.nlm....	Smoldering multiple ...	Progression; Risk factors; Sm...	Oncologist. 2016 Mar;21(3):333-42. dt

Figura 9: Tabela dos artigos no *MySQL Workbench*.

codigo	nome
a00-a09.9	intestinal infectious diseases
a00-b99.9	certain infectious and parasitic diseases
a00.0	cholera due to vibrio cholerae 01, biovar cholerae
a00.1	cholera due to vibrio cholerae 01, biovar eltor
a00.9	cholera, unspecified
a01	typhoid and paratyphoid fevers
a01.0	typhoid fever

Figura 10: Tabela de Enfermidades no *MySQL Workbench*.

#### 4.5 Comparações

Para estabelecer as comparações entre tabelas é selecionada determinada coluna de uma tabela para se comparar com outra de uma diferente. Contudo, as colunas selecionadas nem sempre são iguais. Por exemplo, uma comparação entre a tabela “pmid” e “corpo humano” com o termo sarampo. Primeiro, seleciona-se todos os artigos da tabela pmid citando sarampo. Com esses artigos apenas dessa doença, procura-se no título a partir do método *like* palavras que sejam iguais à base de corpo humano. Por exemplo, numa exposição hipotética, um artigo com título “*Measles causes head issues in newborn babes*”. As palavras *Measles* (rubéola) e *head* (cabeça) estão relacionadas, então a

palavra *head* ganha no contador mais 1. Ao final, as várias partes do corpo humano ganham uma tabela com seu respectivo nome e quantidade; um gráfico com nome e quantidade e tabela com título e link. Seguindo nessa situação hipotética, imagina-se que de 2 mil artigos de sarampo, 500 citem cabeça, 254 fígado e 320 pulmão.

Na Figura 11, está representado esse modelo de extração. No item 1 é a seleção de artigos da base, apenas sobre sarampo. No item 2 é a comparação entre tabelas, apenas que os artigos estão representados como textos e o tabela corpo humano como esqueleto. No item 3 é mostrado os itens selecionados e suas respectivas quantidades, cabeça igual a 1 e braço igual a 1.

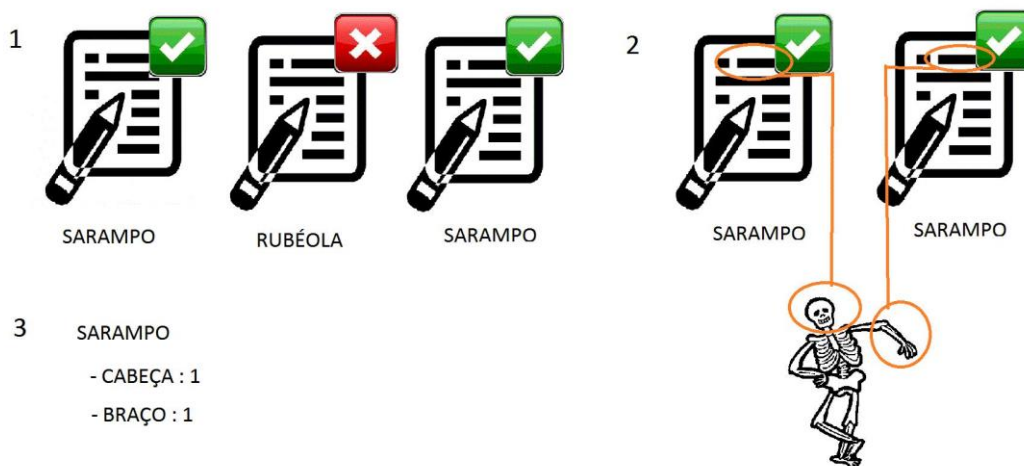


Figura 11: Esquema explicando a seleção e análise de artigos médicos e dados médicos.

Diferentes formas de comparação, recebem diferentes formas de seleção e análise de colunas das tabelas. Na figura 12, é explicitado que sempre duas tabelas são comparadas e essas análises resultam em um gráfico.

