

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO  
FACULDADE DE CIÊNCIAS EXATAS E TECNOLOGIA  
ENGENHARIA BIOMÉDICA

ELLEN CARVALHO DE SOUSA  
FRANCISCO HADDAD AGUERRE  
SOPHIA MARQUES SILVA

**TÉCNICAS COMPUTACIONAIS APLICADAS PARA EXTRAÇÃO E  
CENTRALIZAÇÃO DE INFORMAÇÕES DOS PROTOCOLOS  
CLÍNICOS E DIRETRIZES TERAPÊUTICAS**

SÃO PAULO  
2021

ELLEN CARVALHO DE SOUSA  
FRANCISCO HADDAD AGUERRE  
SOPHIA MARQUES SILVA

**TÉCNICAS COMPUTACIONAIS APLICADAS PARA EXTRAÇÃO E  
CENTRALIZAÇÃO DE INFORMAÇÕES DOS PROTOCOLOS  
CLÍNICOS E DIRETRIZES TERAPÊUTICAS**

Monografia final apresentada à Pontifícia  
Universidade Católica de São Paulo, como  
requisito para a obtenção do título de Bacharel  
em Engenharia Biomédica.

Orientador: Prof. Dr. Lionel Gamarra

SÃO PAULO  
2021

Ellen Carvalho de Sousa  
Francisco Haddad Aguerre  
Sophia Marques Silva

**TÉCNICAS COMPUTACIONAIS APLICADAS PARA EXTRAÇÃO E  
CENTRALIZAÇÃO DE INFORMAÇÕES DOS PROTOCOLOS  
CLÍNICOS E DIRETRIZES TERAPÊUTICAS**

Monografia final apresentada à Pontifícia  
Universidade Católica de São Paulo, como  
requisito para a obtenção do título de Bacharel  
em Engenharia Biomédica.

São Paulo, 23 de novembro de 2021.

**BANCA EXAMINADORA**

---

Prof. Dr. Alexandre Barbosa de Lima

---

Profa. Dra. Fernanda Guerreiro Rossi

---

Prof. Dr. José Jarjura Jorge Junior

## **AGRADECIMENTOS**

A Deus por ter concedido saúde e força para superar as dificuldades.

A Pontifícia Universidade Católica, seu corpo docente, direção e administração pelas oportunidades, suporte e confiança depositada em nós ao longo desses anos e principalmente na última etapa desta jornada.

Ao orientador Dr. Lionel Gamarra, pelo suporte promovido ao longo deste período, pelas suas correções e incentivos.

Aos nossos pais, parentes e amigos pelo amor, incentivo e apoio incondicional.

Ao Prof. Dr. Eduardo Savino pelo suporte, incentivo e contribuição fornecida durante o desenvolvimento do projeto.

Ao Engenheiro Biomédico Sergio Freitas de Siqueira e à Cultivare Prevenção e Promoção de Saúde pelos conteúdos fornecidos para complementação e espaço fornecido para discussões e enriquecimento do trabalho.

E a todos que direta ou indiretamente fizeram parte da nossa formação, o nosso muito obrigado.

## RESUMO

A Medicina Baseada em Evidência propõe uma metodologia para o processo de tomada de decisões médicas, buscando minimizar erros, por não considerar apenas a experiência do profissional, mas também as melhores evidências disponíveis sobre a doença.

Os Protocolos Clínicos e Diretrizes Terapêuticas (PCDT) são documentos nos quais constam evidências para um tratamento efetivo e são desenvolvidos pela Comissão Nacional de Incorporação de Tecnologias no Sistema Único de Saúde (CONITEC). Porém, não há uma centralização das informações contidas nestes protocolos em um repositório para facilitar a consulta de maneira simplificada e intuitiva.

Assim, a proposta do trabalho foi de implementar, através de técnicas computacionais de reconhecimento de texto, um repositório centralizado com informações presentes em PCDTs, buscando promover uma consulta simplificada, intuitiva e ampliando o acesso à informações baseadas em evidências científicas.

A partir da técnica da busca direta é realizada a identificação da posição e extração do trecho de interesse. O trecho extraído é armazenado em um banco de dados. A consulta por uma *query* do termo de busca, inserido pelo usuário, retornará quais PCDTs possuem aquele termo e seu conteúdo extraído é retornado para o usuário.

O desenvolvimento do trabalho demonstrou que é possível a centralização dos conteúdos de interesse presentes no PCDT para facilitar a consulta de maneira simples. A consulta às informações obtidas podem servir de base para aplicação na melhoria na qualidade da assistência e os métodos aplicados para extração e consulta podem também ser aplicados para outros documentos *Portable Document Format* (PDF) com a mesma finalidade: centralização e facilitação da consulta.

**Palavras-chave:** PCDT; MBE; Busca direta; Extração de Texto; Sistematização; Consulta.

## **ABSTRACT**

Evidence-Based Medicine proposes a methodology for the medical decision-making process, minimizing the level of uncertainty and errors, as it does not only consider the professional's experience, but also the best available evidence about that disease.

The Clinical Protocols and Therapeutic Guidelines (PCDT) are documents that contain evidence for an effective treatment and are developed by the National Commission for the Incorporation of Technologies in the Unified Health System (CONITEC). However, there is no centralization of information properly contained in a repository to facilitate consultation in a simplified and intuitive way.

Thus, the proposal of the work was to implement, through computational text recognition techniques, a centralized repository with information present in PCDTs, seeking to promote a simplified, intuitive consultation and expanding access to information based on scientific evidence.

The direct search technique is used to identify the position and extract the stretch of interest. The extracted snippet is stored in a database. The query for a search term query, entered by the user, will return which PCDTs have that term and their extracted content is returned to the user.

The development of the work demonstrated that there is the possibility of centralizing the contents of interest present in the PCDT in order to facilitate the consultation in a simple and intuitive way. The information obtained can serve as a basis for application in improving the quality of care and the methods applied for extraction and consultation can also be applied to other PDF documents with the same purpose: centralization and facilitation of consultation.

**Keywords:** PCDT; MBE; Direct search; Text Extraction; Systematization; Query.

## LISTA DE FIGURAS

Figura 1: Arquitetura dos sistemas especialistas.....	18
Figura 2: Artigos publicados por país acerca de <i>Decision Support and Guideline Based</i> .....	25
Figura 3: Distribuição de artigos por ano.....	26
Figura 4: Distribuição de artigos por área.....	27
Figura 5: Diagrama com os resultados da busca sistemática.....	28
Figura 6: Etapas da metodologia deste projeto.....	31
Figura 7: Portaria de aprovação do PCDT de Acromegalia.....	33
Figura 8: Trecho de interesse: CID-10.....	33
Figura 9: Trecho de interesse: Fármacos.....	34
Figura 10: Técnica de sumarização aplicado à extração do CID-10.....	44
Figura 11: Etapas de extração 1 a 4: ano e número de páginas.....	46
Figura 12: Etapa de extração 5: normalização do conteúdo.....	47
Figura 13: Etapas de extração 6 e 7: extração do CID/fármacos/tempo e armazenamento.....	48
Figura 14: Recorte da tabela PCDT com documentos extraídos e armazenados no BD.....	49
Figura 15: Consulta do termo “G12.0” no PCDT_Atrofia Muscular Espinhal_2019..	50
Figura 16: Consulta do termo “Ibuprofeno” na tabela “PCDT” do BD.....	50
Figura 17: Consulta do fármaco “Ibuprofeno” na Referência Medicamentos.....	51
Figura 18: Interface de Consulta: Page Home.....	52
Figura 19: Interface de Consulta: Page About Us.....	53

Figura 20: Interface de Consulta: Page Consulta aos PCDTs.....	54
Figura 21: Interface de Consulta Page Consulta: Exemplo busca CID G12.0.....	55
Figura 22: Box Plot da dispersão da pontuação obtida por grupo de análise.....	57
Figura 23: Dispersão dos índices por amostra/grupo.....	58

## LISTA DE TABELAS

Tabela 1 – Análise Sistemática da Revisão de Literatura.....	29
Tabela 2 - Recorte da tabela Registros ANVISA de Janeiro de 2021.....	35
Tabela 3 - Recorte da tabela que consolida os resultados de todas as amostras por grupo.....	40
Tabela 4 – Compilação do resultado da coleta de dados para análise estatística....	56
Tabela 5 – Descrição dos dados do índice pela média e desvio padrão dos grupos.....	58
Tabela 6 – Teste de Kruskal-Wallis para avaliação da comparação entre os grupos.....	59
Tabela 7 – Comparações de Post Hoc do índice entre os grupos.....	59
Tabela 8 – Anexo 1: Análise Sistemática I da Revisão de Literatura.....	67

## LISTA DE ABREVIACÕES E SÍMBOLOS

ANS	Agência Nacional de Saúde Suplementar.
ANVISA	Agência Nacional de Vigilância Sanitária.
API	Application Programming Interface.
AVC	Acidente Vascular Cerebral.
CBHPM	Classificação Brasileira Hierarquizada de Procedimentos Médicos.
CDSS	Sistema de Suporte à Decisão Clínica.
CID	Classificação Internacional de Doenças.
CONITEC	Comissão Nacional de Incorporação de Tecnologias.
CUI	Concept Unique Identifier.
DCB	Denominação Comum Brasileira.
DM	Diretriz Médica.
EUA	Estados Unidos da América.
EMA	Agência Europeia de Medicamentos.
FDA	Food and Drug Administration
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo.
IA	Inteligência Artificial.
IDE	Integrated Development Environment.
MBE	Medicina Baseada em Evidência.
MS	Ministério da Saúde.
PCDT	Protocolos Clínicos e Diretrizes Terapêuticas.
PDF	Portable Document Format ou português: Formato de Documento Portátil
PLN	Processamento de Linguagem Natural.
PRISMA	Preferred Reporting Items for Systematic reviews and Meta – Analysis ou português: Itens de relatório para análises sistemáticas e meta-análise.

RI	Recuperação de informação.
SNOMED-CT	Systematized Nomenclature of Medicine - Clinical Terms ou português: Nomenclatura Sistematizada de Medicina – Termos Clínicos
STF	Supremo Tribunal Federal.
SUS	Sistema Único de Saúde.
TUSS	Terminologia Unificada de Saúde Suplementar.
UnB	Universidade de Brasília.
UMLS	Unified Medical Language System.

## SUMÁRIO

<b>RESUMO</b> .....	<b>4</b>
<b>ABSTRACT</b> .....	<b>5</b>
<b>1. Introdução</b> .....	<b>12</b>
1.1. Motivação .....	13
<b>2. Referencial Teórico</b> .....	<b>15</b>
2.1. Medicina Baseada em Evidência .....	15
2.1.2. Protocolos Clínicos e Diretrizes Terapêuticas .....	16
2.2. Sistemas de suporte à tomada de decisão .....	17
2.3. Técnicas de processamento de texto .....	18
<b>3. Objetivos</b> .....	<b>23</b>
3.1. Objetivos Gerais .....	23
3.2. Objetivos Específicos .....	23
<b>4. Revisão Sistemática de Literatura</b> .....	<b>24</b>
4.1. Estratégia de pesquisa .....	24
4.1.1. Compilação dos dados e revisão .....	25
4.1.2. Resultado da busca .....	25
<b>5. Metodologia</b> .....	<b>31</b>
5.1. Seleção da base de dados .....	32
5.1.1. Caracterização das bases de dados .....	32
5.2. Extração do texto no PCDT .....	35
5.3. Armazenamento no Banco de Dados .....	37
5.4. Consulta no Banco de Dados .....	37
5.5. Associação com demais fontes de dados .....	38
5.6. Interface de Consulta .....	38
5.7. Avaliação estatística de desempenho e acuracidade .....	39
<b>6. Resultados e Discussão</b> .....	<b>42</b>
6.1. Seleção da base de dados .....	42
6.2. Extração do texto no PCDT .....	42

6.3. Armazenamento no Banco de Dados .....	49
6.4. Consulta no Banco de Dados.....	49
6.5. Associação com demais fontes de dados .....	51
6.6 Interface de Consulta.....	51
6.7 Avaliação estatística de desempenho e acuracidade .....	56
<b>7. Conclusão .....</b>	<b>60</b>
<b>8. Bibliografia .....</b>	<b>61</b>
<b>9. Anexos .....</b>	<b>67</b>
Anexo 1 – Artigos totais da Revisão Sistemática.....	67
Anexo 2 – Documentos Fonte extraídos pelo sistema.....	74
Anexo 3 – Análise Estatística Grupo 2: Modelo para preenchimento .....	76
Anexo 4 – Análise Estatística Grupo 3: Modelo para preenchimento .....	77
Anexo 5 – Análise Estatística Grupo 4: Padrão Ouro .....	78

# 1. Introdução

## 1.1. Motivação

O caráter de urgência da pandemia de COVID-19 tem desafiado os valores da Medicina Baseada em Evidência (MBE) que segue como premissa a prática criteriosa das evidências científicas para o cuidado do paciente (SACKETT; ROSENBERG, 1996).

O uso de medicamentos como cloroquina e hidroxicloroquina foi apoiado por médicos, cientistas e chefes de Estado, mesmo sem nenhuma comprovação científica acerca da eficiência do tratamento para COVID-19 (FERREIRA; ANDRICOPULO, 2020). A cloroquina é um fármaco com indicação para uso como antimalárico, por outro lado a hidroxicloroquina corresponde a um hidroxilado análogo à cloroquina, porém, com um teor menor de toxicidade, podendo também ser aplicado para o tratamento de lúpus (FERREIRA; ANDRICOPULO, 2020). A situação de urgência e busca por tratamentos simples levou a aplicação de tratamentos sem a eficiência comprovada, sendo até promovido um “*kit covid-19*” composto por azitromicina, ivermectina e cloroquina/hidroxicloroquina como uma medida de tratamento precoce para pacientes que apresentassem sintomas, mesmo sem nenhuma eficiência comprovada (FERREIRA; ANDRICOPULO, 2020).

Um estudo aplicado com 118 profissionais atrelados à área da saúde de Dhaka. O estudo separou 60 pessoas para compor o grupo controle e 58 pessoas compondo o grupo experimental, com a aplicação de 12 mg de ivermectina pelo período de 4 meses. Demonstrou a redução da contaminação, em que o grupo de controle atingiu 73,3% de contaminação, enquanto que o grupo experimental apresentou apenas 6,9%. Porém, o estudo elaborado evidenciou que a amostra aplicada era insuficiente para generalização dos resultados e que seria necessário aplicação em uma escala maior para determinar a eficácia como medida preventiva (FLORÊNCIO; DOURADO, 2021).

A FDA (*Food and Drug Administration*) em 2021 proibiu que a ivermectina fosse aplicada para o tratamento de COVID-19 por não se tratar de um antiviral e

por conta de que seu uso em doses acima da recomendada causa danos como ataxia, bradipneia, ptose, redução da atividade, midríase, coma e morte (SENADO FEDERAL, 2021). A Agência Europeia de Medicamentos (EMA) após a análise das evidências se apresentou contrária ao uso, pois apesar do efeito bloqueador ao SARS-CoV-2 indicados por estudos, este efeito só é obtido com concentrações que são acima do recomendado, o que pode gerar um nível de toxicidade alta para os pacientes (FLORÊNCIO; DOURADO, 2021).

Apesar de existirem estudos bem sucedidos, a quantidade de amostras testadas não são suficientes para generalização dos resultados e as concentrações aplicadas excedem as doses recomendadas, podendo expor os pacientes aos efeitos colaterais e até a morte.

Os PCDT foram desenvolvidos para fornecer aos profissionais da saúde informações relativas ao diagnóstico, tratamento e monitoramento de situações ambulatoriais, tanto como seu risco.

Não há um PCDT para a COVID-19, em razão de não se ter ainda um tratamento especificado, apenas medidas de controle da contaminação como lavar as mãos, manter o distanciamento e uso de máscara. Porém, é necessário compreender a importância da elaboração para evitar que medidas sem comprovação científica sejam aplicadas e exponham pacientes a tratamentos sem eficácia comprovada.

A centralização das informações contidas nos PCDT serviria como uma maneira de promover uma consulta simplificada ao conteúdo com o intuito de tornar o seu conhecimento mais acessível e permitir uma análise crítica de condutas anteriores à sua aplicação, como por exemplo, no processo de recomendação de medicamentos para tratamentos específicos.

## **2. Referencial Teórico**

### **2.1. Medicina Baseada em Evidência**

A MBE se refere à associação entre experiências clínicas e a aplicação de informações científicas a partir de uma análise crítica, objetivando o aumento da qualidade na assistência médica (LOPES, 2000).

A MBE propõe uma metodologia para o processo de tomada de decisões médicas, minimizando o nível de incerteza e de erros, por não considerar apenas a experiência do profissional (NETO, 2008). A metodologia proposta se fundamenta em três pontos: pesquisas por referências científicas, epidemiológicas, revisão sistemática e dados fornecidos pelo paciente; no vínculo estabelecido entre o profissional e o paciente, partindo da experiência do profissional e individualidade do paciente; e por último, nas normas, planos, padrões, protocolos, leis, tempo de atendimento e custos de procedimentos (NETO, 2008).

A prática da MBE tem como um dos fatores que dificultam a sua aplicação o fato de que os protocolos que servem de referência são extensos e de leitura demorada, impossibilitando uma consulta de maneira rápida (MAGHSOUD – LOU et al., 2021). Esse fator afeta o último ponto que fundamenta o método de MBE descrito anteriormente, devido à dificuldade de uma consulta simplificada a cada protocolo. Além disso, há a necessidade de atualizações constantes e, se os documentos não estiverem apropriadamente estruturados, podem provocar dificuldade na compreensão (SEITINGER et al., 2016).

Esses fatores nos levam a considerar que há a necessidade do desenvolvimento de tecnologia capaz de auxiliar o profissional de saúde a centralizar essas informações para facilitar sua consulta em um processo de tomada de decisão e melhorar a prática da MBE.

### **2.1.2. Protocolos Clínicos e Diretrizes Terapêuticas**

A “Comissão Nacional de Incorporação de Tecnologias no Sistema Único de Saúde – Conitec criada pela Lei nº 12.401, de 28 de abril de 2011, dispõe sobre a assistência terapêutica e a incorporação de tecnologia em saúde no âmbito do Sistema Único de Saúde” (CONITEC, 2015).

Dentre as atribuições e responsabilidades desta Comissão está a elaboração dos PCDT, que correspondem a:

“(…) documentos que estabelecem critérios para o diagnóstico da doença ou do agravamento à saúde; o tratamento preconizado, com os medicamentos e demais produtos apropriados, quando couber; as posologias recomendadas; os mecanismos de controle clínico; e o acompanhamento e a verificação dos resultados terapêuticos, a serem seguidos pelos gestores do SUS. Devem ser baseados em evidência científica e considerar critérios de eficácia, segurança, efetividade e custo-efetividade das tecnologias recomendadas” (CONITEC, 2015).

Os protocolos clínicos foram desenvolvidos para fornecer aos médicos informações relativas ao diagnóstico, tratamento e monitoramento de situações ambulatoriais, podendo atuar sob qualquer evento clínico que esteja previamente parametrizado (LANZOLA et al., 2017). Além disso, evidenciam e resumem as evidências científicas para o tema em questão e ainda podem considerar a análise de custo-efetividade de determinada conduta, tanto como seu risco.

Para a elaboração de um PCDT, a CONITEC leva em consideração aspectos como eficácia, acurácia, efetividade, segurança e a avaliação econômica comparativa dos benefícios e dos custos em relação às tecnologias já existentes no SUS.

