

MedicMiner: Plataforma para Extração, Análise e Visualização de Dados Médicos

Davi Leon Costa de Paula Santos¹, Júlia Mori Gomide¹, Lucas Gabriel Silva¹, Thiago Magela Rodrigues Dias¹, André Flávio Clarimundo Rabelo¹

¹Curso Técnico em Informática – Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) – Campus V

Rua Álvares de Azevedo, 400, Divinópolis-MG, 35503-822 – Brazil

davilcostaa@gmail.com, juliamorigomide@gmail.com, lucasgsilva2102@gmail.com, thiagomagela@gmail.com, andresonax@gmail.com

Abstract. *With the growth of the web and the accentuated use of technologies in order to get specific informations, the amount of data has grown exponentially. Thus, search and access to specific content becomes complex, due to the large quantity of irrelevant information and the way it is distributed in different repositories and formats. From this perspective, Data Mining emerges, which through algorithms discovers patterns in large databases. Therefore, the project developed is a web system of medical research with uses data mining techniques. The MedicMiner system aims to plot the graphical visualization of search results and the relationship of medical articles content extracted from the PubMed database.*

Resumo. *Com o crescimento da web e o acentuado uso de tecnologias da informação, o número de dados gerados tem crescido exponencialmente. Assim, a pesquisa e acesso a conteúdos específicos se tornam complexos, devido à grande quantidade de informações irrelevantes e ao modo pelo qual elas estão distribuídas, em diferentes formatos e repositórios. Nessa perspectiva, surge a Mineração de Dados, que, por meio de algoritmos, descobre padrões e regras em grandes bases de dados. Diante disso, o projeto desenvolvido é um sistema web de pesquisas médicas que utiliza técnicas de Mineração de Dados. O sistema MedicMiner tem como objetivo fazer a visualização gráfica do resultado das buscas e relacionamento do conteúdo de artigos médicos extraídos da base da PubMed.*

1. Introdução

A quantidade de dados gerada a cada dia, principalmente, impulsionada por novos serviços e portais disponíveis na *web*, tem contribuído de forma significativa para a difusão do conhecimento. No entanto, devido à distribuição desses dados que, geralmente, estão em formatos distintos, ou em diferentes repositórios, faz com que a análise e obtenção de conhecimento útil, se torne cada vez mais difícil.

Diante disso, novas ferramentas e técnicas se fazem necessárias para que todo o potencial que esses dados possam fornecer seja mais bem explorado.

Dentro desse contexto, encontra-se a Mineração de Dados - *Data Mining*, que é um dos processos KDD (*Knowledge Discovery in Databases*). Esses processos são, basicamente, procedimentos pelos quais se converte uma grande quantidade de dados brutos em

informações úteis (FAYYAD et al., 1996, p. 27-34). A etapa de Mineração de Dados é responsável, principalmente, pela descoberta de padrões, regras e relações que irão propiciar o prognóstico e a correlação entre os dados. Sendo assim, irá auxiliar em tomada de decisões mais precisas e, conseqüentemente, aumentará o nível de confiabilidade das mesmas.

Uma das áreas que mais se beneficiam com a exploração da Mineração de Dados é a área da saúde (SHAMA e MANSOTRA, 2014). Esse campo é um dos quais possui melhor aproveitamento dos dados além de gerar mais efeitos, como poder facilitar no tratamento e controle de doenças.

Logo, o presente trabalho busca extrair dados do portalPubMed¹ que, atualmente, indexa cerca de 28 milhões de artigos científicos da área médica. Após a extração, busca-se fazer o tratamento e análise desses dados, a fim de auxiliar médicos e pesquisadores da área na correlação entre doenças, vacinas, medicamentos, dentre outros.

1.1. Objetivos

Este trabalho tem como objetivo geral propor uma plataforma para a extração, análise e visualização de dados sobre registros de pesquisas da área médica. Para tanto, os seguintes objetivos específicos serão contemplados:

1. *Analisar e estudar o repositório de dados a ser utilizado;*
2. *Realizar a extração dos dados desse repositório (PubMed);*
3. *Fazer o tratamento dos dados e deixá-los no formato desejado para análise;*
4. *Fazer a análise dos dados;*
5. *Fornecer interface gráfica para a visualização dos dados.*

2. Referencial Teórico

Com o passar dos anos, a quantidade de dados gerados na internet tem se tornado cada vez maior. Segundo o Gartner (2017), cerca de 2,2 milhões de terabytes de dados novos são criados todos os dias no mundo. Além disso, estima-se que até 2020 haja um total de 40 trilhões de gigabytes de dados no mundo e, mantendo-se os níveis projetados, nesse mesmo ano serão gerados cerca de 1,7 megabytes de novas informações por segundo para cada habitante do planeta Terra.

Os bancos de dados cumprem sua função de armazenar dados e deixá-los acessíveis, porém, mesmo com sua eficiência, falta a eles capacidade de extrair facilmente conhecimento útil das informações armazenadas. Portanto, faz-se necessário desenvolver sistemas de extração com aplicações especializadas, de acordo com Silberschatz, Korth e Sudarshan (2012). Nesse sentido, surge em 1989, o KDD (*Knowledge Discovery in Databases* ou Descoberta de Conhecimento nas Bases de Dados) que segundo Fayyad (1996) é uma referência ao processo amplo e não trivial de encontrar conhecimento em dados. A descoberta de informação é realizada basicamente em cinco etapas, são elas: Seleção de dados, Pré-Processamento, Formatação, Mineração de dados e Interpretação/avaliação dos dados. Conforme ilustrado na Figura 1.

¹ www.ncbi.nlm.nih.gov/pubmed

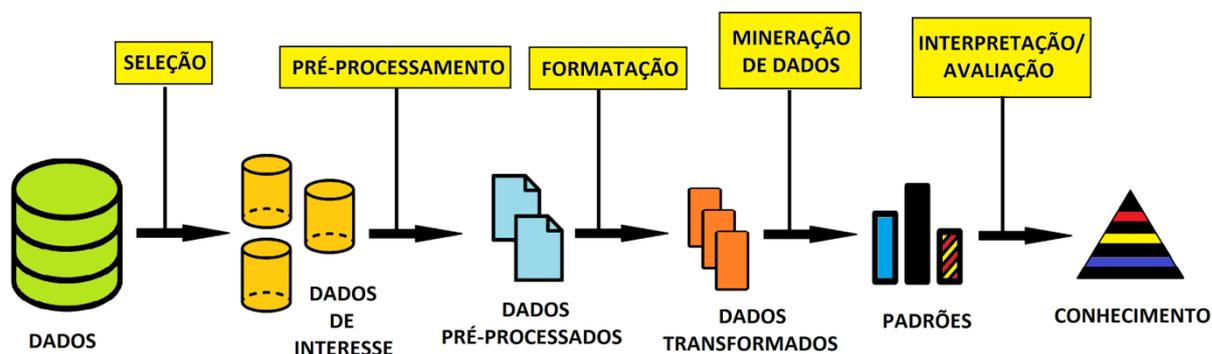


FIGURA 1. Ilustração dos processos de Descoberta de Conhecimento nas Bases de Dados - KDD.

