

ESTUDO DE ANÁLISE QUALITATIVA EM FÓRUM DE DISCUSSÃO

Breno Fabrício Terra Azevedo, IFF, bterra@iff.edu.br
Eliseo Reategui, PGIE/UFRGS, eliseoreategui@gmail.com
Patricia Alejandra Behar, PGIE/UFRGS, pbehar@terra.com.br

Resumo: Este trabalho trata sobre uma proposta para a realização de uma análise das contribuições textuais registradas por alunos em um fórum de discussão. A abordagem desta proposta envolve mineração de textos utilizando grafos. O presente artigo apresenta alguns experimentos realizados em fóruns de discussão e os resultados que podem ser utilizados pelo professor para realizar uma análise qualitativa das contribuições dos alunos.

Palavras-chaves: fórum de discussão, mineração de textos.

Abstract: This paper presents a proposal for the analysis of text contributions posted by the students in a discussion forum. The approach of this proposal involves the use of text mining using graphs. This paper presents some experiments with discussion forums and results that can be used by the teacher to make a qualitative analysis of the student's contributions.

Keywords: discussion forum, text mining.

1. Introdução

O fórum de discussão possibilita a um grupo de pessoas de interesse comum, debater e compartilhar determinadas informações, dúvidas e opiniões (Dornelles, 2001). Nos ambientes virtuais de aprendizagem, os fóruns representam um espaço de interações assíncronas entre os usuários (Behar, 2007).

Segundo Sánchez (2005), o fórum de discussão para fins educacionais em um ambiente on-line é definido como “um espaço de comunicação formado por quadros de diálogo nos quais se inclui mensagens que podem ser classificadas tematicamente. Nestes espaços, os alunos podem realizar contribuições, esclarecer outras, refutar as dos demais participantes, etc., de uma forma assíncrona, sendo possível que as contribuições e mensagens permaneçam todo o tempo à disposição dos demais participantes”. As interações dos alunos nas discussões proporcionam um momento de reflexão sobre os conteúdos educacionais abordados (Palloff e Pratt, 2004).

Este estudo se justifica devido à importância que existe em avaliar qualitativamente as contribuições registradas pelos alunos em um fórum de discussão on-line, isto é, contribuições significativas relativas ao tema em debate. A análise quantitativa é mais simples de ser implementada, como por exemplo, a taxa de participação dos estudantes e o tamanho das discussões. Porém, de acordo com Mazzolini (2007), ela não representa uma forma adequada para retratar as contribuições de um fórum. Entende-se que, para o professor, a análise qualitativa é fundamental para que o mesmo possa realizar um acompanhamento detalhado da construção do conhecimento dos alunos. Além disso, ela permite que o professor possa verificar a real contribuição de cada aluno e observar a própria discussão.

Em um fórum de discussão, o professor sugere um tema a ser discutido pelos alunos e estes apresentam seus relatos. Dependendo do tema proposto, do perfil dos alunos, do

interesse despertado pelo tema, podem-se ter centenas de contribuições. Algumas delas apresentam argumentos favoráveis ao tema proposto e outras apresentam argumentos contrários. Alguns textos redigidos pelos alunos possuem relevância com relação ao assunto tratado, e outros não. Observa-se que as contribuições significativas abordam conceitos importantes relativos ao tema em debate, onde estes conceitos são citados e relacionados entre si.

Para que o professor possa acompanhar minuciosamente todas as contribuições colocadas pelos alunos no debate é necessário um tempo bastante considerável. A importância deste trabalho é proporcionar ao professor a visualização das contribuições relevantes registradas no fórum. Desta forma, o professor poderá observar como os alunos estão interagindo na discussão, visualizar os que colocaram poucos textos significativos no fórum, e estimular aqueles que colocaram muitas contribuições relevantes para interagir com os que colocaram poucas.

Na próxima seção será apresentada uma breve introdução sobre mineração de textos. A seção 3 aborda a metodologia utilizada no trabalho, os resultados encontrados e realiza uma discussão dos mesmos. A seção 4 relata as considerações finais da pesquisa.

2. Mineração de textos

Segundo Feldman (2007), a mineração de textos pode ser definida como um processo intensivo de conhecimento no qual um usuário interage com uma grande quantidade de documentos utilizando ferramentas para análise dos mesmos. O objetivo é extrair informações úteis a partir de coleções de documentos. As informações úteis são identificadas em padrões interessantes nos dados textuais não estruturados.

