

Área: CIÊNCIA DA COMPUTAÇÃO

Título : INDEXAÇÃO AUTOMÁTICA PARA EXTRAÇÃO DE CONHECIMENTO DE BASE DE DADOS MÉDICA RELACIONADA A DOENÇAS PÉPTICAS.

Autor(es): DANIEL DE FAVERI HONORATO, HUEI DIANA LEE, FENG CHUNG WU, RENATO BOBSIN MACHADO, ANTONIO PIETROBOM NETO

E-mail Apresentador: dfaverih@hotmail.com

Orientador(a): HUEI DIANA LEE

E-mail Orientador(a): huei@unioeste.br

Instituição: LABORATÓRIO DE BIOINFORMÁTICA - LABI/UNIOESTE - FOZ. HOSPITAL MUNICIPAL DE PAULÍNIA. INSTITUTO DE TECNOLOGIA EM AUTOMAÇÃO E INFORMÁTICA - ITAI.

Resumo:

A evolução tecnológica tem possibilitado enorme crescimento das bases de dados. Desse modo, torna-se cada vez mais necessário o desenvolvimento de processos para a análise e a compreensão dessas informações. Um desses processos é a Extração de Conhecimento em Base de Dados. Este trabalho está inserido no projeto de Análise Inteligente de Dados aplicada a Doenças Pépticas que está sendo desenvolvido no Laboratório de Bioinformática – Unioeste em parceria com o Hospital Municipal de Paulínia (HMP). Doenças pépticas gastroduodenais apresentam alto índice de incidência na população e despertam cada vez mais interesse da pesquisa nessa área. Neste trabalho foi utilizada uma base de dados relacional obtida no Serviço de Endoscopia Digestiva Alta (EDA) do HMP. Esses dados são provenientes de exames realizados nesse hospital, totalizando 1950 casos descritos por 58 atributos cada um. Neste trabalho foi realizada parte da etapa de pré-processamento a qual foi dividida em duas fases. A primeira fase constituiu na conversão da base de dados original em uma nova base contendo apenas atributos de interesse, especificamente nove atributos, e entre eles, o atributo laudo que armazena o resultado do exame de EDA em formato texto. Ainda nessa fase, realizou-se a aplicação de um algoritmo de pesquisa sobre o atributo laudo com o objetivo de localizar, dentre todos exames realizados, os casos em que os pacientes eram submetidos à biópsia durante o EDA. A partir do atributo laudo também serão extraídos novos atributos para compor a base de dados, a qual será utilizada na etapa de extração de conhecimento. Para que isso seja possível, é necessária a conversão do atributo laudo para um formato adequado. A segunda fase é caracterizada pela construção de um dicionário de palavras. O dicionário está sendo construído com o auxílio de um especialista considerando-se quatro fases: identificação de frases únicas, remoção de stopwords, normalização das palavras por meio da aplicação de stemming e inserção das palavras no dicionário. Após a conclusão dessa fase, o dicionário será usado para auxiliar na transformação do atributo laudo em uma base de dados, que complementada com outros atributos já selecionados, poderá ser utilizada na etapa de extração de conhecimento. As principais contribuições deste trabalho estão relacionadas à redução do tempo na fase de preparação dos dados e na possibilidade da utilização da metodologia proposta em problemas de outros domínios.

Palavras-chave: *ENDOSCOPIA DIGESTIVA ALTA, MINERAÇÃO DE DADOS, PRÉ-PROCESSAMENTO.*

Modalidade de atuação: *INICIAÇÃO CIENTÍFICA - ITAI*