O processo é iniciado com a seleção dos dados de acordo com os objetivos a serem atingidos a partir da exploração dos mesmos. Em seguida, é feito o pré-processamento, que é responsável pelo tratamento e análise de conteúdos inconsistentes. Na etapa de formatação é feita uma preparação desses de acordo com o formato desejado para realizar a mineração. Posteriormente, é feita a mineração dos dados, que realiza a busca de padrões válidos, úteis e compreensíveis. Por fim, é feita a interpretação do resultado obtido, analisando se é o que era objetivado com a realização de todo o processo (NETO, 2010).

Incluso nos processos KDD, destaca-se o método *Data Mining* (Mineração de Dados) que, através de algoritmos, métricas e estatísticas (como regressão, termo e cálculos que são herdados da estatística tradicional) efetiva a descoberta de informação útil para sistemas especialistas e visualização de dados (TAN et al., 2009). Como apontado por Guilherme Ferraz de Arruda (2013), essa técnica pode ser explorada nos mais diversos contextos e apresenta interseção com diversas outras áreas do conhecimento.

Dentre os campos que possuem um histórico de utilização da Mineração de Dados, um que se destaca é a área da saúde. Assim como dito por Shama e Mansotra (2014), essa é uma das áreas na qual um melhor aproveitamento dos dados gera mais impacto, podendo, por exemplo, incrementar o auxílio no controle de epidemias, facilitando diagnósticos e o tratamento de enfermidades.

Inserido nesse contexto, o presente projeto visa fazer a coleta de buscas de artigos científicos. As publicações científicas buscam divulgar a pesquisa para a comunidade, de forma que permita que outros possam utilizá-la e avaliá-la sob outras visões. As revistas, eletrônicas ou impressas, ainda são consideradas como o modo mais rápido e economicamente viável, para os pesquisadores fazerem circular e tornar visíveis os resultados do seu trabalho. Pois, é por meio de uma publicação científica que a sociedade toma conhecimento dos resultados de um trabalho de pesquisa e o que este representa para a coletividade, de acordo com o professor Paulo Roberto Brofman (2012).

Nesse sentido, além da extração, a plataforma MedicMiner fará a análise e visualização gráfica dos resultados das buscas e o relacionamento do conteúdo dos artigos médicos. Eles serão extraídos do repositório da PubMed, uma plataforma mundialmente conhecida que indexa atualmente cerca de 28 milhões de produções médicas e científicas (PUBMED, 2018), mostrando ser uma boa base para a exploração desses dados.

2.1. Trabalhos correlatos

Com o intuito de contextualizar outros trabalhos com a mesma temática, são apresentados exemplos de portais que realizam buscas de dados médicos, são eles: MedlinePlus², Medical Mining³ e SciELO⁴.

2.1.1. MedlinePlus

Medline é a sigla de *Medical Literature Analysis and Retrieval System Online* que, em português, significa Sistema Online de Busca e Análise de Literatura Médica. É uma plataforma sustentada pela Biblioteca Nacional de Medicina dos EUA, a maior biblioteca médica do mundo. Possui ampla informação dos Institutos Nacionais de Saúde e outras fontes confiáveis, além de ser atualizada diariamente.

Segundo informações do próprio *website* (MEDLINE PLUS, 2018), o mesmo oferece conteúdo sobre saúde de forma gratuita. Contém dados sobre medicamentos, tratamentos, significado das palavras, vídeos e ilustrações médicas. Também disponibiliza links para as pesquisas médicas mais recentes sobre determinado assunto e ensaios clínicos sobre uma doença ou condição. Os profissionais de saúde e os consumidores podem depender disso para obter informações atualizadas.

Porém, apesar de relevante, a MedlinePlus não apresenta a facilidade de se poder realizar consultas de múltiplos termos para facilitar pesquisas mais profundas dentro da área.

2.1.2. Medical Mining

O *Medical Mining* é um projeto desenvolvido em 2016 como Trabalho de Conclusão de Curso (TCC) do CEFET-MG (Centro Federal de Educação Tecnológica de Minas Gerais) Campus V em Divinópolis. O portal disponibiliza conteúdos médicos básicos e avançados para os usuários. Com técnicas de mineração de dados, os dados apresentados são analisados e mais completos de acordo com o que foi pesquisado (NUNES, FREITAS e SILVA, 2016).

A plataforma conseguiu atingir os objetivos inicialmente propostos, no entanto, ainda assim é incapaz de entregar resultados mais sofisticados do que tabelas, além de ter gama de pesquisa limitada. Portanto, comparado ao trabalho aqui proposto, é uma ferramenta que por si só é insuficiente.

2.1.3. SciELO

A SciELO (*Scientific Eletronic Library Online*) é uma biblioteca eletrônica de artigos e periódicos científicos brasileiros. O projeto é um resultado da FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo, em parceria com a BIREME - Centro Latino-Americano e com apoio do CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico. Segundo a própria SciELO (2018), o portal tem como objetivo proporcionar um amplo acesso a coleções de periódicos como um todo, aos fascículos de cada título de periódico, assim como aos textos completos dos artigos. A plataforma é constantemente

²www.medlineplus.gov

³www.digddv.cefetmg.br/wp-content/uploads/sites/52/2017/11/Artigo-TCC-Medical-Mining.pdf

⁴www.scielo.br

atualizada de acordo com a publicação de novos conteúdos. É possível acessar periódicos por meio de lista alfabética, lista por assunto ou pela pesquisa de títulos. Os artigos podem ser pesquisados pelo índice de autores, índice de assuntos ou pela pesquisa simples de artigos. Além disso, a SciELO gera relatórios, que podem ser do uso do site, estatísticas de publicação, citação de revistas e co-autoria.

Mesmo bastante utilizada, a SciELO tem como característica exibir resultado de consultas que considera apenas os metadados dos artigos, não permitindo realizar uma pesquisa mais profunda de maneira facilitada.

2.1.4. Análise Comparativa

A Tabela 1 demonstra as funcionalidades principais do MedicMiner e de plataformas similares. Ao se reduzir o escopo de pesquisa para plataformas ligadas aos campos de pesquisa da área da saúde, é possível observar a baixa quantidade de utilidades disponíveis nas ferramentas para uso.

	Pesquisa de termos simples	Pesquisa de múltiplos termos simultaneamente	Geração de relatórios	Geração de gráficos
<i>Medline Plus</i>				
<i>Medical Mining</i>				
<i>SciELO</i>				
<i>MedicMiner</i>				

TABELA 1. Comparação entre o MedicMiner e os trabalhos correlatos.

3. Materiais e métodos

O projeto foi desenvolvido em quatro fases. Na primeira delas foram estudos acerca do tema abordado pelo trabalho. Durante esses estudos, foram tratados assuntos como a mineração de dados, o *Medical Mining* e as maneiras de alterar a ferramenta existente para atender às necessidades da nova plataforma a ser desenvolvida. Para que isso fosse feito, o grupo estudou linguagens de programação como *Python*, PHP e também técnicas de manipulação de banco de dados.

A segunda etapa foi a criação de um site, com interface amigável a usuários pouco familiarizados com a tecnologia, mas que ao mesmo tempo disponha de recursos mais avançados, procurando atender as necessidades do maior número de utilizadores possível. A cor escolhida para ser predominante no site foi a cor verde, que é a cor utilizada em conexão

com a medicina (REZENDE, 2018). Foi feita a página inicial, a de cadastro, login, sobre e pesquisas.

