

Mapeamento de Laudos Médicos de Endoscopia Digestiva Alta Apoiados por Ontologias

Luiz Henrique Dutra da Costa¹, Carlos Andrés Ferrero¹, Hwei Diana Lee¹,
Cláudio Saddy Rodrigues Coy², João José Fagundes², Feng Chung Wu^{1,2}

¹Centro de Engenharias e Ciências Exatas – Universidade Estadual do Oeste do Paraná
Laboratório de Bioinformática – LABI
Parque Tecnológico Itaipu – PTI
Caixa Postal 39, 85869-970 – Foz do Iguaçu, PR, Brasil

²Faculdade de Ciências Médicas – Universidade Estadual de Campinas
Serviço de Coloproctologia
Caixa Postal 6111, 13083-970 – Campinas, SP, Brasil

{lhdc90, anfer86, hueidianalee, wufengchung}@gmail.com

Resumo. *Em ambientes médicos, informações a respeito de dados clínicos de pacientes são frequentemente descritas no formato de laudos textuais. Para analisar essas informações por processos computacionais devem ser transformados para um formato estruturado. Nesse contexto, elaborou-se um método para o mapeamento de laudos para esse formato, o qual foi ampliado pela representação do conhecimento por meio de ontologias. Neste trabalho, esse método foi aplicado inicialmente em 609 laudos de exames de Endoscopia Digestiva Alta. A ontologia construída nesse processo foi utilizada, posteriormente, para mapear outros 3647 laudos. Com base nos resultados, observou-se 85,824% de eficiência de preenchimento das informações contidas nos laudos.*

Abstract. *On medical environments, information about patients' clinical data is frequently described on text reports. The analysis of this information by computer process requires structured data. In this context, it was developed a method to map medical reports into a structured format, which was extended by the ontology-based knowledge representation. In this paper, this method was first applied to a set of 609 High Digestive Endoscopy medical reports. The ontology built on this process was used, lately, to map another set of 3647 medical reports. Based on the results, it is observed 85,824% efficiency rate on filling in information contained on the medical reports.*

1. Introdução

A evolução tecnológica e a disseminação do uso de computadores têm possibilitado o armazenamento de dados nas mais diversas áreas. Para analisar o grande volume de dados produzido, no intuito de extrair algum conhecimento útil, a análise manual se torna inviável e de alto custo, devido ao seu tamanho crescente, tornando-se necessário aplicar métodos computacionais para dar suporte a essa tarefa. Dentre os principais objetivos dessa análise destaca-se a construção de modelos que representem o conhecimento presente nesses dados. Nesse sentido, a Mineração de Dados — MD — engloba métodos que podem ser utilizados para dar apoio a essa tarefa [Han and Kamber 2006].

O processo de MD pode ser dividido em três etapas [Rezende 2003]: pré-processamento, extração de padrões e pós-processamento. A primeira etapa, pré-processamento, tem como objetivo preparar os dados para a etapa seguinte por meio da seleção de dados e atributos e da limpeza dos dados. A segunda etapa, extração de padrões, tem como objetivo construir modelos que representem o conhecimento embutidos nos dados. E por último, a terceira etapa, pós-processamento, tem como objetivo a avaliação e a interpretação dos resultados em conjunto com especialistas de domínio.

Das etapas que envolvem esse processo, a de pré-processamento é a que apresenta maior custo e inclui a adequação dos dados para um formato estruturado, como o formato atributo-valor, que permite a aplicação de métodos para extração de padrões. Na área médica, esses dados se apresentam em diversos formatos [Shortliffe and Barnett 2006] como imagens, formulários e laudos textuais. No entanto, esses formatos, não são adequados para a aplicação de métodos de extração de padrões e o mapeamento manual das informações contidas nesses dados pode apresentar erros, além de ser lento e subjetivo [Lee 2005, Honorato et al. 2005, Cherman et al. 2008, Honorato et al. 2009]. Nesse sentido, torna-se necessária a aplicação de métodos computacionais para mapear informações contidas nesses dados para formatos estruturados.

