ASSOCIAÇÃO EDUCATIVA DO BRASIL - SOEBRAS

FACULDADES INTEGRADAS DO NORTE DE MINAS - FUNORTE

CURSO DE GRADUAÇÃO EM ENGENHARIA BIOMÉDICA

inteligência artificial no apoio à conduta clínica

**ALECK YANN MARQUES DE MATOS MARTINS**

**JONAS RODRIGUES FILHO**

**Marcel veloso campos**

**MONTES CLAROS – MG**

**2017**

**ALECK YANN MARQUES DE MATOS MARTINS**

**JONAS RODRIGUES FILHO**

**MARCEL VELOSO CAMPOS**

**TEXT DATA MINING NO APOIO À CONDUTA CLÍNICA**

Projeto de trabalho de conclusão de curso apresentado à Banca examinadora do Curso de Graduação em Engenharia Biomédica do Instituto de Engenharias como requisito parcial para obtenção do título de bacharel em Engenharia Biomédica.

Professor orientador: Dr. Ms. Marcel Veloso Campos.

**MONTES CLARO – MG**

**2017**

**RESUMO**

**Introdução:** a indubtável necessidade de se trabalhar com quantidade de dados cada vez maiores, vem tornando o processamento e a manipulação destes, exponencialmente mais complexa e morosa. A Inteligência Artificial é um ramo da tecnologia da informação que se propõe a realizar tarefas complexas que outrora seriam realizadas apenas por seres humanos. É evidente que na área de ciências da saúde, a quantidade de artigos, estudos, testes, revistas, substâncias, etc, agiganta-se a cada dia, tornando a utilização de métodos de Inteligência Artificial altamente produtivos e até necessários. No Brasil, a Agência Nacional de Vigilância Sanitária (ANVISA) é o órgão responsável pela análise e liberação de substâncias, assim como homologação de suas bulas. Atualmente o bulário da ANVISA contém o registro de aproximadamente 7912 medicamentos, um número extremamente alto de informações para que um ser humano possa análizar e compreender em tempo hábil. **Objetivo:** Oferecer métodos de pesquisas inteligentes para grandes volumes de dados, oferecendo apoio a profissionais da saúde. **Metodologia:** para efetuar a *Text Data Mining* é preciso um grande conjunto de dados para pesquisa e um conjunto de informações secundária, que serão ultilizados como parâmetro de busca do processo de *Text Data Mining*. Como dado secundário será ultilizado o *International Statistical Classification of Diseases and Related Health Problems (ICD)*, conhecido no Brasil como Código internacional de doenças (CID), que apresenta a listagem aproximada de 12422 variedades de sinais, sintomas, aspectos anormais, queixas, circunstâncias sociais e causas externas para ferimentos ou doenças. O uso de Inteligência Artificial na área da saúde trará importantes ganhos em termos de dinheiro, tempo e vidas e precisa ser ativamente defendido por cientistas de dados e epidemiologistas.

**Palavras-chave:** Text Data Mining. Inteligência Artificial. Brasil. Apoio à Conduta Clínica. Big Data.

**SUMÁRIO**

1 INTRODUÇÃO……………………………………………...…………………...…..4

1.1 Inteligencia Artificial……………..……………………………………………...…...5

1.2 Text Data Mining……….………..…………………………………………………...5

1.2.1 O que não é Text data mining? ……………………………………………………5

1.3 Sistemas Especialistas……………………………………………………………...5

1.4 Conduta Médica……...……………………………………………………………...5

1.5 Bulas…………………..……………………………………………………………...5

1.6 Código Internacional de Doenças…..…………………………………………......5

1.7 Crawler Web…………….…………………………………………………………...5

1.8 Linguagens de Programação e Ambiente de Desenvolvimento…..…………...5

1.9 Objetivo geral………………………………………………………………………...7

1.9.1 Objetivos Específicos………………………………………………………………..7

# 1 introdução

**1.1 Inteligência Artificial (IA)**

A Inteligência Artificial nasceu por volta de 1956 durante uma conferência de verão em Dartmouth College, nos Estados Unidos. Surgiu do anseio de realizar metodologias baseadas na forma como os seres humanos resolviam problemáticas, assim, ultilizando métodos, mecânicmos e algoritmos computacionais que simulam a inteligência humana. (SOUZA; TALON, 2013, p. 60 ).

**1.2 Text data mining (TDM)**

*Text Data Mining* é uma expressão na língua inglesa ligada ciência de dados cuja tradução é mineração de dados textuais. Se trata de métodos que organizam e agregam dados e detectam padrões, anomalias e associaçãoes desejadas ou relevantes. O surgimento do termo *Data mining* surgiu por volta de 1990, quando tomadas as proporções, percebeu-se a necessidade de desenvolver novos métodos computacionais de pesquisa e indexação de informação. *Data mining* se trata da análise do processo conhecido como KDD (*Knowledge Discovery in Databases*), conhecido no Brasil como "Descoberta de Conhecimento em Bases de Dado”.

A tecnologia de mineração de textos vem das técnicas de recuperação de informações, *Machine learning* que é um ramo do estudo de sistemas de Informação inteligentes que por sua vez é uma das aplicações notáveis da Inteligência Artificial e da descoberta tradicional de informações estruturadas, através do uso de bancos de dados e de procedimentos estatísticos. Mineração de textos é um conjunto de métodos usados para navegar, organizar, achar e descobrir informação em bases textuais. Pode ser vista como uma extensão da área de Data Mining, focada na análise de textos. Também é chamada de *Text Data Mining, Knowledge Discovery in Texts*. (BARRETO J. M.: Inteligência Artificial no Limiar do Século XXI, Capítulo 19, 2001.).