A terceira fase foi o desenvolvimento do banco de dados da plataforma. Primeiramente, foi feita a pesquisa de artigos com a palavra “Brazil” e, em seguida, foram baixados cerca de 300 mil publicações científicas com extensão XML. Devido ao grande tamanho do arquivo, o mesmo foi dividido em partes menores para otimizar a extração dos dados. Além disso, foi utilizado um extrator de dados feito com PHP. No final, foi gerada uma base com aproximadamente 80 mil artigos.

A última fase foi a de geração de gráficos. Com a base de dados pronta, foi feito um *script* em JavaScript que gerencia o envio do formulário para um arquivo PHP que realiza consultas ao banco, recebe e trata os dados. Depois de passar pelos devidos processos, tais informações são então retornadas ao primeiro *script* para a plotagem dos gráficos. Os gráficos receberam uma seção separada, destinada apenas à exibição de amostras estáticas, sem a possibilidade de realizar pesquisas.

3.1. PubMed

A PubMed é uma plataforma de acesso gratuito que atualmente possui cerca de 28 milhões de artigos biomédicos publicados na base de dados da MEDLINE. Mantida pela Biblioteca Nacional de Medicina dos EUA, a plataforma oferece dados sobre medicina e saúde, ciências naturais, química, bioengenharia e ciências do comportamento (PUBMED, 2018).

O Gráfico 1 representa a quantidade de artigos publicados na PubMed nos últimos quinze anos.



GRÁFICO 1. Quantidade de artigos publicados por ano na PubMed.

É possível observar o aumento gradativo de publicações de artigos por ano. Em 2003, houve a publicação de cerca de 600 mil artigos, e, em 2017, foram aproximadamente 1,2 milhões, alcançando o dobro de artigos publicados em 2003.

Considerando somente os últimos 14 anos, foram catalisados cerca de 12.820.148 artigos. Porém, também são encontrados artigos publicados há mais de 200 anos, comprovando ser uma fonte de dados muito representativa.

3.2. PHP

A linguagem de programação PHP (SILVA, 2014) foi utilizada, em linhas gerais, para fazer o tratamento dos dados extraídos. Essa linguagem é de uso geral, além de possuir código aberto, sendo assim, extensamente utilizada nas plataformas *web*. A sua simplicidade inicial, aliada aos seus recursos avançados e extensas interações com banco de dados como a classe PDO (*PHP Data Objects*), permite que a plataforma seja implementada de maneira facilitada. Para as funcionalidades que envolvem usuários, foi feito um CRUD (*Create, Read, Update, Delete*) que fosse capaz de gerenciar os recursos e necessidades que os envolvessem (ARRIGONI, 2018). Assim, o PHP foi usado para tratar os dados da pesquisa, recebendo os termos a serem procurados e realizando uma consulta no banco de dados. Em seguida, foi feito um cruzamento de dados para que então os resultados fossem utilizados na plotagem dos gráficos via JavaScript.

3.3. JavaScript

JavaScript é a linguagem de programação da *Web* (FLANAGAN, 2012). É uma linguagem utilizada em diversos *sites* e navegadores, o que a torna presente na maioria das páginas da Internet. Ainda segundo David Flanagan, JavaScript é uma linguagem de alto nível, dinâmica, interpretada e não tipada, conveniente para estilos de programação orientados a objetos e funcionais. Foram aplicados o formato de troca de dados JSON e os *frameworks* Chart.js e Materialize⁵ na plataforma desenvolvida. O JSON (*JavaScript Object Notation*) é um padrão leve para troca de dados, que é efetivo para trocar dados entre aplicações implementadas com o uso de linguagens modernas de programação (SILVA, 2009). Já a plotagem dos gráficos resultantes das pesquisas realizadas foram feita utilizando Chart.js, que possui recursos para a visualização de gráficos diversos e interativos (TAVARES et al., 2016).

3.4. Python

Por fim, o tratamento dos dados e a criação de grafos para a visualização dos resultados das pesquisas foi feita utilizando a linguagem de programação *Python* (MENEZES, 2014). Essa linguagem é interpretada, orientada a objeto e de alto nível, sendo assim, atrativa para análise irá contemplar uma série de atividades, parametrizadas pelos usuários. Foram empregadas três bibliotecas *Python*: *PyMySQL*, *NetworkX* e *Matplotlib*. A primeira é uma interface *Python* para o *MySQL*, e foi usada para manipular as informações armazenadas no banco de dados. Já as duas restantes, foram utilizadas em conjunto para a plotagem e renderização dos grafos, para que enfim fossem exibidos na plataforma.

⁵www.materializecss.com

4. Descrição das Funcionalidades

Nessa seção, são exibidas as telas da plataforma MedicMiner e suas respectivas funcionalidades, permitindo assim, uma visão ampliada do sistema.

4.1. Página inicial MedicMiner – Parte Superior

A página inicial do site conta com uma interface amigável e instrutiva como mostra a Figura 2. No canto superior esquerdo está a logo do projeto e, na direita, está o menu, com os itens: *Home*, *Pesquisa*, *Histórico*, *Cadastro* e *Login*. Logo abaixo da imagem com o título e subtítulo, há uma descrição da plataforma e, mais abaixo, um exemplo de pesquisas já feitas no *site*.



FIGURA 2. Parte superior da página inicial do MedicMiner.

4.2. Página inicial MedicMiner – Parte Inferior

Na Figura 3 é mostrada a parte inferior da página inicial do MedicMiner. Nela é mostrada dois exemplos de pesquisas gráficas feitas na plataforma.