Para o mapeamento de informações contidas em laudos médicos textuais é necessária a construção de algoritmos capazes de compreender o que está representado nos termos descritos nas sentenças do laudo médico [Friedman and Johnson 2006]. Para auxiliar na tarefa de compreensão dos termos presentes nos laudos médicos podem ser utilizadas estruturas específicas para representação de conhecimento [Chute 2005, Revere and Fuller 2005, Bodenreider and Burgun 2005, Cherman et al. 2008, Honorato et al. 2009]. Uma das formas é por meio de ontologias, que constituem uma representação formal dos conceitos de determinado domínio e de seus relacionamentos [Gruber 1993, Carvalheira 2007].

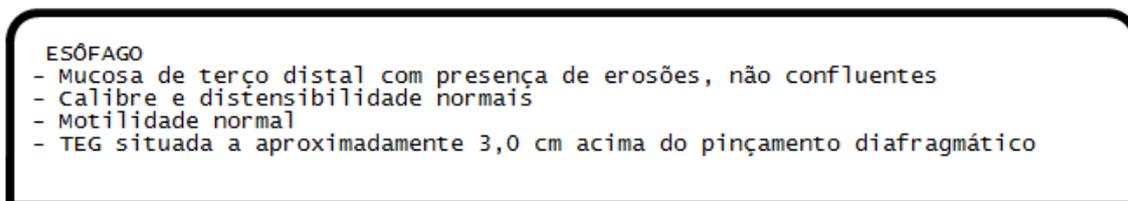
O objetivo deste trabalho consiste em aplicar um método de mapeamento para um conjunto de 609 laudos médicos de exames de Endoscopia Digestiva Alta — EDA — e construir uma ontologia, para posteriormente aplicar o método e a ontologia construída em diferentes conjuntos de laudos de EDA, que não tenham sido considerados na construção da ontologia. Este trabalho constitui parte do projeto 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.

Este trabalho está organizado da seguinte maneira: na Seção 2 são descritos os métodos e as ferramentas utilizadas para o desenvolvimento deste trabalho; na Seção 3 são apresentados os resultados e discutidas as principais contribuições do trabalho; na Seção 4 são apresentados a conclusão e os trabalhos futuros.

2. Material e Método

Em hospitais e clínicas médicas, frequentemente, os especialistas descrevem as informações no formato de laudos descritos textualmente. As sentenças que constituem esses laudos representam, em linguagem natural, as observações a respeito da saúde do

paciente a partir da anamnese ou da realização de exames complementares. O exame de EDA é um exame complementar que é descrito dessa forma pelo especialista. Na Figura 1, é apresentado um exemplo de laudo médico de EDA.



ESÔFAGO
- Mucosa de terço distal com presença de erosões, não confluentes
- Calibre e distensibilidade normais
- Motilidade normal
- TEG situada a aproximadamente 3,0 cm acima do pinçamento diafragmático

Figura 1. Exemplo de parte de um laudo de EDA utilizado

Como mencionado, para que as informações registradas nesses laudos possam ser analisadas por métodos computacionais que permitam a extração de informações relevantes e a construção de modelos preditivos é estritamente necessário que estejam representadas em um formato estruturado. O formato comumente utilizado é o atributo-valor, o qual é constituído por uma tabela na qual as linhas representam os laudos e as colunas os aspectos relevantes que podem ser encontradas nos laudos e as células representam os valores dessas informações em cada exame.

O mapeamento das informações contidas nos laudos para esse formato pode ser realizado de modo manual, porém, quando considerada uma grande quantidade de laudos esse processo pode tornar-se demorado e influenciado por fatores subjetivos dos que realizam essa tarefa. Desse modo, no intuito de reduzir o tempo consumido para a realização dessa tarefa e a subjetividade envolvida na transcrição das informações, foi desenvolvido um método de mapeamento de laudos médicos [Honorato et al. 2008].

