



Mineração de texto e o processamento de linguagem natural em sumários de alta hospitalar

Text mining and natural language processing in discharge summaries

Hugo Bulegon¹
Claudia Maria Cabral Moro²

Descritores: Processamento de linguagem Natural; Mineração de textos; Narrativas clínicas; Sumários de alta

RESUMO

Objetivo: O objetivo deste trabalho foi fazer uma revisão da literatura sobre Processamento de Linguagem Natural e Mineração de Textos, apresentando suas principais características, baseando-se em exemplos relevantes desta área. **Método:** A revisão foi realizada no PUBMED considerando-se trabalhos publicados nos últimos dez anos (1999 a junho de 2010). **Resultados:** Foram selecionados treze artigos que descrevem a recuperação de informações em sumários de alta. Também foram relacionadas as principais ferramentas para PLN e Mineração de Textos. **Conclusão:** Tanto o PLN quanto a Mineração de Textos apresentaram resultados expressivos nesta pesquisa, caracterizando-se como métodos efetivos para auxiliar o processo de extração de informação de narrativas clínicas.

Descriptors: Natural language Processing; text mining; Clinical narratives; Discharge summaries

ABSTRACT

Objective: The purpose of this paper was to review the literature on Natural Language Processing and Text Mining and to present their main characteristics based on relevant examples. **Methods:** The review was based on PUBMED and papers published in the last ten years (1999 to June 2010) were examined. **Results:** There were selected 13 thirteen articles that described information retrieval applied to discharge summaries. **Conclusion:** The main tools for NLP and Text Mining were also described. Both NLP and Text Mining showed significant results and can be effective methods to help extracting information from clinical narratives.

Autor Correspondente: **Claudia Maria Cabral Moro**
e-mail: c.moro@pucpr.br

¹ Mestrando do Programa de Pós-Graduação da Pontifícia Universidade Católica do Paraná - PPGTS/PUCPR Informática em Saúde. Curitiba (PR), Brasil.

² Professora Titular do Programa de Pós-Graduação da Pontifícia Universidade Católica do Paraná - PPGTS/PUCPR na área de Informática em Saúde. Curitiba (PR), Brasil.

Artigo recebido: 24/03/2010
Aprovado: 24/06/2010

INTRODUÇÃO

Durante as atividades de atendimento aos pacientes geralmente são gerados registros de dados contendo informações em formato de textos. Estes documentos são denominados narrativas clínicas. Na maioria das vezes não são estruturados, isto é, os textos são elaborados de forma livre. Alguns exemplos de narrativas são as evoluções, as descrições de cirurgia, os laudos de exames, os sumários de internação e de alta⁽¹⁾. Considerando-se que estes documentos podem conter dados importantes para a assistência, acompanhamento do paciente, gerenciamento de instituições e para a vigilância epidemiológica, atualmente, existem diversas iniciativas de identificar e extrair informações relevantes das narrativas clínicas⁽²⁾. No processo de identificação de informações são utilizados principalmente métodos e técnicas de Processamento de Linguagem Natural - PLN ou Mineração de Textos - *Text Mining* são⁽³⁾.

O PLN surgiu devido à necessidade de compreensão automática e comunicação em geral do ser humano com o computador. Trata-se de um mecanismo criado não só para extrair as informações de textos, mas também pode facilitar a entrada de dados nos sistemas e a estruturação dos dados⁽⁴⁾.

O PLN está dividido em quatro etapas: análise morfológica, análise sintática, análise semântica e análise pragmática; que são realizadas nesta mesma ordem. A análise morfológica é responsável por definir artigos, substantivos, verbo e adjetivos armazenando-os em um dicionário (*Thesaurus*). Depois de construído o dicionário, a análise sintática faz uso dele procurando mostrar relacionamento entre elas e em um segundo momento, verifica sujeito, predicado, complementos nominais e verbais, adjuntos e apostos⁽⁴⁾. Na análise semântica, ocorre o encontro de termos ambíguos, de sufixos, afixos, ou seja, questões dos significados associados aos morfemas componentes de uma palavra, o sentido real da frase ou palavra. Para a junção e visualização de todas as etapas, a análise pragmática faz a junção de todo o mecanismo e mostra visualmente o resultado. Para este caso, existem alguns algoritmos que apresentam em forma de árvore todos os passos até chegar à conclusão, que é a identificação dos termos de uma sentença.