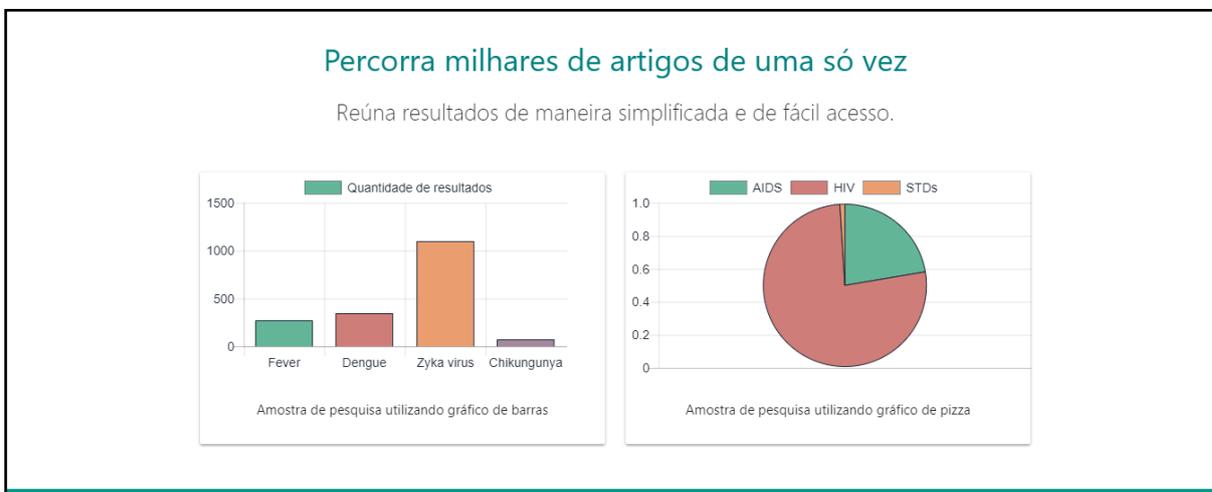
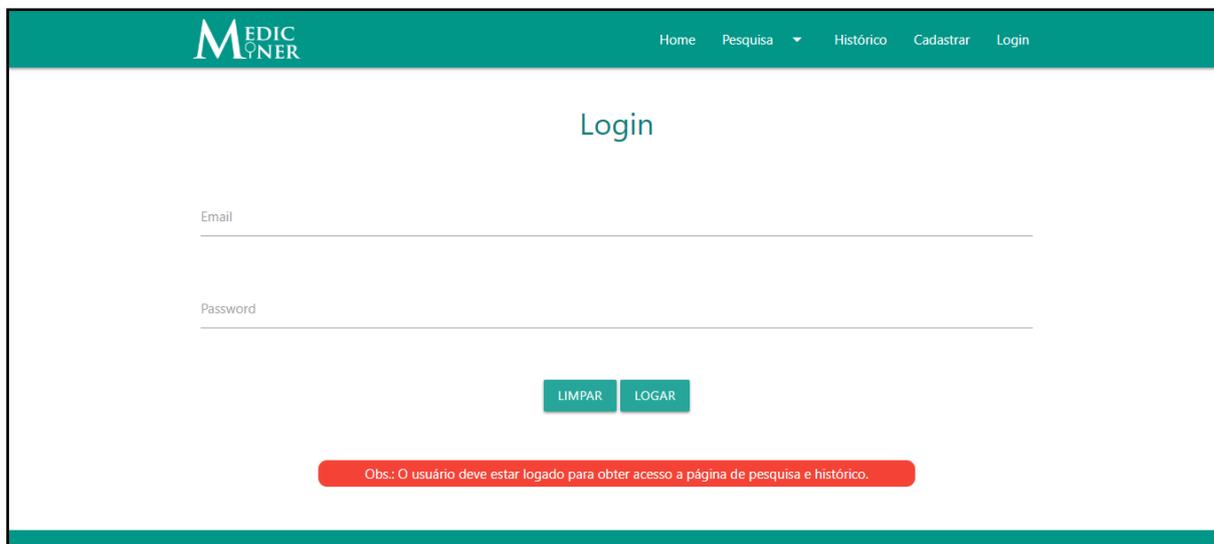


FIGURA 3. Parte inferior da página inicial do MedicMiner.

4.3. Interface de Login

Para logar no *site* o usuário deverá ter feito antes o cadastro. Para entrar é necessário o *e-mail* e a senha previamente registrados. Na Figura 4 é mostrada a interface de *login* do *site*.

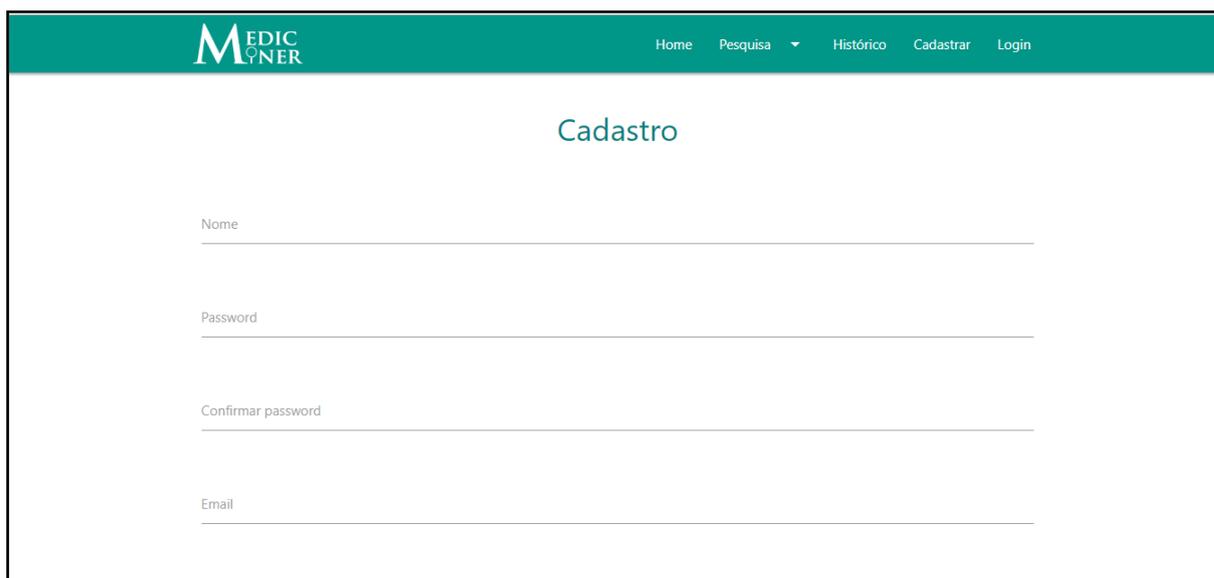


The screenshot shows the login page of the MedicMiner website. At the top, there is a teal navigation bar with the MedicMiner logo on the left and menu items 'Home', 'Pesquisa', 'Histórico', 'Cadastrar', and 'Login' on the right. The main content area is white and features the title 'Login' in teal. Below the title are two input fields: 'Email' and 'Password'. Underneath these fields are two teal buttons labeled 'LIMPAR' and 'LOGAR'. At the bottom of the form area, there is a red notification bar with the text: 'Obs.: O usuário deve estar logado para obter acesso a página de pesquisa e histórico.'

FIGURA 4. Interface de Login do MedicMiner.

4.4. Interface de Cadastro

Na página de cadastro (Figura 5) o usuário deverá registrar seus dados pessoais, são eles: Nome, *Password* (senha), Confirmar *password* (confirmar senha), E-mail, Instituição (caso seja um pesquisador ou estudante), Data de Nascimento e Telefone.



The screenshot shows the registration page of the MedicMiner website. It has the same teal navigation bar as the login page. The main content area is white and features the title 'Cadastro' in teal. Below the title are four input fields: 'Nome', 'Password', 'Confirmar password', and 'Email'. The 'Password' and 'Confirmar password' fields are vertically aligned, as are the 'Nome' and 'Email' fields.

FIGURA 5. Interface de cadastro.

4.5. Página de Pesquisa

Para realizar pesquisas no *site* o usuário deve ter cadastro e estar *logado*. Caso, o mesmo clique na aba de pesquisa ou histórico sem ter feito o *login*, ele será redirecionado para a página de *login*.

4.5.1. Página de Pesquisa – Interface Inicial

Na Figura 6 mostra que é possível fazer dois tipos de pesquisas: de Palavras-chave em um artigo único e Pesquisa gráfica. Na primeira, o usuário deve digitar as palavras que espera encontrar em um mesmo artigo. Na segunda, o mesmo deve digitar os termos e escolher o tipo de gráfico para fazer a visualização. Nos dois casos, o usuário deve inserir termos em inglês, visto que a base de dados e seus artigos são todos na língua inglesa. Em seguida, serão mostrados os resultados das pesquisas feitas.