2.1. Método Proposto

O método proposto em [Honorato et al. 2005] é semi-automático e interativo e consiste na identificação de padrões textuais dentro dos laudos em conjunto com especialistas do domínio para, posteriormente, mapear as informações contidas nos laudos por meio de um algoritmo computacional para uma base de dados estruturada. Na Figura 2 é ilustrado o método proposto constituído de duas fases.

Na primeira fase do método são realizadas duas etapas, a identificação de padrões e a construção do dicionário do domínio, as quais são desenvolvidas junto com especialistas do domínio. A primeira etapa é composta pelas seguintes tarefas:

Identificação de frases únicas: nesta tarefa são identificadas todas as frases contidas nos laudos e eliminadas todas frases repetidas, de modo a selecionar um exemplar de cada frase, formando um Conjunto de Frases Únicas - CFU;

Definição da lista de *stopwords*: nesta tarefa é definida uma lista de palavras que serão filtradas dos laudos para reduzir novamente o CFU. Essas palavras são definidas em conjunto com especialistas como sendo não-importantes para o mapeamento das informações, como artigos, conjunções e preposições;

Construção do arquivo de padronização: nesta tarefa é construído um arquivo de padronizações no qual são definidas palavras ou expressões que serão substituídas por outras equivalentes a fim de unificar a descrição dos termos;

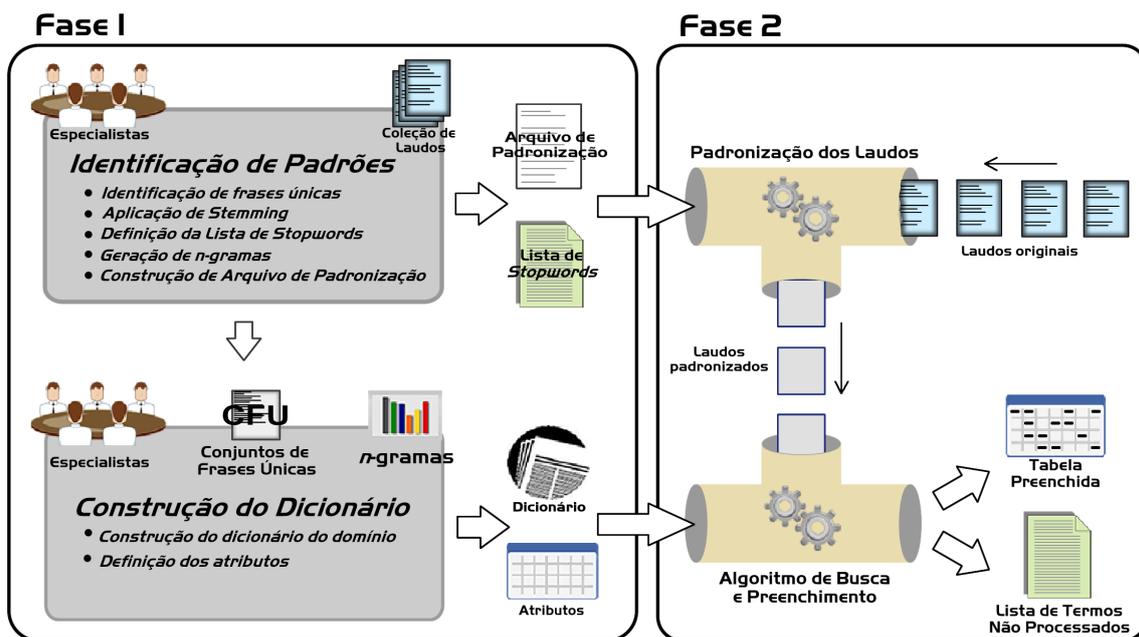


Figura 2. Método de mapeamento de laudos médicos para tabela atributo-valor

Geração de n -gramas: nesta tarefa é realizada a geração de n -gramas, a qual permite identificar as unidades terminológicas de maior frequência nos laudos.