As ferramentas do PLN também podem ter um complemento além de extrair informações, pois podem ser aplicadas para auxiliar a construir textos livres que podem ser oriundos de digitação ou de transformações como gravações de voz⁽⁵⁾.

A mineração de textos muitas vezes utilizada como sinônimo de PLN é justificado pelo grau de relacionamento entre eles, mas basicamente pode ser dividida em três sub-áreas: Classificação de textos, onde é usada para a indexação de documentos, filtragem de documentos, extração da informação e algoritmos de aprendizado de máquina supervisionado

(como por exemplo, Bayes), o Agrupamento de Textos, que envolve em geral técnicas de aprendizado não supervisionado como *clustering* hierárquico, redes SOM e a Extração da Informação, além de engenharia do conhecimento e aprendizado de máquina⁽⁶⁾. As ferramentas de PLN geralmente são utilizadas como base para o processo de Mineração de Textos.

O objetivo deste trabalho foi fazer uma revisão da literatura sobre PLN e Mineração de Textos, apresentando suas características, diferenças e relacionamentos, baseando-se em exemplos aplicados às narrativas clínicas.

MÉTODOS

A revisão de literatura foi realizada no PUBMED selecionando-se artigos que descrevem a recuperação de informação em narrativas clínicas, primeiramente pelo termo *Natural Language Processing*. A pesquisa foi realizada compreendendo o período entre 1999 e junho de 2010. A extração de informação em sumários de alta foi um assunto que apareceu com maior frequência. A revisão foi refinada utilizando-se os termos "*Discharge Summary and Text Mining*" e "*Discharge Summary and Natural Language Processing*".

Na Tabela 1 são apresentados os resultados da pesquisa bibliográfica. Artigos que foram selecionados em mais de um termo; foram considerados somente uma vez.

Para a seleção dos artigos foi verificada no resumo a forma da extração de informação. Quando o conteúdo não era suficiente para a compreensão, optou-se por realizar a leitura completa do trabalho. Não foram utilizados artigos com conteúdo similar. O termo *Natural Language Processing Tool* (Ferramentas para Processamento de Linguagem Natural) foi adicionado, pois durante a seleção dos artigos surgiram alguns exemplos apresentados a utilização ferramentas de PLN.

As ferramentas auxiliam em etapas da extração e tem grande importância nos resultados. O termo *Natural Language Processing and SNOMED or ICD* foi utilizado porque durante a seleção de artigos nos outros termos, apareceram alguns resultados com este assunto. Sabendo da importância que a nomenclatura SNOMED e a classificação CID - Classificação Internacional de Doenças (*ICD*) têm na padronização e estruturação de informações na área de saúde, entendeu-se que deveria ser feito uma revisão neste contexto para complementar o estudo.

Dos treze artigos selecionados, três foram descritos de forma mais detalhada para explicar o PLN e Mineração de Textos aplicados em narrativas clínicas.

Apesar de muitos artigos terem sido selecionados, somente foram utilizados os que descrevem as etapas do PLN e de Mineração de Textos contendo diferentes procedimentos. Aqueles que eram semelhantes, apenas foram citados junto ao que apresentou e justificou de melhor forma seus métodos.

Tabela 1 – Seleção de artigos

Termos	Encontrados	Selecionados	Utilizado(s)
Discharge Summary and Text Mining	113	26	3
Discharge Summary and Natural Language Processing	126	19	3
Natural Language Processing Tool	566	9	1
Natural Language Processing and SNOMED or ICD	943	12	2
Natural Language Processing	1068	11	4
Total	2816	77	13

Ao final estão relacionadas as principais ferramentas para PLN e Mineração de Textos utilizadas para a identificação de informações em narrativas clínicas.

