

# Uso de mineração de texto como ferramenta de avaliação da qualidade informacional em laudos eletrônicos de mamografia\*

*Text mining as a tool for assessment of informational quality of electronic mammographic reports*

Paulo Roberto Barbosa Serapião<sup>1</sup>, Kátia Mitiko Firmino Suzuki<sup>2</sup>, Paulo Mazzoncini de Azevedo Marques<sup>3</sup>

**Resumo** **OBJETIVO:** Investigação do uso da técnica de mineração de texto como forma de avaliar a qualidade informacional de laudos eletrônicos de mamografia, tendo como parâmetro de qualidade a adesão ao léxico BI-RADS®. **MATERIAIS E MÉTODOS:** Foram extraídos 22.247 laudos de mamografia do banco de dados do sistema de informação em radiologia do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, no período de janeiro de 2000 até junho de 2006. Foram realizados dois experimentos, um buscando-se verificar a utilização mais correta dos termos do léxico – experimento 1 (especificidade do método de mineração), e outro buscando-se verificar toda e qualquer tentativa de uso ou alusão ao léxico – experimento 2 (sensibilidade do método de mineração). **RESULTADOS:** Experimento 1: variação entre 11% e 61% de laudos contendo termos do léxico em sua conclusão, distribuída de forma aleatória ao longo do tempo, a partir do ano de 2001. Experimento 2: variação entre 44% e 100% de laudos que se referem de alguma forma ao léxico em sua conclusão. **CONCLUSÃO:** Os resultados indicam um bom potencial da aplicação da ferramenta de mineração de texto para a avaliação da qualidade das informações contidas em laudos eletrônicos de mamografia. *Unitermos:* Mamografia; BI-RADS; Teoria da informação; Informática médica.

**Abstract** **OBJECTIVE:** To investigate the utilization of text mining technique for evaluating the informational quality of electronic mammographic reports considering adherence to the BI-RADS® lexicon as a quality parameter. **MATERIALS AND METHODS:** A total of 22,247 mammography reports of the period between January, 2000 and June, 2006 were collected from the radiology information database of Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, SP, Brazil. Two experiments were undertaken – experiment 1 to evaluate the accuracy in the adoption of the lexicon terms (text mining method specificity), and experiment 2 to identify all and any attempt to utilize or refer to the lexicon (text mining method sensitivity). **RESULTS:** Experiment 1: variation between 11% and 61% in reports including lexicon terms in their conclusion, randomly distributed over time since 2001. Experiment 2: variation between 44% and 100% in reports that somehow refer to the lexicon in their conclusion. **CONCLUSION:** Results indicate a good potential for text mining tool application for assessing the quality of information included in electronic mammography reports. *Keywords:* Mammography; BI-RADS; Information theory; Medical information technology.

Serapião PRB, Suzuki KMF, Azevedo-Marques PM. Uso de mineração de texto como ferramenta de avaliação da qualidade informacional em laudos eletrônicos de mamografia. *Radiol Bras.* 2010;43(2):103–107.

\* Trabalho realizado no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (HCFMRP-USP), Ribeirão Preto, SP, Brasil.

1. Bacharel, Doutorando junto ao Programa de Pós-Graduação em Clínica Médica da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP-USP), Ribeirão Preto, SP, Brasil.

2. Analista de Sistemas, Mestre, Doutoranda junto ao Programa de Pós-Graduação em Clínica Médica da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP-USP), Ribeirão Preto, SP, Brasil.

3. Doutor, Engenheiro Eletrônico, Docente do Centro de Ciências das Imagens e Física Médica (CCIFM) do Departamento de Clínica Médica da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP-USP), Ribeirão Preto, SP, Brasil.

Endereço para correspondência: Dr. Paulo Mazzoncini de Azevedo Marques. FMRP-USP, Departamento de Clínica Médica, Avenida dos Bandeirantes, 3900. Ribeirão Preto, SP, Brasil, 14049-900. E-mail: pmarques@fmrp.usp.br

Recebido para publicação em 1/10/2009. Aceito, após revisão, em 24/2/2010.