Os sistemas de mineração de textos baseiam-se em rotinas de pré-processamento, algoritmos para descoberta de padrões, e elementos para apresentação dos resultados. Nestes sistemas, as operações de pré-processamento baseiam-se na identificação e extração de características representativas dos documentos em linguagem natural. Essas operações são responsáveis por transformar dados não estruturados, armazenados em coleções de documentos, em uma estrutura expressa em um formato intermediário (Feldman, 2007; Hearst, 1999; Tan, 1999).

A mineração de textos explora técnicas e metodologias provenientes das áreas de recuperação de informação, extração de informação, e linguística computacional de corpus. Para extrair informações úteis, são descobertas características relevantes nos documentos, cujas mais utilizadas são: caracteres, palavras, termos e conceitos. Podem ser realizadas abordagens híbridas para a geração de representações de documentos baseadas em características (Feldman, 2007). Por exemplo, pode-se primeiro extrair termos do texto, então adaptá-los ou normalizá-los, comparando-os a uma lista de tópicos relevantes (conceitos), extraídos através de uma categorização.

As tecnologias empregadas na mineração incluem extração de informação, rastreamento de tópicos, produção de sumários, categorização de textos, agrupamento de textos, links conceituais, visualização de informação, análise de perguntas e respostas (Fan, 2006).

Uma técnica bastante comum utilizada é a representação das características de um documento através de um modelo de espaço vetorial. Nesta técnica, cada termo do documento torna-se uma característica dimensional. O valor de cada dimensão pode indicar o número de vezes que o termo aparece no texto, ou pode indicar o peso do termo a ser considerado, como por exemplo, a quantidade de documentos nos quais o termo aparece. No entanto, essa técnica descarta informações importantes, como por exemplo, a ordem nos quais os termos aparecem, onde os mesmos aparecem, e a

proximidade entre eles (Schenker, 2003).

Os grafos são construções matemáticas importantes e efetivas para realizar a modelagem de relacionamentos e de informação estrutural. Os grafos são utilizados em diversos tipos de problemas, incluindo ordenação, compressão, análise de tráfego, alocação de recursos, entre outros. Como os grafos retêm mais informações do que vetores de características atômicas simples, eles representam uma modelagem valiosa, e podem ser utilizados para a representação de textos (Schenker, 2003).

O software Sobek (Lorenzatti, 2007) é um exemplo de utilização de mineração de textos baseado em métodos estatísticos. Esse software permite construir um grafo a partir de informações estatísticas obtidas do texto, onde os vértices e as arestas possuem informações referentes ao número absoluto e relativo de ocorrência dos termos (vértices) e associações (arestas) em determinado documento. O grafo obtido a partir do documento representa uma rede dos conceitos que foram trabalhados dentro do texto. Macedo (2008, 2009), utiliza o Sobek para analisar as contribuições dos alunos na construção de textos coletivos. Sua pesquisa tem como foco apoiar o trabalho do professor, identificando através da visualização dos grafos problemas nos textos envolvendo coesão, coerência, entre outros. Este trabalho utiliza o Sobek para gerar os grafos das contribuições textuais produzidas pelos alunos no fórum. A partir destes grafos, foi feita uma análise dos mesmos para identificar se os textos possuem relevância, ou se tem conexão, integração entre os conceitos obtidos, com o tema do debate.

Pesquisas têm sido realizadas com o intuito de auxiliar a análise de fóruns de discussão. Gerosa (2003) propôs um acompanhamento das discussões através de aspectos relacionados à forma de estruturação do discurso e à categorização das mensagens. A estruturação permite observar o encadeamento hierárquico da discussão para ter indícios do aprofundamento da discussão. A categorização das mensagens auxilia na identificação das características de cada tipo de mensagem.

Os trabalhos de Bassani e Behar(2006), Li(2008), apresentam uma metodologia para analisar os fóruns de discussão envolvendo o mapeamento dos fluxos de interação entre os participantes.

Dringus e Ellis (2005) conduziram uma das primeiras pesquisas na utilização de mineração de dados e de textos para analisar fóruns de discussão. Eles integraram a mineração no processo de consulta do fórum e desenvolveram um conjunto de ferramentas para auxiliar a visualização dos dados relacionados com datas, ritmo de contribuição, e sequências de trocas de mensagens.