Mineração de textos é um conjunto de métodos usados para navegar, organizar, achar e descobrir informação em bases textuais. Pode ser vista como uma extensão da área de Data Mining, focada na análise de textos. Também é chamada de Text Data Mining, Knowledge Discovery in Texts. (PASSOS, 2006, p. 2).

Nos últimos anos a ciência de dados encontrou fortes indícios de que a próxima grande fronteira da epidemiologia será a análise de grandes bancos de dados (*Big data*). O crescimento do número de estudos multicêntricos, a quantidade de dados disponíveis, tem criado uma demanda por novas formas de análise de dados complexos e desestruturados – um conjunto de técnicas conhecido como *Data mining*. (CHIAVEGATTO FILHO, 2015, p. 326).

**1.2.1 O que não é Text data mining?**

*Data data mining* é uma metodologia de de busca diferente de um mecanismo de busca comum. Na busca o usuário já sabe o que quer encontrar. As tecnologias usadas em mineração de textos ajudam os usuários a descobrirem informações desconhecidas. *O Text data mining* é diferente de análise de constituintes, pois não é necessário formalizar toda a construção sintática do texto. *Data mining* não se pretende simular o comportamento humano. Text d*ata mining* não é mineração de dados, pois trabalha com textos utiliza algoritmos de mineração de dados além de outros convenientes. (PASSOS, 2006, p. 2).

**1.2.2 O que é Text data mining?**

*Data data mining* é uma metodologia de de busca diferente de um mecanismo de busca comum. Na busca o usuário já sabe o que quer encontrar. As tecnologias usadas em mineração de textos ajudam os usuários a descobrirem informações desconhecidas. *O Text data mining* é diferente de análise de constituintes, pois não é necessário formalizar toda a construção sintática do texto. *Data mining* não se pretende simular o comportamento humano. Text d*ata mining* não é mineração de dados, pois trabalha com textos utiliza algoritmos de mineração de dados além de outros convenientes. (PASSOS, 2006, p. 2).

**1.3 SISTEMAS ESPECIALISTAS (SE)**

**1.4 CONDUTA MÉDICA (CM)**

**1.5 BULAS**

**1.6 CÓDIGO INTERNACIONAL DE DOENÇAS (CID)**

**1.7 CRAWLER WEB**

**1.8 LINGUAGENS DE PROGRAMAÇÃO E AMBIENTE DE DESENVOLVIMENTO**

* Linguagem de desenvolvimento web: Hypertext Preprocessor (PHP) em sua versão 7.0.13.
* Servidor síncrono: Apache em sua versão 2.4.23 Handler.
* Banco de dados SQL relacional: MySQL 5.0.12.

# Biblioteca para analisar modelo de objeto de documento de páginas HTML: PHP Simple HTML DOM Parser 5.0.

# Framework de rastreamento web: PHPCrawl 0.8. Biblioteca PHP para analisar arquivos PDF e extrair elementos textuais: PDF parser 0.9.10.

# Versionamento Git através do https://GitHub.com.

* Ambiente UNIX: Linux Ubuntu 16.05.
* Editor de Textos: Atom 1.14.3.

**1.9 Objetivos**

**1.9.1 Objetivo geral**

Oferecer métodos de pesquisas inteligentes para o apoio de profissionais da saúde.

**1.9.2 Objetivos específicos**

* Adquirir todas as bulas registradas na ANVISA através de métodos de Data mining e salvá-los em um banco de dados SQL.
* Adquirir todos os Códigos Internacionais de Doenças através de pesquisa em dados disponíveis na comunidade de desenvolvimento GitHub e salvá-los em um banco de dados SQL.
* Ultilizar uma b[iblioteca PHP para analisar arquivos PDF e extrair elementos textuais dos dados obtidos da ANVISA.](http://www.pdfparser.org/)
* Criação de algoritmos de busca e análise de padrões ultilizado em bulas.
* Pesquisas inteligentes para relacionamentar áspectos e índices importantes das bulas e os sinais, sintomas, aspectos anormais, queixas, circunstâncias sociais e causas externas para ferimentos ou doenças listados no CID.

**1.9 Justificativa**

A medida que a quantidade de informação cresce, a capacidade de entendimento e controle sobre esses dados é drasticamente reduzida, tornando necessário o uso de algoritmos e técnicas computacionais para análise de dados.

*Coiera*(1998, p. 4) médico australiano especializado em Informática Médica descreve que:

O intelecto humano é formado por um conjunto complexo de fenômenos, e a Inteligência Artificial pode implementá-lo e quando um caso é complexo, raro, ou a pessoa que está fazendo o diagnóstico é inexperiente, um sistema especialista pode ajudar a achar um diagnóstico mais provável e seguro, baseado nos dados do paciente.

Existem várias pesquisas tentando aplicar a inteligência artificial na área médica através dos sistemas de apoio à decisão médica, no entanto, estes sistemas ainda não estão sendo utilizados de forma natural, por não se preocuparem com heurísticas de usabilidade. Solucionar problemas e oferecer interfaces de usuários amigáveis é também um grande desafio para a Inteligência Artificial, uma vez que se tratam sempre de métodos de algoritmos que envolvem cálculos complexos, tornando a usabilidade para usuário final inviável. (Westphal, 2003, p. 172 apud SOUZA; TALON, 2013, p. 60).

A utilização de métodos de *Text Data Mining* em saúde permite que possamos associar dados de diferentes áreas da saúde, auxiliando de forma evidente a decisão clínica de um profissional da saúde.

A indexação de bulas e cids de maneira inteligente atravéz de algoritmos precisa ser investigada, criando algoritmos que possam auxiliar de forma extremamente rápida e confiável a conduta clínica.