The screenshot displays the MedicMiner search interface. At the top, there is a navigation bar with the MedicMiner logo and links for Home, Pesquisa, Histórico, and Cadastro. The user is logged in as 'Olá, Lucas'. The main content area is divided into two sections. The first section, 'Palavras-chave em artigo único', prompts the user to enter keywords for a single article search, with an example 'Ex.: AIDS, HPV, Cancer' and a 'PESQUISAR' button. The second section, 'Pesquisa gráfica', prompts the user to enter keywords and select a chart type from a dropdown menu, also with an example 'Ex.: AIDS, HPV, Cancer' and a 'PESQUISAR' button.

FIGURA 6. Interface de Pesquisa do MedicMiner.

4.5.2. Página de Pesquisa - Gráfico de Pizza

Foi feita uma pesquisa com as palavras-chave “vaccine”, ”hvp”, “cancer” como pode ser visto na Figura 7.

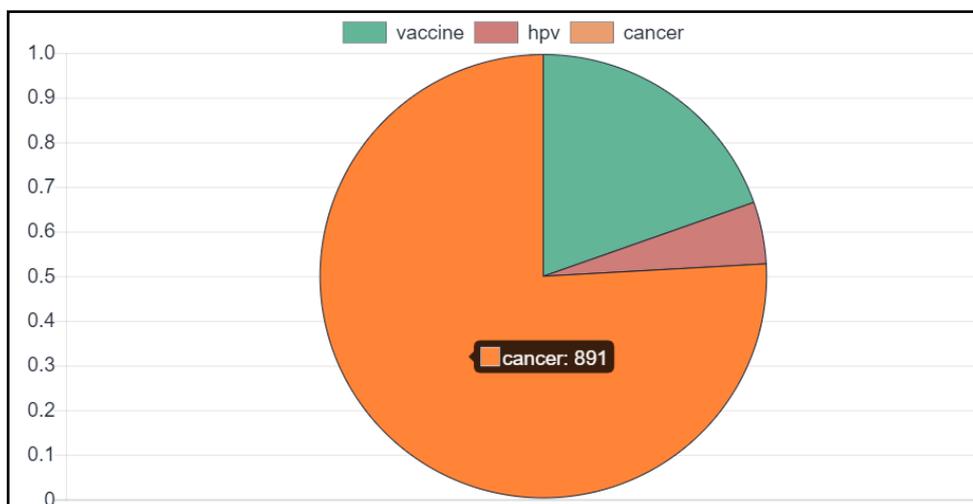


FIGURA 7. Gráfico de “Pizza” resultante de pesquisa realizada na plataforma.

4.5.3. Página de Pesquisa - Gráfico de Barras

Foi realizada uma outra pesquisa com os termos “fever”, ”dengue”, “chikungunya”. Nessa pesquisa o resultado foi mostrado em gráfico de barras como mostra a Figura 8.

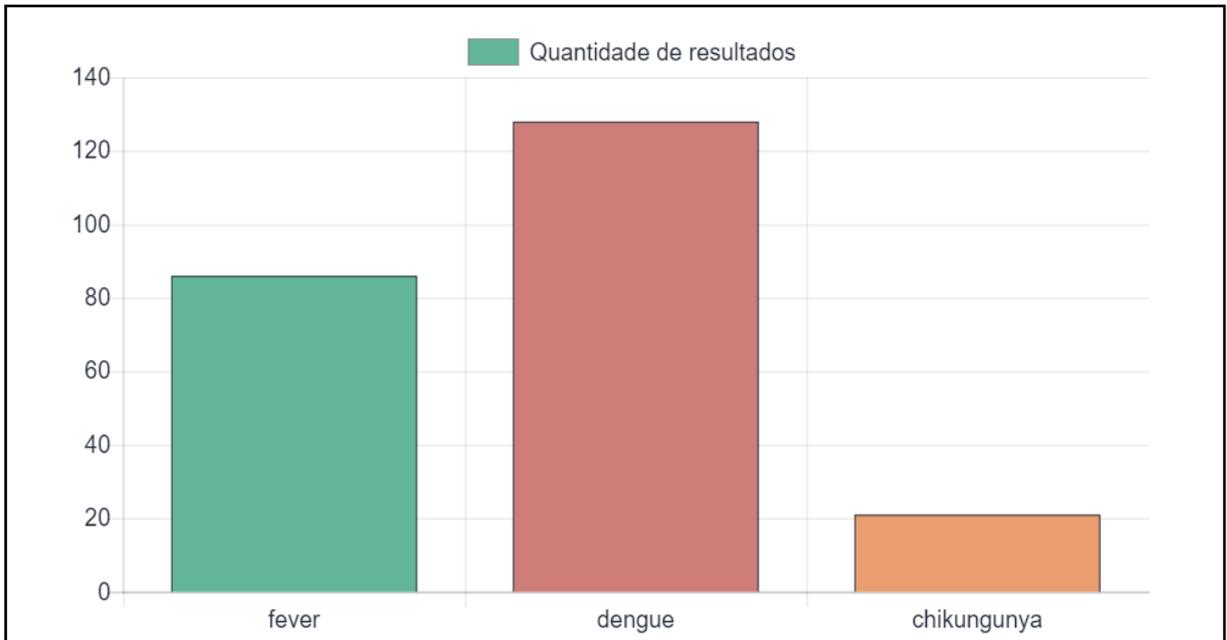


FIGURA 8. Gráfico de “Barra” resultante de pesquisa realizada na plataforma.

4.5.4. Página de Pesquisa - Gráfico de Área Polar

Para demonstrar resultados em forma de gráfico de “área polar”, foi feita uma pesquisa com os termos “mosquito”, ”rainforest”, “fever”. Na Figura 9 é possível ver o que foi obtido.

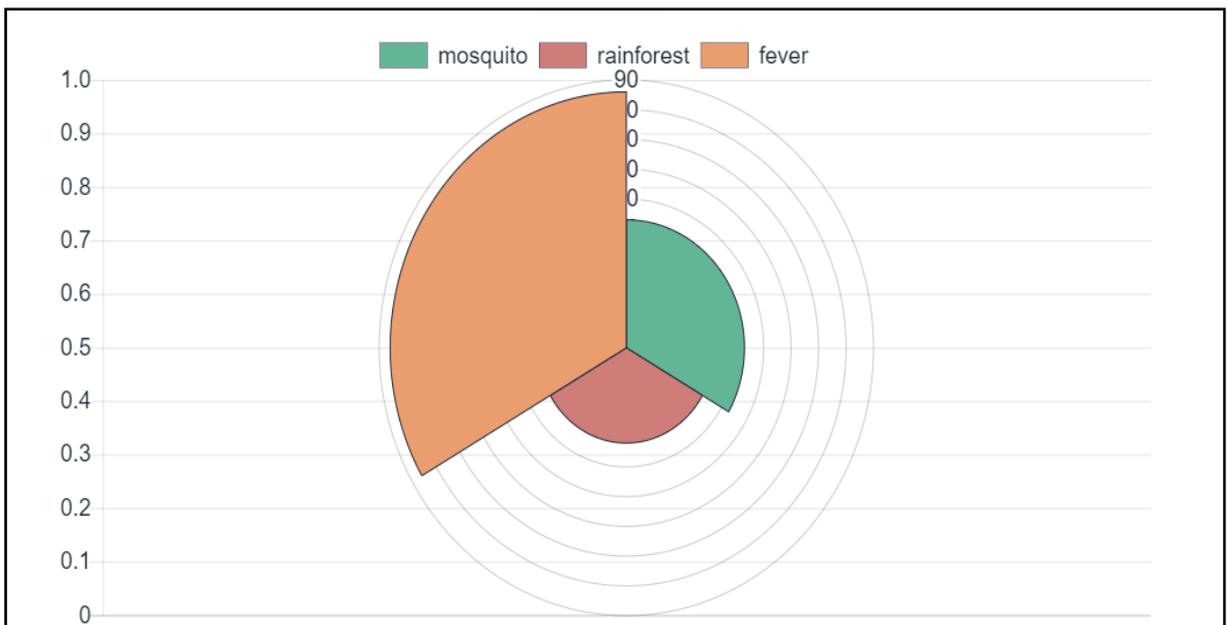


FIGURA 9. Gráfico de “Área Polar” resultante de pesquisa realizada na plataforma.