Outra pesquisa para analisar fóruns de discussão é apresentada por Lin (2009). O trabalho utiliza mineração de textos e propõe um sistema para classificação dos gêneros das contribuições, como: anúncio, pergunta, explicação, interpretação, conflito, afirmação, entre outros. Este sistema pode ser utilizado para facilitar o processo de codificação da análise do conteúdo de um fórum.

No trabalho desenvolvido por Chen (2008), são apresentadas algumas características que podem ser consideradas para medir tópicos de alta qualidade em fóruns de discussão. Baseado nestas características foi proposto um modelo para reconhecê-las que envolve: a extração de sinais de séries temporais que contêm indícios entre tópicos de alta e não alta qualidade, o diagnóstico de características destes sinais utilizando a transformada de pacote wavelet, e em seguida o reconhecimento dos tópicos de alta qualidade utilizando uma rede neural back-propagation.

3. Análise qualitativa em fóruns de discussão: um estudo piloto

Para analisar as contribuições textuais produzidas pelos alunos nos fóruns, foi utilizado um software de mineração de textos chamado Sobek (Lorenzatti, 2007), cujo objetivo é gerar um grafo que representa os conceitos e suas associações em um texto. A partir dos grafos gerados, foi realizado manualmente um estudo dos mesmos, para analisar quais contribuições foram relevantes de acordo com o tema proposto.

Para direcionar as atividades deste trabalho definiu-se o coeficiente de relevância temática (CRT) de uma contribuição textual, cujo objetivo é analisar o quanto um texto é relevante a um determinado tema de discussão. O coeficiente de relevância temática indica o grau de relevância da contribuição textual de acordo com o tema do fórum.

O coeficiente de relevância pode ser calculado a partir da seguinte fórmula:

- NC: número de conceitos relevantes utilizados no texto
- NA: número de associações entre os conceitos relevantes utilizados no texto

$$CRT = NC + NA$$

Para o cálculo do NC, os termos considerados como equivalentes semânticos também são considerados como conceitos relevantes. Os prefixos, sufixos, e plurais dos conceitos relevantes citados no texto são convertidos, para também serem considerados como conceitos relevantes. Com o CRT, os grafos que possuem mais conceitos importantes relativos ao tema, com mais associações entre os conceitos, possuem maior coeficiente de relevância temática.

A metodologia empregada envolveu:

- a) A escolha de um fórum de discussão.
- b) A definição dos conceitos importantes relativos ao tema do fórum, bem como a definição da associação entre estes conceitos.
- c) A definição de possíveis termos que podem ser considerados como equivalentes semânticos aos conceitos envolvidos.
- d) A definição do valor mínimo do coeficiente de relevância temática que será considerado para análise.
- e) A coleta das contribuições textuais produzidas pelos alunos no fórum.
- f) Para cada contribuição textual, foi gerado o seu grafo conceitual com o Sobek.
- g) A partir de cada grafo gerado, foi calculado o coeficiente de relevância temática da contribuição textual.
- h) A partir da organização realizada com os grafos gerados, as contribuições textuais relativas aos mesmos foram organizadas.
- i) A partir dos grafos gerados foi realizada uma análise da quantidade de contribuições relevantes colocadas por cada aluno no fórum.

Para avaliar a metodologia proposta e analisar os resultados, foram realizados três experimentos, com fóruns de discussão distintos.

O primeiro experimento foi realizado em um fórum de discussão do Ambiente ROODA, na disciplina “Tópicos Especiais Z1” da turma do Doutorado em Informática na Educação, que ocorreu no primeiro semestre de 2008. O tema do fórum foi

“Comunidades Virtuais”. O fórum contou com a participação de 05 alunos, que realizaram 25 contribuições textuais.

Para este fórum, o professor definiu como importantes, os seguintes conceitos: “comunidades”, “virtuais”, “comunidades virtuais”, “aprendizagem”, “comunidades virtuais de aprendizagem”, “educação”, “relações sociais”. As associações relevantes entre os conceitos ficaram estabelecidas da seguinte maneira: o conceito “comunidades virtuais de aprendizagem” está relacionado com os conceitos “comunidades”, “comunidades virtuais”, “aprendizagem”, “educação”. Os termos “CVs” e “CVAs” foram definidos como equivalentes semânticos aos termos “comunidades virtuais” e “comunidades virtuais de aprendizagem”, respectivamente.

Foi definido o valor mínimo 01 para o coeficiente de relevância temática (CRT) a ser considerado na análise. O coeficiente de relevância temática foi calculado a partir do grafo gerado em cada contribuição textual. Como um exemplo de cálculo do CRT, pode-se observar a mensagem abaixo extraída do fórum, postada pelo aluno C.