RESULTADOS

Os trabalhos envolvendo PLN e Mineração de Textos em Narrativas Clínicas estão descritos nos resultados deste artigo e foram colocados sucintamente na Tabela 2 para dar uma visão geral do conteúdo que foi encontrado.

O artigo *“Using regular expressions to extract information on pacemaker implantation procedures from clinical reports”* descreve o PLN e a Mineração de Textos como uma ferramenta que pode auxiliar no complemento de informação após extrair os dados clínicos obtidos em texto livre de relatórios cirúrgico, relacionados com a estimulação cardíaca e desfibrilação. Esta extração

possui uma série de regras e expressões regulares (derivado da Linguagem Regular identifica cadeias de caracteres de interesse) para extrair informação dos pacientes, dos dispositivos e procedimentos dos médicos⁽¹⁵⁾. Para validar o método e criar um padrão ouro, especialistas fizeram a extração manual da base e comparam os resultados. Um cardiologista especialista em marca-passos definiu uma lista de seis categorias de dados relevantes em ambas as fontes: Nome do paciente, nome do cardiologista, modelo do dispositivo, fabricante, número de série e dispositivo de simulação.

O sistema extraiu informações incluindo dados numéricos, texto e de combinação de números e caracteres, com sensibilidade maior que 90% e um valor preditivo superior a 95%. Apresentou uma precisão superior à conseguida com a extração manual.

Bui e colaboradores⁽¹⁶⁾ definiram regras em seu

Tabela 2 – Artigos apresentados nos resultados

Artigo	Autores	Assunto/Objetivo
A case report using SNOMED CT for grouping Adverse Drug Reactions Terms ⁽⁷⁾	Alecu, I., Bousquet, C., Jaulent M.	Extração de Informação contida em texto utilizando estrutura do SNOMED.
A Study of Abbreviations in Clinical Notes ⁽⁸⁾	Xu, H., Stetson, P.D. E Friedman, C.	Extração de acrônimos, abreviaturas e outros.
A System for Automated Lexical Mapping ⁽⁹⁾	Sun J. Y. E Sun Y.	Extração de informação contida em texto e padronizando em formatos médicos.
Automatic construction of rule-based ICD-9-CM coding systems ⁽¹⁰⁾	Farkas, R. E Szarvas G.	Extração de informação padronizada contida em texto utilizando estrutura do CID 9.
Automated Detection of Adverse Events Using Natural Language Processing of Discharge Summaries ⁽¹¹⁾	Melton. G. B., E Hripcsak. G.,	Extração de Informação contida em texto.
Automated Knowledge Acquisition from Clinical Narrative Reports ⁽¹²⁾	Wang, X., Chused. A., Elhadad, N., Friedman. C. E Markatou, M.	Extração de informação contida em texto.
Lessons Extracting Diseases from Discharge Summaries ⁽¹³⁾	Long, W	Extração de informação contida em texto.
Manual Annotation of Colonoscopy Videos: A First Step towards Automation ⁽⁴⁾	Groen, P. C., Tavanapong, W., Oh E J. Wong, J.	Extração de Informação contida em áudio e construção do Sumário.
Summarising Complex ICU Data in Natural Language ⁽¹⁴⁾	Hunter, J., Freer, Y., Gatt, A., Logie, R., McIntosh, N., Meulen, M. V. D. M., Portet, F., Reiter. E., Sripada, S. E Sukes, C.	Extração de informação contida em uma interface interpretadora de sinais de um monitor de UTI neonatal.
Using regular expressions to extract information on pacemaker implantation procedures from clinical reports ⁽¹⁵⁾	Rosier, A., Burgun, A., Mabo P.	Extração de informação contida em texto.

método para extrair e combinar relações entre o HIV e os medicamentos da doença baseados em resumos disponíveis na base Pubmed. As regras foram criadas a partir de um estudo do texto e da sua classificação, a ferramenta utilizada não foi relatada, apenas citado que foi utilizada uma ferramenta NLP. Os pesquisadores procuraram por palavras chaves para criar regras, e encontraram os principais termos como “*resistance*”, “*responsive*” e qual era sua classificação semântica.

