

Utilização de ontologias para auxiliar no mapeamento de laudos médicos para bases de dados estruturadas: estudo de caso em laudos de Endoscopia Digestiva Alta

Luiz H.D. Costa¹, Carlos A. Ferrero¹, Huei D. Lee¹, Cláudio S.R. Coy², João J. Fagundes², Feng C. Wu^{1,2}

¹Laboratório de Bioinformática – LABI,

Universidade Estadual do Oeste do Paraná – UNIOESTE, Parque Tecnológico Itaipu – PTI

²Departamento de Moléstias do Aparelho Digestivo – DMAD, Serviço de Coloproctologia Faculdade de Ciências Médicas – FCM, Universidade Estadual de Campinas – UNICAMP

1. Objetivos

Dados médicos comumente se apresentam na forma de Laudos Médicos – LM. No entanto, para que seja possível a aplicação de métodos computacionais para Extração de Conhecimento – EC, é necessário que estejam representados num formato estruturado. Neste trabalho, que tem como objetivo a construção de ontologias para auxiliar no processo de mapeamento de conhecimento médico, é apresentado um estudo de caso que consiste na utilização de uma ontologia para auxiliar no mapeamento de LM textuais de Endoscopia Digestiva Alta – EDA – para Bases de Dados – BD.

2. Material e Método

O Método de Mapeamento – MM – dos LM textuais para BD estruturadas [1] é dividido em duas fases. Na fase (1) é realizada, em conjunto com especialistas, a identificação de padrões nos LM, a geração de n -gramas e do arquivo de padronização. Em seguida, com base nos artefatos gerados anteriormente e em conjunto com especialistas, são definidos os atributos da BD e é construído um Dicionário do Conhecimento – DC – do domínio dos LM. Na fase (2), as padronizações são aplicadas aos LM, os quais são processados por um Algoritmo de Busca e Preenchimento – ABP – o qual, com base no DC, identifica os termos e preenche, os atributos da BD.

O DC é representado como uma classificação hierárquica dos termos presentes nos LM. Ele consiste em uma lista de locais, na qual cada local possui uma lista de características relacionadas ao local e cada característica pode possuir uma lista de subcaracterísticas. Cada relação local-característica ou local-característica-subcaracterística apresenta explicitamente o atributo e o valor a ser preenchido no mesmo. Porém, o DC não é completo o suficiente, pois limita a representação dos relacionamentos entre os termos.

Desse modo, foi proposta a construção de uma ontologia para representação do domínio

relacionado aos termos dos LM. Essa ontologia substitui o DC no MM. Desse modo, na fase (2), foi desenvolvido um novo ABP para mapear o conteúdo dos LM padronizados para os atributos definidos, com base na ontologia.

A nova abordagem foi aplicada às informações relacionadas ao esôfago do conjunto de 609 LM textuais de EDA realizados no período de 2000 a 2005 no Hospital Municipal de Paulínia.

3. Resultados e Discussão

A nova abordagem do MM aplicada ao conjunto de 609 LM de EDA permitiu o mapeamento correto de 100% dos atributos esperados, em concordância com a abordagem que utiliza o DC [1].

O mapeamento do conhecimento relacionado a LM de EDA por meio da ontologia, por sua vez, permite uma representação com mais relações e mais completa do que com o DC. Essa representação pode ser facilmente expandida e melhorada, permitindo um mapeamento mais completo do conjunto de LM, assim como pode ser utilizada por outras tarefas para EC.

4. Conclusão e Trabalhos Futuros

Neste trabalho, constatou-se que o uso de ontologias, no MM de LM textuais para BD estruturadas, além de ser tão eficaz quanto a utilização do DC, representa melhor o domínio relacionado aos termos dos LM. Como trabalhos futuros destacam-se: a melhoria da capacidade de mapeamento de LM de EDA por meio da expansão da ontologia; e a aplicação da nova abordagem a LM de outras áreas, como Coloscopia e Manometria Anorretal.

5. Referências Bibliográficas

- [1] Honorato DDF, Cherman EA, Lee HD, Monard MC, Wu FC. Construction of an Attribute-Value Representation for Semi-structured Medical Findings Knowledge Extraction. CLEI Electronic Journal. 2008;11(2):1–12.