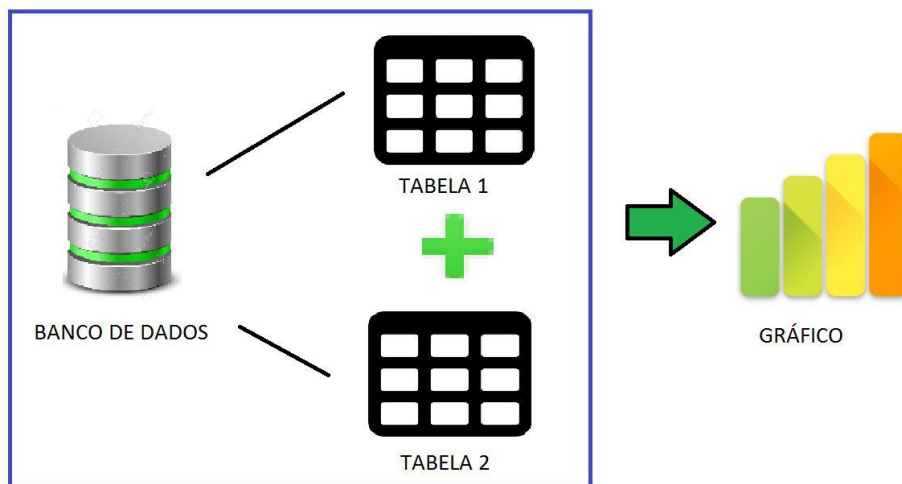


Figura 12: Esquema explicando o funcionamento básico da comparação entre tabelas.

## 4.6 Interface

A interface é composta por uma página principal, com 3 opções de menu no canto superior esquerdo (*Home*, *Sobre*, *Glossário* e *Histórico*). A Figura 13 mostra a *Home Page* do site.

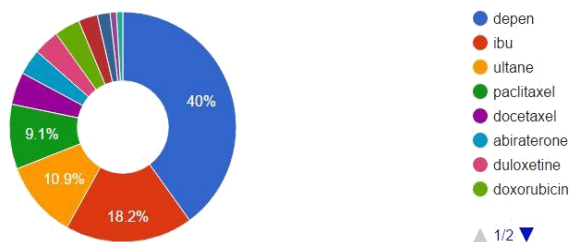


Figura 13: Página Principal do Site.

A página *Home* descreve as funcionalidades do site com um pequeno tutorial de como utilizar suas funções e dois campos centrais para serem preenchidos. Um deles é o nome da enfermidade a ser pesquisada e o outro é o campo no qual será escolhido a forma da análise, entre 5 opções. São elas: outras enfermidades relacionadas, medicamentos, vacinas, data das publicações, parte do corpo humano afetadas e localidades mais citadas. Em seguida, estará o botão de confirmação “Buscar” abaixo, que irá efetuar a análise no banco de dados e mostrar o resultado em um gráfico e duas tabelas. Os gráficos podem ser de vários modelos, organizados na seguinte ordem : Gráfico de barras horizontais, de setores, de linhas, de barras, de barras verticais e mapa-múndi, respectivamente as 5 opções do site. Já a primeira tabela têm as mesmas informações que o gráfico, apenas é apresentado de forma diferente em duas colunas, geralmente, com o nome e quantidade de citações nos artigos com valores decrescentes. A outra tabela apresenta o título dos artigos envolvidos na pesquisa e os links para lê-los. A Figura 14 apresenta a interface dessa área do site.



cancer



### Resumo

Nome	Quantidade de citações
depen	44

Figura 14: Resultado em forma de gráfico da consulta de medicamentos.

A área Sobre apresenta informações dos desenvolvedores do site, incluindo os respectivos contatos pessoais. Além de um pequeno recado pessoal, objetivos de estar criando essa plataforma e agradecimentos aos que contribuíram com o projeto, como pode ser visto na Figura 15.