Depois analisaram palavras próximas a elas como “*responsive*”, “*associated*”, analisaram sua classificação e criaram regras para definir que aquela palavra poderia ser um termo que estavam procurando⁽¹⁶⁾.

Outros pesquisadores desenvolveram uma ferramenta baseada em PLN nomeada como MedEx que extrai medicamentos em narrativas clínicas. Foram testados 50 sumários de alta e a ferramenta atingiu 93,2% de acerto nesta extração⁽¹⁷⁾.

Liao e colaboradores⁽¹⁸⁾ testaram regras utilizando PLN e compararam sua eficiência com um algoritmo definido como regressão logística para classificar pacientes com artrite reumatóide. Analisaram informações com mais de 29 mil pacientes para criar regras a serem utilizadas com o PLN e 500 sumários foram treinados para criar regras utilizando o algoritmo. Com o PLN obtiveram 94% de acerto na extração e o algoritmo alcançou um resultado inferior a 88% .

A informação também pode ser extraída através de vídeos, a partir da aquisição do áudio, este é transformado em texto. O conteúdo auxilia na construção do sumário, assim como sugerido por Groen et al⁽⁵⁾.

Summarising Complex ICU Data in Natural Language é um trabalho que envolve a interpretação da interface de forma automática, criando sumários automaticamente através da interpretação de sinais de um monitor neonatal localizado em uma base de dados⁽¹⁴⁾. A motivação para o desenvolvimento deste trabalho foi a percepção de que os especialistas despendem 5% do seu tempo de atendimento para avaliar o monitor do paciente que acompanha variáveis fisiológicas (frequência cardíaca, por exemplo) e outras variáveis como configuração do equipamento e resultados de laboratório. Neste trabalho, o primeiro passo foi a construção de uma ontologia descrevendo todas as anotações e eventos inferidos. A seguir é realizada a extração de termos através da interpretação dos sinais fisiológicos e são inseridos os conceitos médicos em relação ao sinal. Ao final é realizada a seleção das estruturas com os eventos requeridos em forma de árvore, e é feita a transformação deste conteúdo em um texto. Segundo Hunter e colaboradores, os sumários resultantes não são tão bons quanto os gerados pelos especialistas.

Em registros clínicos, frequentemente são encontrados acrônimos, mas nem sempre eles estão entre os termos médicos. A partir desta informação, foi desenvolvido um modelo para a construção de anotações clínicas eliminando os chamados falsos-

acrônimos. A detecção destes acrônimos ocorreu em textos livres e foram colocados em um dicionário (*Thesaurus*). Foram usados quatro métodos para a detecção: o primeiro faz uma comparação no inventário e utiliza *Tokens* para buscar as abreviaturas no texto, onde o desconhecido é descartado. O segundo utiliza um método heurístico baseado em regras que utiliza informações relativas da palavra de formação, tais como letras maiúsculas, caracteres alfabéticos e numéricos funcionando em conjunto com o inventário. No terceiro e quarto método, é formado uma árvore de decisão com o algoritmo J48, onde se realiza a análise pragmática do conteúdo. O método obteve uma precisão de 91,4%. A construção do inventário usou duas fontes de conhecimento: o *Unified Medical Language System (UMLS)* e a base de dados do MEDLINE. Na Tabela 3, algumas das abreviaturas que foram encontradas⁽⁸⁾.

Uma alternativa para evitar parte dessas abreviações pode ser a codificação de elementos encontrados no texto para o CID e o SNOMED. Não é possível construir um padrão único de todas as áreas da saúde, pois todos os casos precisam ser avaliados para obter mais exatidão no processo, mas já há casos na radiologia, terminologias biomédicas, reações de drogas que conseguiram bons resultados^(7,10).