Mensagem: “Acredito que as CVAs podem oferecer um recurso valioso à educação, pois elas possibilitam um mecanismo de construção coletiva de conhecimento, através das discussões realizadas pelos membros da comunidade.”

O grafo gerado a partir da mensagem anterior pode ser visualizado na figura 1.

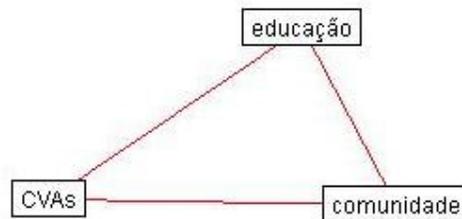


Figura 1 – Grafo gerado a partir de uma mensagem do fórum do experimento 1.

Para o grafo anterior, o CRT é calculado da seguinte maneira:

$$NC = 3$$

$$NA = 3$$

$$CRT = NC + NA = 6$$

Foi realizada uma análise da quantidade de contribuições relevantes colocadas por cada aluno no fórum. Os resultados são apresentados na tabela 1.

Tabela 1 - Quantidade de contribuições relevantes do primeiro experimento.

Aluno	Quantidade de contribuições relevantes
01	32%
02	24%
03	16%
04	12%
05	4%
Contribuições não relevantes	12%

Através da tabela 1, percebeu-se que a utilização da metodologia proposta permitiu verificar que a maior parte das contribuições feitas pelos estudantes era relevante.

O segundo e terceiro experimentos foram realizados em um fórum de discussão do Ambiente ROODA, na disciplina “Seminário Integrador VII – B” de uma turma do Curso de Graduação em Pedagogia, que ocorreu no segundo semestre de 2009. O tema do fórum foi “Aprender com os outros”. A turma foi dividida em dois grupos, e cada grupo contribuiu em um fórum de discussão. O primeiro fórum contou com a participação de 27 alunos, que realizaram 45 contribuições textuais. O segundo fórum contou com a participação de 29 alunos, que realizaram 67 contribuições textuais.

Para estes fóruns, o professor definiu como importantes, os seguintes conceitos: “aprender”, “aprendizagem”, “conhecimento”, “construção do conhecimento”, “Piaget”, “afeto”, “afetividade”, “estruturas cognitivas”, “descentração”, “relações afetivas”, “experiências afetivas”, “laços afetivos”, “cooperação”, “colaboração”. Foram estabelecidas as associações relevantes entre estes conceitos.

Foi definido o valor mínimo 01 para o coeficiente de relevância temática (CRT) a ser considerado na análise. O coeficiente de relevância temática foi calculado a partir do grafo gerado em cada contribuição textual. Foi realizada uma análise da quantidade de contribuições relevantes colocadas por cada aluno do primeiro e segundo grupos. Os resultados são apresentados nas tabelas 2 e 3, respectivamente.

Tabela 2 - Quantidade de contribuições relevantes do primeiro grupo.

Aluno	Quantidade de contribuições relevantes
01	4,4%
02	4,4%
03	8,8%
04	6,6%
05	8,8%
06	2,2%
07	8,8%
08	2,2%
09	2,2%
10	2,2%
11	2,2%
12	2,2%
13	2,2%
14	2,2%
15	4,4%
16	4,4%
17	2,2%
18	2,2%
19	2,2%
20	4,4%
21	4,4%
22	4,4%
23	2,2%

24	2,2%
25	2,2%
26	2,2%
27	2,2%
Contribuições não relevantes	0%

Tabela 3 - Quantidade de contribuições relevantes do segundo grupo.

Aluno	Quantidade de contribuições relevantes
01	5,97%
02	5,97%
03	1,49%
04	2,98%
05	5,97%
06	1,49%
07	8,95%
08	2,98%
09	2,98%
10	2,98%
11	2,98%
12	2,98%
13	2,98%
14	1,49%
15	1,49%
16	2,98%
17	1,49%
18	1,49%
19	1,49%
20	1,49%
21	1,49%
22	2,98%
23	1,49%
24	1,49%
25	1,49%
26	1,49%
27	2,98%
28	1,49%
29	1,49%
Contribuições não relevantes	20,89%

A partir dos resultados apresentados nas tabelas 1, 2, 3, percebe-se que a metodologia empregada apóia a avaliação qualitativa, na medida em que ela verifica se conceitos importantes de serem tratados estão efetivamente sendo considerados nos textos dos

estudantes. Em todas as tabelas pode-se verificar a quantidade de contribuições relevantes ao tema do fórum produzidas por cada aluno, e a quantidade total de contribuições não relevantes.