Desde o início da pandemia de COVID-19, a CONITEC tem acompanhado estudos científicos publicados em âmbito nacional e internacional sobre tecnologias para prevenção, tratamento e diagnóstico da doença, de modo a elaborar o mais breve possível diretrizes para o tratamento da COVID, visto que o relatório final de um PCDT pode levar de seis meses a um ano para ser concluído (CONITEC, 2021).

## 2.2. Sistemas de suporte à tomada de decisão

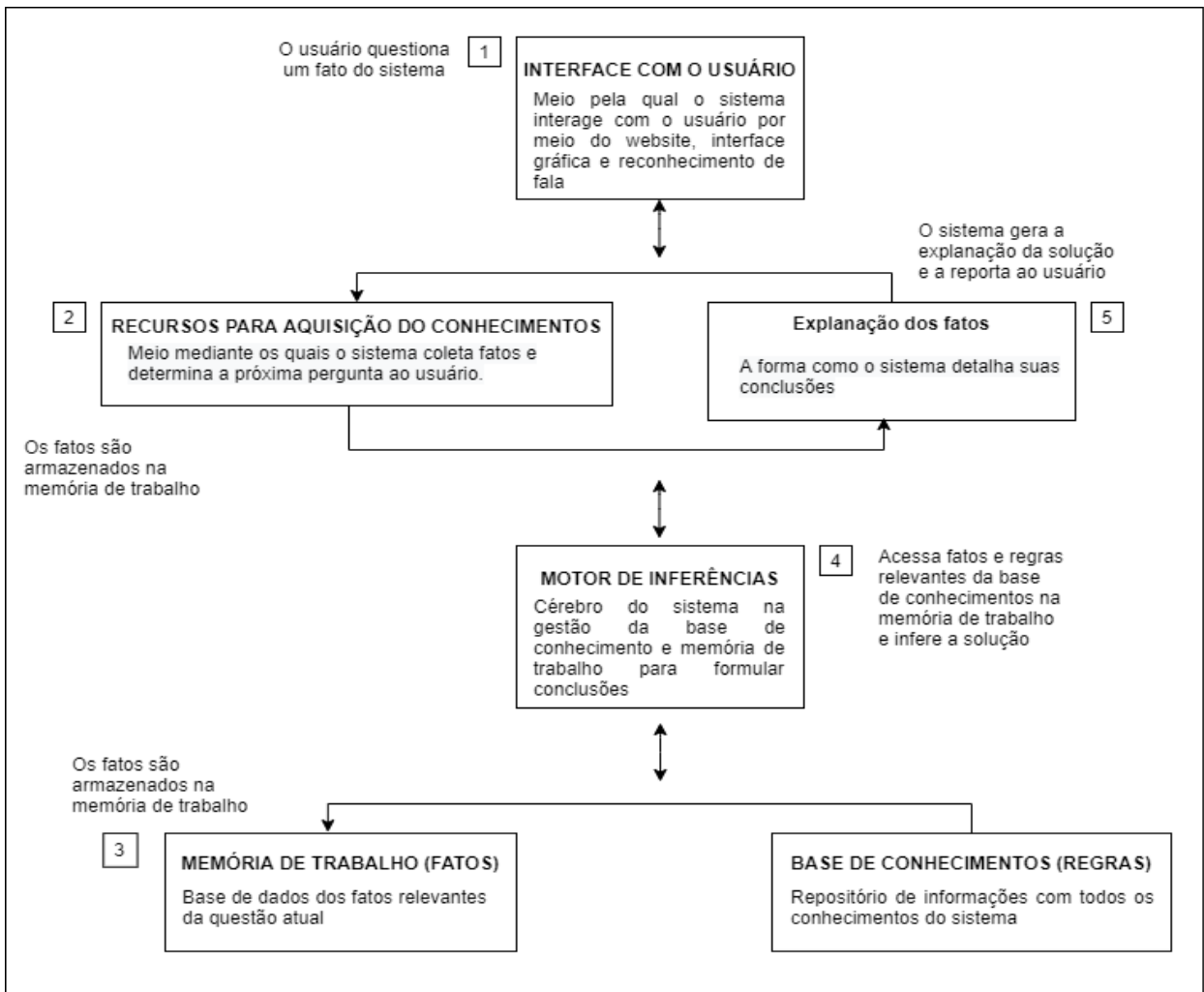
Os sistemas de apoio à decisão médica correspondem, na prática, à consolidação de evidências científicas disponíveis e parametrizadas sobre diversas doenças do cotidiano. Não obstante, já se estuda o impacto da não utilização de diretrizes no cuidado à saúde do paciente em comparação com os resultados clínicos de casos em que se aplica o suporte à tomada de decisão, sendo que aqueles que aplicaram o sistema de suporte à decisão obtiveram melhores resultados (AHAMED et al., 2016).

Dessa forma, há evidência científica de que sistemas de suporte à decisão clínica (CDSS) permitem o processamento de dados extensos e apresentação do conhecimento de maneira estruturada ao usuário final, a partir de diretrizes práticas de cuidado à saúde (SEITINGER et al., 2016).

A capacidade de auxiliar no processo de tomada de decisão – com informações precisas –, enquadram no que diz respeito à literatura, em sistemas especialistas, um dos principais domínios de pesquisa da Inteligência Artificial (IA) (GAVRANIC et al., 2021). No entanto, segundo AKABANE, aconselhar os usuários do sistema com informações adequadas, capacitar decisões assistenciais, interpretar dados de entrada e sugerir opções alternativas consistem diretamente na capacitação destes sistemas. No entanto, os sistemas especialistas são incapazes de produzir resultados precisos a partir de bases de dados inadequadas ou substituir a tomada de decisão humana, garantindo a plena característica de colaborar com as ações humanas (AKABANE, 2018).

Para elucidar a estrutura básica deste conceito, a arquitetura dos sistemas especialistas de uma maneira geral consiste em: **base de conhecimento**, **motor de interface** e a **interface com o usuário** (AKABANE, 2018), de acordo com a Figura 1 a seguir.

**Figura 1: Arquitetura dos sistemas especialistas.**



Fonte: Adaptado (AKABANE, 2018).

### 2.3. Técnicas de processamento de texto

Nas últimas décadas, houve um aumento considerável da quantidade de informação distribuída e armazenada, a um nível que torna inviável o processamento manual de cada informação obtida (LADEIRA, 2010).

Uma alternativa para resolver a problemática do acúmulo de informação, se dá pela aplicação de técnicas de tecnologias da informação para o processamento do acervo (LADEIRA, 2010). A mineração de dados é uma das formas que reúne uma série de disciplinas, tais como aprendizagem de máquina, estatística e linguística computacional, com a finalidade de promover a extração de informação de um conjunto de dados por meio de um código (ANDRADE, 2015). Com a

mineração de textos, é possível realizar uma análise e extração de textos, frases ou palavras de um grande volume de documentos, permitindo ampliar a compreensão do conteúdo (MORAIS, 2007).

Os textos em geral se apresentam de forma não estruturada, exigindo um pré-processamento para estruturá-los (ANDRADE, 2015). As técnicas de Processamento de Linguagem Natural (PLN) permitem esse pré-processamento por promoverem uma compreensão dos textos e dos dados de entrada no sistema para sua estruturação e aplicação em códigos de processamento (ANDRADE, 2015).

Um exemplo no Brasil, destaca-se o sistema VICTOR desenvolvido pela Universidade de Brasília (UnB) em parceria com o Supremo Tribunal Federal (STF), em que a partir de métodos de PLN e aprendizado de máquina, realiza o reconhecimento de padrões semânticos nos textos processuais e classificando de maneira automática, o que hoje demanda uma análise manual, otimizando diretamente na produtividade daqueles que a realizam (SILVA, 2018).

Logo, viabilizar técnicas para o desenvolvimento de tecnologias com PLN, na atualidade, incide diretamente no conceito de visão computacional, em que se atribui ao computador a capacidade de operar imagens ou textos, processando operações de classificação, identificação de entidades e reconhecendo estruturas nos textos (AKABANE, 2018).

Já o processo de mineração de textos possui duas abordagens, sendo ela a semântica e estatística, podendo ser aplicado separadamente ou em conjunto (MORAIS, 2007).

A abordagem semântica tem como objetivo a identificação da funcionalidade e da importância do termo no texto, aplicando técnicas de PLN (FERNEDA, 2003).

A abordagem estatística analisa a frequência do termo de busca no documento, tendo como resultado uma simplificação de toda a informação. Por não se ter capacidade computacional suficiente para o desenvolvimento de modelos estatísticos mais ricos, a simplificação pode não ser uma descrição exata do conteúdo abordado no texto (ANDRADE, 2015).

A preparação dos dados dentro da mineração de textos compõe uma etapa posterior à definição da abordagem, corresponde a seleção dos textos que servirão

de base para a consulta e a identificação da parte de interesse (MORAIS, 2007). O processo de preparação de dados envolve a recuperação de informação (RI). A área de RI apresenta os modelos necessários para a representação dos textos em tópicos específicos (FERNEDA, 2003).

A partir dos Sistemas de Recuperação de Informações Textuais (SRI Textual) é possível indexar e recuperar textos a partir de uma consulta por um termo, e este deve se adequar ao objetivo da busca (FERNEDA, 2003).

A consulta é estabelecida por meio de um mecanismo de Análise de Relevância através de uma função de similaridade que possibilita a comparação entre o termo da busca e o conteúdo presente no documento (MORAIS, 2007). Dentre os métodos que podem ser aplicados para a recuperação de informação há: **booleano**, **espaço vetorial**, **probabilístico**, **difuso (fuzzy)**, **busca direta**, **aglomeração (clusters)**, **lógico** e o **conceitual** (MORAIS, 2007).

O **método booleano** se refere à aplicação de operadores (*and*, *or* e *not*) para unir, descrever intersecções e remover partes de um conjunto de palavras. O uso dos operadores permite retornar à intersecção com o termo de consulta (ou trecho de interesse). Com possibilidade de ordenação a partir do nível de correlação com os termos de busca (SOARES, 2016).

O **modelo espaço vetorial** é a representação dos documentos como um vetor de termos que possuem pesos que determinam o grau de importância. Quando um termo não está contido no documento é atribuído peso zero, através da fórmula são identificados a importância dos termos de busca, sendo geralmente atrelado a frequência aparição no documento (REZENDE et al., 2011).

O **método probabilístico** se estabelece pelo cálculo da probabilidade do documento possuir relevância a consulta, a base matemática para o modelo probabilístico se dá pelo Método Bayesiano (SOARES, 2016)

O **modelo difuso** aplica, assim como o método probabilístico, um vetor de termos com os graus de importância, porém, não há conjunto vazio somente a atribuição de um nível de importância inferior, podendo apresentar um certo nível de incerteza que busca ser tratada pela teoria difusa (MORAIS, 2007).

A **busca direta** se refere a consulta de *strings* de maneira direta no texto, como resultado são identificadas todas as ocorrências dos termos no documento. O método é simples, porém, não é aconselhada a sua aplicação para um conjunto muito extenso de textos (SOARES, 2016).

No **método de aglomeração (*clusters*)** ocorre a identificação do conteúdo dos documentos, a indexação e os agrupamentos de conteúdos similares. Os conteúdos são considerados similares de acordo com a quantidade de palavras em comum e pela frequência, o resultado da busca se restabelece pelo retorno do grupo que apresenta similaridade com o termo de busca (REZENDE et al., 2011).

O **método lógico** se estabelece na aplicação de modelos matemáticos para realizar o processo de consulta do texto. O modelo lógico exige uma construção adequada da lógica predicativa com a incorporação da semântica para o processo de busca, fazendo com o que o mecanismo possua a capacidade de determinar a relevância do conteúdo presente no texto para definir se há correlação com o termo de busca (FERNEDA, 2003).

O **modelo conceitual** diferente dos métodos e modelos anteriores, não aplica uma consulta baseada na busca de termos idênticos ao da consulta, a correlação é estabelecida através da análise da morfologia do termo de busca e do termo presente no conteúdo do texto levando em consideração o contexto do conteúdo e o termo de busca (MORAIS, 2007). O conjunto de palavras associadas ao contexto do conteúdo do documento deve ser desenvolvido com a atribuição de pesos que variam com o contexto, quanto mais próximo o termo está do contexto mais alto será o peso atribuído. Os descritores devem ser construídos com cautela, pois a escolha incorreta levará a resultados inadequados da busca. A aplicação do modelo conceitual é limitada, pois ocorre certo nível de incerteza e ambiguidade devido aos problemas relacionados à sinonímia e polissemia (MORAIS, 2007).

A **indexação** é o processo que sucede a escolha do método, se refere ao processo de identificação de características da fonte em que se deseja consultar e a aplicação características para a definição de índices (FERNEDA, 2003). A indexação possui como etapas para a sua construção a identificação de termos, remoção de

*stopwords*, normalização, padronização do vocabulário e seleção de termos relevantes para a consulta (FERNEDA, 2003).

A identificação de termos se divide na identificação de termos simples e compostos. Para os termos simples são utilizados analisadores léxicos que realizam a localização dos termos. Pode ser realizado anteriormente ao processo a normalização dos termos com a remoção de caracteres maiúsculo, substituição de caracteres especiais, erros ortográficos, ajuste das tabulações e padronização dos tipos dos dados. A identificação de termos compostos corresponde a identificação de um conjunto de termos que pode ser feita pela análise da frequência de repetição do conjunto de termos no texto ou através da aplicação de um dicionário de expressões (REZENDE et al., 2011).

A remoção de *stopwords* é a remoção dos termos que não apresentam relevância para o processo de análise do texto por não afetarem o contexto do que o texto transmite permitindo a sua remoção para facilitar o processamento (REZENDE et al., 2011). A normalização morfológica também pode ser aplicada ao processo se refere a redução dos termos ao seu radical, eliminando sufixos, prefixos, gênero, número e grau, promovendo a diminuição do tamanho do índice (REZENDE et al., 2011).

A definição da relevância dos termos pode ser feita por meio da análise da frequência dos termos, análise da estrutura do documento ou por meio da posição sintática do termo (MORAIS, 2007).

Por fim, o texto é selecionado e deve ser realizada a análise dos resultados, verificando a eficiência e eficácia das partes que foram extraídas como resultado da consulta (REZENDE et al., 2011).

As técnicas de processamento de texto podem ser aplicadas com finalidades diversas para a facilitação de processos. O trabalho busca realizar o emprego de técnicas computacionais de processamento para a extração de informações contidas em PCDT específicos. O objetivo é a construção de um repositório centralizado com estas informações promovendo uma consulta simplificada e intuitiva, e simultaneamente contribuir na promoção da MBE.

## **3. Objetivos**

### **3.1. Objetivos Gerais**

Desenvolver e aplicar técnicas computacionais de reconhecimento de texto na extração de informações contidas dos PCDTs e centralizar em um repositório acessível de maneira simplificada, facilitando o processo de execução da MBE.

### **3.2. Objetivos Específicos**

- Compreender os conceitos de medicina baseada em evidência, sistemas de suporte a tomada de decisão e técnicas de processamento de texto;
- Realizar uma revisão sistemática acerca do estado da arte sobre sistemas de apoio à decisão utilizando diretrizes médicas;
- Selecionar as bases de dados relevantes para o desenvolvimento do projeto;
- Desenvolver código que extraia o conteúdo de interesse, centralize e apresente a consulta a partir dos termos de busca;
- Otimizar o código, avaliar sua acurácia e desenvolver a interface de consulta.

## 4. Revisão Sistemática de Literatura

A revisão de literatura foi realizada de forma sistemática incluindo tópicos relevantes para a compreensão e desenvolvimento do projeto, incluindo os conceitos: medicina baseada em evidência e sistemas de suporte à decisão clínica correlatos a técnicas computacionais de processamento de linguagem natural, preferencialmente aplicados às Diretrizes ou Protocolos Clínicos utilizados.

### 4.1. Estratégia de pesquisa

A busca na literatura foi realizada em março de 2021 e as publicações correspondem ao período delimitado de 2015 a 2021 com a utilização das palavras chave Decision Support System e Guideline-Based, que poderiam ser relatadas pelo autor tanto no título, no *abstract*, como na palavra chave daquela publicação. Foram retiradas da busca qualquer *review* (revisão de literatura, revisão sistemática e demais) pois o objetivo se concretiza na aplicação de sistemas de suporte à tomada de decisão utilizando Diretrizes Médicas (DM).

Além dos elementos gramaticais, foi delimitada na pesquisa apenas as publicações em inglês, com a busca efetivada também através das subáreas do repositório, sendo as escolhidas: Médica, Computacional, Saúde e Enfermagem.

Aplicando ao banco eletrônico de artigos científicos SCOPUS, a estratégia de busca se baseia na aplicação do método booleano a partir da concatenação das seguintes palavras-chaves: ( TITLE- ABS- KEY ( "Decision Support System" ) AND TITLE- ABS- KEY ( "Guideline- Based" ) AND NOT TITLE- ABS- KEY ( review ) ) AND ( LIMIT- TO ( PUBYEAR , 2020 ) OR LIMIT- TO ( PUBYEAR , 2019 ) OR LIMIT- TO ( PUBYEAR , 2018 ) OR LIMIT- TO ( PUBYEAR , 2017 ) OR LIMIT- TO ( PUBYEAR , 2016 ) OR LIMIT- TO ( PUBYEAR , 2015 ) ) AND ( LIMIT- TO ( DOCTYPE , "ar" ) ) AND ( LIMIT- TO ( SUBJAREA , "MEDI" ) OR LIMIT- TO ( SUBJAREA , "COMP" ) OR LIMIT- TO ( SUBJAREA , "HEAL" ) OR LIMIT- TO ( SUBJAREA , "NURS" ) ) AND ( LIMIT- TO ( LANGUAGE , "English" ) ) .

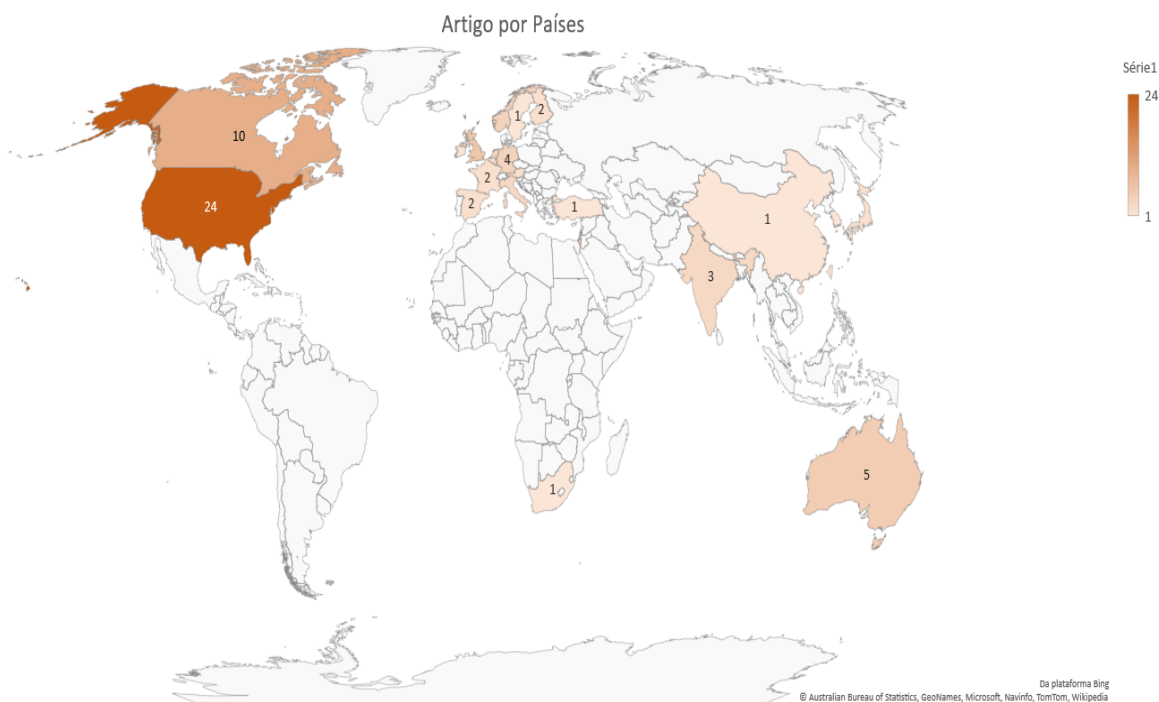
### 4.1.1. Compilação dos dados e revisão

Os artigos foram compilados (Anexo 1) e analisados baseando no nível de correlação com o escopo e de maneira independente. Após o um consenso, houve a extração dos dados relevantes e análise dos artigos que serão apresentados nos resultados de busca.