## INTRODUÇÃO

A radiologia é uma especialidade de intenso relacionamento com outras especialidades médicas. Isto decorre da própria configuração de sua prática, que busca estabelecer elementos de conhecimento diagnóstico por meio de imagens, fomentando a decisão médica de outras áreas. Neste contexto, o relatório clinicoradiológico é a base do relacionamento do radiologista com as demais especialidades, configurando-se como elemento-chave que materializa a interpretação de sua percepção a res-

peito de determinado exame<sup>(1)</sup>. Atualmente, com o crescente desenvolvimento, implementação e uso de sistemas eletrônicos de informação em saúde, a investigação de modelos e padrões que otimizem os processos de criação, gravação, armazenamento e recuperação da informação clínica ganha destaque no meio acadêmico-científico<sup>(2)</sup>. Um dos pontos críticos nesse contexto diz respeito à necessidade de se registrar os sintomas e condutas diagnósticas médicas em uma forma única, seguindo uma lógica que possa ser repetida inúmeras vezes, de forma objetiva, pelos diferentes atores do

processo de atenção à saúde. Este aspecto é particularmente importante quando se considera que existem estudos que apontam que a baixa qualidade de um relatório clínico pode favorecer a ocorrência do erro médico<sup>(3)</sup>.

Uma maneira possível para se alcançar a inclusão das informações em documento eletrônico de forma objetiva é o uso de um padrão informacional. Com isso, pode-se modelar o conteúdo e a estrutura da informação, levando-se em consideração também as necessidades e particularidades específicas das diferentes áreas do conhecimento<sup>(4)</sup>. Os padrões informacionais, utilizados em grande escala na medicina são, na verdade, modelos de representação do conhecimento que visam à organização dos relacionamentos entre informações (conceitos e termos) de um domínio, proporcionando uma manipulação e recuperação eficaz da informação. Os modelos mais usados são: as ontologias, as taxonomias e os tesouros. Basicamente, as ontologias têm como objetivo descrever um domínio de conhecimento contendo relações conceituais semânticas<sup>(5)</sup>. As taxonomias, de outra forma, são sistemas categóricos de organização e representação do conhecimento, bem como sistemas classificatórios de coisas e seres<sup>(6)</sup>. Os tesouros, por sua vez, são vocabulários controlados que servem para melhorar a efetividade da informação registrada e sua recuperação por meio de sistemas humanos e/ou automáticos (eletrônicos)<sup>(7)</sup>. No contexto dos exames mamográficos, o Breast Imaging Reporting and Data System (BI-RADS<sup>®</sup>) pode ser considerado como uma taxonomia voltada para a organização lexical. É um conjunto de termos utilizados para a descrição da avaliação da existência de câncer nas mamas e a categorização de condutas condizentes com os achados diagnosticados pelo médico, o que simplifica e facilita a ação de transcrever a situação da paciente<sup>(8)</sup>. Vários estudos têm apontado a elevada acurácia do BI-RADS como sistema para auxiliar os médicos na descrição das lesões mamárias e na tomada de condutas<sup>(9-13)</sup>. O BI-RADS foi desenvolvido pelo Colégio Americano de Radiologia (ACR), sendo atualmente aceito como padrão pela comunidade médica.

O Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universi-

dade de São Paulo (HCFMRP-USP) possui, desde o ano de 1999, um Sistema de Informação em Radiologia (*Radiology Information System – RIS*) em funcionamento no ciclo de ensino/assistência que possibilita a geração de exames eletronicamente em texto livre<sup>(14)</sup>. Durante seus dez anos de funcionamento, o RIS-HCFMRP acumulou aproximadamente um milhão de relatórios radiológicos, relativos aos diversos tipos de exames que são realizados no Serviço de Radiodiagnóstico do HCFMRP. Todavia, desde o início de seu funcionamento, a qualidade informacional da base de dados do RIS-HCFMRP ainda não foi mensurada de forma objetiva. De um ponto de vista prático, a qualidade do relatório radiológico pode ser aferida a partir da verificação do uso de padrões informacionais, como o BI-RADS, por exemplo.