E a segunda etapa, construção do dicionário, é composta pelas seguintes tarefas:

Construção do dicionário de domínio: nesta tarefa é construído o dicionário, que consiste em uma estrutura hierárquica de locais, em que cada local possui uma lista de características que podem ser encontradas nesse local e, em que cada característica, por sua vez, pode possuir uma lista de subcaracterísticas. Cada local representa uma região anatômica, as características representam as possíveis anormalidades observadas no exame em determinado local e as subcaracterísticas representam possíveis especificações das características. Cada sequência de local e característica ou local, característica e subcaracterística está diretamente relacionada com os atributos que constituem a base de dados, bem como com os valores que podem ser preenchidos para cada atributo;

Definição da tabela atributo-valor: nesta tarefa, com base em interações com especialistas e nas tarefas da fase identificação de frases únicas, são definidos os atributos que irão compor a tabela atributo-valor e os possíveis valores que cada atributo pode assumir na tabela.

A partir do desenvolvimento dessas duas etapas, na segunda fase, é realizada a padronização dos laudos. São utilizados o arquivo de padronização e a lista de *stopwords* construídos na primeira fase, no intuito de generalizar determinadas informações contidas nos laudos abstraído parte do conteúdo. Para processar essas informações foi construído o algoritmo de busca e preenchimento que, a partir dos laudos padronizados e com base no dicionário construído e na tabela de atributos definida na primeira fase. São identificadas as informações para cada laudo e armazenadas na base de dados. Esse processo, ao final, produz uma lista contendo todos os termos que não foram processados pelo algoritmo.

```
esofago
esofago_inferior erosao_sim nao confluentes
calibre normal
distensibilidade normal
motilidade normal
teg gi
```

Figura 3. Exemplo de parte de um laudo de EDA padronizado

2.2. Representação de Conhecimento por meio de Ontologias

As ontologias são representações de conhecimento por meio da definição de categorias (classes), propriedades (atributos) e relações entre indivíduos, estes que representam um dado objeto existente no domínio de conhecimento. Em [Uschold and Gruninger 1996] foi proposto um método para a construção de ontologias, constituído de três etapas:

Definição de escopo e objetivo: nesta etapa são definidos o domínio e o grau de especificidade da ontologia, os requisitos que devem ser atendidos e as ferramentas e a linguagem para representar o conhecimento;

Construção: nesta etapa é realizada a construção da ontologia com base nos conceitos que constituem parte do escopo definido durante a primeira etapa, considerando também os relacionamentos existentes entre esses conceitos;

Avaliação: nesta etapa, a partir da especificação inicial dos requisitos que a ontologia deve atender, é avaliada a expressividade e a consistência da ontologia. Caso algum requisito não tenha sido satisfeito retorna-se à etapa de construção da ontologia. Este processo é iterativo e interativo e deve prosseguir até que todos os requisitos sejam atendidos.

A ferramenta computacional utilizada para a construção da ontologia foi o *software* Protégé¹, que consiste em um dos ambientes mais utilizados para esse fim, e a linguagem de descrição das ontologias construídas com o Protégé foi a *Web Ontology Language* — OWL. Neste trabalho utiliza-se a sublinguagem OWL DL (*Description Logic*) que inclui todas as construções da OWL, porém com restrições de uso que garantem a completude e a computabilidade da ontologia [McGuinness and van Harmelen 2004].

Com base no formato dos laudos de EDA, os principais conceitos representados no dicionário foram determinados como classes, conforme apresentado na Figura 4.