### 4.1.2. Resultado da busca

Os artigos foram encontrados aplicando a estratégia de pesquisa supracitada. Dos 63 artigos encontrados, o mapa abaixo (Figura 2) apresenta a distribuição desta totalidade nos países do mundo.

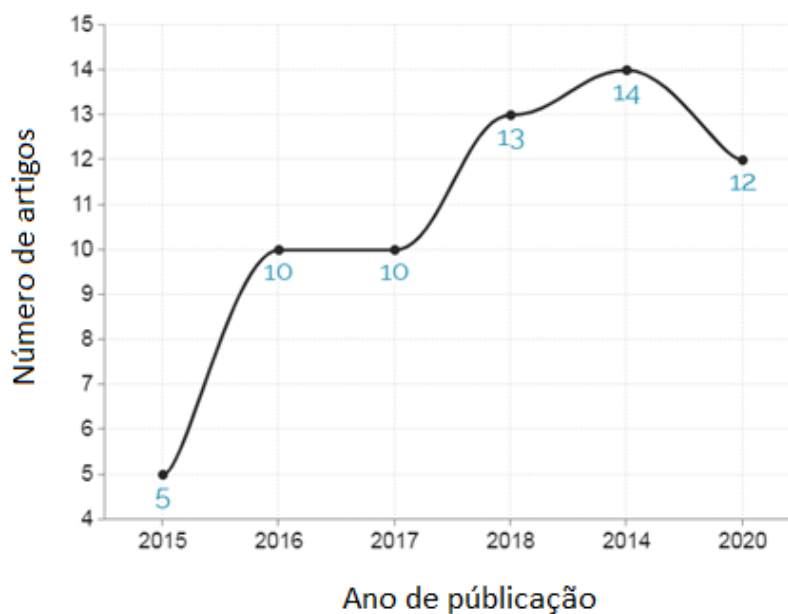
**Figura 2:** Distribuição e quantificação dos artigos encontrados por país.



Fonte: Scopus.

Na Figura 3 apresentamos a distribuição destas publicações nos últimos 5 anos, nota-se que a grande maioria se consolidou a partir do ano de 2017- 2018 e em ascensão, se comparado aos anos de 2015 e 2016.

**Figura 3:** Distribuição dos artigos encontrados por ano de publicação.

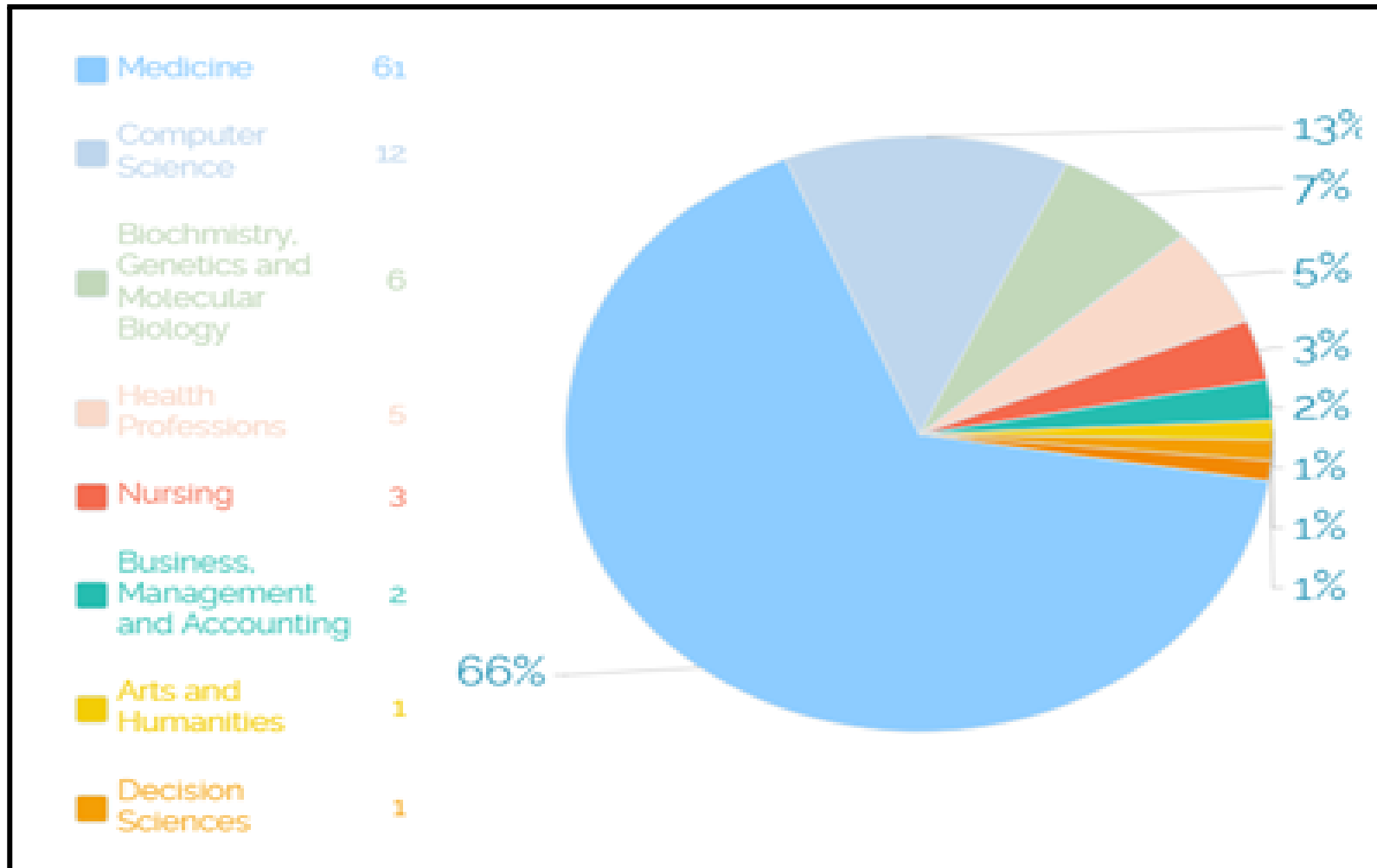


Fonte: Scopus.

Dado a variedade das áreas dos artigos científicos publicados encontrados na Revisão Sistemática, foram selecionados apenas aqueles que o conteúdo apresentava aplicação ou eram relativos aos sistemas de auxílio à tomada de decisões clínicas, tanto como artigos que relacionassem a aplicação de técnicas computacionais para o processamento de protocolos clínicos.

Por fim, a Figura 4 abaixo apresenta a distribuição temática da totalidade dos artigos encontrados, em que pelo menos 60% se referem a temas restritos à medicina.

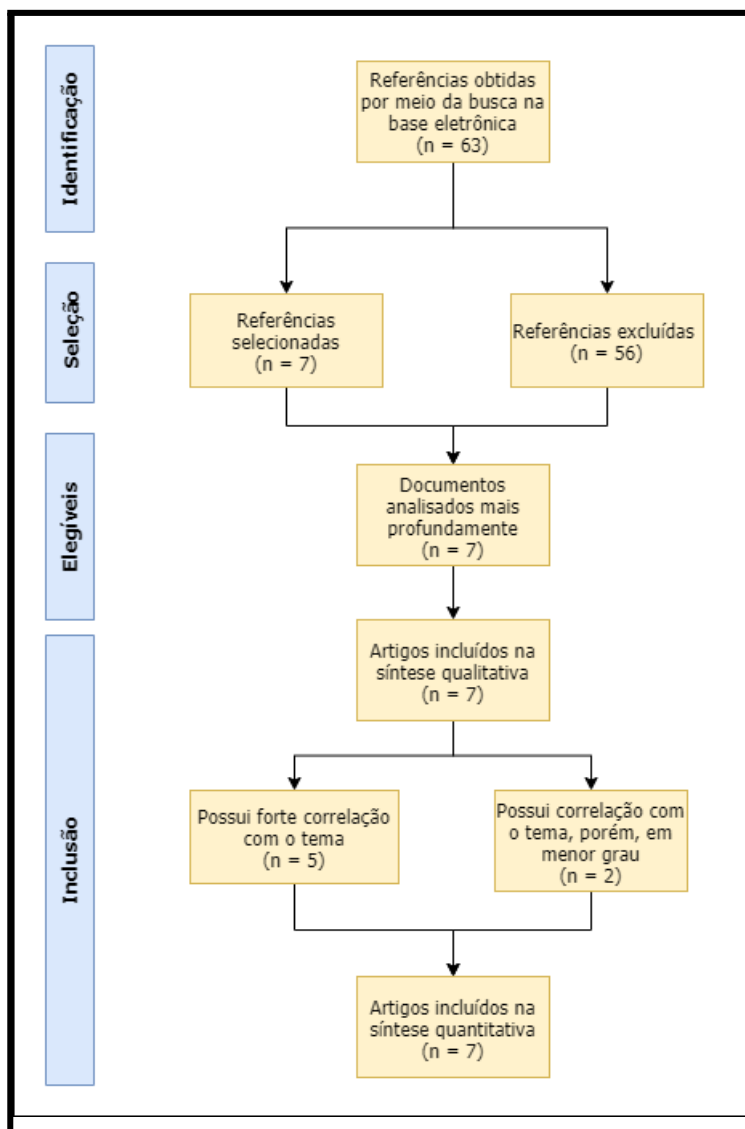
Figura 4: Distribuição de artigos por área obtidos através da Revisão Sistemática.



Fonte: Scopus

A totalidade dos artigos foram avaliados e foram atribuídas notas de 0 a 5, sendo 0 com baixa correlação teórica/técnica e 5 com alta correlação teórica/técnica ao assunto proposto. O total de artigos encontrados foi disposto em um diagrama com as respectivas inclusões e exclusões (Figura 5).

**Figura 5:** Diagrama com os resultados da busca sistemática.



**Fonte:** Autoria Própria.

De modo a analisar e sistematizar os artigos incluídos na síntese quantitativa, adaptamos a metodologia PRISMA, que por definição, corrobora na avaliação de outras revisões sistemáticas, mas também pode ter aplicação na análise e avaliação de publicações e intervenções (GALVÃO, 2015) como colocado na tabela 1.

**Tabela 1:** Análise Sistemática da Revisão de Literatura.

Referência	Ano	Uso de DM	Temática da DM	PLN aplicado a DM	Bases de dados utilizadas	Técnicas utilizadas	Usuário Final	País
Becker, M., Böckmann, B., et al. (2020)	2020	Sim	Diretrizes Alemãs de Câncer Colorretal	Não	Unified Medical Language System (UMLS) - Concept Unique Identifier (CUI), CID 10	Business Process Model and Notation (BPMN)	Médicos	Alemanha
Becker, M., Kasper, S., Böckmann, B., et al. (2019)	2019	Sim	Câncer Colorretal, Câncer de Cólon e Câncer Retal	Não	Unified Medical Language System (UMLS), CID 10	Aprendizado supervisionado ou aprendizado a partir de exemplos - Weka e o Naïve Bayes e códigos SVN	Médicos	Alemanha
Seitinger, A., Rappelsberger, A., et al. 2016	2016	Sim	Dermatologia	Não	Arden Syntax	Representational State Transfer (REST), Simple Object Access Protocol (SOAP), Client Uniform Resource Locator (cURL) e PHP scripting language	Médicos	NI
Lanzola, G., Bossi, P., et al. (2017)	2017	Sim	NI	Não	SNOMED CT e ICD9-CM	Alium - PROforma	Paciente /Médico	Alemanha
Maghsoud-Lou, E., Christie, S., et al. (2017)	2017	Sim	Neurologia	Não	NA	Spinal Problem E-Referral (SPER) system - Spinal Condition Consultation Protocol (SCCP)	Paciente /Médico/ Gestor	Canadá
Ahamed, T., Lederman, R., Bosua, R., et al. (2016)	2016	Sim	AVC, Abstinência alcoólica e Cetoacidose Diabética	Não	NI	Delphi Study	Enfermeiros	Austrália
Ennis, J., Gillen, D., Rubenstein, et al. (2015)	2015	Sim	Diretrizes sobre complicações de funções renais	NI	NA	NI	Médicos	EUA

**\*NA:** Não se aplica; **\*\*NI:** Não informado; **PLN:** Processamento de Linguagem Natural; **DM:** Diretriz Médica; **AVC:** Acidente Vascular Cerebral; **UMLS:** Unified Medical Language System; **CUI:** Concept Unique Identifier; **CID:** Classificação Internacional de Doenças; **SNOMED-CT:** Systematized Nomenclature of Medicine - Clinical Terms; **BPMN:** Business Process Model and Notation; **CURL:** Client Uniform Resource Locator; **SPER:** Spinal Problem E-Referral; **SOAP:** Simple Object Access Protocol; **SCCP:** Spinal Condition Consultation Protocol; **EUA:** Estados Unidos da América.

**Fonte:** Autoria Própria.

Vale ressaltar que, de todas as publicações selecionadas com o score 5 possuem a utilização de Diretrizes Médicas, cada uma com uma temática clínica de interesse, mas que nenhuma aplica técnicas de PLN para corroborar no entendimento destes documentos propriamente.

Analisando as bases de dados que estas publicações citaram, 43% utilizam algum repositório de termos médicos (sejam eles UMLS, CUI ou SNOMED CT) e a Classificação Internacional de Doenças (seja a versão do CID- 9 ou CID- 10).

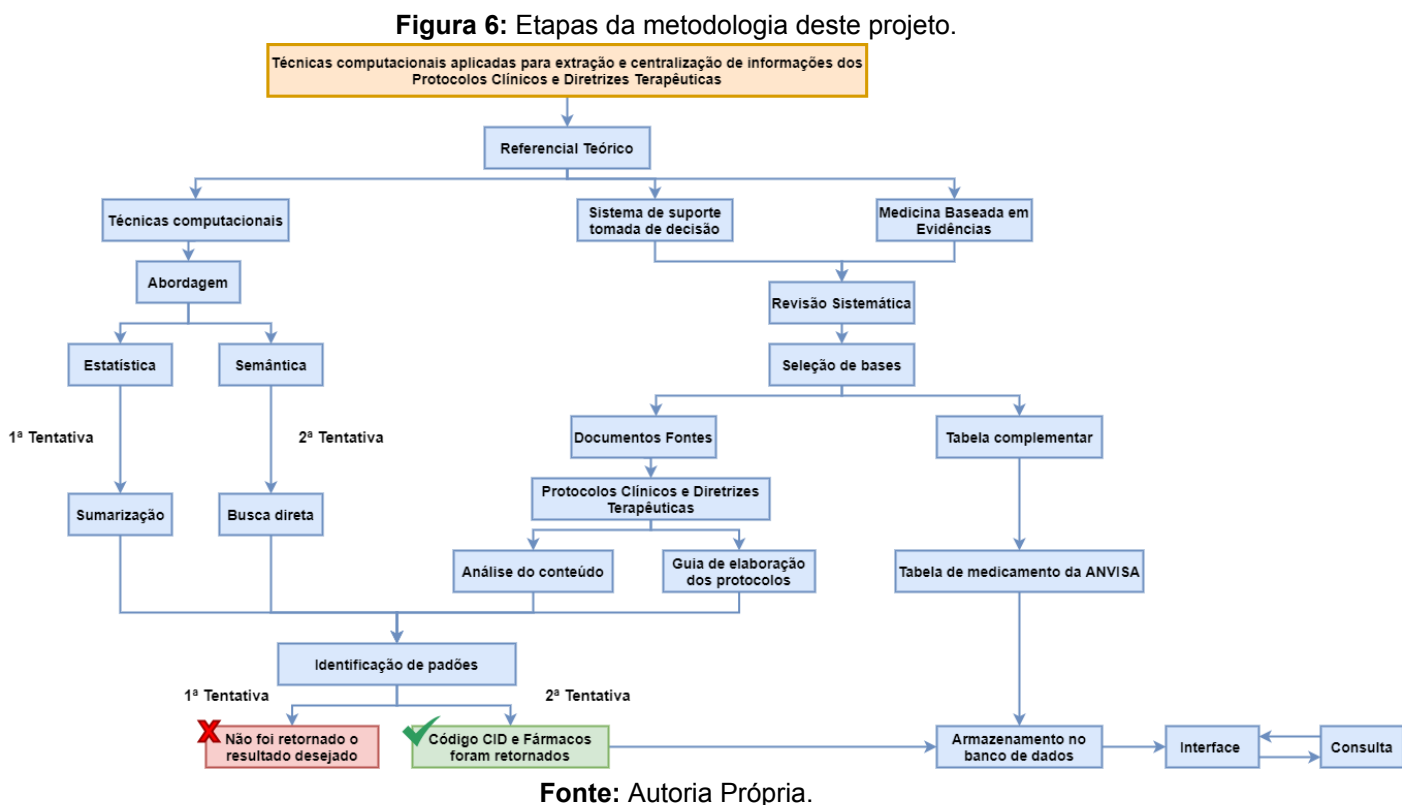
Agora, analisando o campo Usuário Final – a que público se direciona o desenvolvimento proposto –, 28% declaram que será direcionado a pacientes e os que direcionam apenas para os Médicos/Enfermeiros corresponde a 72%. Dentre os arquivos analisados, apenas 14% declara que a aplicação também se adequa a profissionais que atuam na Gestão Hospitalar.

## 5. Metodologia

De modo a caracterizar a Metodologia proposta, foram definidas as seguintes variáveis de desfecho: Informações a serem extraídas dos PCDTs (Doença, Ano, CIDs, Fármacos, Número de Páginas), Armazenamento (em Banco de Dados) e a Consulta (a partir da Doença, do CID ou Fármaco).

Por tanto, o desenvolvimento deste projeto obedece a Figura 6 a seguir, que sintetiza as etapas desde o processo de:

- 1– Seleção das Base de Dados;
- 2– Extração do conteúdo do PCDT;
- 3– Armazenamento no Banco de Dados (BD);
- 4– Consulta no BD;
- 5– Correlação com demais bases previamente selecionadas;
- 6– Interface da Consulta;
- 7– Análise do Resultado da Consulta.



O desenvolvimento do código foi a partir da linguagem de programação *Python 3.9* com o *Visual Studio* como *IDE*, como banco de dados para armazenamento foi utilizado o *MySQL* e a biblioteca *Streamlit* para o desenvolvimento da interface.