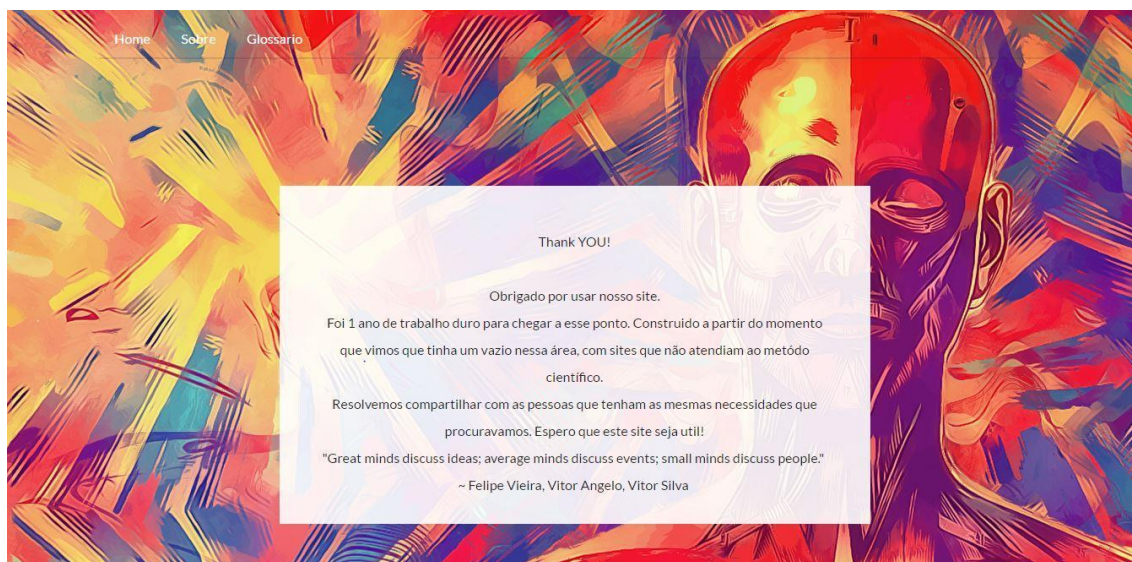


Figura 15: Layout da interface da área Sobre do site.

A área Glossário apresenta uma pequena enciclopédia de variadas doenças. Portanto, contém informações desde do resumo da doença, até alguns dos sintomas e profilaxia, sendo uma área a parte do portal, que pode ser consultada independentemente das

outras áreas. A Figura 16 demonstra o modo que o site exibe as informações do glossário.



Figura 16: Área referente ao Glossário com o menu lateral de a à z.

As informações são exibidas com o banco de dados já preenchido, utilizando a técnica de mineração de dados. No entanto, algumas doenças recebem informações adicionais de forma manual, inseridas manualmente. Todas as inclusões e forma base são incluídas as devidas fontes, atendendo a um método científico, visando ser o menos falacioso possível. A ideia é que os dados sejam atualizados com o tempo, a medida que os integrantes do grupo estudem mais sobre o assunto, aprendam mais sobre medicina e, conseqüentemente, aumentem o acervo.

#### 4.7 TRABALHOS FUTUROS

Como a atual estrutura de banco de dados é constituída de tabelas que não são ligadas por chaves estrangeiras, as consultas se tornam mais demoradas para serem concluídas pelo grande volume de dados nas tabelas, entre milhares a milhões de itens. Uma forma de superar tal problema é remodelando o banco de dados, de tal forma que acelere o processo das análises, conforme é observado na Figura 17.