Estudos<sup>(15-19)</sup> apontam que mineração de texto (*text mining*) é uma técnica adequada para manipulação automática de grandes volumes de dados, pertencente ao campo da ciência da computação, cientificamente ligada ao desenvolvimento de ferramentas de recuperação automática da informação. O método básico consiste em explorar e identificar termos relevantes em um grupo textual ou documental, bem como estabelecer padrões textuais e desenvolver grupos temáticos de assuntos pela frequência de aparecimento de termos no domínio a ser analisado<sup>(18,19)</sup>. Com base no resultado da mineração de texto, é possível identificar com segurança os termos que fazem parte de um determinado conjunto de relatórios.

O presente trabalho teve por objetivo investigar a viabilidade da utilização da técnica de mineração de texto como forma de avaliar a qualidade informacional dos laudos eletrônicos de mamografia, tendo como parâmetro de qualidade a adesão ao léxico BI-RADS.

## MATERIAIS E MÉTODOS

Foram extraídos 22.247 laudos de mamografia do banco de dados do RIS-HCFMRP, compreendidos no período de janeiro de 2000 até junho de 2006. Esses laudos foram organizados em recortes temporais semestrais e exportados para tabelas contendo em cada linha informações refe-

rentes a um laudo e, para cada laudo (linha) uma coluna contendo o número de identificação do laudo no RIS (identificador unívoco do laudo), uma coluna contendo a data do laudo (data na qual o laudo foi revisado e salvo como definitivo no RIS), uma coluna contendo o texto do campo Descrição e uma coluna contendo o texto do campo Conclusão. Como o HCFMRP é um hospital universitário, o laudo é digitado primeiramente pelo residente e depois revisado e aprovado ou corrigido pelos docentes ou radiologistas contratados. Após esse ciclo, o laudo recebe o rótulo de definitivo. Nome da paciente ou qualquer tipo de identificação dos especialistas e/ou residentes foram excluídos do processo de mineração. O trabalho foi aprovado pelo Comitê de Ética em Pesquisa do HCFMRP.

A mineração de texto foi feita utilizando-se um conjunto de ferramentas computacionais voltadas para esse fim comercializadas pela Provalis Research (SimStat, WordStat e QDAMiner) e envolveu duas etapas: identificação e contagem de termos presentes nos campos Descrição e Conclusão dos laudos, e avaliação de palavras-chaves dentro de contexto (*key word on context*), que permite visualizar palavras-chaves dentro de seu texto de origem, possibilitando também a identificação dessa origem, ou seja, a identificação do laudo no qual a palavra-chave foi encontrada. O resultado da primeira etapa permite visualizar as palavras (termos) presentes nos laudos, bem como estabelecer sua frequência dentro da amostra processada de documentos, sem, porém, associar o termo com seu texto de origem. Pode-se também definir uma frequência mínima de repetição para que os termos sejam contabilizados. Como a proposta dessa etapa foi de mapear o conteúdo da base de dados, definiu-se o limiar 1,0 de frequência, que é o mais baixo possível e possibilita a recuperação de todos os termos existentes. Os termos encontrados e suas respectivas frequências podem ser visualizados em ordem alfabética. A partir dos resultados da primeira etapa definiu-se um conjunto de palavras-chaves que foram processadas na segunda etapa usando a ferramenta de contexto. A ferramenta de contexto associa cada palavra-chave ao seu texto de origem, permitindo que laudos que contenham mais de uma

palavra-chave, ou palavras-chaves repetidas, sejam contabilizados somente uma vez. Ou seja, possibilita a contagem do número de laudos do banco de dados que contenham uma ou mais palavras-chaves. A mineração de texto foi aplicada de forma independente nos campos referentes à Descrição e Conclusão dos laudos.