### **5.1. Seleção da base de dados**

A primeira etapa consiste na seleção da base de dados e para isso, foi segmentado da seguinte forma:

**a) Documentos Fonte:** PCDTs públicos (disponibilizados pela CONITEC), em formato pdf e realizados de acordo com o Guia de Elaboração dos Protocolos (MINISTÉRIO DA SAÚDE, 2019) e com até 6 anos de publicação;

**b) Referência Medicamentos:** Tabela pública, disponibilizada pela ANVISA, denominada de “Preço Máximo ao Consumidor”. Foi editada, restando apenas os seguintes campos de interesse:

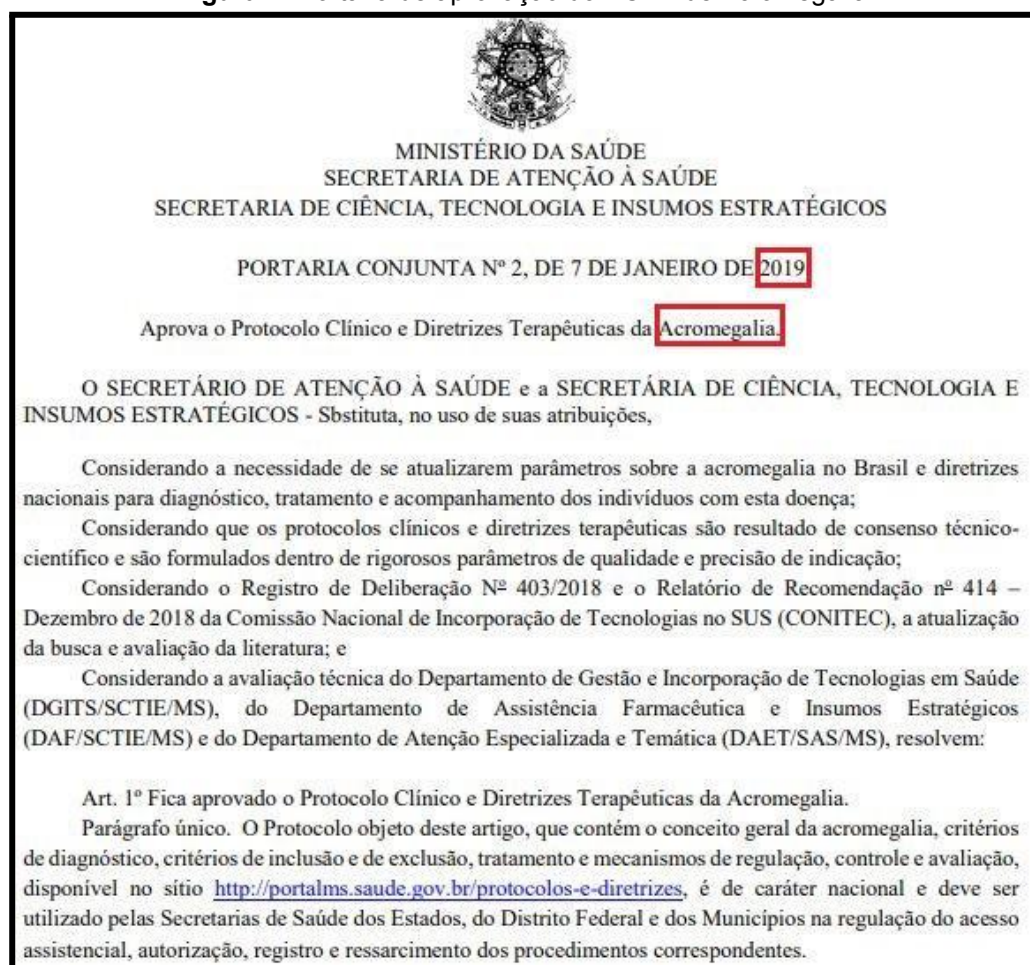
- Registro ANVISA;
- Princípio Ativo;
- Laboratório;
- Produto;
- Apresentação;
- Classe terapêutica;

Os arquivos supracitados foram armazenados localmente e, nos tópicos adiante, serão descritos sua utilização no programa em desenvolvimento.

#### **5.1.1. Caracterização das bases de dados**

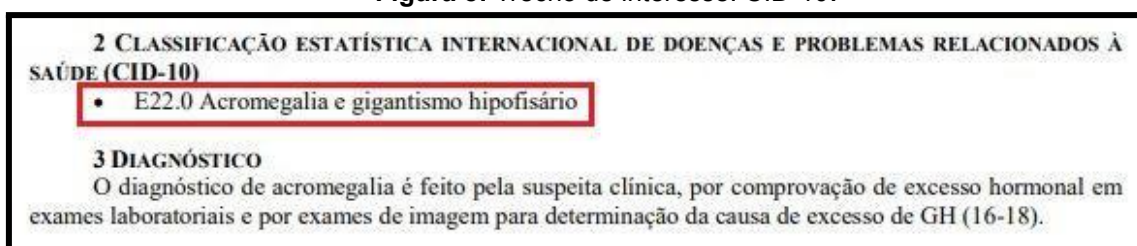
Analisando a caracterização dos Documentos Fonte, estes possuem diversos tópicos com informações relevantes para a assistência em saúde. A figura 7 demonstra a temática referente ao PCDT e o ano. Na figura 8 é evidenciado o trecho de interesse referente ao CID e na figura 9 o trecho de interesse referente ao fármaco.

**Figura 7:** Portaria de aprovação do PCDT de Acromegalia.



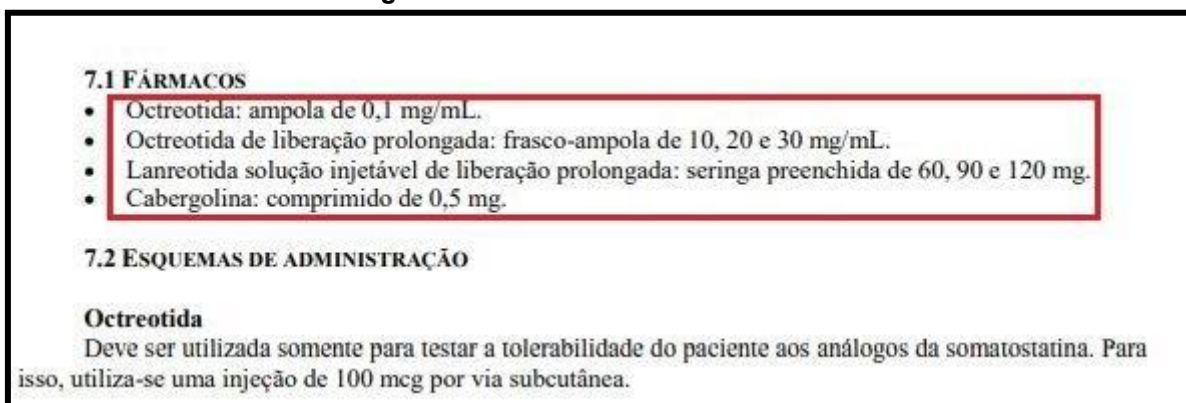
**Fonte:** PCDT\_Acromegalia\_2019

**Figura 8:** Trecho de interesse: CID-10.



**Fonte:** PCDT\_Acromegalia\_2019

**Figura 9:** Trecho de interesse: Fármacos.



**Fonte:** PCDT\_Acromegalia\_2019

Foram extraídos da CONITEC (às 16:55 do dia 26/04/2021) 112 documentos fonte que atendiam integralmente ao primeiro e segundo critérios, anteriormente colocados.

Analisando o segundo critério de elegibilidade dos documentos fonte: a concordância com o Guia de Elaboração dos Protocolos Clínicos (MINISTÉRIO DA SAÚDE, 2019), no conteúdo interno deverá conter dentre outras informações, os seguintes tópicos de interesse para extração de informação:

- Classificação Estatística Internacional de Doenças e Problemas Relacionadas à Saúde (CID– 10): *“foi conceituada para padronizar e catalogar as doenças e problemas relacionados à saúde, tendo como referência a Nomenclatura Internacional de Doenças, estabelecida pela Organização Mundial de Saúde”*. (DATASUS, 2008).
- Tratamento Farmacológico: *“mencionar os medicamentos que serão abordados pelo PCDT utilizando o nome da substância ativa de acordo com a Denominação Comum Brasileira (DCB).”* (MINISTÉRIO DA SAÚDE, 2019).

No entanto, os capítulos e categorias do CID-10 não costumam ser abordados diretamente nos PCDTs, sendo apenas utilizados para consultar informações armazenadas no banco, que será escopo de desenvolvimento da interface de consulta.

Caracterizando a base de dados Referência de Medicamentos, foi selecionada a versão datada do mês de janeiro de 2021, disponibilizada e baixada através do site da ANVISA (ANVISA, 2021). Após o download, as colunas foram editadas para conter apenas os registros de interesse, como demonstra a Tabela 2.

**Tabela 2:** Recorte da tabela Registros ANVISA de Janeiro de 2021.

	A	AL	AM	AN	AO	AP
1	REGISTRO ANVISA	Princípio Ativo (1/21)	Laboratório (1/21)	Produto (1/21)	Apresentação (1/21)	Classe Terapêutica (1/21)
2	1002001390017	DIENOGESTE;VALERATO DE ESTRADIOL	SCHERING DO BRASIL QUÍMICA E	CHARMELIE	COM REV A + COM REV B + COM REV C + CO	G3A4 - PREPARAÇÕES CONTRACEPTIVAS TRIFÁSICAS
3	1002001390025	DIENOGESTE;VALERATO DE ESTRADIOL	SCHERING DO BRASIL QUÍMICA E	CHARMELIE	COM REV A + COM REV B + COM REV C + CO	G3A4 - PREPARAÇÕES CONTRACEPTIVAS TRIFÁSICAS
4	1002900020141	HIDROCLOROTIAZIDA;MALEATO DE ENA	MERCK SHARP & DOHME FARMAI	CO-RENITEC	20 MG + 12,5 MG COM CT BL AL/AL X 30	C9B1 - INIBIDORES DA ECA ASSOCIADOS A ANTI-HIF
5	1002900020158	HIDROCLOROTIAZIDA;MALEATO DE ENA	MERCK SHARP & DOHME FARMAI	CO-RENITEC	10 MG + 25 MG COM CT BL AL/AL X 30	C9B1 - INIBIDORES DA ECA ASSOCIADOS A ANTI-HIF
6	1002900050014	MONTELUCASTE	MERCK SHARP & DOHME FARMAI	SINGULAIR	5 MG COM MAST CT 3 BL AL PLAS INC X 10	R3J2 - ANTIASMÁTICOS/DPOC ANTILEUCOTRIENOS
7	1002900050022	MONTELUCASTE	MERCK SHARP & DOHME FARMAI	SINGULAIR	10 MG COM REV CT 3 BL AL PLAS INC X 10	R3J2 - ANTIASMÁTICOS/DPOC ANTILEUCOTRIENOS
8	1002900050057	MONTELUCASTE	MERCK SHARP & DOHME FARMAI	SINGULAIR	4 MG COM MAST CT BL AL PLAS INC X 10	R3J2 - ANTIASMÁTICOS/DPOC ANTILEUCOTRIENOS
9	1002900050065	MONTELUCASTE	MERCK SHARP & DOHME FARMAI	SINGULAIR	4 MG COM MAST CT 3 BL AL PLAS INC X 10	R3J2 - ANTIASMÁTICOS/DPOC ANTILEUCOTRIENOS
10	1002900050073	MONTELUCASTE	MERCK SHARP & DOHME FARMAI	SINGULAIR	5 MG COM MAST CT BL AL PLAS INC X 10	R3J2 - ANTIASMÁTICOS/DPOC ANTILEUCOTRIENOS
11	1002900050081	MONTELUCASTE	MERCK SHARP & DOHME FARMAI	SINGULAIR	10 MG COM REV CT BL AL PLAS INC X 10	R3J2 - ANTIASMÁTICOS/DPOC ANTILEUCOTRIENOS
12	1002900050111	MONTELUCASTE	MERCK SHARP & DOHME FARMAI	SINGULAIR	4 MG GRAN SOL CT 30 ENV AL/PLAS X 500 M	R3J2 - ANTIASMÁTICOS/DPOC ANTILEUCOTRIENOS
13	1002900050121	MONTELUCASTE	MERCK SHARP & DOHME FARMAI	SINGULAIR	4 MG GRAN SOL CT 10 ENV AL/PLAS X 500 M	R3J2 - ANTIASMÁTICOS/DPOC ANTILEUCOTRIENOS
14	1002900070066	LOSARTANA POTÁSSICA	MERCK SHARP & DOHME FARMAI	COZAAR	100 MG COM REV CT BL AL PVC/PE/PLVDC BC	C9C - ANTAGONISTAS DA ANGIOTENSINA II PUROS

Fonte: Autoria Própria.

## 5.2. Extração do texto no PCDT

O processo de extração dos trechos de interesse dos PCDT teve duas tentativas antes da definição da estratégia. As tentativas referentes foram descritas a seguir.

### 1ª Tentativa: Sumarização

Em um momento inicial a estratégia era de gerar como resultado um texto simplificado que promovesse um resumo do conteúdo contido no PCDT, do qual deveria conter o CID-10 e os fármacos. As etapas executadas foram o pré-processamento, cálculo da frequência das palavras, frequência proporcional dos termos, tokenização das sentenças, nota para cada uma das sentenças, ordenação das que apresentaram maior nota e construção do resumo.

O texto foi pré-processado de modo a conversão de todos os caracteres em letras minúsculas, remoção caracteres especiais, pontuações e *stopwords* (palavras presentes no texto, mas que não geram prejuízo na compreensão do texto).

O segundo passo foi o cálculo da frequência dos termos no texto. Após o cálculo da frequência, foram atribuídos pesos para a frequência das palavras de modo a estabelecer seu respectivo nível de relevância. O cálculo da nota a ser

atribuída corresponde à divisão do valor da frequência do termo que está analisando pelo valor do termo que aparece com maior frequência. Quanto maior o valor obtido como nota, maior a importância do termo.

A tokenização se estabelece pela separação do texto em sentenças. A separação em sentenças facilita a atribuição de notas para cada uma das sentenças de acordo com o peso a que foi atribuído para cada termo.

A partir das notas atribuídas para cada sentença, as sentenças são ordenadas de acordo com o seu nível de importância. Por fim, a construção do resumo é realizada retomando as *stopwords*, a disposição do texto em maiúsculas e minúsculas e a estrutura original, mas, contendo apenas as sentenças com maior peso.

## **2ª Tentativa: Busca direta**

O processo de extração por meio da busca direta consistiu na aplicação dos subtítulos como índices para a busca e identificação dos trechos de interesse. O arquivo é lido e com identificação das páginas através da biblioteca *PyPDF2*. O trecho é convertido com a biblioteca *Unicode*, da qual promove um texto sem acentuação ou caracteres especiais, facilitando o processo de comparação de termos.

O método *find()* é utilizado realizando a busca dos índices e retornando o valor correspondente a posição do termo no texto. Através da identificação de dois pontos, um onde inicia um subtítulo e o outro ponto equivalente ao próximo subtítulo, promove a localização do trecho de interesse de extração. No caso, os índices aplicados até o momento foram: “CID-10”, “DIAGNOSTICO”, “FARMACO” e “ESQUEMAS”.

Foram observadas algumas variações em alguns dos subtítulos como a pluralização de alguns termos ou a supressão de alguns caracteres. O ajuste do código para abranger estas variações permitiu um bom funcionamento do processo de extração.

Após a extração, é realizada uma análise manual do conteúdo extraído antes do armazenamento do conteúdo no banco de dados, com o intuito de verificar se o conteúdo esperado foi extraído.

### 5.3. Armazenamento no Banco de Dados

Após a validação do conteúdo extraído, este é armazenado em um campo do tipo *TEXT* em um banco MySQL. Para isso, foi utilizada a biblioteca *mysql.connector* que permite a alimentação do banco de dados a partir de um programa em Python.

O banco de dados foi constituído de diversas tabelas, mas principalmente a tabela “PCDT” que foi alimentada com os conteúdos extraídos dos protocolos, teve a seguinte estrutura:

- ID – primary key, not null, auto increment;
- Nome\_Arquivo – VARCHAR ();
- Doença – VARCHAR()
- Ano – YEAR ();
- CID – TEXT ();
- Fármaco – TEXT ();
- Duração - time
- Páginas - Int

As demais fontes de dados foram correlatas ao processamento e associadas à consulta do conteúdo armazenado do PCDT, que gera o resultado final.

### 5.4. Consulta no Banco de Dados

O processo de consulta ao conteúdo armazenado se estabelece por meio da interface da tela de “Consulta”. O processo para a plotagem dos resultados na tela se dá pelo emprego de uma query aplicada aos campos em que possam conter o conteúdo de busca, o *FULLTEXT* é um mecanismo que possibilita a ordenação dos campos do tipo *TEXT* por similaridade semântica, facilitando o processo de busca, a função *MATCH()* permite determinar as colunas em que foram realizadas as consultas e *AGAINST()* promove qual o termo consultado e o formato que deve ser

apresentado para definir o nível de relevância, o retorno é armazenado em um campo denominado *score* e a busca só retornará resultados que possuam *score* acima de 0, ou seja, que possuam relevância com o termo de busca.

### **5.5. Associação com demais fontes de dados**

A associação das informações (extraídas dos PCDTs e armazenadas no banco de dados) com as demais fontes de dados obedecem ao tópico 5.1.1 Caracterização da Base de Dados, de modo a conter apenas os tópicos relevantes a serem associados.

A tabela de medicamentos, a correlação ocorreu através do termo de busca da tabela principal com o campo Princípio Ativo da Referência de Medicamentos. As correlações efetuadas compuseram tabelas auxiliares permitindo a visualização da informação a partir da consulta ao banco.

### **5.6. Interface de Consulta**

A interface de consulta foi desenvolvida buscando auxiliar na visualização das informações retornadas no processo de consulta em função do termo de busca. A estrutura foi feita aplicando a biblioteca *Streamlit*, a página é composta por um menu do qual é possível selecionar as páginas que se deseja acessar, sendo elas, “*Home*”, “*About Us*” e “*Consulta*”. A *Home* possui um descritivo geral acerca do projeto, em *About Us* é exibido os membros da equipe e uma breve descrição e na página “*Consulta*” há um campo de busca sendo retornado o resultado da consulta. Os métodos aplicados para escrita da página foram `title()` para dar títulos às páginas, `header()` para os subtítulos e `write()` para os textos da página; o `selectbox()` permite selecionar a página e o condicional determina qual o conteúdo será exibido conforme a página selecionada conteúdo selecionado.

## **5.7. Avaliação estatística de desempenho e acuracidade**

Para avaliar a estatística do desempenho e a acuracidade do sistema desenvolvido, foram estabelecidas duas perguntas nortearam a análise, são elas:

- 1) O sistema realiza a extração e armazenamento preciso das informações pré especificadas?
- 2) O desempenho da extração e armazenamento pelo sistema é, comparativamente, superior aos outros grupos?

