



## **Ontologias Como Apoio Ao Mapeamento De Laudos Médicos De Endoscopia Digestiva Alta Para Bases De Dados Estruturadas**

Luiz Henrique Dutra da Costa (PIBIC/Unioeste), Carlos Andres Ferrero, Cláudio Saddy Rodrigues Coy, João José Fagundes, Feng Chung Wu, Huei Diana Lee (Orientadora), e-mail: hueidianalee@gmail.com.

Universidade Estadual do Oeste do Paraná/ Centro de Engenharias e Ciências Exatas / Laboratório de Bioinformática / Curso de Ciência da Computação / Foz do Iguaçu, PR.

### **Ciências Exatas e da Terra, Ciência da Computação.**

**Palavras-chave:** Mineração de Dados, Representação do Conhecimento, Aprendizado de Máquina.

#### **Resumo:**

A aplicação de técnicas de Análise Inteligente de Dados auxilia na tarefa de análise de grandes conjuntos de dados, que necessitam estar em um formato estruturado. No entanto, em diversas áreas, como a área médica, os dados frequentemente se encontram em um formato pouco estruturado. Neste trabalho foram utilizadas ontologias para apoiar no mapeamento de laudos médicos textuais para bases de dados estruturadas, que era realizado utilizando uma outra estrutura específica, denominada de dicionário, para representação do conhecimento referente aos laudos. Comparando o conjunto de atributos preenchidos pelas duas abordagens, o novo método apresentou uma equivalência de 100% em relação ao outro método para o conjunto de 609 laudos e 99,976% para o conjunto de 3647 laudos. Além disso, com a nova abordagem é possível expandir facilmente os termos e os atributos mapeados a fim de mapear informações, as quais não eram mapeadas com o método não apoiado por ontologias.

#### **Introdução**

O avanço tecnológico tem possibilitado o armazenamento de grandes repositórios de dados. Para a análise desses dados, no entanto, o uso de técnicas de análise manual pode não prover resultados completos, pois, com o crescimento da quantidade e da complexidade dos dados, essas técnicas podem ser propensas a erros e/ou subjetividade.

Uma forma de tornar a tarefa de análise de dados mais eficiente é por meio da aplicação de técnicas de Análise Inteligente de Dados como, por exemplo, as associadas ao processo de Mineração de Dados (MD) (Fayyad et al., 1996; Solange, 2003). No entanto, é necessário que os dados estejam em um formato adequado para que essas técnicas possam ser utilizadas.

Na área médica, é comum que os dados sejam armazenados de modo semiestruturado ou desestruturado, como em imagens e laudos médicos. Para que o processo de MD possa ser aplicado a esses dados é

necessário que eles estejam em um formato estruturado, por exemplo, a tabela atributo-valor (TAV).

Este trabalho tem como objetivo a utilização de ontologias para auxílio ao mapeamento de laudos médicos de Endoscopia Digestiva Alta (EDA) e faz parte do projeto de Análise Inteligente de Dados desenvolvido em parceria entre o Laboratório de Bioinformática (LABI) da Universidade Estadual do Oeste do Paraná (UNIOESTE) / Foz do Iguaçu, o Laboratório de Inteligência Computacional (LABIC) da Universidade de São Paulo (USP) / São Carlos e o Serviço de Coloproctologia da Faculdade de Ciências Médicas da Universidade Estadual de Campinas — UNICAMP/Campinas.

## Materiais e métodos

Os laudos de EDA, assim como outros laudos médicos, são descritos por meio de texto semiestruturado ou desestruturado, composto por sentenças que descrevem as observações do médico em relação ao paciente.

O método de mapeamento de laudos médicos para bases de dados estruturadas (Honorato et al., 2005), desenvolvido no LABI, é executado em duas fases, conforme apresentado na Figura 1.