4.5.5. Página de Pesquisa - Gráfico de Área Polar: Visualizando Resultados

Na Figura 10 foi feita a pesquisa com os termos “aids”, ”hiv”, “blood”.Nela, é mostrado que ao colocar a seta em cima da área do gráfico referente a um termo específico (no caso a palavra-chave “blood”), é visualizada a quantidade de artigos em que a palavra aparece.

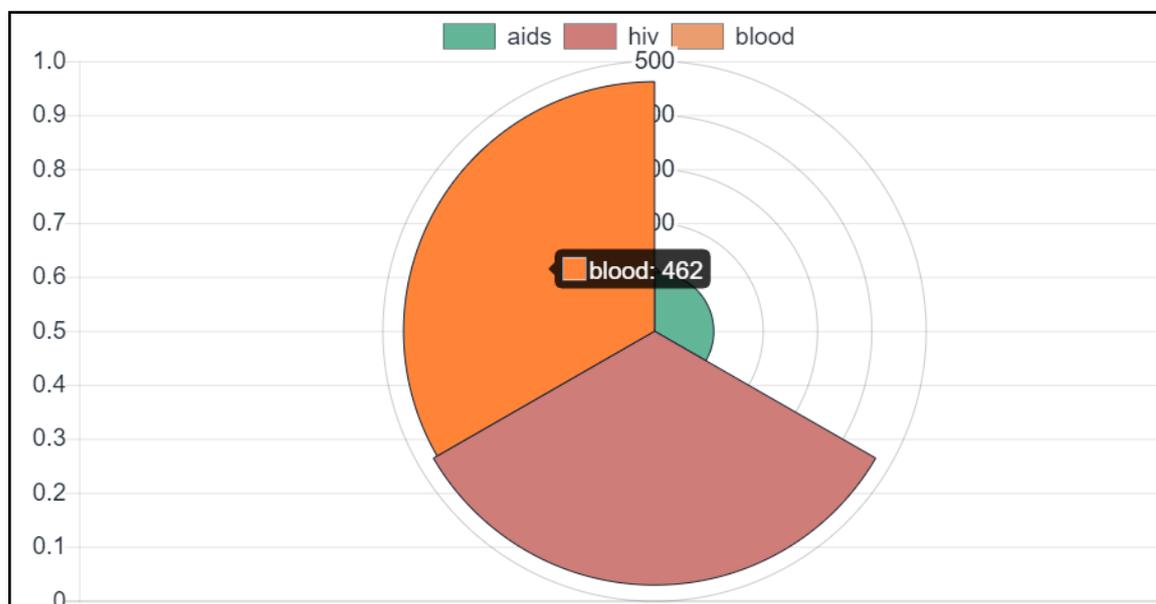


FIGURA 10. Gráfico de “Área Polar” destacando a quantidade de artigos com o termo “blood”.

4.5.6. Página de Pesquisa - Gráfico de Área Polar: Exclusão de termo

Na Figura 11 é mostrado como é possível retirar determinado termo do gráfico já gerado. Para isso, basta clicar em cima da palavra e ela será excluída da pesquisa e será gerado um novo gráfico. Caso o usuário queira colocar o termo novamente nos resultados, é necessário apenas clicar novamente.

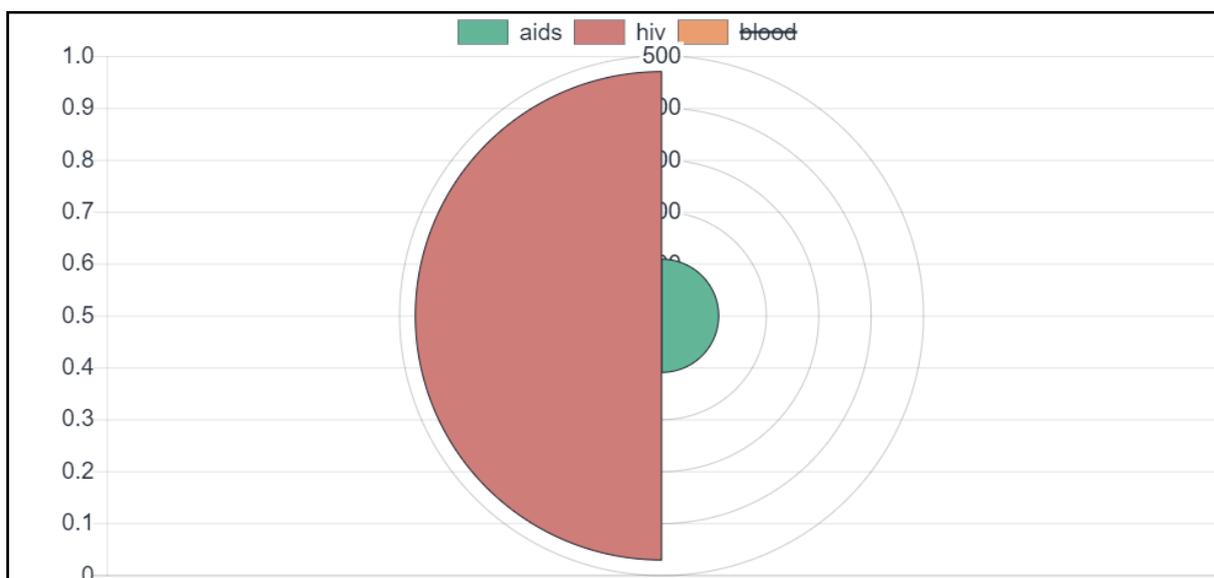


FIGURA 11. Gráfico de “Área Polar” com a exclusão do termo “blood” da pesquisa.

4.6. Página de Lista dos links de Artigos Encontrados

Na seção de pesquisa de “Palavras-chave um artigo único”, foram inseridos os termos “blood” e “bone”. Em seguida, o sistema abre uma nova guia com os resultados encontrados, como mostra a Figura 12. Para acessar cada artigo, basta o usuário clicar no link com o nome da publicação.



FIGURA 12. Listagem de Artigos Encontrados com os termos “blood” e “bone”.

4.7. Página de Grafos

Na parte de pesquisa do menu, o usuário pode selecionar a opção “Grafos”. Ao clicar, ele será redirecionado para uma aba em que são mostrados três exemplos de grafos já gerados como mostra a Figura 13.



FIGURA 13. Página de Grafos.