De modo a responder estatisticamente estas perguntas, a metodologia a seguir, busca comparar o resultado obtido através de diferentes formatos de extração e armazenamento das mesmas informações, para isso consiste:

- a) Seleção da amostra: um conjunto de 20 PCDTs analisados;
- b) Definição dos grupos de pesquisa e distribuição das amostras, respectivamente:
  - i) Grupo 1: Sistema - código desenvolvido, totalidade da amostra: 20 PCDTs;
  - ii) Grupo 2: Voluntários - 5 voluntários aleatórios, anonimizados e sem experiência neste processo - cada um com 4 PCDTs;
  - iii) Grupo 3: Autores - integrantes do grupo e desenvolvedores do projeto - 20 PCDTs.
  - iv) Grupo 4: Padrão Ouro - consenso dos autores sobre o que está correto para cada item de verificação, amostra total: 20 PCDTs.
- c) Realizar a coleta de dados dos PCDTs em cada grupo (mensurando os tempos respectivos), a partir dos seguintes itens de verificação:
  - i) a Doença;
  - ii) o Ano de edição;
  - iii) a lista de CIDs relacionados;
  - iv) a lista de Fármacos relacionados;
  - v) o Número de Páginas.

Após as etapas citadas, vale explicitar os diferentes formatos de armazenamento: o G1 realizou diretamente no BD e tanto G2 como G3 tiveram as coletas de dados armazenadas em tabelas excel, seus tempos preenchidos manualmente e posteriormente as respectivas pontuações, de acordo com os formatos em anexo. (Anexo 3 - Análise Estatística: Modelo para preenchimento Grupo 2 e Anexo 4 - Análise Estatística: Modelo para preenchimento Grupo 3).

Todos os grupos de coleta de dados (G1, G2 e G3) serão comparados com os conteúdos registrados através do Grupo 4: Padrão Ouro para a adequada pontuação.

A valorização do acerto será dada para cada um dos itens de verificação, a partir do consenso estabelecido, sendo: 2- Correto; 1- Parcialmente Correto; e 0- Incorreto. Dessa forma, o acerto completo das informações daquele PCDT corresponderia a pontuação 10 e, se caso, as informações fossem totalmente incorretas teriam pontuação 0.

Posteriormente, foram comparados estatisticamente os acertos e os tempos dos três primeiros grupos para realizar a extração/armazenamento de cada documento. Para isso, foi preenchida a tabela abaixo, representada pela Tabela 3.

**Tabela 3:** Recorte da tabela que consolida os resultados de todas as amostras por grupo.

Grupo/PCDT	pontuacao_total	tempo_total (s)
Voluntario/1		
Voluntario/2		
Voluntario/3		
Voluntario/4		
Voluntario/5		

**Fonte:** Autoria Própria.

Após reunir e pontuar todos os dados obtidos por cada grupo na análise dos 20 PCDTs, foi determinado o valor ponderado de cada registro que constitui na divisão entre o tempo total e a pontuação obtida. De modo a hierarquizar os resultados, os dados ponderados foram divididos pelo maior elemento ponderado, consolidando cada amostra em um índice de 0 a 1.

Dessa forma, os índices foram representados pela média e desvio padrão. Após análise de normalidade e igualdade de variância pelos testes de Shapiro Wilk e Levene respectivamente foi realizado análise de ANOVA para comparação entre os três grupos, sendo posteriormente realizado a análise Post Hoc para os resultados significativos. Adotamos neste estudo um nível de significância de  $p < 0,05$  sendo todos os testes realizados pelo software JASP V.14.01.

## **6. Resultados e Discussão**

### **6.1. Seleção da base de dados**

Após o processo de *download* cada um dos arquivos foi analisado quanto a sua estrutura interna, verificando variações semânticas que os subtítulos poderiam apresentar.

Em meio a seleção, foram descartados todos aqueles que não possuíam os subtítulos que são aplicados como índices (“CID-10”, “DIAGNOSTICO”, “FARMACO” e “ESQUEMAS”), dada a impossibilidade de identificação do trecho de interesse de extração. Também foram descartados aqueles que possuíam estruturas internas diferentes dos documentos mais recentes, possuindo por exemplo informações no rodapé, pois ainda não foi implementada a função que identifica e realiza a limpeza dos mesmos.

Após a seleção dos documentos adequados os mesmos foram separados de acordo com a sua especialidade médica e seu respectivo ano, para que fossem renomeados de maneira padronizada para o seu armazenamento no banco de dados.

O número de páginas de cada documento também foi levantado, de modo a permitir análises posteriores relativas ao tempo que se levaria para o processamento de cada arquivo. Mesmo se tratando de informações de extração que estão geralmente nas partes iniciais, o número de páginas foi levantado para avaliação.

### **6.2. Extração do texto no PCDT**

Os resultados referentes ao processo de extração dos trechos de interesse dos PCDTs, como mencionado na Metodologia, foram necessárias algumas tentativas antes da definição da estratégia aplicada no momento. Os resultados obtidos estão descritos a seguir.

### **1ª Tentativa: Sumarização**

A sumarização foi aplicada para o Protocolo de Tabagismo nas páginas em que continham o CID-10 e os Fármacos.

Os resultados obtidos e evidenciados na Figura 10 não corresponderam às expectativas, pois era de interesse obter como resultado um texto simplificado do qual contivesse principalmente o código CID. No entanto, dado que a relevância é estabelecida pela frequência do termo no documento e o, analisando apenas o código CID-10, este não se repete no texto, logo não é tido como relevante e, portanto, não é retornado.

Conclui-se que, como a extração das informações relativas ao CID não foram executadas, impossibilitando a associação com a base do CID-10, motivou o desenvolvimento a partir da aplicação de outras técnicas.

**Figura 10:** Técnica de sumarização aplicado à extração do CID-10.

```
[nltk_data] Downloading package punkt to
[nltk_data]   C:\Users\user\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]   C:\Users\user\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
['apresentado, no ano anterior, pelo menos três dos critérios a seguir: • Desejo forte e compulsivo para consumir a substância (fissura ou craving); • dificuldade para controlar o uso (início, término e níveis de consumo); • estado de abstinência fisiológica diante da suspensão ou redução, caracterizado por síndrome de abstinência e consumo da mesma substância ou similar, com a intenção de aliviar ou evitar sintomas de abstinência (reforço negativo); • evidência de tolerância, ou seja, necessidade de doses crescentes da substância para obter os efeitos produzidos anteriormente com doses menores; • abandono progressivo de outros prazeres em detrimento do uso de substâncias psicoativas; • aumento do tempo empregado para conseguir ou consumir a substância ou recuperar-se de seus efeitos; • persistência no uso apesar das evidentes consequências, como câncer pelo uso do tabaco, humor deprimido ou perturbações das funções cognitivas relacionada com a substância.', 'O diagnóstico de dependência da nicotina é clínico e deve ser feito por meio de avaliação procedida pelo profissional da saúde que atende o usuário de tabaco, seguindo os critérios diagnósticos descritos acima.', 'Composto de seis perguntas, com escores que variam de zero a dez, ele indica que o tabagista poderá experimentar sintomas desconfortáveis da síndrome de abstinência, quando o resultado encontrado for acima de 6 pontos (26).']
Press any key to continue . . .
```

**Fonte:** Autoria Própria.

## **2ª Tentativa: Busca direta**

Já a aplicação da busca direta demonstrou bons resultados permitindo a extração do conteúdo de interesse, sendo apenas necessário o ajuste do código para incluir as variações dos subtítulos.

Nesta primeira versão, o processo de extração foi subdividido em 7 etapas para facilitar a compreensão do código.

A Etapa 1 (Figura 11) corresponde ao processo de entrada dos dados, a primeira informação de entrada é do nome do arquivo que se deseja extrair, todos os documentos foram previamente parametrizados com o tema do PCDT e o respectivo ano de publicação, segue exemplo “PCDT\_Tema\_Ano”.

A Etapa 2 (Figura 11) se refere à segmentação do nome do arquivo para a definição do ano do PCDT.

Na etapa 3 (Figura 11) o documento é identificado e aberto por meio da biblioteca PyPDF2. A Etapa 4 (Figura 11), o número de páginas do documento é definido para ser aplicado como parâmetros do *range* da aplicação.

Na Etapa 5 (Figura 12) a normalização dos dados com a remoção das quebras de linha e acentuações.

A Etapa 6 (Figura 13) se estabelece pelo início do processo de extração, em que são identificadas as posições dos índices CID-10, Diagnóstico, Fármacos e Esquemas. A posição de dois índices é retornada, um anterior ao trecho a que se deseja extrair e outro posterior, a partir da posição é possível selecionar o intervalo e extrair apenas o trecho de interesse. A duração do processo de extração é verificada nesta etapa sendo retornada em segundos.

Anterior a Etapa 7 (Figura 13) o conteúdo extraído é analisado manualmente para a identificação de falhas no processo de extração, tais como conteúdo incompleto ou vazio. E por fim, na Etapa 7 o processo é iniciado verificando se a conexão com o banco foi estabelecida, seguida do armazenamento do conteúdo e encerramento do processo.

Figura 11: Etapas de extração 1 a 4: ano e número de páginas.

# Extração do conteúdo do PCDT

## Etapa 1: Entrada do nome do arquivo pelo usuário

Digite o nome do arquivo a que se deseja extrair:

PCDT\_Acromegalia\_2019

Buscar

## Etapa 2: Identificação do ano do arquivo

O nome do arquivo é separado em três partes para a obtenção do ano do arquivo

**Nome do arquivo:** PCDT\_Acromegalia\_2019

**Resultado da separação:**

```
▼ [
  0 : "PCDT"
  1 : "Acromegalia"
  2 : "2019"
]
```

**Ano do arquivo:** 2019

**Doença:** Acromegalia

## Etapa 3: Identificação e leitura do documento

O arquivo é identificado na pasta local e é realizado a sua abertura utilizando a biblioteca PYPDF2

## Etapa 4: Definição do número de páginas

**Número de páginas:** 32

Fonte: Autoria Própria.

**Figura 12:** Etapa de extração 5: normalização do conteúdo.

## **Etapa 5: Normalização do conteúdo**

É realizado a remoção dos quebras de linhas e acentuações

**Conteúdo sem alteração:**

ANEXO

P ROTOCOLO C LÍNICO E D IRETRIZES T E

**Remoção de quebras de linha:**

ANEXO

P ROTOCOLO C LÍNICO E D IRETRIZES T E

**Remoção de acentuações:**

ANEXO

P ROTOCOLO C LÍNICO E D IRETRIZES T E

**Fonte:** Autoria Própria.

**Figura 13:** Etapas de extração 6 e 7: extração do CID/fármacos/tempo e armazenamento.

### **Etapa 6: Início do processo de extração**

As páginas são lidas até que sejam identificados os índices 'CID-10', 'Diagnostico', 'Farmacos' e 'Esquemas'. Com identificação da posição de dois índices, um anterior e outro posterior ao conteúdo:

**Posição do 'CID-10' no arquivo:** 3062

**Posição do 'Diagnostico' no arquivo:** 3118

**O conteúdo do CID-10 extraído:**

E22.0 Acromegalia e gigantismo hipofisario 3

**Posição do 'Farmaco' no arquivo:** 4

**Posição do 'Esquema' no arquivo:** 257

**O conteúdo do Farmaco extraído:**

Octreotida: ampola de 0,1 mg/mL. Octreotida de liberacao prolongada: frasco-ampola de 10, 20 e 30 mg/mL. Lanreotida solucao injetavel de liberacao prolongada: seringa preenchida de 60, 90 e 120 mg. Cabergolina: comprimido de 0,5 mg. 7.2

**Duração da extração:** 0.507262s

### **Etapa 7: Conexão com o banco de dados**

Conexao estabelecida com sucesso

Conteúdo armazenado com sucesso

Conexao encerrada

**Fonte:** Autoria Própria.

No tópico a seguir estão explicitados o conteúdo que foi extraído e armazenado no banco de dados a partir da aplicação do método de busca direta.

### 6.3. Armazenamento no Banco de Dados

O conteúdo extraído é analisado manualmente anterior ao processo de armazenamento no banco de dados, com o intuito de garantir que os campos serão preenchidos com o conteúdo adequado. Em caso de não estarem de acordo, os arquivos são separados para execução de testes unitários.

A Figura 14 a seguir demonstra os PCDT e os conteúdos que são armazenados na tabela PCDT.

Figura 14: Recorte da tabela PCDT com documentos extraídos e armazenados no BD.

id	nome_arquivo	doenca	ano	CID	farmaco
1	PCDT_Acromegalia_2019	Acromegalia	2019	E22.0 Acromegalia e gigantismo hipofisario 3	Octreotida: ampola de 0,1 mg/mL.
2	PCDT_Aнемia Hemolítica Autoimune_2018	Anemia Hemolítica Autoimune	2018	-D59.0 Anemia hemolítica autoimune induzida por droga - D59....	- Prednisona: comprimidos de 5 mg
3	PCDT_Aнемia_2017	Anemia	2017	- N 18.0 Doença renal em estadio final - N18.8 Outra insuficienci...	- Sulfato ferroso: comprimidos de 40
4	PCDT_Artrite Psoriaca_2021	Artrite Psoriaca	2021	- M070 Artropatia psoriasica interfalangiana distal - M072 Esp...	- Ibuprofeno: comprimidos revestid
5	PCDT_Atrofia Muscular Espinhal_2019	Atrofia Muscular Espinhal	2019	G12.0 - Atrofia muscular espinal infantil tipo I (Werdnig-Hoffman...	- Nusinersena: 2,4 mg/mL solucao in
6	PCDT_Autismo_2016	Autismo	2016	Fazem parte dos TEA, os seguintes diagnosticos [1,2]: - F84.0...	- Risperidona: solucao oral de 1 mg/l
7	PCDT_Colangite Biliar_2019	Colangite Biliar	2019	K74.3 Cirrose biliar primaria (Colangite destrutiva nao-supurativ...	- Acido ursodesoxicolico: comprimid
8	PCDT_Deficiencia Biotinidase_2018	Deficiencia Biotinidase	2018	- E88.9 Disturbio metabolico nao especificado	- Biotina: capsulas de 2,5 mg. As caç
9	PCDT_Diabete Insipido_2018	Diabete Insipido	2018	E23.2 Diabete insipido 3.	- Desmopressina: 0,1 mg/mL (100 mc
10	PCDT_Doença Falciforme_2018	Doença Falciforme	2018	- D57.0 Anemia falciforme com crise - D57.1 Anemia falciforme ...	- Hidroxiureia: capsulas de 500 mg;
11	PCDT_Doencade Wilson_2018	Doencade Wilson	2018	- E83.0 Disturbios do metabolismo do cobre. 3	- Penicilamina: capsula de 250 mg. -
12	PCDT_Espondilite Ancilosante_2018	Espondilite Ancilosante	2018	- M45 Espondilite ancilosante - M46.8 Outras espondilopatias infi...	- Ibuprofeno: comprimidos revestid
13	PCDT_Glaucoma_2018	Glaucoma	2018	H 40.1 Glaucoma primario de angulo aberto; H 40.2 Glaucoma ...	- Timolol: solucao oftalmica a 0,5%.

Fonte: Autoria Própria.

O total de PCDT testados foram 39, sendo que 32 apresentaram sucesso para extração como evidenciado na Figura acima, e 7 foram separados para análise devido conteúdo estar incompleto ou vazio em meio ao processo, exigindo análise mais aprofundada para identificação do ajuste necessário.

### 6.4. Consulta no Banco de Dados

No processo de consulta no banco de dados, foi pesquisado o termo Ibuprofeno, fármaco do qual possuía duas ocorrências no repositório, sendo: no PCDT de Artrite Psoríase e no PCDT de Espondilite Ancilosante, ambos obtiveram score de 1.3136, como mostram as Figura 15 e a Figura 16, que demonstram o resultado obtido a partir da consulta do código CID G12.0 no banco de dados, retornando score 6.2828, com o PCDT de Atrofia Muscular Espinhal.

**Figura 15:** Consulta do termo “G12.0” no PCDT\_Atrofia Muscular Espinhal\_2019.

	Termo	Tema	ano	CID	farmaco	Score
▶	G12.0	Atrofia Muscular Espinhal	2019	G12.0 - Atrofia muscular espinal infantil tipo I (Werdnig-Hoffman...	Nusinersena: 2,4 mg/mL solucao injetavel - Frasco-ampola (5 mL). 7.1	2.384146213531494

**Fonte:** Aatoria Própria.

**Figura 16:** Consulta do termo “Ibuprofeno” na tabela “PCDT” do BD.

	Termo	Tema	ano	CID	farmaco	Score
▶	Ibuprofeno	Artrite Psoriaca	2021	- M070 Artropatia psoriasica interfalangiana distal - M072 Esp...	- Ibuprofeno: comprimidos revestidos de 200, 300 e 600 mg; suspensa...	1.5451436042785645
	Ibuprofeno	Espondilite Ancilosante	2018	- M45 Espondilite ancilosante - M46.8 Outras espondilopatias infl...	- Ibuprofeno: comprimidos revestidos de 200 mg, 300 mg e 600 mg, so...	1.5451436042785645

**Fonte:** Aatoria Própria.

## 6.5. Associação com demais fontes de dados

Após a identificação de um *score* acima de 0 (zero), que se refere a uma alta probabilidade do termo consultado estar presente em algum registro no banco, se realiza a busca do mesmo termo consultado, nesse caso o princípio ativo Ibuprofeno na tabela Referência Medicamentos, buscando da mesma forma a correlações com a base de fármacos vigentes pela ANVISA, como mostra a Figura 17.

