

Metodologia de Mapeamento de Laudos Endoscópicos

Everton Alvares Cherman¹, Huei Diana Lee¹, Daniel de Faveri Honorato²,
Renato Bobsin Machado¹, Feng Chung Wu^{1,3}

¹Laboratório de Bioinformática — LABI,
Universidade Estadual do Oeste do Paraná — UNIOESTE,
Parque Tecnológico Itaipu — PTI

²Universidade de São Paulo — USP

³Universidade Estadual de Campinas — UNICAMP

1. Objetivos

Processos computacionais que podem auxiliar na extração e análise de conhecimento, como o de Descoberta de Conhecimento em Bases de Dados (DCBD), necessitam que os dados estejam no formato atributo-valor [1]. Neste trabalho é proposta uma melhoria na metodologia para o mapeamento de Laudos Médicos (LM) para uma Base de Dados (BD) estruturada apresentada em [2], bem como um estudo de caso dessa metodologia proposta.

2. Material e Métodos

Para a realização deste trabalho foram utilizados 610 LMs de Endoscopia Digestiva Alta (EDA) contendo informações relacionadas ao estômago, as quais foram mapeadas para uma BD estruturada utilizando a metodologia composta por duas fases. Na primeira fase é construído um dicionário do domínio do conhecimento por meio de reuniões com os especialistas com o auxílio de um arquivo de padronização e de um Conjunto de Frases Únicas (CFU) gerado a partir de todas as frases contidas nos LMs. A estrutura do dicionário consiste em locais e características associadas a esses locais. Neste trabalho foi incluído o nível de subcaracterísticas, possibilitando uma descrição mais detalhada das características que compõem os locais. Na segunda fase, com base no dicionário definido na fase anterior, é realizado o mapeamento das informações extraíndo os atributos e seus valores por meio de um novo algoritmo adequado ao aumento do nível de detalhamento dos atributos. Desse modo, cada atributo da BD pode ser formado pela relação local e característica ou local, característica e subcaracterística.

3. Resultados e Discussão

O CFU extraído a partir dos 610 LMs representou 4,91% do total de 5213 frases originais, verificando uniformidade na descrição dos LMs. A BD foi composta por 1081 atributos, incluindo a definição de atributos estabelecidos como importantes pelos especialistas e que não foram encontrados nesses 610 LMs. O algoritmo utilizado para o mapeamento foi adequado ao aumento do nível de detalhamento das informações, permitindo em conjunto com o dicionário construído, que todas as informações descritas corretamente fossem mapeadas integralmente para a BD.

4. Conclusões

Neste trabalho, foi proposta uma melhoria na estrutura do dicionário e no algoritmo de preenchimento da BD. O dicionário definido poderá ser usado para mapear novos conjuntos de LMs com informações mais detalhadas, acelerando o processo de mapeamento e evitando a subjetividade do preenchimento manual da BD.

5. Referências Bibliográficas

- [1] Fayyad, U. M., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, vol. 17:37–54.
- [2] Honorato, D. D. F., et al. (2005). Uma metodologia para auxiliar no processo de construção de bases de dados. In *Anais do V Encontro Nacional de Inteligência Artificial, XXV Congresso da Sociedade Brasileira de Computação*, pages 593–601, Porto Alegre, RS.