4.8. Visualização de Exemplos de Grafos

A Figura 14 representa um exemplo de grafo que o usuário pode visualizar na plataforma.

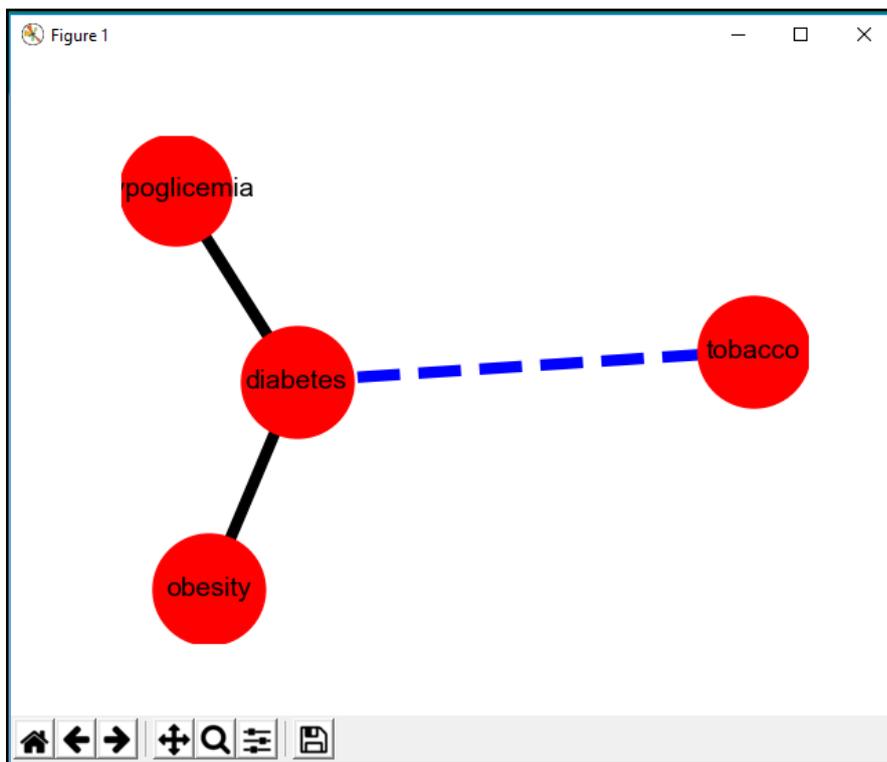


FIGURA 14. Exemplo de Grafo disponível na plataforma.

4.9. Página de Histórico

Quando são feitas pesquisas de palavras-chave em artigo único são armazenados os termos inseridos e a data em que as inserções foram feitas, como mostra a Figura 15. O usuário pode visualizar os resultados novamente bem como apagá-los do histórico.

Termos	Data		
bone, blood	2018-10-24	Pesquisar	Excluir
dengue, fever	2018-10-24	Pesquisar	Excluir
vaccine, fever	2018-10-29	Pesquisar	Excluir
cancer, blood	2018-10-29	Pesquisar	Excluir
bone, calcium	2018-10-29	Pesquisar	Excluir
vaccine, blood	2018-10-30	Pesquisar	Excluir

FIGURA 15. Listagem de Artigos Encontrados com os termos “blood” e “bone”.

5. Considerações Finais

O projeto desenvolvido visa possibilitar a realização de pesquisas especializadas, utilizando técnicas de mineração de textos que facilite a recuperação de dados com a combinação de parâmetros que não é possibilitada pelas interfaces de pesquisa disponível no portal da PubMed.

Logo, ao utilizar o portal MedicMiner responsável por realizar extração do conjunto de artigos do repositório da PubMed é possível realizar a combinação de diversas variáveis e com isso, obter resultados complexos com baixo custo computacional e em uma interface de fácil utilização. Diante da estratégia adotada, é possível realizar pesquisas com múltiplos parâmetros e visualizar a representação desses resultados de forma dinâmica e interativa. Os objetivos inicialmente propostos foram alcançados e, acredita-se que a plataforma auxilia significativamente em pesquisas mais precisas no conjunto de artigos que descrevem resultados de estudos nas áreas médicas.

Como trabalho futuro, espera-se a geração de grafos dinâmicos de colaborações científicas e de termos para identificar também de forma visual o relacionamento entre diversos elementos de análises, como por exemplo, co-ocorrência de doenças, sintomas e medicamentos, e ainda, a colaboração científica entre pesquisadores, institutos de pesquisa e países.

Referências

FAYYAD, Usama; PIATE TSKI-SHAPIRO, Gregory; SMYTH, Padhraic. **The KDD process for extracting useful knowledge from volumes of data.** Communications Of TheAcm: New York, v. 39, n. 11, p. 27-34, 1996.

SHAMA, A; MANSOTRA, V. **Emerging applications of data mining for healthcare management - A critical review.** 2014 International Conference on Computing for Sustainable Global Development (INDIACom), New Dehli. p. 377-382. 2014.

NUNES, Felipe Vieira; FREITAS, Vitor Silva de; SILVA, Vitor Angelo. **Medical Mining.** 2016. 19f. Trabalho de Conclusão de Curso – Centro Federal de Educação Tecnológica – CEFET-MG – Divinópolis, Minas Gerais – Brasil, 2016.

Menezes, N. N. C. **Introdução à Programação com Python: Algoritmos e lógica de programação para iniciantes.** 2ª edição. Editora Novatec, 17 de junho de 2014.

SILVA, J. **PHP na Prática.** 1ª Edição. Editora Elsevier, 2014.

ARRUDA, Guilherme Ferraz de. **Mineração de dados em redes complexas: estrutura e dinâmica.** 2013. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, University of São Paulo, São Carlos, 2013. doi:10.11606/D.55.2013.tde-25062013-085958. Acesso em: 17 de maio de 2018.

MEDLINE PLUS. **About MedlinePlus.** 2018. Disponível em: <<https://medlineplus.gov/aboutmedlineplus.html>>. Acesso em: 28 de março de 2018.

SERVIÇO DE DOCUMENTAÇÃO E DIVULGAÇÃO. SDD. **Qual a diferença entre Medline® e PubMed®?** 2010. Disponível em: <<https://sddinforma.wordpress.com/2010/09/14/qual-a-diferenca-entre-medline%C2%AE-e-pubmed%C2%AE/>>. Acesso em: 31 de março de 2018.

SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. **Sistema De Banco De Dados.** 6ª. ed. Rio de Janeiro: Editora Elsevier, 2012. 781 p. Disponível em: <<https://books.google.com.br/books?id=1FBaDwAAQBAJ&printsec=frontcover&dq=sistema+de+banco+de+dados&hl=pt-BR&sa=X&ved=0ahUKEwiv5uLhx-rdAhWDFJAKHXbCBQkQ6AEIKDAA#v=onepage&q&f=false>>. Acesso em: 15 maio 2018.

SCIELO - **Sobre esse site.** 2018. Disponível em: <<http://www.scielo.br/?lng=pt#about>>. Acesso em: 18 de maio de 2018.

ARRIGONI, Ricardo. **CRUD com PHP PDO**. Disponível em: <<https://www.devmedia.com.br/crud-com-php-pdo/28873>>. Acesso em: 17 de maio de 2018.