**Figura 17:** Consulta do fármaco “Ibuprofeno” na Referência Medicamentos.

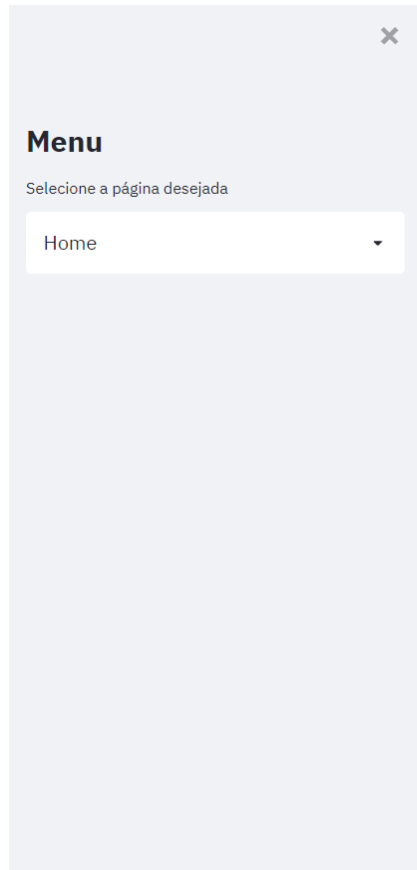
Termo	principio_ativo	registro_anvisa	laboratorio	produto	apresentacao	classe_terapeutica	score
Ibuprofeno	IBUPROFENO ARGININA	1008401480063	ZAMBON LABORATÓRIOS FARMACÊUTICOS LT...	SPIDUFEN	1155 MG GRAN SOL CT 10...	M1C - AGENTES ANTI-REUMÁTIC...	4.8862605094
Ibuprofeno	IBUPROFENO;ARGININA	1008401480081	ZAMBON LABORATÓRIOS FARMACÊUTICOS LT...	SPIDUFEN	770 MG GRAN SOL CT 6 E...	N2B2 - ANALGÉSICOS NÃO NAR...	4.8862605094
Ibuprofeno	IBUPROFENO;ARGININA	1008401480098	ZAMBON LABORATÓRIOS FARMACÊUTICOS LT...	SPIDUFEN	770 MG GRAN SOL CT 20 ...	N2B2 - ANALGÉSICOS NÃO NAR...	4.8862605094
Ibuprofeno	IBUPROFENO ARGININA	1008401480144	ZAMBON LABORATÓRIOS FARMACÊUTICOS LT...	SPIDUFEN	1155 MG GRAN SOL CT 2 ...	M1C - AGENTES ANTI-REUMÁTIC...	4.8862605094
Ibuprofeno	IBUPROFENO ARGININA	1008401480179	ZAMBON LABORATÓRIOS FARMACÊUTICOS LT...	SPIDUFEN	770 MG GRAN SOL CT 10 ...	N2B2 - ANALGÉSICOS NÃO NAR...	4.8862605094
Ibuprofeno	IBUPROFENO;ARGININA	1008401480233	ZAMBON LABORATÓRIOS FARMACÊUTICOS LT...	SPIDUFEN	770 MG GRAN SOL CT 10 ...	N2B2 - ANALGÉSICOS NÃO NAR...	4.8862605094
Ibuprofeno	IBUPROFENO;ARGININA	1008401480268	ZAMBON LABORATÓRIOS FARMACÊUTICOS LT...	SPIDUFEN	1155 MG GRAN SOL CT 10...	M1C - AGENTES ANTI-REUMÁTIC...	4.8862605094
Ibuprofeno	IBUPROFENO ARGININA	1008401480322	ZAMBON LABORATÓRIOS FARMACÊUTICOS LT...	SPIDUFEN	770 MG COM REV CT BL A...	N2B2 - ANALGÉSICOS NÃO NAR...	4.8862605094
Ibuprofeno	IBUPROFENO	1017500210012	CATALENT BRASIL LTDA.	CAPSFEN	600 MG CAP GEL MOLE CT...	M1A1 - ANTIRREUMÁTICOS NÃO...	4.8862605094
Ibuprofeno	IBUPROFENO	1017500210020	CATALENT BRASIL LTDA.	CAPSFEN	600 MG CAP GEL MOLE CT...	M1A1 - ANTIRREUMÁTICOS NÃO...	4.8862605094
Ibuprofeno	IBUPROFENO	1017500210039	CATALENT BRASIL LTDA.	CAPSFEN	600 MG CAP GEL MOLE CT...	M1A1 - ANTIRREUMÁTICOS NÃO...	4.8862605094
Ibuprofeno	IBUPROFENO	1017500210047	CATALENT BRASIL LTDA.	CAPSFEN	600 MG CAP GEL MOLE CT...	M1A1 - ANTIRREUMÁTICOS NÃO...	4.8862605094
Ibuprofeno	IBUPROFENO	1017500210055	CATALENT BRASIL LTDA.	CAPSFEN	600 MG CAP GEL MOLE CT...	M1A1 - ANTIRREUMÁTICOS NÃO...	4.8862605094

Fonte: Autoria Própria.

## 6.6 Interface de Consulta

O usuário deve realizar a seleção da página que deseja acessar. Em *Home*, contém um descritivo sobre o projeto contendo o Título, Problemática, Métodos de extração e Análise estatística, Figura 18. Em *About Us*, há uma imagem contendo todos os membros da equipe, ano de formação e a motivação para a realização do projeto, Figura 19. Na página de consulta inicialmente há uma barra de pesquisa para realização da busca dos termos, Figura 20, quando o termo é retornado os campos de fármaco, CID, ano, tema, termo e *score*, conforme a Figura 21.

**Figura 18:** Interface de Consulta: Page Home.



## Técnicas computacionais aplicadas para extração e centralização de informações dos Protocolos Clínicos e Diretrizes Terapêuticas

A proposta do trabalho é implementar, através de técnicas computacionais de reconhecimento de texto, um repositório centralizado com informações presentes em PCDTs, buscando promover uma consulta simplificada, intuitiva e ampliando o acesso à informações baseadas em evidências científicas.

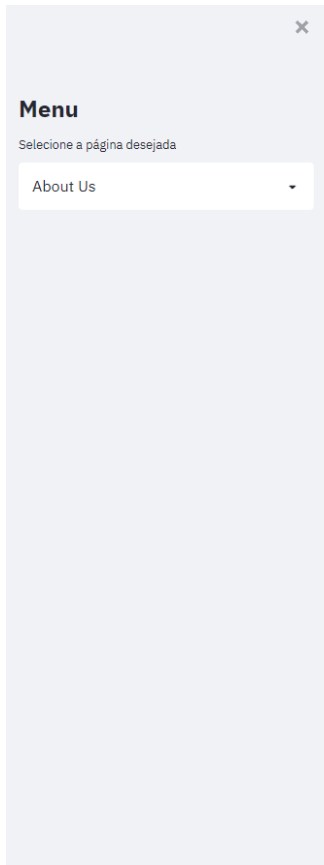
### Medicina Baseada em Evidências

A Medicina Baseada em Evidência (MBE) propõe uma metodologia para o processo de tomada de decisões médicas, minimizando o nível de incerteza e de erros, por não considerar apenas a experiência do profissional, mas também as melhores evidências disponíveis sobre a doença.

### Protocolos Clínicos e Diretrizes Terapêuticas

Os Protocolos Clínicos e Diretrizes Terapêuticas (PCDT) são documentos nos quais constam evidências de eficácia, segurança e conformidade, assegurados e desenvolvidos pela Comissão Nacional de Incorporação de Tecnologias no Sistema Único de Saúde (CONITEC). Porém, não há uma centralização das informações contidas nestes protocolos em um repositório para facilitar a consulta de maneira simplificada e intuitiva.

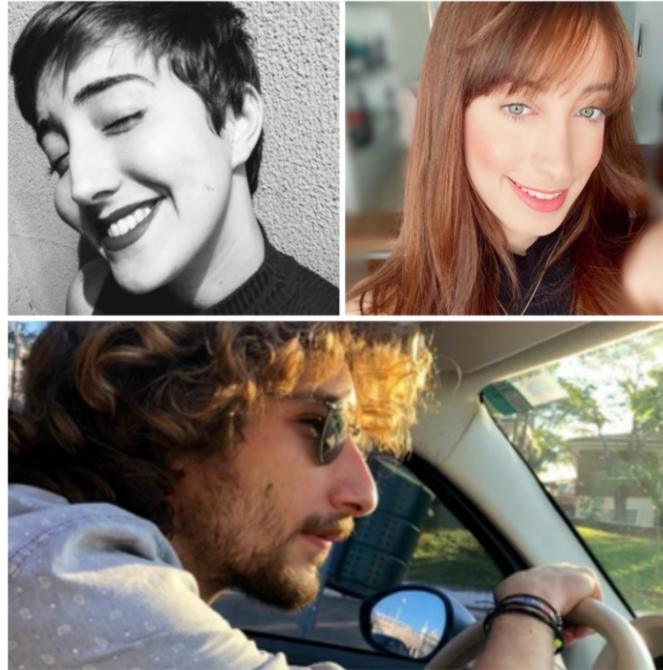
**Fonte:** Autoria Própria.



**Figura 19:** Interface de Consulta: Page About Us.

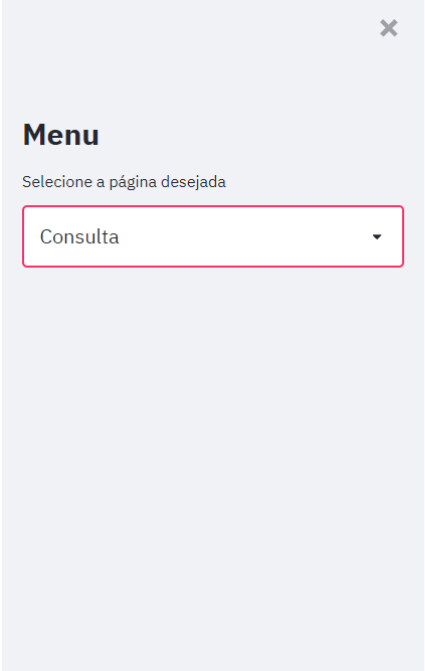
## About Us

A equipe é composta por Francisco Haddad Aguerre, Ellen Carvalho de Sousa e Sophia Marques Silva. Somos estudantes de Engenharia Biomédica do 5º, iniciamos esse projeto movidos pela tentativa de tornar a assistência ao paciente mais segura, eficiente e baseada por meio da aplicação das informações extraídas dos documentos e sua consulta de maneira simplificada pelos profissionais em meio ao processo de tomada de decisão.



**Fonte:** Autoria Própria.

**Figura 20:** Interface de Consulta: Page Consulta aos PCDTs.



## Consulta aos PCDTs

Digite o termo de busca (fármaco ou CID):

Buscar

**Fonte:** Autoria Própria.

**Figura 21:** Interface de Consulta Page Consulta: Exemplo busca CID G12.0.



## Consulta aos PCDTs

Digite o termo de busca (fármaco ou CID):

### Resultado da consulta:

	Termo de Interesse	Tema	CID
0	G12.0	Atrofia Muscular Espin...	G12.0 - Atrofia muscul... Nusinersena

**Fonte:** Autoria Própria.

## 6.7 Avaliação estatística de desempenho e acuracidade

Os resultados obtidos após a coleta de dados dos itens de interesse foram pontuados de acordo com o método proposto para cada um dos PCDTs analisados. Foram compilados na tabela 4 abaixo, a duração do processo de extração e armazenamento de cada um dos grupos, acompanhado da pontuação obtida.

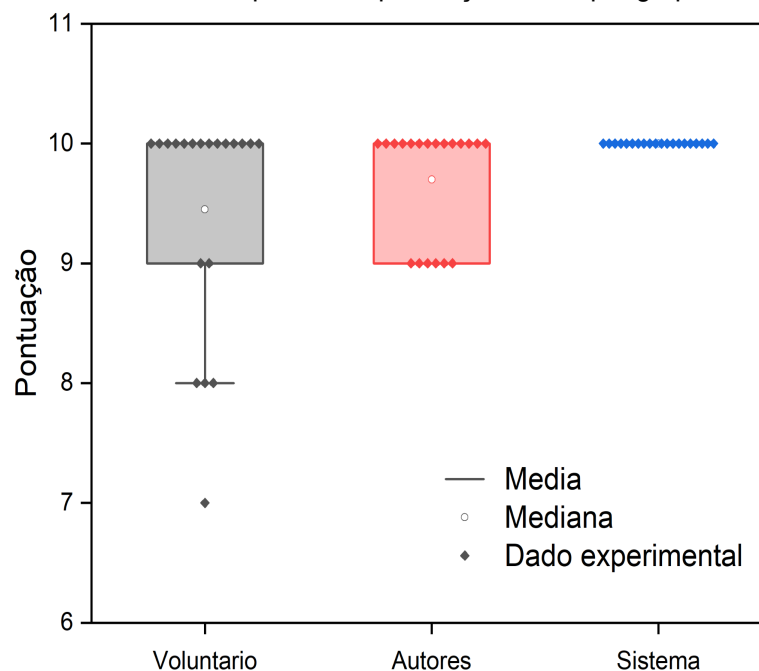
**Tabela 4:** Compilação do resultado da coleta de dados para análise estatística.

COMPILAÇÃO DOS RESULTADOS									
PCDTs / GRUPO		G1: Sistema		G2: Voluntários		G3: Autores			
id	nome_arquivo	duracao	pontuação_total	voluntários	duracao	pontuação_total	autores	pontuação_total	
1	PCDT_Acromegalia_2019	00:00:02	10	A	00:55:49	10	F	00:42:15	10
2	PCDT_Anemia Hemolítica Autoimune_2018	00:00:04	10		01:00:37	10		00:36:33	9
3	PCDT_Anemia na Doença Renal Crônica_2017	00:00:01	10		02:58:21	8		00:38:11	9
4	PCDT_Artrite Psoriásica_2021	00:00:03	10	B	01:22:14	10	F	00:56:55	10
5	PCDT_Atrofia Muscular Espinhal Sq tipo I_2019	00:00:02	10		02:12:58	9		00:42:25	9
6	PCDT_Comportamento Agressivo no Transtorno do Espectro do Autismo_2016	00:00:02	10		01:27:32	10		00:39:29	9
7	PCDT_Colangite Biliar Primária_2019	00:00:01	10	C	01:04:33	10	G	00:33:47	9
8	PCDT_Deficiência Biotinidase_2018	00:00:01	10		01:29:10	9		00:39:32	10
9	PCDT_Diabetes Insípido_2018	00:00:01	10		01:06:56	10		00:32:32	10
10	PCDT_Doença Falciforme_2018	00:00:02	10	D	01:00:02	10	G	00:41:05	10
11	PCDT_Doença de Wilson_2018	00:00:01	10		00:37:01	10		00:32:05	10
12	PCDT_Espondilite Ancilosante_2018	00:00:02	10		00:45:47	10		00:33:27	10
13	PCDT_Glaucoma_2018	00:00:06	10	E	00:52:07	10	H	00:39:36	10
14	PCDT_Hemangioma Infantil_2019	00:00:03	10		00:50:57	10		00:40:32	10
15	PCDT_Hiperprolactinemia_2015	00:00:01	10		00:50:24	10		00:31:52	10
16	PCDT_Hipoparatiroidismo_2016	00:00:01	10	E	01:38:05	8	H	00:35:29	10
17	PCDT_Hormônio de Crescimento - Hipopituitarismo_2018	00:00:04	10		01:37:22	7		00:43:20	9
18	PCDT_Homocistinúria Clássica_2020	00:00:01	10		01:04:30	10		00:38:26	10
19	PCDT_Parkinson_2017	00:00:02	10	E	00:51:02	8	H	00:43:47	10
20	PCDT_Púrpura Trombocitopênica_2019	00:00:03	10		00:51:39	10		00:41:20	10

Fonte: Autoria Própria.

A dispersão da pontuação obtida através da análise de cada PCDT por grupo será apresentada na Figura 22 abaixo através do gráfico de box plot.

**Figura 22:** Box Plot da dispersão da pontuação obtida por grupo de análise.



**Fonte:** Autoria Própria.

Como apresentado, os Voluntários possuem alta variabilidade de pontuação, fazendo com que possua a menor mediana do estudo. Por outro lado, os Autores obtiveram mais solidez nos resultados, pelo fato de já ter alguma experiência com a temática, resultando no segundo melhor resultado evidenciado pela localização da mediana. Por fim, conclui-se através deste gráfico, que o Sistema obteve excelente acurácia no que se refere a exatidão da extração proposta, sendo possível responder a primeira pergunta norteadora desta análise.

Desta forma foi possível obter a porcentagem de correlação destas informações, sendo: os Voluntários 94,5%, Autores 97,05% e Sistema 100%. É possível perceber que dentre os três grupos do estudo, o sistema acertou corretamente todas as informações (100%), seguida dos autores (97,05%) e dos voluntários (94,5%), mostrando que o sistema desenvolvido apresenta uma melhor performance em comparação às pessoas.

De modo a analisar comparativamente os resultados intergrupos, como colocado na metodologia desta análise, foram adotados os índices em cada PCDT

analisado, na tabela 5 abaixo segue, por grupo respectivamente: a média, o desvio padrão e o número de amostras analisadas.

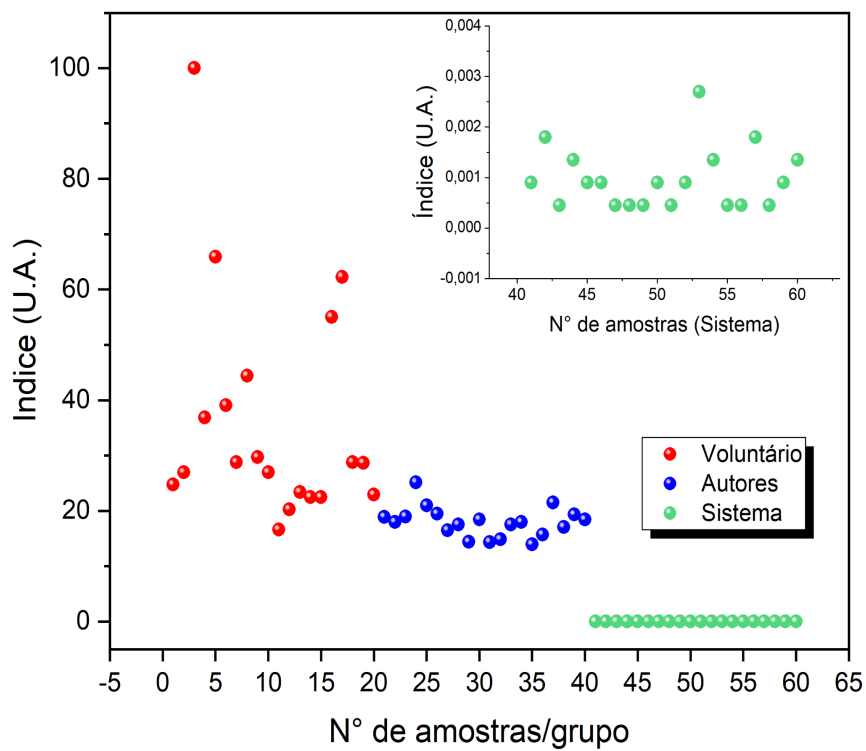
**Tabela 5:** Descrição dos dados do índice pela média e desvio padrão dos grupos

Grupos	média	Desvio padrão	N
Autores	17,959	2,713	20
Sistema	9,662e -4	6,060e -4	20
Voluntário	36,320	20,433	20

Fonte: Autoria Própria.

Abaixo, a Figura 23 consiste na dispersão dos dados analisados em gráfico de box plot.

**Figura 23:** Dispersão dos índices por amostra/grupo.



Fonte: Autoria Própria.

Após verificado que não há uma distribuição normal dos dados foi aplicado o teste de ANOVA não paramétrica (Kruskal-Wallis) e foi verificado uma diferença significativa entre os grupos ( $p < 0,001$ ) como mostrado na tabela 6.

**Tabela 6:** Teste de Kruskal-Wallis para avaliação da comparação entre os grupos

Factor	Estatística	df	p
Grupos	49,904	2	< ,001

**Fonte:** Autoria Própria.

Na análise de Post Hoc comparando os grupos dois a dois verificamos que todas as comparações foram significativas ( $p < 0,001$ ), conforme mostrado na tabela 7 a seguir.

**Tabela 7:** Comparações de Post Hoc do índice entre os grupos.

		Diferença da média	Erro padrão	t	p
Voluntário	Autores	18,362	3,763	4,879	< ,001
	Sistema	36,319	3,763	9,651	< ,001
Autores	Sistema	17,958	3,763	4,772	< ,001

**Fonte:** Autoria Própria.

Conclui-se, portanto, que os as amostras comparadas apresentaram diferença significativa, comprovando as hipóteses de que o sistema realiza a extração e armazenamento preciso das informações e apresenta um desempenho superior aos demais grupos da amostra, vale ressaltar que, a diferença da média entre o grupo de Voluntários e o Sistema é praticamente o dobro da relação Voluntário x Autor.

## 7. Conclusão

O trabalho procurou analisar a possibilidade de aplicação de técnicas computacionais para a extração e consulta de informações contidas nos PCDTs, além da associação com a tabela de medicamentos da ANVISA, para o desenvolver um repositório centralizado com estas informações buscando promover uma consulta simplificada e intuitiva, e facilitar a aplicação do método de MBE. Os códigos não foram adicionados ao trabalho devido ao interesse de emissão de patente para obter propriedade da autoria legal do mesmo.

Os resultados preliminares confirmaram a possibilidade de realizar a extração e consulta, necessitando apenas de alguns ajustes para abranger as variações que podem ocorrer nos subtítulos e a realização dos testes unitários para facilitar a identificação de falhas no código ou variações no documento para documento. Além disso, a partir da análise estatística foi evidenciado que o sistema em relação à comparação aos Autores e aos Voluntários apresentou relevância significativa quanto ao tempo de execução da extração, comprovando a eficiência da aplicação.