Existem também situações onde o PLN pode surgir como uma alternativa para auxiliar na identificação de informações na prevenção de diagnósticos. Dados recentes sugerem que existem taxas significativas da falta de detecção de grandes pólipos e cânceres. O trabalho neste segmento consiste em captar digitalmente e armazenar o vídeo completo arquivo gerado durante colonoscopia, também captando sinais de áudio. Os vídeos são captados de forma anônima durante a rotina endoscópica prática, onde os pacientes assinam um termo para a colaboração. A linguagem natural atua para converter o áudio presente no vídeo em narrativas permitindo a criação de uma biblioteca onde os diagnósticos e procedimentos são encontrados e classificados, que serão usados futuramente para treinar um sistema que está sendo desenvolvido de forma automatizada⁽⁴⁾.

Na Tabela 4, quatro trabalhos são apresentados para demonstrar que o PLN e a Mineração de Textos são mecanismos que conseguem obter resultados expressivos em se tratando de extração de informação em narrativas clínicas.

Nesta mesma tabela, é possível observar ainda que o artigo “Automated Detection of Adverse Events Using Natural Language Processing of Discharge Summaries” fez uso de uma ferramenta chamada *MedLEE*. Trata-se de uma ferramenta criada utilizando conceitos do PLN que auxilia na extração de informação. A Tabela 5 apresenta mais detalhes das suas funcionalidades e também outras ferramentas que têm algum relacionamento com o PLN ou a Mineração de Textos.

Tabela 3 – Abreviaturas: exemplos

Tipo de abreviatura	Exemplo	Frequência
Acronímico	BP – Pressão Arterial	50,00%
Palavras Abreviadas	PT – Paciente	32,00%
Contração	t/d/a – tabaco, drogas ou álcool	9,00%
Outros	Etoh-álcool	9,00%

Adaptado de A Study of Abbreviations in Clinical Notes⁽⁸⁾**Tabela 4** – Características dos artigos em PLN e Mineração de Textos

Artigo	Objetivo / Método	Resultado / Conclusão
A System for Automated Lexical Mapping ⁽⁹⁾	Fazer um mapeamento léxico de bases de dados com o objetivo de padronizar o vocabulário médico. Normalizar e realizar um algoritmo de alinhamento inspirado no DNA para formatos como o UMLS e o LOINC.	O algoritmo de alinhamento utilizado foi capaz de mapear entre 57% a 78% de todas as bases de dados. O mapeamento léxico pode facilitar a interação de dados e é feita de forma mais rápida que a manual.
Automated Detection of Adverse Events Using Natural Language Processing of Discharge Summaries ⁽¹¹⁾	Verificar se o PLN consegue detectar e de maneira eficaz diversos eventos contidos em sumários de alta. Construção de um padrão-ouro e utilização do MedLEE.	De 57.452 sumários de alta, o sistema identificou 1.590 eventos em 1.461 casos. De forma manual 704 eventos em 652 casos. O PLN foi considerado eficaz. O intervalo de confiança foi 95%
Automated Knowledge Acquisition from Clinical Narrative Reports ⁽¹²⁾	Desenvolver métodos automatizados de aquisição e descoberta do conhecimento médico incorporados em narrativas clínicas. Busca focada em doenças e sintomas. Utiliza o MedLEE para extrair essas informações.	A precisão do método foi de 92% e o estudo mostrou que a aquisição e a extração do conhecimento são eficazes. Pode ser generalizado para detectar outras associações clínicas como, por exemplo, entre doenças e medicamentos.
Lessons Extracting Diseases from Discharge Summaries ⁽¹³⁾	Utilizar a ferramenta desenvolvida para analisar seus resultados. Sumários de alta são utilizados no teste.	O sistema conseguiu extrair 93% do desejado. Os conceitos perdidos foram ocasionados por ambiguidades ou palavras/frases separadas. A extração de conceitos médicos de resumos de alta com PLN é eficaz e com potencial.