Nessa figura, *Thing* é a classe principal que corresponde a todos os indivíduos, ela contém todas as outras classes. As classes *Atributo* e *Valor_de_Atributo* estão diretamente envolvidas com a tabela atributo-valor, em que a classe *Atributo* refere-se aos atributos da tabela atributo-valor e o atributo *Valor_de_Atributo* refere-se aos possíveis valores que pode assumir cada atributo. Ambas as classes produzem subclasses as quais não são demonstradas na Figura 4, pois constituem apenas agrupamentos de atributos e valores de atributos afins. A classe *Termo* corresponde aos termos presentes nos laudos padronizados e consiste em dois tipos de termos principais: *Regiao*, que refere-se aos termos das porções anatômicas e *Observacao*, que refere-se aos termos que representam características (a classe *Caracteristica*) ou anormalidades (a classe *Anormalidade*) observadas pelo médico e descritas nos laudos.

¹<http://protege.stanford.edu/>

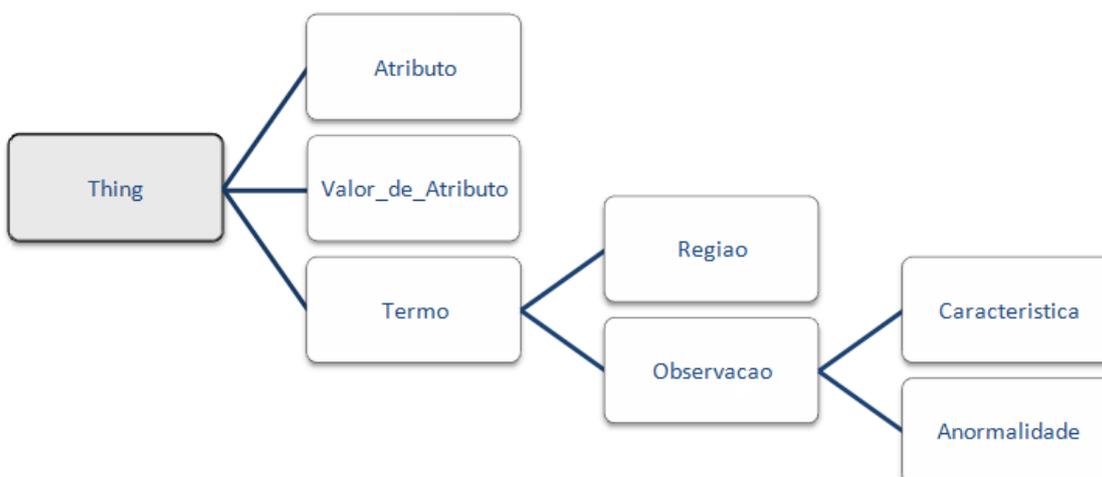


Figura 4. Representação geral dos conceitos da ontologia construída na primeira iteração

2.3. Algoritmo de Mapeamento

A partir da representação ontológica dos conceitos relacionados ao exame de EDA envolvidos nos laudos foi construído um algoritmo para o mapeamento das informações presentes nos laudos para uma base de dados estruturada. Desse modo, o método utilizado para desenvolvimento deste trabalho considera a representação ontológica mencionada no lugar do dicionário do conhecimento construído na segunda etapa do método proposto em [Honorato et al. 2005].

Esse algoritmo de busca e preenchimento foi desenvolvido na linguagem de programação Java² com o auxílio de duas *Application Programming Interfaces* — API — para processamento da ontologia: o *Pellet Reasoner*³ e a *OWL API*⁴. As etapas principais do algoritmo para o mapeamento de cada laudo são:

Divisão do laudo em sentenças: Nesta etapa o laudo é dividido em um conjunto de sentenças;

Processamento das sentenças: Nesta etapa os termos são identificados e a cada termo é vinculada uma propriedade, conforme sua classificação definida na ontologia, e os termos não identificados são adicionados à uma lista de termos não processados. Em seguida o conjunto de propriedades mapeadas é associado segundo as regras de associação definidas e são identificados quais atributos devem ser preenchidos, assim como os valores que serão preenchidos. Essa etapa então é repetida em cada sentença do laudo até que todas tenham sido processadas. Cria-se então um conjunto de atributos e seus respectivos valores para esse conjunto de sentenças.