Os objetivos futuros são a ampliação das bases a serem associadas ao conteúdo extraído dos PCDTs, aprimorar o processo de extração adotando outras tecnologias e aplicar as informações centralizadas para uma possível aplicação de um sistema de suporte à tomada de decisão. Além do sistema, sugere-se aprimorar também a análise estatística realizando a comparação indicando a porcentagem de acerto de cada item analisado, além de aumentar a quantidade de amostras apreciadas por grupo.

Por fim, este trabalho realiza a intersecção de diversos conceitos relevantes para a saúde pública como a medicina baseada em evidência e os sistemas de suporte à decisão, atrelado à técnicas computacionais de reconhecimento de texto. Como evidenciado através da revisão sistemática, este trabalho corresponde ao estado da arte da aplicação destes conceitos no Brasil. Caracteriza, portanto, um importante avanço na extração, armazenamento e consulta de informações médicas baseadas em evidências científicas.

## 8. Bibliografia

AHAMED, T.; LEDERMAN, R.; BOSUA, R.; VERSPOOR, K.; BUNTINE, W.; HART, G. (2016). **Towards a Methodology for Nursing– Specific Clinical Decision Support Systems (CDSS)**. Journal of Decision Systems, 25(sup1), p. 23–34. doi:10.1080/12460125.2016.1187387.

AKABANE, Getulio K. **Gestão estratégica das tecnologias cognitivas: conceitos, metodologias e aplicações**. [S. l.]: Editora Érica, ed. 1, p.1–296, 2018. ISBN 978–8536527864.

ANVISA. **Listas de preços de medicamentos**, 08 mai. 2021. Disponível: <<https://www.gov.br/anvisa/pt-br/assuntos/medicamentos/cmmed/precos>>. Acesso: 25 mai. 2021.

Becker, M.; Böckmann, B.; Jöckel, K.– H.; Stuschke, M.; Paul, A., Kasper, S., & Virchow, I. (2020). **Mapping Patient Data to Colorectal Cancer Clinical Algorithms for Personalized Guideline-Based Treatment**. Applied Clinical Informatics, v.11, n.02, p. 200–209. doi:10.1055/s–0040–1705105

Becker, M., Kasper, S., Böckmann, B., Jöckel, K.– H., & Virchow, I. (2019). **Natural Language Processing of German Clinical Colorectal Cancer Notes for Guideline– Based Treatment Evaluation**. International Journal of Medical Informatics. doi:10.1016/j.ijmedinf.2019.04.022

BRASIL. **Lei nº 12.401**. Altera a Lei no 8.080, de 19 de setembro de 1990, para dispor sobre a assistência terapêutica e a incorporação de tecnologia em saúde no âmbito do Sistema Único de Saúde – SUS, p. 80– 81, 2011.

CONITEC. **A comissão**, 13 abr. 2015. Disponível em: <<http://conitec.gov.br/entenda-a-conitec-2>>. Acesso em: 16 mar. 2021.

DATASUS. **Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde**. 2008. Disponível em: <<http://www2.datasus.gov.br/DATASUS/index.php?area=060203>>. Acesso em: 28 mar. 2021.

Ennis, J., Gillen, D., Rubenstein, A., Worcester, E., Brecher, M. E., Asplin, J., & Coe, F. (2015). **Clinical decision Support improves physician guideline adherence for laboratory monitoring of chronic kidney disease: a matched cohort study**. BMC Nephrology, v. 16, n. 01. doi:10.1186/s12882-015-0159-5.

FERNEDA, Edberto. **Recuperação de Informação: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação**. Orientador: Profª Drª Johanna Wilhelmina Smit. Tese (Doutorado em Ciência da Comunicação) - Universidade de São Paulo, USP, p. 1-147, 2003.

FERREIRA, Leonardo L. G.; ANDRICOPULO, Adriano D. **Medicamentos e tratamentos para a Covid-19**. Estudos Avançados, v.34, n. 100, p. 7-27, 2020.

FLORÊNCIO, Valéria; DOURADO, Péricles; VIEIRA, Luciana; LIMA, Alessandra. **Tratamento precoce da COVID-19**. Subsecretaria de Saúde Gerência de Informações Estratégicas em Saúde CONECTA-SUS, Goiás, p.1-6, 2021.

Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). **Pesquisadores no Brasil publicam 56% dos artigos científicos originados na América Latina: número de publicações de cientistas do Estado de São Paulo supera o de todos os países da região, exceto Brasil**, 2011. Disponível em: <<http://www.fapesp.br/indicadores/boletim3.pdf>>.

FOX, John. **Cognitive systems at the point of care: the CREDO program**. Journal of Biomedical Informatics, v. 68, p.83– 95, 2017.

GALLOTTI, Renata Mahfuz Daud. **Eventos adversos: o que são?** Rev. Assoc. Med. Bras., São Paulo , v. 50, n. 2, p. 114, Abr. 2004 . Disponível em <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0104-42302004000200008](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-42302004000200008&lng=en&nrm=iso)&lng=en&nrm=iso>. Acesso em: 07 mai. 2021. <https://doi.org/10.1590/S0104-42302004000200008>.

GALVÃO, Tais Freire; PANSANI, Thais de Souza Andrade; HARRAD, David. **Principais itens para relatar Revisões sistemáticas e Meta- análises: A recomendação PRISMA**. Epidemiologia e Serviços de Saúde [online]. 2015, v. 24, n. 2 [Acessado 8 Junho 2021] , p. 335–342. Disponível em: <<https://doi.org/10.5123/S1679-49742015000200017>>. Acesso em: 16 mar. 2021.

GAVRANIC dos Reis, Sueli Nascimento; AGUERRE, Pedro. **A inteligência artificial como apoio à área de gestão de pessoas: do recrutamento à demissão**. Impactos da inteligência artificial na gestão de pessoas / Organização de Myrt Thânia de Souza Cruz e João Pinheiro de Barros Neto. – São Paulo: Tikibooks, 2020. p. 174-185; II. Disponível em: <[https://pedroaguerre.files.wordpress.com/2020/05/livro-impactos-da-inteligc38ancia-artificial-na-gestc383o-de-pessoas\\_puc-adm.pdf](https://pedroaguerre.files.wordpress.com/2020/05/livro-impactos-da-inteligc38ancia-artificial-na-gestc383o-de-pessoas_puc-adm.pdf)>. Acesso em: 10 out. 2021.

GONÇALVES, João Manuel de Campos. **Utilização de técnicas de data mining na previsão do plano terapêutico em medicina intensiva**. 2012. 104 f. Dissertação (Mestrado) - Curso de Engenharia e Gestão de Sistemas de Informação, Escola de Engenharia, Universidade do Minho, Braga, 2012.

LADEIRA, Ana Paula. **Processamento de linguagem natural: caracterização da produção científica dos pesquisadores brasileiros**. Doutorado em Ciências da Informação: Universidade Federal de Minas Gerais, p. 1– 262, 2010.

LANZOLA, G.; BOSSI, P.; QUAGLINI, S.; ZINI, E. M. (2017). **An Environment for Guideline Based Decision Support Systems for Outpatients Monitoring**. *Methods of Information in Medicine*, v.56, n. 4, p.283– 293. doi:10.3414/me16– 01– 0142.

LOPES, A.A. **Medicina Baseada em Evidências: a arte de aplicar o conhecimento científico na prática clínica**. *Rev Ass Med Brasil*, Salvador, v. 46, n. 3, p. 285– 288, 2000.

MAGHSOUD– LOU, E. *et al.* **Protocol– Driven Decision Support within e– Referral Systems to Streamline Patient Consultation, Triaging and Referrals from Primary Care to Specialist Clinics**. *Journal of Medical Systems*, Springer, v. 41, p. 1– 1, 2017. DOI <https://doi.org/10.1007/s10916– 017– 0791– 7>. Disponível em: <<https://link.springer.com/article/10.1007/s10916– 017– 0791– 7>>. Acesso em: 29 mar. 2021.

Maia Alves de Andrade, Patrícia Helena. **Aplicação de Técnicas de Mineração de Textos para Classificação de Documentos: um Estudo da Automatização da Triagem de Denúncias na CGU**. Patrícia Helena Maia Alves de Andrade; orientador Marcelo Ladeira; co-orientador Rommel Novaes Carvalho. -- Brasília, p. 1-65, 2015.

MINISTÉRIO DA SAÚDE. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. Departamento de Gestão e Incorporação de Tecnologias em Saúde. **Guia de elaboração de protocolos clínicos e diretrizes terapêuticas: delimitação do escopo**. Ministério da Saúde, Secretaria de Ciência, Tecnologia e Insumos Estratégicos, Departamento de Gestão e Incorporação de Tecnologias em Saúde. – 2. ed. – Brasília: Ministério da Saúde, 2019. 28 p. : il. Modo de acesso: World Wide Web: <[http://bvsms.saude.gov.br/bvs/publicacoes/guia\\_elaboracao\\_protocolos\\_delimitacao\\_escopo\\_2ed.pdf](http://bvsms.saude.gov.br/bvs/publicacoes/guia_elaboracao_protocolos_delimitacao_escopo_2ed.pdf)>. ISBN 978– 85– 334– 2699– 3>.

MORAIS, Edison Andrade Martins. **Mineração de textos**. Dissertação de Mestre em Ciência da Computação, Instituto de Informática da Universidade Federal de Goiás. 2007.

NETO, José Antônio Chehuen *et al.* **Percepção da aplicabilidade da Medicina Baseada em Evidência**. HU Revista, Juiz de Fora, v. 34, n. 1, p. 33– 39, 2008.

REZENDE, Solange *et al.* **O uso da Mineração de Textos para Extração e Organização Não Supervisionada de Conhecimento**. Revista de Sistemas de Informação da FSMA, Embrapa Informática Agropecuária, ano 2011, v. 1, n. 7, p. 7-21., 3 jan. 2011.

SANTOS, Ronnie *et al.* **Técnicas de Processamento de Linguagem Natural Aplicadas ao Processo de Mineração de Textos: resultados preliminares de um mapeamento sistemático**. Revista de Sistemas e Computação, Salvador, v.4, n. 2, p. 116– 125, 2014.

SEITINGER, A.; RAPPELSBERGER, A.; LEITICH, H.; BINDER, M.; ADLASSNIG, K.– P. (2016). **Executable medical guidelines with Arden Syntax—Applications in dermatology and obstetrics**. Artificial Intelligence in Medicine. doi:10.1016/j.artmed.2016.08.003.

SENADO FEDERAL (BRASÍLIA). **Senador Renan Calheiros. CPI - COVID: Relatório Final**, Brasília, DF.: Senado Federal, v. 1, n. 1, p. 1-1288, 26 out. 2021. Disponível em: <https://g1.globo.com/politica/cpi-da-covid/noticia/2021/10/19/leia-a-integra-da-minuta-do-relatorio-da-cpi-da-covid.ghtml>. Acesso em: 31 out. 2021.

SILVA, Nilton Correia da *et al.* **Document type classification for Brazil's supreme court using a Convolutional Neural Network**. Departamento de Ciência da

Computação University of Brasília. In International Conference on Forensic Computer Science and Cyber Law (ICoFCS), p. 1-4, 2018.

SOARES, Vanessa da Silva. **Mineração de textos para identificar perfis de satisfação de clientes**. Trabalho de Conclusão de Curso em Curso de Ciência da Computação da Universidade de Santa Cruz do Sul, p.1-56, 2016.

## 9. Anexos

### Anexo 1 – Artigos totais da Revisão Sistemática

**Tabela 8 – Análise Sistemática I da Revisão de Literatura.**

ID	Título	Ano	Autor	Nota	Fonte	Data
1	Predicting Optimal Hypertension Treatment Pathways Using Recurrent Neural Networks	2020	Ellen	1	Scopus	29/03/2021
2	Evaluation of a National Comprehensive Cancer Network Guidelines– Based Decision Support Tool in Patients With Non– Small Cell Lung Cancer: A Nonrandomized Clinical Trial	2020	Ellen	1	Scopus	29/03/2021
3	Qualitative analysis of clinicians' perspectives on the use of a computerized decision aid in the treatment of psychotic disorders	2020	Ellen	3	Scopus	29/03/2021
4	The Integrated Tracking, Referral, and Electronic Decision Support, and Care Coordination (I– TREC) program: scalable strategies for the management of hypertension and diabetes within the government healthcare system of India	2020	Ellen	3	Scopus	29/03/2021
5	Atrial Fibrillation Screen, Management, and Guideline– Recommended Therapy in the Rural Primary Care Setting: A Cross– Sectional Study and Cost– Effectiveness Analysis of eHealth Tools to Support All Stages of Screening	2020	Ellen	1	Scopus	29/03/2021
6	Evaluation of Point– of– Care Decision Support for Adult Acne Treatment by Primary Care Clinicians	2020	Ellen	1	Scopus	29/03/2021
7	Towards a goal– oriented methodology for clinical– guideline– based management recommendations for patients	2020	Ellen	4	Scopus	29/03/2021

	with multimorbidity: GoCom and its preliminary evaluation						
8	Clinical decision trees Support systematic evaluation of multidisciplinary team recommendations	2020	Ellen	1	Scopus	29/03/2021	
9	Development and efficacy of a computerized decision Support system for osteoporosis management in the community	2020	Ellen	1	Scopus	29/03/2021	
10	Integrated Management Program Advancing Community Treatment of Atrial Fibrillation (IMPACT– AF): A cluster randomized trial of a computerized clinical decision Support tool	2020	Ellen	1	Scopus	29/03/2021	
11	Implementation of an ontological reasoning to Support the guideline– based management of primary breast cancer patients in the DESIREE project	2020	Ellen	3	Scopus	29/03/2021	
12	Mapping Patient Data to Colorectal Cancer Clinical Algorithms for Personalized Guideline– Based Treatment	2020	Ellen	5	Scopus	29/03/2021	
13	Decision Analysis Model for Prehospital Triage of Patients With Acute Stroke	2019	Ellen	1	Scopus	29/03/2021	
14	Clinical requirements of future patient monitoring in the intensive care unit: Qualitative study	2019	Ellen	1	Scopus	29/03/2021	
15	Development and dissemination of clinical decision Support across institutions: standardization and sharing of refugee health screening modules	2019	Ellen	1	Scopus	29/03/2021	
16	eHealth tools to provide structured assistance for atrial fibrillation screening, management, and guideline– recommended therapy in metropolitan general practice: The AF– SMART study	2019	Ellen	2	Scopus	29/03/2021	

17	Utilizing Precision Medicine to Estimate Timing for Surgical Closure of Traumatic Extremity Wounds	2019	Ellen	1	Scopus	29/03/2021
18	ERegTime, efficiency of health information management using an electronic registry for maternal and child health: Protocol for a time– motion study in a cluster randomized trial	2019	Ellen	1	Scopus	29/03/2021
19	Detection of Cardiac arrhythmia using fuzzy logic	2019	Ellen	1	Scopus	29/03/2021
20	CDs in a learning health care system: Identifying physicians' reasons for rejection of best– practice recommendations in pneumonia through computerized clinical decision Support	2019	Ellen	3	Scopus	29/03/2021
21	A computerised decision Support system for cardiovascular risk management 'live' in the electronic health record environment: development, validation and implementation—the Utrecht Cardiovascular Cohort Initiative	2019	Ellen	2	Scopus	29/03/2021
22	Standards– Based Clinical Decision Support Platform to Manage Patients Who Meet Guideline– Based Criteria for Genetic Evaluation of Familial Cancer	2019	Francisco	2	Scopus	29/03/2021
23	Implementing cardiovascular disease prevention guidelines to translate evidence– based medicine and shared decision making into general practice: Theory– based intervention development, qualitative piloting and quantitative feasibility	2019	Francisco	2	Scopus	29/03/2021
24	Retrospective evaluation of a clinical decision Support tool for effective computed tomography angiography utilization in urgent brain imaging of suspected TIA/minor stroke in the emergency department	2019	Francisco	1	Scopus	29/03/2021

25	Natural language processing of German clinical colorectal cancer notes for guideline-based treatment evaluation	2019	Francisco	5	Scopus	29/03/2021
26	Reducing waste: A guidelines-based approach to reducing inappropriate Vitamin D and TSH testing in the inpatient rehabilitation setting	2019	Francisco	2	Scopus	29/03/2021
27	Not a magic pill: A qualitative exploration of provider perspectives on antibiotic prescribing in the outpatient setting	2018	Francisco	3	Scopus	29/03/2021
28	Development of a tailored intervention with computerized clinical decision Support to improve quality of care for patients with knee osteoarthritis: Multi-method study	2018	Francisco	3	Scopus	29/03/2021
29	Selecting Test Cases from the Electronic Health Record for Software Testing of Knowledge-Based Clinical Decision Support Systems	2018	Francisco	3	Scopus	29/03/2021
30	Decision Support for assessment of left ventricular diastolic function	2018	Francisco	3	Scopus	29/03/2021
31	Executable medical guidelines with Arden Syntax—Applications in dermatology and obstetrics	2016	Francisco	5	Scopus	29/03/2021
32	Effectiveness of a standardized electronic admission order set for acute exacerbation of chronic obstructive pulmonary disease	2018	Francisco	1	Scopus	29/03/2021
33	CardioVAI: An automatic implementation of ACMG-AMP variant interpretation guidelines in the diagnosis of cardiovascular diseases	2018	Francisco	3	Scopus	29/03/2021
34	LifeMap: towards the development of a new technology in sudden cardiac death risk stratification for clinical use	2018	Francisco	1	Scopus	29/03/2021

35	An Electronic Health Record–Based Strategy to Address Child Tobacco Smoke Exposure	2018	Francisco	3	Scopus	29/03/2021
36	Screening for cervical cancer in primary care a decision analysis for the us preventive services task force	2018	Francisco	1	Scopus	29/03/2021
37	Pediatric decision Support using adapted Arden Syntax	2018	Francisco	3	Scopus	29/03/2021
38	Diabetes– related behavior change knowledge transfer to primary care practitioners and patients: Implementation and evaluation of a digital health platform	2018	Francisco	4	Scopus	29/03/2021
39	An environment for guideline–based decision Support systems for outpatients monitoring	2017	Francisco	5	Scopus	29/03/2021
40	Impact of Gram stain results on initial treatment selection in patients with ventilator–associated pneumonia: A retrospective analysis of two treatment algorithms	2017	Francisco	1	Scopus	29/03/2021
41	Derivation and validation of the Personal Support Algorithm: An evidence– based framework to inform allocation of personal Support services in home and community care	2017	Francisco	1	Scopus	29/03/2021
42	Reconciliation of multiple guidelines for decision Support: a case study on the multidisciplinary management of breast cancer within the DESIREE project	2017	Francisco	1	Scopus	29/03/2021
43	Automated home monitoring and management of patient–reported symptoms during chemotherapy: results of the symptom care at home RCT	2017	Sophia	1	Scopus	29/03/2021
44	CAC Score Improves Coronary and CV Risk Assessment Above Statin Indication by ESC and AHA/ACC Primary Prevention Guidelines	2017	Sophia	1	Scopus	29/03/2021

45	Protocol– Driven Decision Support within e– Referral Systems to Streamline Patient Consultation, Triaging and Referrals from Primary Care to Specialist Clinics	2017	Sophia	5	Scopus	29/03/2021
46	Diagnosis and treatment in chronic pancreatitis: an international survey and case vignette study	2017	Sophia	4	Scopus	29/03/2021
47	Summary of the proceedings of the international forum 2016: “Imaging referral guidelines and clinical decision Support – how can radiologists implement imaging referral guidelines in clinical routine?”	2017	Sophia	3	Scopus	29/03/2021
48	Templates as a method for implementing data provenance in decision Support systems	2017	Sophia	3	Scopus	29/03/2021
49	Automating Performance Measures and Clinical Practice Guidelines: Differences and Complementarities	2016	Sophia	1	Scopus	29/03/2021
50	Development of the Parkland– UT Southwestern Colonoscopy Reporting System (CoRS) for evidence– based colon cancer surveillance recommendations	2016	Sophia	3	Scopus	29/03/2021
51	An architecture for a continuous, user– driven, and data– driven application of clinical guidelines and its evaluation	2016	Sophia	1	Scopus	29/03/2021
52	The Office Guidelines Applied to Practice program improves secondary prevention of heart disease in Federally Qualified Healthcare Centers	2016	Sophia	1	Scopus	29/03/2021
53	Implementing shared decision making in federally qualified health centers, a quasi– experimental design study: The Office– Guidelines Applied to Practice (Office– GAP) program	2016	Sophia	1	Scopus	29/03/2021
54	A fuzzy framework for encoding uncertainty in clinical decision– making	2016	Sophia	1	Scopus	29/03/2021

55	Computational prediction of multidisciplinary team decision-making for adjuvant breast cancer drug therapies: A machine learning approach	2016	Sophia	2	Scopus	29/03/2021
56	Improving physicians' performance with a stroke CDSS: A cognitive fit design approach	2016	Sophia	4	Scopus	29/03/2021
57	Guideline-based decision Support has a small, non-sustained effect on transthoracic echocardiography ordering frequency	2016	Sophia	1	Scopus	29/03/2021
58	Towards a Methodology for Nursing-Specific Clinical Decision Support Systems (CDSS)	2016	Sophia	5	Scopus	29/03/2021
59	A clinical decision Support system for the diagnosis of probable migraine and probable tension-type headache based on case-based reasoning	2015	Sophia	1	Scopus	29/03/2021
60	A multiple-scenario assessment of the effect of a continuous-care, guideline-based decision Support system on clinicians' compliance to clinical guidelines	2015	Sophia	1	Scopus	29/03/2021
61	Solving the interoperability challenge of a distributed complex patient guidance system: A data integrator based on HL7's Virtual Medical Record standard	2015	Sophia	1	Scopus	29/03/2021
62	Planning for action: The impact of an asthma action plan decision Support tool integrated into an Electronic Health Record (EHR) at a large health care system	2015	Sophia	3	Scopus	29/03/2021
63	Clinical decision Support improves physician guideline adherence for laboratory monitoring of chronic kidney disease: A matched cohort study	2015	Sophia	5	Scopus	29/03/2021

Fonte: Autoria Própria.