Tabela 5 – Ferramentas de PLN e Mineração de Texto

Ferramenta	Tipo de Ferramenta	Descrição / Funcionalidades
CoGroom ⁽¹⁹⁾	Processamento em Linguagem Natural	Realiza a análise morfológica, sintática e semântica e pode ser acoplado ao OpenOffice, em português. Versão gratuita. http://cogroosourceforge.net/
Dt-Search ⁽²⁰⁾	Mineração de Textos	Possui conceito de stemming com a possibilidade de construir um thesaurus de acordo com a necessidade, em inglês. Versão paga/DEMO. http://www.dtsearch.com
GATE ⁽²¹⁾	Mineração de Textos/ Processamento em Linguagem Natural	Pode ser usado como um plugin e possui conceitos de categorização, stopwords, stemming e análise semântica, em inglês, Francês, Alemão, Espanhol, Italiano, Chinês, entre outros. Versão gratuita. http://www.gate.ac.uk
MedLEE ⁽²²⁾	Processamento em Linguagem Natural	Possui conceito de categorização e análise sintática, em inglês. Versão DEMO. http://www.lucid.cpmc.columbia.edu/medlee
OpenNLP ⁽²³⁾	Processamento em Linguagem Natural	Etiquetador, e agrupador, algoritmo de aprendizado, em inglês. Versão gratuita. http://opennlp.sourceforge.net
TextAnalyst ⁽²⁴⁾	Mineração de Textos	Possui conceitos de stopwords, stemming, sumarização, clustering, em inglês. Versão paga/DEMO. http://www.megaputer.com/textanalyst.php
Weka ⁽²⁵⁾	Mineração de Dados / Textos / Processamento em Linguagem Natural	Possui entre outros conceitos, clustering e o poder de auxiliar na análise pragmática com algoritmos de classificação pré-definidos como, por exemplo, o J48. Versão gratuita. http://www.cs.waikato.ac.nz/ml/weka
WizDoc ⁽²⁶⁾	Mineração de Textos	Possui conceito de stopwords, stemming, thesaurus com corretor e indexação. Versão paga. http://www.searchtools.com/tools/wizdoc.html

DISCUSSÃO

A mineração de textos pode ser dividida em três sub-áreas: Classificação de textos, Agrupamento de Textos e a Extração da Informação e as ferramentas de PLN geralmente são utilizadas como base para o processo de Mineração de Textos. O PLN está dividido

em quatro etapas: análise morfológica, análise sintática, análise semântica e análise pragmática; que são realizadas nesta mesma ordem e que formam uma composição da Mineração de Textos.

Tanto o PLN quanto a Mineração de Textos apresentaram resultados expressivos na pesquisa e podem ser utilizados para extrair informação de

narrativas clínicas. As ferramentas existentes podem auxiliar na extração de informações das narrativas clínicas, onde existem algumas opções de ferramentas que podem auxiliar na extração.

O número de trabalhos envolvendo estas duas áreas

vem crescendo, isso se deve principalmente a dois fatos: a necessidade do texto como entrada de dados no prontuário, que apresentam informações importantes; e os resultados positivos que o PLN e a Mineração de Textos vêm conseguindo alcançar na recuperação deste conteúdo.