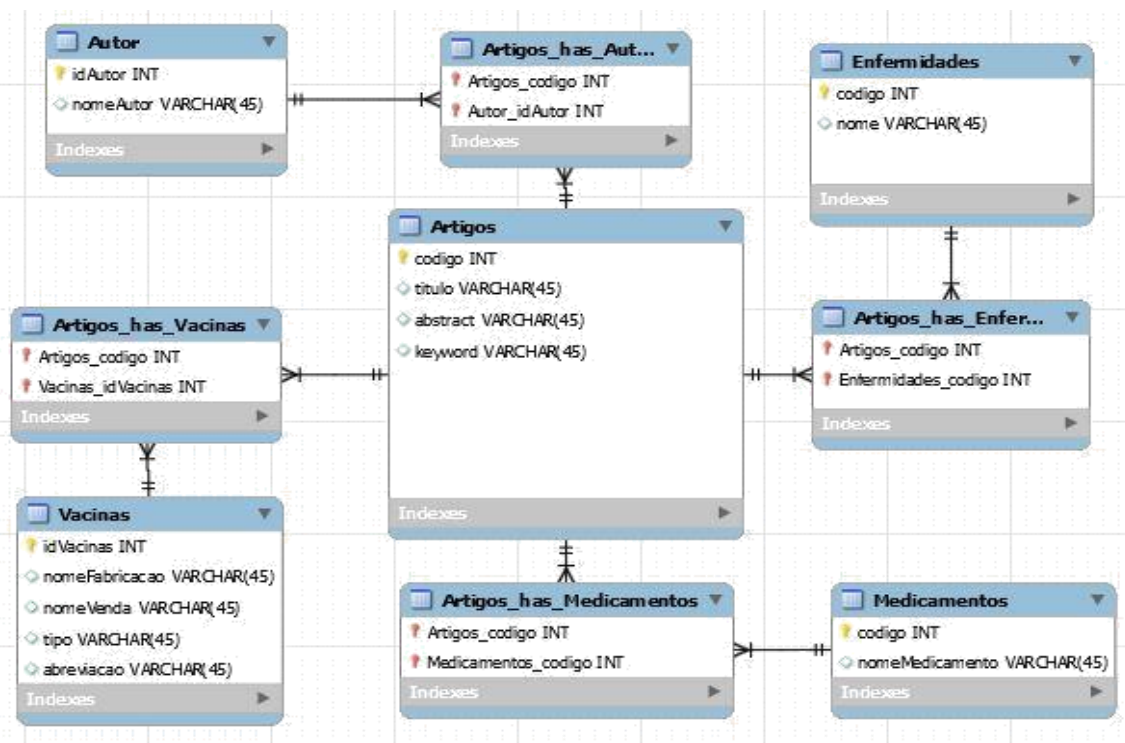


Figura 17: Futuro modelo do banco de dados do site *Medical Mining*.

## 5. RESULTADOS

O portal além de oferecer um glossário de doenças de A á Z com suas definições, possibilita também ao usuário fazer pesquisas sobre doenças e fazer comparações com medicamentos, vacinas, partes do corpo humano afetadas pela doença pesquisada, os países que mais aparecem em artigos científicos da medicina sobre determinada doença e outras enfermidades relacionadas a ela, apresentando gráfico afim de facilitar a visualização e entendimento de resultados encontrados por leigos. Abaixo pode-se observar exemplo dessa comparação, feita após a pesquisa do tema “Pneumonia”:

**Enfermidades :** Asma com 26 citações. Tuberculose com 24. Pleura com 16. Disfagia com 16.

Fibrose cística com 12. Miosite com 12 e enfisema com 11 e outros resultados.

**Tendência de Pesquisa:** 2012 com 1 artigo. 2014 com 21 artigos. 2015 com 2688 artigos e 2016 com 22276.

**Corpo humano:** Secreção com 407 artigos, ouvido com 268, costela com 63 e braço com 51.

**Localidades Mais afetadas:** Mali com 54 artigos, China com 52, Japão com 41, Irã com 40 e Oman com 36 .



## 6. REFERÊNCIAS

- [1] Zika vírus no Brasil. Disponível em: <<http://exame.abril.com.br/ferramentas/zika-virus-e-microcefalia-a-evolucao-do-surto-no-brasil/>>. Acesso em 10 de novembro de 2016.
- [2] Read by QxMD. Disponível em: <<http://www.qxmd.com/apps/read-by-qxmd-app>>. Acesso em 29 de maio de 2016.
- [3] READ BY QxMD. Disponível em: <<https://play.google.com/store/apps/details?id=com.qxmd.readbyqxmd&hl=pt>>. Acesso em 30 de outubro de 2016.
- [4] PUBMED. Disponível em: <<https://pt.wikipedia.org/wiki/PubMed>>. Acesso em 30 de maio de 2016.
- [5] Luiz Eduardo Borges: Phyton. Disponível em: <[https://ark4n.files.wordpress.com/2010/01/python\\_para\\_desenvolvedores](https://ark4n.files.wordpress.com/2010/01/python_para_desenvolvedores)>. Acesso em 30 de outubro de 2016.
- [6] PYTHON. Disponível em: <<https://www.python.org/>>. Acesso em 30 de outubro de 2016.
- [7] Marcelino Pereira dos Santos Silva: Phyton. Disponível em: <[docplayer.com.br/8954638-Mineracao-de-padres-de-mudanca](http://docplayer.com.br/8954638-Mineracao-de-padres-de-mudanca)>. Acesso em 30 de outubro de 2016.
- [8] Mineração de dados segundo Eliana e Delcio Lago. Disponível em: <<https://www.cremesp.org.br/?siteAcao=BibliotecaCRM>>. Acesso em 30 de outubro de 2016.
- [9] MEDLINE. Disponível em: <<https://www.medline.com/>>. Acesso em 30 de outubro de 2016.
- [10] ATC Code. Disponível em: <[www.atccode.com](http://www.atccode.com)>. Acesso em 03 de junho de 2016.
- [11] CDC. Disponível em: <<http://www.cdc.gov/az/a.html>>. Acesso em 05 de junho de 2016.