## Anexo 2 – Documentos Fonte extraídos pelo sistema

Extraídos com êxito
PCDT_Acromegalia_2019
PCDT_Anemia Aplástica, Mielodisplasia e Neutropenias Constitucionais - Uso de Fatores Estimulantes e Crescimento de Colônias de Neutrófilos_2016
PCDT_Anemia na Doença Renal Crônica_2017
PCDT_Aplasia Pura Adquirida Crônica de Série Vermelha_2016
PCDT_Artrite Psoriaca_2021
PCDT_Atrofia Muscular Espinhal 5q tipo I_2019
PCDT_Colangite Biliar Primária_2019
PCDT_Comportamento Agressivo no Transtorno do Espectro do Autismo_2016
PCDT_Deficiencia Biotinidase_2018
PCDT_Diabete Insípido_2018
PCDT_Diabete Melito Tipo 1_2019
PCDT_Dislipidemia_2019
PCDT_Doença Crohn_2017
PCDT_Doença Falciforme_2018
PCDT_Doencade Wilson_2018
PCDT_Endometriose_2016
PCDT_Esclerose Sistêmica_2017
PCDT_Espondilite Ancilosante_2018
PCDT_Glaucoma_2018
PCDT_Hemangioma Infantil_2019

PCDT\_Hemoglobinúria Paroxística Noturna\_2019

PCDT\_Hiperprolactinemia\_2015

PCDT\_Hipoparatiroidismo\_2016

PCDT\_Homocistinúria Clássica\_2020

PCDT\_Mucopolissacaridose do tipo I\_2018

PCDT\_Mucopolissacaridose Tipo IV A\_2019

PCDT\_Parkinson\_2017

PCDT\_Púrpura Trombocitopênica\_2019

PCDT\_Retocolite Ulcerativa\_2020

PCDT\_Síndrome de Barré\_2020

Necessário realização de teste unitário

PCDT\_AdenocarcinomaEstomago\_2018

PCDT\_Artrite\_2019

PCDT\_DegeneracaoMuscular\_2018

PCDT\_DoençaPompe\_2020

PCDT\_Epilepsia\_2018

PCDT\_EscleroseLateralAmiotrofica\_2020

PCDT\_EscleroseMultipla\_2019

### Anexo 3 – Análise Estatística Grupo 2: Modelo para preenchimento

GRUPO 2: Voluntários														
voluntário	arquivo	id	ano	pontuacao_ano	doenca	pontuacao_doenca	cids	pontuacao_cid	farmacos	pontuacao_farmacos	numero_paginas	pontuacao_n_paginas	tempo_total	pontuacao_total
A	PCDT_Acromegalia_2019	1												
A	PCDT_Anemia Hemolitica Autoimune_2018	2												
A	PCDT_Anemia na Doença Renal Crônica_2017	3												
A	PCDT_Artrite Psoriaca_2021	4												
B	PCDT_Atrofia Muscular Espinhal 5q tipo I_2019	5												
B	PCDT_Comportamento Agressivo no Transtorno do Espectro do Autismo_2016	6												

## Anexo 4 – Análise Estatística Grupo 3: Modelo para preenchimento

GRUPO 3: Autores														
Autor	arquivo	id	ano	pontuacao_ano	doenca	pontuacao_doenca	cids	pontuacao_cid	farmacos	pontuacao_farmacos	numero_paginas	pontuacao_n_paginas	tempo_total	pontuacao_total
F	PCDT_Acromegalia_2019	1												
F	PCDT_Anemia Hemolitica Autoimune_2018	2												
F	PCDT_Anemia na Doença Renal Crônica_2017	3												
F	PCDT_Artrite Psoriaca_2021	4												
F	PCDT_Atrofia Muscular Espinhal 5q tipo I_2019	5												
F	PCDT_Comportamento Agressivo no Transtorno do Espectro do Autismo_2016	6												

## Anexo 5 – Análise Estatística Grupo 4: Padrão Ouro

GRUPO 4: PADRÃO OURO						
id	nome_arquivo	doenca	ano	cid	farmaco	paginas
1	PCDT_Acromegalia_2019	Acromegalia	2019	E22.0 Acromegalia e gigantismo hipofisario	Octreotida: ampola de 0,1 mg/mL. Octreotida de liberacao prolongada: frasco-ampola de 10, 20 e 30 mg/mL. Lanreotida solucao injetavel de liberacao prolongada: seringa preenchida de 60, 90 e 120 mg. Cabergolina: comprimido de 0,5 mg.	32
2	PCDT_Anemia Hemolitica Autoimune_2018	Anemia Hemolitica Autoimune	2018	- D59.0 Anemia hemolitica autoimune induzida por droga - D59.1 Outras anemias hemoliticas autoimunes	- Prednisona: comprimidos de 5 mg e 20 mg. - Metilprednisolona: ampola de 500 mg. - Ciclofosfamida: frasco-ampola de 200 e 1.000 mg ou drageas de 50 mg. - Ciclosporina: capsula de 10, 25, 50 e 100 mg e solucao oral com 100 mg/mL em frasco de 50 mL. - Imunoglobulina humana: frasco com 0,5, 1, 2,5, 3, 5 e 6 g. - Acido folico: comprimido de 5 mg.	18
3	PCDT_Anemia na Doença Renal Crônica_2017	Anemia na Doença Renal Crônica	2017	- N 18.0 Doença renal em estadio final - N18.8 Outra insuficiencia renal cronica	- Sulfato ferroso: comprimidos de 40 mg, solucao oral de 25 mg/mL e xarope de 5 mg/mL; - Sacarato de hidroxido ferrico: solucao injetavel de 100 mg, frasco de 5 mL.	25
4	PCDT_Artrite Psoriaca_2021	Artrite Psoriaca	2021	- M070 Artropatia psoriatica interfalangiana distal - M072 Espondilite psoriatica - M073 Outras artropatias psoriaticas.	- Ibufrofeno: comprimidos revestidos de 200, 300 e 600 mg; suspensao oral de 50 mg/mL. - Naproxeno: comprimidos de 250 e 500 mg. - Prednisona: comprimidos de 5 e 20 mg. - Metilprednisolona: frasco de 40 mg/2 mL. 13 - Ciclosporina: capsulas de 10, 25, 50 e 100 mg; solucao oral 100 mg/mL. - Leflunomida: comprimidos de 20 mg. - Sulfassalazina: comprimidos de 500 mg. - Metotrexato: comprimidos de 2,5 mg; solucao injetavel (frasco com 2 mL) - Adalimumabe: solucao injetavel de 40 mg. - Etanercepte: solucao injetavel de 25 e 50 mg. - Infliximabe: po para solucao injetavel de 100 mg/10 mL - Golimumabe: solucao injetavel de 50 mg. - Secuquinumabe: solucao injetavel com 150mg/mL. - Certolizumabe Pegol: solucao injetavel 200 mg/mL. - Tofacitinibe: comprimidos de 5 mg.	40
5	PCDT_Atrofia Muscular Espinhal 5q tipo I_2019	Atrofia Muscular Espinhal 5q tipo I	2019	G12.0 - Atrofia muscular espinal infantil tipo I (Werdnig-Hoffman)	Nusinersena: 2,4 mg/mL solucao injetavel - Frasco-ampola (5 mL).	29
6	PCDT_Comportamento Agressivo no Transtorno do Espectro do Autismo_2016	Comportamento Agressivo no Transtorno do Espectro do Autismo	2016	Fazem parte dos TEA, os seguintes diagnosticos [1,2]: - F84.0 Autismo infantil - F84.1 Autismo atipico - F84.3 Outro transtorno desintegrativo da infancia - F84.5 Sindrome de Asperger - F84.8 Outros transtornos invasivos do desenvolvimento	- Risperidona: solucao oral de 1 mg/mL (para doses que exigem fracoes de 0,5 mg); comprimidos de 1, 2 e 3 mg.	62
7	PCDT_Colangite Biliar Primária_2019	Colangite Biliar Primária	2019	K74.3 Cirrose biliar primaria (Colangite destrutiva nao-supurativa cronica)	- Acido ursodesoxicolico: comprimido de 50 mg, 150 mg e 300 mg.	17
8	PCDT_Deficiencia Biotinidase_2018	Deficiencia Biotinidase	2018	- E88.9 Disturbio metabolico nao especificado	- Biotina: capsulas de 2,5 mg. As capsulas devem ser integralmente deglutidas por criancas maiores (escolares e adolescentes) ou ter seu conteudo retirado e administrado a lactentes, pre-escolares ou pacientes com disturbios de degluticao juntamente com leite materno, formula infantil ou leite. A seguir, sem lavar a colher ou o dispensador, deve-se adicionar nova quantidade do liquido utilizado para que seja administrado todo o conteudo da capsula. Apresentacoes liquidas nao sao recomendadas devido a baixa solubilidade da biotina em solucao aquosa (25).	15
9	PCDT_Diabete Insipido_2018	Diabete Insipido	2018	E23.2 Diabete insipido	Desmopressina: 0,1 mg/mL (100 mcg/mL) com aplicacao nasal (frasco de 2,5 mL em solucao ou spray) e comprimidos de 0,1 e 0,2 mg.	25
10	PCDT_Doença Falciforme_2018	Doença Falciforme	2018	- D57.0 Anemia falciforme com crise - D57.1 Anemia falciforme sem crise - D57.2 Transtornos falciformes heterozigoticos duplos NOTA: A Talassemia Beta (D56.1) e Outras talassemias (D56.8) incluem-se nos transtornos heterozigoticos duplos.	- Hidroxiureia: capsulas de 500 mg; - Fenoximetilpenicilina potassica: po para solucao oral 80.000 UI/mL; - Benzilpenicilina benzatina: po para suspensao injetavel de 600.000 UI e 1.200.000 UI; - Estolato de eritromicina: suspensao oral de 25 e 50 mg/mL e comprimidos de 500 mg.	26
11	PCDT_Doencade Wilson_2018	Doencade Wilson	2018	- E83.0 Disturbios do metabolismo do cobre.	- Penicilamina: capsula de 250 mg. - Trientina: capsula de 250 mg. - Piridoxina: comprimido de 40 mg. - Sulfato de zinco heptahidratado: comprimido mastigavel 10 mg; xarope 4 mg/mL; solucao injetavel 200 mcg/mL.	15

17	PCDT_Espondilite Ancilosa_2018	Espondilite Ancilosa	2018	- M45 Espondilite anquilosante - M46.8 Outras espondilopatias inflamatórias especificadas.	- Ibufreno: comprimidos revestidos de 200 mg, 300 mg e 600 mg, solucao oral de 50 mg/mL - Naproxeno: comprimidos de 250 mg e 500 mg. - Sulfasalazina: comprimidos de 500 mg. - Metilprednisolona: frasco de 40 mg/2 mL - Metotrexato: comprimidos de 2,5 mg; frasco-ampola com 50 mg/2 mL - Adalimumabe: solucao injetavel de 40 mg. Etanercepte: frasco ampola ou seringa preenchida de 25 mg e 50 mg - Infliximabe: frasco-ampola com 100 mg/10 ml - Golimumabe: seringa preenchida com 50 mg. Aprovado apenas para pacientes adultos. - Certolizumabe pegol: solucao injetavel com 200mg/mL com caneta aplicadora. Aprovado apenas para pacientes adultos. - Secuquinumabe: solucao injetavel com 150mg/ml com caneta aplicadora. Aprovado apenas para pacientes adultos.	74
13	PCDT_Glaucoma_2018	Glaucoma	2018	H 40.1 Glaucoma primario de angulo aberto; H 40.2 Glaucoma primario de angulo fechado; H 40.3 Glaucoma secundario a traumatismo ocular; H 40.4 Glaucoma secundario a inflamacao ocular; H 40.5 Glaucoma secundario a outros transtornos do olho; H 40.6 Glaucoma secundario a drogas; H 40.8 Outro glaucoma; Q15.0 Glaucoma congenito.	Timolol: solucao oftalmica a 0,5%. Dorzolamida: solucao oftalmica a 2%. Brinzolamida: suspensao oftalmica a 1%. Brimonidina: solucao oftalmica a 0,2%. Latanoprost: solucao oftalmica a 0,005%. Travoprost: solucao oftalmica a 0,004%. Bimatoprost: solucao oftalmica a 0,03%. Pilocarpina: solucao oftalmica a 2%. Acetazolamida: comprimido de 250 mg. Manitol: solucao intravenosa a 20%.	48
14	PCDT_Hemangioma Infantil_2019	Hemangioma Infantil	2018	D18.0 Hemangioma de qualquer localizacao.	Propranolol: comprimidos de 10 mg e 40 mg; Prednisona: comprimidos de 5 mg e 20 mg; Prednisolona: solucao oral de 1 mg/mL e 3 mg/mL; Alfaferona: frasco-ampola de 3.000.000 UI, 5.000.000 UI e 10.000.000 UI.	29
15	PCDT_Hiperprolactinemia_2015	Hiperprolactinemia	2015	E22.1 Hiperprolactinemia.	Bromocriptina: comprimidos de 2,5 mg. Cabergolina: comprimidos de 0,5 mg.	14
16	PCDT_Hipoparatiroidismo_2016	Hipoparatiroidismo	2016	- E20.0 - Hipoparatiroidismo idiopatico - E20.1 - Pseudohipoparatiroidismo - E20.8 - Outro hipoparatiroidismo - E89.2 - Hipoparatiroidismo pos-procedimento.	- Carbonato de calcio: comprimidos de 500 mg de calcio - Carbonato de calcio + colecalciferol: comprimidos de 500 mg de calcio + 200 UI - Carbonato de calcio + colecalciferol: comprimidos de 500 mg de calcio + 400 UI - Carbonato de calcio + colecalciferol ou fosfato de calcio tribasico + colecalciferol: comprimidos de 600 mg de calcio + 400 UI - Gliconato de calcio 10% solucao injetavel ampola com 10 mL - Cloreto de calcio 10% ampola com 10mL - Alfalcidol: capsulas de 0,25 mcg e 1 mcg - Calcitriol: capsulas de 0,25 mcg.	5
17	PCDT_Hormônio de Crescimento - Hipopituitarismo_2018	Hormônio de Crescimento - Hipopituitarismo	2018	E23.0 Hipopituitarismo.	Somatropina injetavel: 4 UI, 12UI, 15UI, 16UI, 18UI, 24UI e 30UI. Na formula de conversao, 1 mg equivale a 3 UI. Ha apresentacoes comerciais com volumes de diluente diferentes para a mesma dose de hormonio, o que devera ser observado na prescricao e orientacao ao paciente.	24
18	PCDT_Homocistinúria Clássica_2020	Homocistinúria Clássica	2020	- E72.1 Disturbios do metabolismo dos aminoacidos que contem enxofre homocistinúria.	- Acido acetilsalicílico: comprimido de 100 e 500 mg - Acido folico: comprimido de 5 mg - Cianocobalamina (vitamina B12): solucao injetavel de 1.000 mcg - Cloridrato de piridoxina (vitamina B6): comprimido de 40 e 100 mg - Formula metabolica isenta de metionina.	20
19	PCDT_Parkinson_2017	Parkinson	2017	- G20 Doenca de Parkinson.	Levodopa/carbidopa: comprimidos de 200/50 mg e 250/25 mg. Levodopa/benserazida: comprimidos ou capsulas de 100/25 mg ou comprimido de 200/50 mg. Comprimidos de 100/25 mg sao disponiveis na forma dispersivel, ou seja, de diluicao em agua. Bromocriptina: comprimidos de 2,5 mg. Pramipexol: comprimidos de 0,125, 0,25 e 1 mg. Amantadina: comprimidos de 100 mg. Biperideno: comprimidos de 2 mg e comprimidos de liberacao controlada de 4 mg. Triexifenidil: comprimidos de 5 mg. Selegilina: comprimidos de 5 mg e dragea ou comprimido de 10 mg. Tolcapona: comprimidos de 100 mg. Entacapona: comprimidos de 200 mg. Clozapina: comprimidos de 25 mg e 100 mg. Rasagilina: comprimidos de 1mg.	27
20	PCDT_Púrpura Trombocitopênica_2019	Púrpura Trombocitopênica Idiopática	2019	D69.3 Purpura trombocitopenica idiopatica.	Prednisona: comprimidos de 5 mg e 20 mg; Dexametaco: ampolas de 4 mg/mL e comprimidos de 4 mg; Metilprednisolona: ampolas de 500 mg; Imunoglobulina humana intravenosa: ampolas com 0,5 g, 1,0 g, 2,5 g, 3 g, 5 g e 6 g; Ciclofosfamida: drageas de 50 mg; Azatioprina: comprimidos de 50 mg; Eltrombopague: comprimidos revestidos de 25 mg e 50 mg; Danazol: capsulas de 100 mg e 200 mg; Vincristina: frasco-ampola de 1mg/mL.	48