Após a execução dessas três etapas para todos os laudos, obtêm-se uma tabela atributo-valor contendo as informações mapeadas a partir dos laudos e também uma lista de termos não-mapeados. É importante ressaltar que o algoritmo é de complexidade linear em relação ao número total de termos a serem mapeados, isto é, o tempo de execução do

²<http://java.sun.com/>

³<http://clarkparsia.com/pellet/>

⁴<http://owlapi.sourceforge.net/>

algoritmo é linearmente dependente da quantidade de laudos a serem processados e da quantidade de termos a serem processados em cada laudo.

2.4. Delineamento Experimental

Para o desenvolvimento deste trabalho foram utilizados dois conjuntos de laudos médicos. O primeiro conjunto de laudos — CL1 — consiste de 609 laudos, realizados no período de 2000 a 2005 no Hospital Municipal de Paulínia — HMP — e o segundo conjunto — CL2 — consiste de 3647 laudos, realizados no mesmo período também no HMP. A partir disso, no Passo 1, foi aplicada a primeira fase do método proposto nos laudos contidos em CL1 e construídos os arquivos de padronização. Com base nos laudos contidos em CL1 e nos arquivos de padronização foi construída uma ontologia para representar o conhecimento do domínio descrito pelas frases analisadas e o algoritmo que permite mapear as informações contidas nos laudos para o formato estruturado, com base na ontologia construída.

No Passo 2, os arquivos de padronização construídos na fase anterior são utilizados para padronizar o conjunto de laudos CL2. Após a padronização, a ontologia e o algoritmo foram utilizados para mapear as informações contidas nesses laudos, não analisados previamente, para uma base de dados estruturada no formato atributo-valor. Desse modo, é realizado um estudo comparativo para observar o desempenho dos arquivos de padronização e da ontologia no mapeamento do CL1 e CL2 e verificar a capacidade de generalização do método para o mapeamento de informações que não foram utilizadas durante a primeira etapa do método proposto.

3. Resultados Parciais e Discussão

A aplicação do método anterior preenchendo 85,279% dos valores dos atributos da tabela atributo-valor. O método foi aplicado a um conjunto de 3647 laudos de EDA, apenas na secção de esôfago, e preencheu 85,284% dos valores dos atributos da tabela atributo-valor. Ao comparar os resultados verifica-se a presença de problemas em um conjunto de 12 laudos dos 3647 aos quais o método foi aplicado, esse conjunto é subdividido em três subconjuntos:

O primeiro subconjunto — SCL1 —, composto de 3 laudos, apresentou preenchimento errôneo no valor do atributo *teg* pois continha frases com construções não encontradas no conjunto de 609 laudos e, portanto, não são contempladas pelo algoritmo e pela ontologia. O preenchimento desses três atributos repercutiu no aumento percentual de valores preenchidos, pois esse atributo não era preenchido para esses laudos com o método anterior.

O segundo subconjunto — SCL2 —, composto de 8 laudos, apresentou inconsistência quanto ao preenchimento do atributo *calibre_esofago*, pois esses laudos continham frases com uma combinação de palavras não considerada no método anterior. Essa combinação altera a contextualização de duas palavras, a qual pode ocasionar o preenchimento errôneo do atributo *calibre_esofago*. Essa mudança de contexto também não era considerada na padronização realizada no método anterior.

O terceiro subconjunto — SCL3 —, composto de 1 laudo, apresentou preenchimento alterado do valor do atributo *esofagite_ulcerada_inferior* em relação ao método anterior, porém foi identificado que não era realizado o mapeamento correto desse caso

específico. Esse erro é ocasionado pela ocorrência de duas descrições relacionadas a um mesmo atributo: uma específica ao atributo e outra, mais geral, a qual engloba mais atributos.

