Elaboração de um dicionário para a representação de achados radiográficos

Roberto A. Rocha; Stanley M. Huff; Peter J. Haug

Department of Medical Informatics, University of Utah School of Medicine.

614 North Medical Plaza, Salt Lake City, UT 84112, EUA.

E-mail: Roberto.Rocha@m.cc.utah.edu

Resumo - Os autores resumem o processo de elaboração de um dicionário para a representação de achados radiográficos de tórax. O dicionário é construído a partir de diferentes fontes de termos, incluindo vocabulários médicos e laudos de raios-X de tórax. A relevância de cada fonte é estimada a partir das proporções com que elas podem ser encontradas na versão final do dicionário.

Abstract - The authors summarize the process for creating a dictionary to represent chest radiology findings. The dictionary is built from different sources of terms, including medical vocabularies and chest X-ray reports. The relevance of each source is estimated using the proportions with which they can be found in the final edition of the dictionary.

Introdução

O registro médico eletrônico é considerado um dos grandes desafios para a Informática Médica, atraindo pesquisadores e profissionais da área¹. Dentre as tecnologias necessárias para a implementação do prontuário eletrônico, o processamento da linguagem médica tem sido identificado como um tópico prioritário².

Uma das estratégias propostas é a criação de modelos que representam a linguagem médica utilizando a estrutura semântica dos diversos tipos de documentos encontrados no prontuário médico³. Estes modelos semânticos transformam a linguagem médica utilizada em cada domínio em uma linguagem formal, permitindo que a informação representada se torne acessível a diversos sistemas computadorizados.

O desenvolvimento destes modelos semânticos é um processo complexo³. Uma das etapas iniciais deste desenvolvimento é a criação de um dicionário, onde os termos encontrados no domínio que está sendo estudado são criteriosamente identificados e classificados⁴.

Metodologia

Considerando a existência de vários vocabulários médicos, o objetivo deste estudo é determinar a utilidade de três diferentes vocabulários na criação de um dicionário para a representação de achados radiográficos de tórax. Os vocabulários analisados foram o PTXT⁵, o dicionário de dados do ILIAD⁶, e o "Metathesaurus" do Projeto "Unified Medical Language System" (UMLS)⁷. Somente os segmentos destes vocabulários considerados

relevantes à radiologia torácica foram incluídos neste estudo.

Além dos três vocabulários mencionados, a terminologia empregada em duas coleções de laudos foi também incluída no processo de construção do dicionário. O dicionário foi organizado sob a forma de um banco de dados onde cada conceito foi associado a um número, permitindo no caso de sinônimos que vários termos e palavras fossem associados a um mesmo conceito.

Considerando as características do modelo semântico para o qual o dicionário estava sendo desenvolvido, o processo de revisão manual dos termos buscou a identificação de conceitos individuais ("atômicos"), ou seja, frases longas e descritivas foram separadas em componentes menores⁴.

Após a elaboração do dicionário, rotinas especializadas foram empregadas para comparar a edição final do dicionário com as diversas fontes de termos utilizadas.

Resultados

A versão final do dicionário possuía 4496 itens, associados a 3216 conceitos, com uma "granularidade" (mediana do número de palavras em cada item) igual a 1. Com relação aos vocabulários utilizados como fontes de termos para o dicionário, o segmento do PTXT continha 444 itens e granularidade igual a 2, o segmento do dicionário de dados do ILIAD possuía 186 itens e granularidade igual a 7, e o segmento do Metathesaurus continha 4584 itens e granularidade igual a 2. Com relação às coleções de laudos analisadas, a primeira coleção continha 135 laudos, resultando em 734 frases com uma granularidade

igual a 8, e a segunda coleção continha 453 laudos, resultando em 1831 frases com uma granularidade igual a 9.

A versão final do dicionário continha 33.33% dos termos encontrados no PTXT, 1.08% dos termos encontrados no ILIAD, e que 8.75% dos termos encontrados no Metathesaurus. Esta versão do dicionário possuía 89.69% das palavras encontradas no PTXT, 97.72% das palavras encontradas no ILIAD, e 43.46% das palavras encontradas no Metathesaurus. Com relação as coleções de laudos, esta versão do dicionário continha 0.41% das frases e 93.41% das palavras encontradas na primeira coleção, e 0% das frases e 71.86% das palavras encontradas na segunda coleção.

Discussão e Conclusões

Os resultados obtidos confirmam que a terminologia necessária para a representação de achados radiográficos não costuma estar disponível em vocabulários médicos⁸. O segmento do PTXT utilizado neste estudo é uma exceção, pois inclui termos criados especificamente para a representação de achados radiográficos. O mesmo se aplica ao dicionário de dados do ILIAD, cujo segmento corresponde aos termos utilizados nos "frames" da base de conhecimento que analisam dados radiográficos. Normalmente, as melhores fontes de termos para a construção de dicionários de domínios específicos são coleções representativas dos documentos encontrados nestes domínios³.

Outro problema identificado com este estudo está relacionado com a granularidade dos termos. Enquanto que os vocabulários médicos costumam apresentar termos com baixa granularidade, as frases encontradas em textos descritivos costumam ser muito complexas. Portanto, apesar de documentos representativos do domínio em estudo serem as melhores fontes, estes documentos precisam ser processados antes que o seu conteúdo possa ser utilizado.

A construção de dicionários para sistemas clínicos é um processo demorado e caro, apesar de sucessos recentes na sua geração automática³. A existência de vocabulários médicos geralmente facilita o processo, porém raramente os vocabulários possuem a abrangência e o nível de detalhe desejados^{3,4}.

Referências Bibliográficas

- ¹Dick, R. S.; Steen, E. B. *The Computer-based Patient Record: An Essential Technology for Health Care*. Washington, D.C.: National Academy Press, 1991.
- ²McCray, A. T.; et al. Concepts, Knowledge, and Language in Health-Care Information Systems. *Methods of Information in Medicine*, v. 34, p. 1-4, 1995.
- ³Evans, D. A.; et al. Toward a Medical-concept Representation Language. *Journal of the American Medical Informatics Association*, v. 1, p. 207-17, 1994.
- ⁴Rocha, R. A.; et al. Designing a Controlled Medical Vocabulary Server: The VOSER Project. *Computers and Biomedical Research*, v. 27, p. 472-507, 1994.
- ⁵Kuperman, G. J.; et al. *HELP: A Dynamic Hospital Information System*. New York, NY: Springer-Verlag, 1991.
- ⁶Warner, H. R.; et al. ILIAD as an Expert Consultant to Teach Differential Diagnosis, apresentado no "Twelfth Symposium on Computer Applications in Medical Care", Washington, D.C., p. 371-376, 1988.
- ⁷National Library of Medicine. *UMLS Knowledge Sources, fourth experimental edition*. Bethesda, MD: National Library of Medicine, 1993.
- ⁸Friedman, C. The UMLS Coverage of Clinical Radiology, apresentado no "Sixteenth Symposium on Computer Applications in Medical Care", Baltimore, MD, p. 309-13, 1992.

Agradecimentos:

Roberto A. Rocha é bolsista do CNPq.