REFERÊNCIAS

- Grossman E, Cardoso MHCA. As narrativas em medicina: contribuições à prática clínica e ao ensino médico. *Rev. Bras. Educ. Méd.* 2006;30(1):6-14.
- Popowich F. Using text mining and natural language processing for health care claims processing. *SIGKDD Explorations*. 2005;7(1):59-66.
- Hersh W. Evaluation of biomedical text-mining systems: lessons learned from information retrieval. *Brief Bioinform.* 2005;6(4):344-56.
- Mooney RJ, Bunescu R. Mining knowledge from text using information extraction. *SIGKDD Explorations*. 2005;7(1):3-10.
- Groen PC, Tavanapong W, Oh J, WONG J. Manual annotation of colonoscopy videos: A first step towards automation. *AMIA Annu Symp Proc.*, 2006:901.
- Hotho A, Nürnberger A, PAAß G. A brief survey of text mining. *Journal for Computational Linguistics and Language Technology*. 2005;20(1):19-62.
- Alecú I, Bousquet C, Jaulent M. A case report: using SNOMED CT for grouping Adverse Drug Reactions Terms. *BMC Med Inform.* 2008;8(Suppl 1):S4.
- Xu H, Stetson PD, Friedman C. A study of abbreviations in clinical notes. *AMIA Annu Symp Proc.* 2007:821-5.
- Sun JY, Sun Y. A system for automated lexical mapping. *J Am Med Inform Assoc.* 2006;13(3):334-43.
- Farkas R, Szarvas G. Automatic construction of rule-based ICD-9-CM coding systems. *BMC Bioinformatics*. 2008;9(Suppl 3):S10.
- Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc.* 2005;12(4):448-57.
- Wang X, Chused A, Elhadad N, Friedman C, Markatou M. Automated knowledge acquisition from clinical narrative reports – AMIA. *Annu Symp Proc.* 2008: 783-7.
- Long W. Lessons extracting diseases from discharge summaries - AMIA Annu Symp Proc. 2007: 478-82.
- Hunter J, Freer Y, Gatt A, Logie R, Mcintosh N, Meulen MVDM, Portet F, Reiter E, Sripada S, Sukes C. Summarising complex ICU data in natural language. *AMIA Annu Symp Proc.* 2008: 1225.
- Rosier A, Burgun A, Mabo P. Using regular expressions to extract information on pacemaker implantation procedures from clinical reports. *AMIA Annu Symp Proc.* 2008:81-5.
- Bui QC, Nualláin BO, Boucher CA, Sloom PM. Extracting causal relations on HIV drug resistance from literature. *BMC Bioinformatics*; 2010;11(Suppl 1):S21.
- Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. *J Am Med Inform Assoc.* 2010;17(1):19-24.
- Liao KP, Cai T, Gainer V, Goryachev S, Zeng-Treitler Q, Raychaudhuri S, Szolovits P, Churchill S, Murphy S, Kohane I, Karlson EW, Plenge RM. Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care Res (Hoboken)*. 2010;62(3).
- CoGrOO. Corretor gramatical acoplável ao OpenOffice.org [homepage on the internet]. São Paulo. [atualizada em 2010 Jun 8; citada 2010 Jun 18]. Disponível em: <http://cogroo.sourceforge.net>.
- Dt-Search Corporation. Dt-Search. Tulip Hill Terrace. [homepage on the internet]. EUA. [cited 2010 Jun 18]. Available from: <http://www.dtsearch.com>.
- GATE. General Architecture for Text Engineering. [homepage on the internet] Inglaterra. [cited 2010 Jun 18]. Available from: <http://www.gate.ac.uk>.
- Friedman C. MedLEE. A medical language extraction and encoding system. [homepage on the internet]. Nova Iorque, EUA – Columbia University; [updated 2006 Nov 17; cited 2010 Jun 18]. Available from: <http://zellig.cpmc.columbia.edu/medlee>.
- Morton, T. OpenNLP. [homepage on the internet]. Austin; [update 2008 nov 28; cited 2010 Jun 18]. Available from: <http://opennlp.sourceforge.net>.
- Text Analyst, Megaputer. [homepage on the internet]. Bloomington, EUA. [update 2007 Fev 5; cited 2010 Jun 18]. Available from: <http://www.megaputer.com/textanalyst.php>.
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA Data Mining Software: An Update; *SIGKDD Explorations*. 2009;11(1):10-18.
- WizDoc. [homepage on the internet]. Tel Aviv, Israel. [update 2007 Fev 15; cited 2010 Jun 18]. Available from: <http://www.wizsoft.com/default.asp>