No campo Descrição foi aplicada somente a primeira etapa da mineração, com o objetivo de se verificar a uniformidade de ortografia dos termos utilizados. Dois experimentos foram realizados na base de laudos no campo Conclusão, utilizando-se os resultados da etapa de contextualização: um buscando-se verificar a utilização precisa do léxico, restringindo-se a possibilidade de variação ortográfica – experimento 1 (privilegiando-se a especificidade do método de mineração para detecção do uso do padrão), e outro buscando-se verificar toda e qualquer tentativa de uso ou alusão ao léxico – experimento 2 (privilegiando-se a sensibilidade do método de mineração para detecção do uso do padrão). No experimento 1 considerou-se como palavra-chave apenas o termo BI-RADS, com suas possíveis variantes de escrita com letra minúscula ou maiúscula, uma vez que o programa utilizado não faz essa distinção. No experimento 2, um conjunto maior de palavras-chaves foi utilizado, buscando-se exemplificar possíveis termos e variações associados ao uso do léxico em texto livre. As palavras-chaves utilizadas no experimento 2, selecionadas a partir dos resultados obtidos na etapa 1 de mineração, foram: BI, BIR, BIRADA, BIRADAS, BIRADS, BIRARDS, BIRAS, BIRDAS, BIRDS, BIRRADAS, BIRRADS, BRADS, CAT, CATEG, CATEGIRA, CATEGORA, CATEGORAIA,

CATEGORIA, CATEGORIAI, CATEGORIA CATEORIA, CATERIA, CATERO-RIA, RADAS, RADES, RADS.

**RESULTADOS**

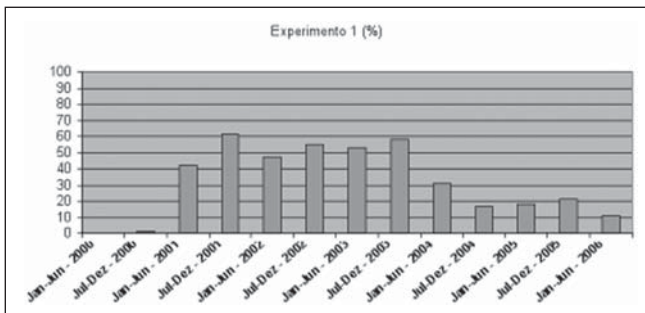
A mineração de texto mostrou que na redação das descrições dos 22.247 laudos foram utilizados 4.435 termos. Destes, uma parcela estatisticamente significativa estava com a grafia incorreta (21%; n = 934). Com isso, o número de termos grafados corretamente utilizados para a descrição das mamografias foi de 3.501. Considerando os termos descritivos médicos, os que tiveram o maior número de erros de grafia dentro do conjunto de dados analisado foram: microcalcificação, lipossustituído, monomórfica, multiductal, nódulo, pleomórfica, Bi-Rads, parênquima, puntiforme, linfonodo e assimétricos. O termo microcalcificação possui o maior número de erros, formando

um conjunto de 36 formas diferentes de escrita no conjunto de dados estudado.

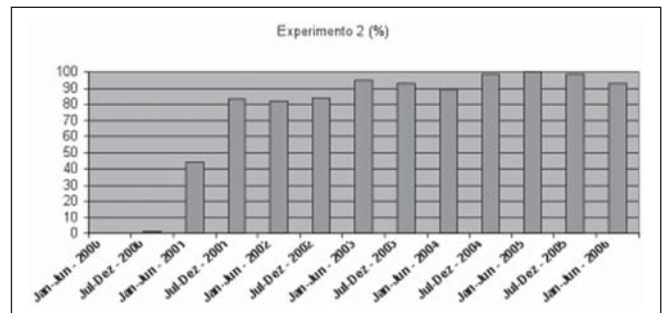
No campo Conclusão, os resultados do experimento 1 (especificidade) mostraram uma variação entre 11% e 61% de laudos contendo termos do léxico em sua conclusão, a partir do ano de 2001. Essa variação, aparentemente, foi aleatória, apresentando oscilação em relação à adesão ou não ao uso do léxico ao longo do tempo, conforme pode ser visto na Tabela 1. Os resultados do experimento 2 (sensibilidade) mostraram uma variação entre 44% e 100% de laudos que se referem de alguma forma ao léxico em sua conclusão, com evidente crescimento de seu uso a partir do ano de 2001 (Tabela 1). As Figuras 1 e 2 apresentam os gráficos em porcentagem do uso do BI-RADS correspondentes aos resultados da mineração de texto nos experimentos 1 e 2, respectivamente. A Figura 3 mostra uma comparação entre os resultados dos

**Tabela 1** Resultados da mineração de texto em valores absolutos e em porcentagem de laudos que utilizam a categorização BI-RADS na conclusão, para os experimentos 1 e 2.