BROFMAN, P. R. **A importância das publicações científicas**. Cogitare Enfermagem, Curitiba, v. 17, n. 3. 419-421, 2012.

TAN, Pang-Ning ; STEINBACH, Michael ; KUMAR, Vipin . **Introdução ao DataMining: Mineração de Dados**. Rio de Janeiro: Ciência Moderna, 2009. 928 p. Disponível em: <<http://www.uokufa.edu.iq/staff/ehsanali/Tan.pdf>>. Acesso em: 31 maio 2018.

PUBMED. **Home PubMed**. 2018. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/>>. Acesso em: 31 maio 2018.

NETO, Marco Aurélio Silva et al. **Técnicas de mineração visual de dados aplicadas aos dados de instrumentação da barragem de Itaipu**. 2010. 721-734 p. (Programa de Pós-Graduação em Métodos Numéricos em Engenharia) UFPR, São Carlos, 2010. 17. Disponível em: <<http://www.scielo.br/pdf/gp/v17n4/a07v17n4.pdf>>. Acesso em: 01 jun. 2018.

REZENDE, Joffre M. de. **O SÍMBOLO DA MEDICINA: TRADIÇÃO E HERESIA**. Disponível em: <<http://www.psiquiatriageral.com.br/educacaomedica/simbolo.htm>>. Acesso em: 02 out. 2018.

FLANAGAN, David. **JavaScript: O guia definitivo**. 6ª. ed. São Paulo: Bookman, 2012. 1080 p. Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=lang_pt&id=zWNYDgAAQBAJ&oi=fnd&pg=PR1&dq=javascript&ots=Izxky-ObfJ&sig=RTb5SAQtc8EVsckSdviOvyCueg0#v=onepage&q&f=false>. Acesso em: 02 out. 2018.

SILVA, Maurício Samy. **Ajax com JQuery: Requisições Ajax com a Simplicidade de JQuery**. São Paulo: Novatec, 2009. 328 p. Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=lang_pt&id=iFGiDQAAQBAJ&oi=fnd&pg=PA23&dq=json+javascript&ots=o_IQhO6kZl&sig=rfkvW359BTf7IRN-6oxOKveh9p8#v=onepage&q=json%20javascript&f=false>. Acesso em: 02 out. 2018.

TAVARES, Thales Nicolai et al. **Desenvolvimento de uma ferramenta para monitoramento de dispositivos em rede sem fio**. 2016. 6 p. Anais do EATI (Encontro Anual de Tecnologia da Informação)- Colégio Técnico Industrial de Santa Maria – Universidade Federal de Santa Maria (UFSM), Santa Maria - RS, 2016. Disponível em: <<http://eati.info/eati/2016/assets/anais/Longos/171.pdf>>. Acesso em: 02 out. 2018.

GARTNER. **Big Data e Inteligência Artificial: números e estatísticas impressionantes**. 2017. Disponível em: <<http://www.bigdatabusiness.com.br/os-grandes-e-impressioantes-numeros-de-big-data/>>. Acesso em: 07 nov. 2018.

APÊNDICE 1– Diagrama de Casos de Uso do MedicMiner

A Figura 13 representa o Diagrama de Casos de Uso do MedicMiner. Na plataforma, todas as funcionalidades (Casos de Uso) são executadas apenas se o usuário estiver *logado* no sistema.

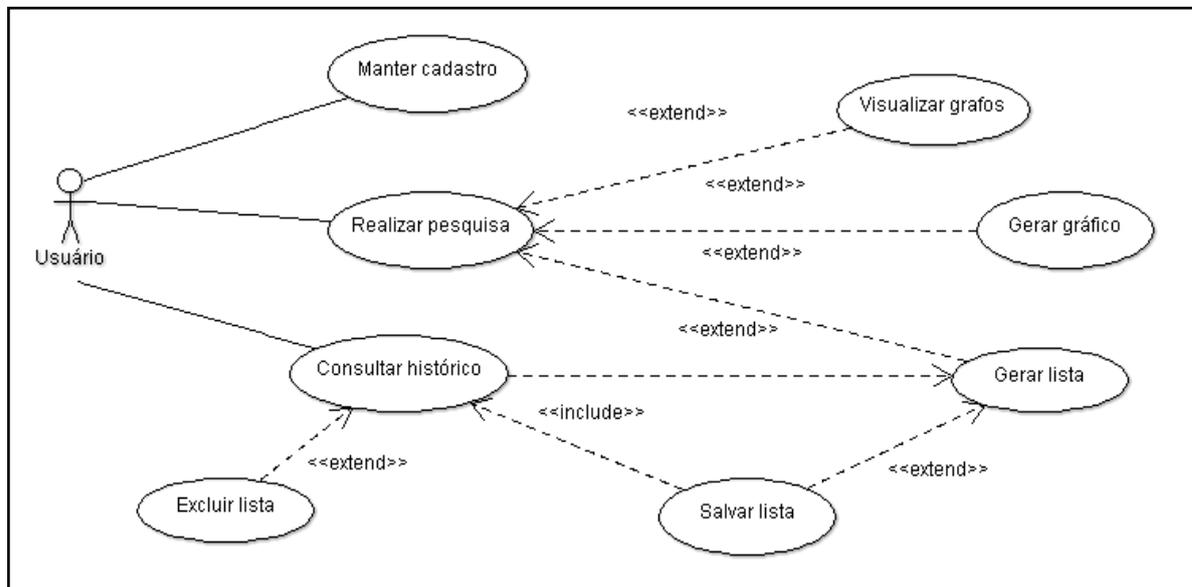


FIGURA 13. Diagrama de Casos de Uso do MedicMiner.

APÊNDICE 2– Diagrama Entidade-Relacionamento do MedicMiner

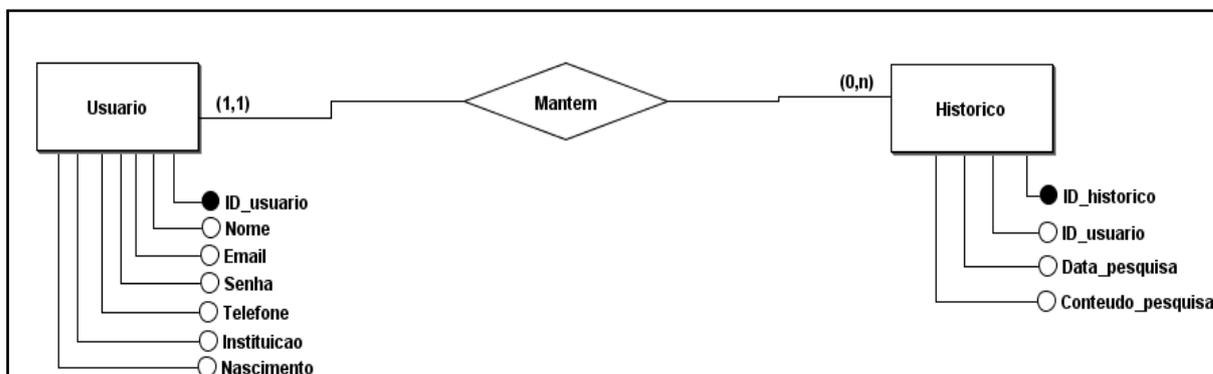


FIGURA 14. Diagrama Entidade-Relacionamento do MedicMiner.