- [12] BCG. Disponível em: < <http://www.minhavidade.com.br/saude/tudo-sobre/16655-vacina-bcg>>. Acesso em 05 de junho de 2016.
- [13] HPV: causas, sintomas e tratamento. Disponível em: <<http://www.minhavidade.com.br/saude/temas/hpv>>. Acesso em 05 de junho de 2016.
- [14] VacinaHPVemmeninos.Disponivelem:<<http://g1.globo.com/bemestar/noticia/2016/10/ministerio-da-saude-anuncia-vacinacao-contrahpv-para-meninos>>. Acesso em 07 de novembro de 2016.
- [15] ICD(CentroInternacionaldeDoenças).Disponivelem:<<http://en.centralx.com/icd-10/p/a.html> >. Acesso em 06 de maio de 2016.
- [16] O que é Bot. Disponível em: <<http://olhardigital.uol.com.br/noticia/entenda-de-uma-vez-por-todas-o-que-e-um-bot-e-como-ele-funciona/>>. Acesso em 10 de novembro de 2016.
- [17] IDE Pycharm. Disponível em: <<https://www.jetbrains.com/pycharm>>. Acesso em 10 de novembro de 2016.
- [18] Como funciona as bibliotecas *Beautysoup*. Disponível em: <<http://wiki.python.org.br/BeautifulSoup>>. Acesso em 10 de novembro de 2016.
- [19] Selenium: o que você deveria saber. Disponível em: <[pythonclub.com.br/selenium-parte-1.html](http://pythonclub.com.br/selenium-parte-1.html)>. Acesso em 06 de novembro de 2016.
- [20] O que é PhantomJS. Disponível em: <<http://imasters.com.br/framework/conhecendo-o-phantomjs/>>. Acesso em 06 de novembro de 2016.
- [21] Definição de SQL like. Disponível em: <[www.1keydata.com](http://www.1keydata.com) > SQL > Comandos SQL>. Acesso em 10 de novembro de 2016.

## 7. APÊNDICE

A seguir, na Figura 18, o diagrama caso de uso que demonstra as ações do usuário dentro do sistema:

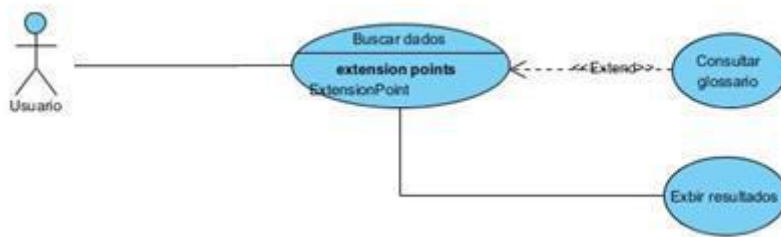


Figura 18: Diagrama casos de uso da plataforma *Medical Mining*.

## Buscar dados

Ator principal : Usuário

-

Resumo : Esse caso de uso permite o usuário a buscar dados com termo a ser pesquisado e área a ser analisada

Pré – condição : Não há

Pós – condição : O usuário terá acesso aos resultados da pesquisa com gráficos e tabelas Fluxo principal

- 1 – Informar termo
- 2 – Informar área
- 3 – Apertar buscar
- 4 – Analisar dados
- 5 – Consultar banco de dados e fazer relação

Fluxo secundário

Consultar glossário

- 1 - Consultar doença desejada

2 – Buscar no BD e mostrar resultado

Exibir resultado

1 – Buscar dados e mostrar resultados

2 – Exibir gráficos e tabelas