Período	Total de laudos por semestre	Experimento 1 Número de laudos com conclusão BI-RADS (porcentagem de laudos)	Experimento 2 Número de laudos com conclusão BI-RADS (porcentagem de laudos)
Janeiro-junho/2000	1.437	0 (0%)	0 (0%)
Julho-dezembro/2000	1.564	1 (< 1%)	1 (< 1%)
Janeiro-junho/2001	1.409	604 (42%)	621 (44%)
Julho-dezembro/2001	1.686	1.034 (61%)	1.391 (83%)
Janeiro-junho/2002	1.714	816 (47%)	1.403 (82%)
Julho-dezembro/2002	2.012	1.097 (55%)	1.631 (84%)
Janeiro-junho/2003	1.746	1.115 (53%)	1.667 (95%)
Julho-dezembro/2003	1.942	925 (58%)	1.801 (93%)
Janeiro-junho/2004	1.560	550 (31%)	1.391 (89%)
Julho-dezembro/2004	2.064	411 (17%)	2.052 (99%)
Janeiro-junho/2005	2.033	358 (18%)	2.033 (100%)
Julho-dezembro/2005	1.585	377 (21%)	1.568 (99%)
Janeiro-junho/2006	1.495	160 (11%)	1.390 (93%)



**Figura 1.** Gráfico de barras mostrando a adoção do BI-RADS ao longo do tempo junto ao Serviço de Radiodiagnóstico do HCFMRP, considerando a mineração com base no termo BI-RADS (experimento 1).



**Figura 2.** Gráfico de barras mostrando a adoção do BI-RADS ao longo do tempo junto ao Serviço de Radiodiagnóstico do HCFMRP, considerando a mineração com base na lista ampliada de termos (experimento 2).

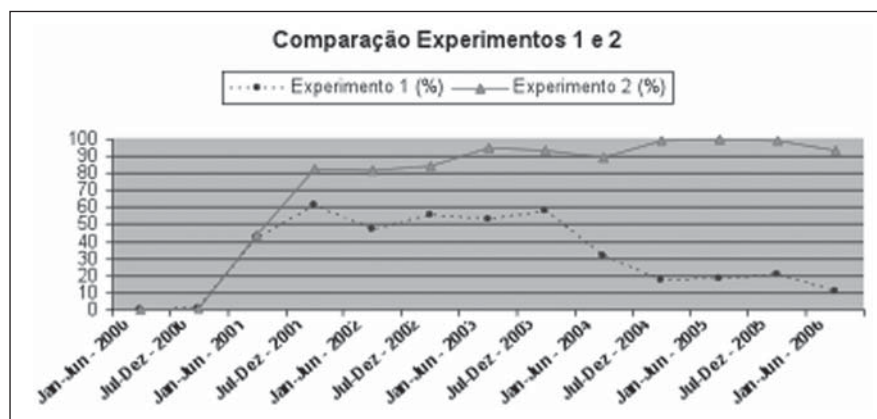


Figura 3. Gráfico de linhas mostrando a comparação entre os resultados dos experimentos 1 e 2.

experimentos 1 e 2, com relação à porcentagem de laudos que utilizam o BI-RADS em sua conclusão ao longo do tempo.

## DISCUSSÃO

Os resultados obtidos pela mineração de texto no campo Descrição explicitam uma variação significativa na forma de redação (erros de ortografia) dos termos descritivos contidos no BI-RADS. Porém, uma avaliação posterior desses erros através da ferramenta *palavra-chave no contexto* evidenciou a possibilidade efetiva de compreensão do conteúdo do campo Descrição. Pode-se concluir, com isso, que provavelmente existe baixa possibilidade de confusão em relação à interpretação diagnóstica devido à presença dessas variantes.