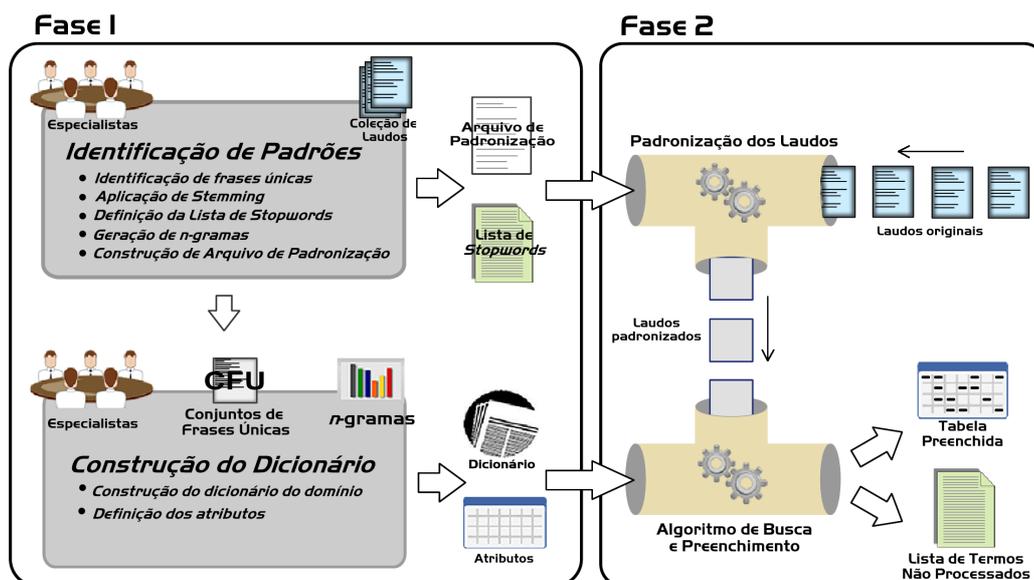


Figura 1 – Método de mapeamento de laudos médicos.

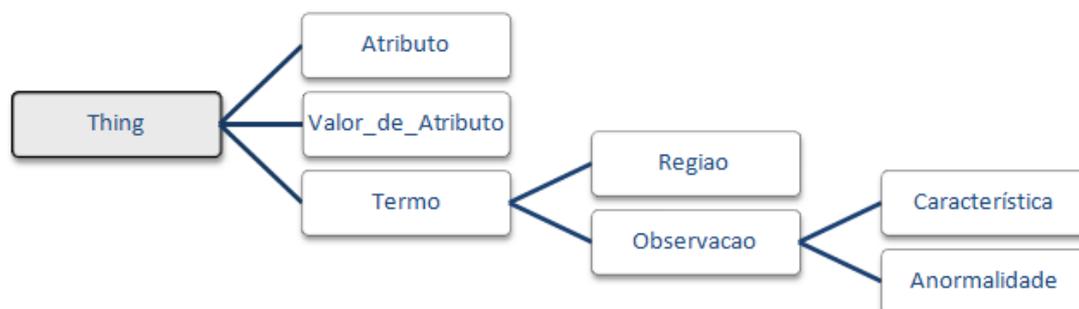
Na Fase 1 é realizada a identificação de padrões por meio de: (1) **Identificação das frases únicas** que compõe o conjunto de laudos; (2) **Aplicação de Stemming** para a redução das palavras ao seu radical a fim de reduzir o conjunto de frases únicas; (3) **Definição da lista de stopwords** contendo as palavras que não são relevantes para o mapeamento das informações; (4) **Geração de n-gramas** a fim de identificar os termos mais frequentes nos laudos; (5) **Construção do Arquivo de Padronização (AP)** contendo as padronizações que serão aplicadas aos laudos. Em seguida é realizada a definição dos atributos da TAV e a construção do dicionário do conhecimento do domínio, no qual é definido um conjunto de regras que

determinam o mapeamento de frases do laudo para os atributos da TAV. Cada uma dessas regras descreve um conjunto de “local – característica” ou “local – característica – subcaracterística” que define o preenchimento dos atributos da TAV.

Na Fase 2 é realizada, primeiramente, a padronização dos laudos que serão mapeados por meio da remoção de *stopwords* e da aplicação das padronizações definidas no AP. Em seguida os laudos padronizados são processados pelo Algoritmo de Busca e Preenchimento, que utiliza o dicionário do conhecimento como apoio para mapear os laudos para a TAV.