Da totalidade de valores preenchidos nos atributos da base de dados, 99,976% dos valores foram preenchidos com os mesmos valores obtidos com o método anterior. Os valores que foram preenchidos com outros valores são os supracitados nos subconjuntos SCL1, SCL2 e SCL3.

Na primeira aplicação do novo método proposto a secção de esôfago de cada laudo de um conjunto de 609 laudos havia apresentado preenchimento 100% equivalente ao do método anterior aplicado para os mesmos 609 laudos. Ao aplicarmos o novo método proposto, utilizando a mesma ontologia construída para o conjunto de 609 laudos e o mesmo arquivo de padronizações, a um outro conjunto de 3647 laudos obtivemos uma alta taxa de preenchimento, sem a realização de nenhuma adição à ontologia ou ao arquivo de padronização.

A nova abordagem ao método de mapeamento, além de ser capaz de mapear a quase totalidade dos valores de atributos do método anterior, a substituição do dicionário de conhecimento por uma ontologia proporcionou a possibilidade de representar melhor o conhecimento relacionado aos laudos e à tarefa de mapeamento para um formato estruturado. A flexibilidade da representação da ontologia na linguagem OWL permite que novas classes, termos e relações sejam adicionadas. Dessa maneira, é possível expandir o potencial de representação da ontologia em relação ao domínio, sendo possível eliminar a necessidade de certas padronizações, o que preserva mais o conteúdo do laudo original. Com base nesse conteúdo mais completo torna-se possível a adição de novos atributos à tabela atributo-valor, o que permite que mais informação seja mapeada a partir dos laudos.

Além disso, as ontologias em linguagem OWL conta com diversas ferramentas de apoio, como motores de inferência e bibliotecas de código aberto para as mais diversas tarefas. A ontologia pode ser ampliada para mapear uma maior parcela do domínio dos laudos de EDA, como as outras secções do laudo, e também pode ser utilizada para outras tarefas baseadas em conhecimento que sejam aplicadas ao mesmo domínio, como a MD.

4. Conclusão e Trabalhos Futuros

Neste trabalho, o novo método proposto para o mapeamento de laudos médicos textuais para bases de dados estruturadas, apoiado por uma ontologia, construída anteriormente para um conjunto de 609 laudos de EDA, foi aplicado, utilizando a mesma ontologia, a um outro conjunto de 3647 laudos de EDA. A aplicação desse método obteve uma taxa de preenchimento de 85,284% dos valores dos atributos da base de dados, dentre os quais, 99,976% foram preenchidos com os mesmos valores obtidos na aplicação do método anteriormente proposto, sendo que este obteve taxa de preenchimento de 85,279%.

O pré-processamento de laudos médicos e o mapeamento das informações contidas nesses laudos para bases de dados estruturadas constitui uma das tarefas de maior custo quando é de interesse a mineração de dados nesse formato. A construção de ontologia, além de permitir representar de modo mais completo o conhecimento embutido nos dados, possibilitou a generalização do mapeamento de laudos desse domínio não considerados para construção da ontologia. Desse modo, foram reduzidos tanto o tempo para

o mapeamento de laudos, quanto o tempo para a construção de uma ontologia específica para os 3647 laudos de EDA.

Como trabalhos futuros será realizada a expansão da ontologia de modo que sejam mapeadas mais informações da secção de esôfago dos laudos de EDA, de modo a aumentar a gama de dados estruturados extraídos dos laudos. Também será realizada a aplicação do método em outros domínios da área médica, como para laudos de Manometria Anorretal e Coloscopia. A ontologia também será expandida e adaptada para a aplicação em diversas tarefas baseadas em conhecimento que possam ser aplicadas ao domínio de EDA.