Os resultados obtidos pela mineração de texto no campo Conclusão no experimento 2 evidenciam a crescente adesão ao uso do léxico junto ao Serviço de Radiodiagnóstico ao longo dos anos. Porém, a comparação com os resultados obtidos no experimento 1 deixa claro que ainda é necessário algum trabalho no sentido da uniformização da ortografia e sintaxe na redação da conclusão dos laudos. A partir da comparação desses resultados pode-se concluir que, aparentemente, a partir do ano de 2001 existe uma franca tendência de adoção do BI-RADS no Serviço de Radiodiagnóstico e que, no princípio, os radiologistas devem ter seguido com mais cuidado a ortografia e sintaxe do léxico. Também que, com o passar dos anos, houve efetiva adoção do padrão, porém, com diminuição dos cuidados relativos à forma de escrita dos

termos, principalmente a partir de meados de 2003. Da mesma forma como ocorre no campo Descrição, essas variações de ortografia e sintaxe aparentemente não devem impactar de forma negativa na rotina de atendimento das pacientes, no que se refere ao diagnóstico e conduta. Desse ponto de vista, o entendimento do conteúdo dos laudos por parte dos médicos especialistas, no contexto da área de conhecimento, tende a minimizar, ou até mesmo eliminar, as possíveis confusões que poderiam ser geradas por um uso não uniforme do BI-RADS no que se refere à ortografia e sintaxe. Por exemplo, uma conclusão contendo “Categoria” não deixa dúvidas de que se trata de um achado benigno. Porém, sob um ponto de vista epidemiológico e de gestão em saúde, a existência de uma grande variação na ortografia e sintaxe utilizadas pode dificultar bastante a obtenção de informações corretas para planejamentos e tomadas de decisões. Além disso, no que concerne à formação de recursos humanos, é sempre recomendável a busca pelo ideal.

O BI-RADS, além de um vocabulário, é também um instrumento de padronização informacional dos laudos de mamografia. Sua apropriação na prática cotidiana já foi apontada por outros estudos como limitada nos mesmos aspectos apresentados neste artigo<sup>(20,21)</sup>. Uma possível solução para essa problemática seria o desenvolvimento de documentos eletrônicos estruturados, utilizando o léxico como mapa conceitual e, eventualmente, ampliado com elementos da prática local/institucional. A técnica de mineração de texto, nesse caso, poderia ser muito bem aproveitada, já que o índice de

frequência de termos utilizados nos laudos pode servir de base para a construção de uma estrutura inicial, possibilitando a construção de vocabulários e desenvolvimento de padrões informacionais mesmo em exames ainda não cobertos por um léxico.

## CONCLUSÃO

Os resultados obtidos indicam um bom potencial da aplicação da ferramenta de mineração de texto para a avaliação da qualidade das informações contidas em laudos eletrônicos de mamografia. Particularmente no caso aqui descrito, a mineração de texto evidenciou que existe franca aderência ao uso do BI-RADS ao longo do tempo no Serviço de Radiodiagnóstico do HCFMRP. Evidenciou também que existe grande variação de grafia e sintaxe no uso do léxico, provavelmente resultante do fato de o RIS utilizar texto livre no laudo radiológico, e que a distribuição dessa variação ao longo do tempo parece ser aleatória. A partir da comparação dos resultados obtidos com a mineração de texto é possível se identificar as variações que ocorrem com maior frequência e se buscar a implementação de ações corretivas visando ao uso otimizado do léxico.

## Agradecimentos

À Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) e à Fundação de Apoio ao Ensino Pesquisa e Assistência (Faepa) do HCFMRP, pelo financiamento do estudo. Ao Prof. Dr. Jorge Elias Júnior, do Centro de Ciências das Imagens e Física Médica da FMRP, por sua valiosa colaboração na discussão e interpretação dos resultados.