Neste trabalho propõe-se a utilização de uma ontologia para substituir o dicionário do conhecimento na tarefa de representar o conhecimento relacionado aos laudos e ao seu processo de mapeamento. A ontologia é utilizada para representar os conceitos e os objetos do domínio de mapeamento dos laudos e dos atributos da TAV, ou seja, representará os termos dos laudos, suas relações entre si e com os atributos da TAV, os quais também são representados na ontologia.

A ontologia é construída utilizando a linguagem OWL<sup>1</sup> em conjunto com o editor de ontologias Protégé<sup>2</sup>. Esse método foi aplicado à seção de esôfago de um conjunto de 609 laudos médicos de EDA, realizados no Hospital Municipal de Paulínia no período de 2000 a 2005. A estrutura geral dos conceitos da ontologia construída é apresentada na Figura 2.



**Figura 2** – Estrutura geral dos conceitos da ontologia construída.

Posteriormente, utilizando a mesma ontologia, a Fase 2 do método foi aplicada a outro conjunto de 3647 laudos de EDA, realizados no Hospital Municipal de Paulínia no período de 2000 a 2005.

## Resultados e Discussão

Com base no conjunto de atributos preenchidos pela primeira abordagem, o novo método apresentou uma equivalência de 100% para o conjunto de 609 laudos e 99,976% para o conjunto de 3647 laudos. Em relação à taxa de preenchimento dos atributos da TAV, como um todo, o mapeamento do conjunto de 3647 laudos realizado com método apoiado por ontologias preencheu 85,284% dos atributos da TAV e o realizado com o método do dicionário preencheu 85,279%.

<sup>1</sup> <http://www.w3.org/2004/OWL/>

<sup>2</sup> <http://protege.stanford.edu/>



Em relação aos atributos preenchidos da TAV que diferem entre os dois métodos foram identificados: três laudos com mapeamento errôneo de um atributo, pois apresentavam um padrão que não estava contido na ontologia; um laudo apresentou preenchimento de um atributo que não era interpretado corretamente pois existem duas descrições para o mesmo atributo, uma mais geral e outra mais específica, sendo necessário o desenvolvimento de regras para que o método dê prioridade às descrições específicas e depois às gerais.

A ontologia também auxilia na representação do conhecimento relacionado aos laudos de EDA. Essa ontologia poderá ser expandida a fim de conter conceitos não considerados anteriormente e possibilita a expansão dos atributos mapeados a fim de obter uma base de dados mais completa e refinada, evitando a perda de informação no mapeamento dos laudos.

### **Conclusões**

O novo método proposto neste trabalho, que utiliza ontologias como apoio para o processo de mapeamento de laudos médicos para bases de dados estruturadas, mostrou-se eficiente, tendo um nível muito similar de eficácia em comparação com o método que utilizava um dicionário de conhecimento. Uma das vantagens dessa nova abordagem é que a ontologia construída auxilia no mapeamento do conhecimento relacionado à estrutura dos laudos médicos, que poderá ser utilizada em estudos e aplicações futuras.

Essa representação também poderá ser facilmente expandida a fim de contemplar uma gama maior de informações a respeito dos laudos e a respeito do domínio relacionado à EDA. Além disso, esse método poderá ser aplicado no mapeamento de outros tipos de laudos, como laudos de coloscopia e manometria anorretal.

### **Agradecimentos**

À Unioeste pelo auxílio por meio de bolsa PIBIC.

### **Referências**

Fayyad, U. M.; Shapiro, G. P.; Smyth, P. From data mining to knowledge discovery: An overview. *American Association for Artificial Intelligence*, Menlo Park, CA, 1996, p. 1–30.

Honorato, D. F.; Lee, H. D.; Monard, M. C.; Wu, F. C.; Machado, R. B.; Neto, A. P.; Ferrero, C. A. Uma metodologia para auxiliar no processo de construção de bases de dados estruturadas a partir de laudos médicos. In *Anais do Encontro Nacional de Inteligência Artificial*, São Leopoldo/ RS, Brasil, 2005, p. 593–601.



Rezende, S. O. Sistemas Inteligentes: Fundamentos e Aplicações. Barueri, Brasil: Editora Manole, 2003.