5. Agradecimentos

A Universidade Estadual do Oeste do Paraná — UNIOESTE —, pelo auxílio na realização deste trabalho por meio do Programa Institucional de Bolsas de Iniciação Científica — PIBIC.

Referências

- Bodenreider, O. and Burgun, A. (2005). Biomedical ontologies. In Chen, H., Fuller, S. S., Friedman, C., and Hersh, W., editors, *Medical Informatics*, chapter 8, pages 211–236. Springer.
- Carvalho, L. C. D. C. (2007). Método semi-automático de construção de ontologias parciais de domínio com base em textos. Master's thesis, Escola Politécnica da Universidade de São Paulo, São Paulo, SP.
- Cherman, E. A., Spolaôr, N., Lee, H. D., Costa, L. H. D., Fagundes, J. J., Coy, C. S. R., and Wu, F. C. (2008). Metodologia de mapeamento computacional de informações médicas: Aplicação em laudos de coloscopia e manometria anorretal. *Revista Brasileira de Coloproctologia*, 57^o Congresso Brasileiro de Coloproctologia, 29:42–42.
- Chute, C. G. (2005). Medical concept representation. In Chen, H., Fuller, S. S., Friedman, C., and Hersh, W., editors, *Medical Informatics*, chapter 6, pages 163–182. Springer.
- Friedman, C. and Johnson, S. B. (2006). Natural language and text processing in biomedicine. In Shortliffe, E. H. and Cimino, J. J., editors, *Biomedical Informatics*, chapter 2, pages 312–343. Springer.
- Gruber, T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing. *Formal Ontology in Conceptual Analysis and Knowledge Representation*.
- Han, J. and Kamber, M. (2006). *Data Mining: Concepts and Techniques*, chapter 10.4 Text Mining. Morgan Kaufmann.
- Honorato, D. D. F., Cherman, E. A., Lee, H. D., Monard, M. C., and Wu, F. C. (2008). Construction of an attribute-value representation for semi-structured medical findings knowledge extraction. *CLEI Electronic Journal*, 11(2):1–12.
- Honorato, D. D. F., Lee, H. D., Monard, M. C., Wu, F. C., Machado, R. B., Neto, A. P., and Ferrero, C. A. (2005). 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*, pages 593–601, São Leopoldo, RS, Brasil.

- Honorato, D. D. F., Monard, M. C., Lee, H. D., Neto, A. P., and Wu, F. C. (2009). Avaliação de um método de mapeamento de laudos médicos para uma representação estruturada: estudo de caso com laudos de endoscopia digestiva alta. In *Anais do IX Workshop de Informática Médica (WIM), XXIX Congresso da Sociedade Brasileira de Computação (CSBC)*, pages 1–10, Bento Gonçalves, RS, Brasil.
- Lee, H. D. (2005). *Seleção de atributos importantes para a extração de conhecimento de bases de dados*. Tese de doutorado, ICMC-USP, São Carlos, SP, Brasil.
- McGuinness, D. L. and van Harmelen, F. (2004). *OWL Web Ontology Language Overview*. Disponível em: <http://www.w3.org/TR/owl-features/>. Acesso em: 20 jul 09.
- Revere, D. and Fuller, S. S. (2005). Characterizing biomedical concept relationships. In Chen, H., Fuller, S. S., Friedman, C., and Hersh, W., editors, *Medical Informatics*, chapter 7, pages 183–210. Springer.
- Rezende, S. O. (2003). *Sistemas Inteligentes - Fundamentos e Aplicações*. Manole, Barueri-SP, Brasil, 1 edition.
- Shortliffe, E. H. and Barnett, G. O. (2006). Biomedical data: Their acquisition, storage, and use. In Shortliffe, E. H. and Cimino, J. J., editors, *Biomedical Informatics*, chapter 2, pages 46–79. Springer.
- Uschold, M. and Gruninger, M. (1996). Ontologies: principles, methods and applications. *Knowledge Engineering Review*, 11:93–155.