## REFERÊNCIAS

1. Reiner BI, Knight N, Siegel EL. Radiology reporting, past, present, and future: the radiologist's perspective. *J Am Coll Radiol*. 2007;4:313-9.
2. Shortliffe EH, Perreault LE, Wiederhold G, et al. *Medical informatics: computer applications in health care and biomedicine*. 2nd ed. New York: Springer; 2003.
3. Fitzgerald R. Error in radiology. *Clin Radiol*. 2001;56:938-46.
4. Serapião PRB, Azevedo-Marques PM. Applications in text-mining for the medical information architecture: constructing a representation system of knowledge for clinical practice. III Latin American Medical Informatics Congress, 2008, Buenos Aires. INFOLAC2008, 2008.

5. Rubin DL, Noy NF, Musen MA. Protégé: a tool for managing and using terminology in radiology applications. *J Digit Imaging*. 2007;20(Suppl 1):34–46.
6. Beam C. Interpretation error in mammography: taxonomy and measurement. *Semin Breast Dis*. 2003;6:153–7.
7. National Information Standards Organization. Guidelines for the construction, format, and management of monolingual thesauri. Bethesda: National Information Standards Organization; 2003.
8. Camargo Júnior HSA. BI-RADS®-ultra-som: vantagens e desvantagens dessa nova ferramenta de trabalho. *Radiol Bras*. 2005;38:301–3.
9. Vieira AV, Toigo FT. Predição de malignidade em pacientes das categorias 4 e 5 BI-RADS™. *Radiol Bras*. 2004;37:25–7.
10. Kestelman FP, Souza GA, Thuler LC, et al. Breast Imaging Reporting and Data System – BI-RADS®: valor preditivo positivo das categorias 3, 4 e 5. Revisão sistemática da literatura. *Radiol Bras*. 2007;40:173–7.
11. Roveda Junior D, Piato S, Oliveira VM, et al. Valores preditivos das categorias 3, 4 e 5 do sistema BI-RADS em lesões mamárias nodulares não-palpáveis avaliadas por mamografia, ultra-sonografia e ressonância magnética. *Radiol Bras*. 2007;40:93–8.
12. Melhado VC, Alvares BR, Almeida OJ. Correlação radiológica e histológica de lesões mamárias não-palpáveis em pacientes submetidas a marcação pré-cirúrgica, utilizando-se o sistema BI-RADS. *Radiol Bras*. 2007;40:9–11.
13. Nascimento JHR, Silva VD, Maciel AC. Acurácia dos achados ultrassonográficos do câncer de mama: correlação da classificação BI-RADS® e achados histológicos. *Radiol Bras*. 2009;42:235–40.
14. Azevedo-Marques PM, Caritá EC, Benedicto AA, et al. Integração RIS/PACS no Hospital das Clínicas de Ribeirão Preto: uma solução baseada em “web”. *Radiol Bras*. 2005;38:37–43.
15. Kahn CE Jr, Rubin DL. Automated semantic indexing of figure captions to improve radiology image retrieval. *J Am Med Inform Assoc*. 2009;16:380–6.
16. Huang Y, Lowe HJ, Klein D, et al. Improved identification of noun phrases in clinical radiology reports using a high-performance statistical natural language parser augmented with the UMLS specialist lexicon. *J Am Med Inform Assoc*. 2005;12:275–85.
17. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19:453–73.
18. Konchady M. Text mining application programming. Boston: Charles River Media; 2006.
19. Wives LK. Utilizando *conceitos* como descritores de textos para o processo de identificação de conglomerados (*clustering*) de documentos [tese de doutorado]. Porto Alegre: Universidade Federal do Rio Grande do Sul; 2004.
20. Godinho ER, Koch HA. Submissão às recomendações do BI-RADS™ por médicos e pacientes: análise preliminar de 3.000 exames realizados em uma clínica particular. *Radiol Bras*. 2004;37:21–3.
21. Vieira AV, Toigo FT. Classificação BI-RADS™: categorização de 4.968 mamografias. *Radiol Bras*. 2002;35:205–8.