Observou-se que a versão atual do Sobek, utilizada nos experimentos pilotos, realiza perfeitamente a mineração de um texto e a geração do grafo representativo dos conceitos encontrados no mesmo. No entanto, as outras atividades desenvolvidas nos experimentos foram executadas manualmente.

Para permitir ao Sobek oferecer todos os recursos necessários ao professor, citados na metodologia proposta neste trabalho, será necessário aprimorá-lo e criar novas funcionalidades como:

- a) Realizar o stemming¹ de palavras.
- b) Permitir a leitura de um conjunto de conceitos que possuam equivalência semântica.
- c) Permitir a identificação de perguntas feitas pelos alunos.
- d) Permitir a leitura automática dos textos provenientes de um fórum de discussão.
- e) Permitir a definição do valor mínimo do coeficiente de relevância temática que será considerado para análise.
- f) Executar o cálculo do coeficiente de relevância temática para cada grafo gerado a partir dos textos produzidos pelos alunos.
- g) A partir dos grafos gerados, realizar uma análise da quantidade de contribuições relevantes colocadas por cada aluno no fórum.
- h) Gerar um relatório visual para o professor, contendo informações sobre as contribuições relevantes de um tema de discussão. Estas informações disponibilizadas ao professor indicarão a quantidade de contribuições relevantes de cada aluno e a quantidade de contribuições não relevantes.

Com base nos resultados deste trabalho, verificou-se que a metodologia apresentada pode auxiliar o professor a realizar diversas atividades relacionadas ao fórum, como:

- Analisar quais contribuições textuais necessitam de intervenção.
- Visualizar os alunos que colocaram apenas algumas contribuições relevantes no fórum para oferecer-lhes maior auxílio.
- Estimular os alunos que colocaram muitas contribuições relevantes para interagir com os alunos que colocaram poucas.

A partir das atividades citadas, o professor pode direcionar seu auxílio aos alunos que registraram poucas contribuições relevantes. O professor pode também, estimular um processo de aprendizagem colaborativa, sugerindo a interação entre os alunos que mais registraram textos relevantes com os alunos que redigiram poucos.

4. Considerações finais

A partir dos experimentos iniciais realizados neste trabalho, observou-se que mineração de textos utilizando a representação através de grafos, pode ser utilizada na criação de indicadores de relevância das postagens de um fórum. Através desta análise, o professor

¹ *Stemming* é a redução de um conjunto de palavras ao mesmo *stem*, através da retirada dos prefixos e sufixos, permanecendo apenas a raiz das palavras (Spark-Jones, 1997).
V. 7 Nº 3, dezembro, 2009

tem a sua disposição uma ferramenta para auxiliá-lo a acompanhar as contribuições relevantes produzidas pelos alunos em um fórum de discussão.

No desenvolvimento das novas funcionalidades do Sobek, também pretende-se implementar uma ordenação das contribuições relevantes de um tema de discussão, baseado nos grafos gerados. Com esta ordenação, o professor poderá visualizar quais foram as contribuições textuais mais relevantes colocadas no fórum. Esta informação pode fornecer ao professor outro indicador para que seja possível identificar quais alunos estão trabalhando melhor os conceitos de um determinado tema de um fórum de discussão.

Com o aprimoramento do Sobek, o mesmo poderá ser adaptado a diferentes Ambientes Virtuais de Aprendizagem e, inclusive para trabalhar com vários formatos de fóruns de discussão. A utilização da metodologia proposta neste trabalho, bem como a implementação das novas funcionalidades no Sobek fazem parte de uma pesquisa de Doutorado em desenvolvimento².

Referências bibliográficas

BASSANI, P. S.; BEHAR, P. A. Análise das interações em ambientes virtuais de aprendizagem: uma possibilidade para avaliação da aprendizagem em EAD. **Revista Novas Tecnologias na Educação (RENOTE)**. Porto Alegre, v.4, n.1, jul. 2006.

BEHAR, P. A.; BERNARDI, M.; SOUZA, A. P. F. de Castro; KELLEN, K. ROODA: desenvolvimento, implementação e validação de um AVA para UFRGS. In: **TALLER INTERNACIONAL DE SOFTWARE EDUCATIVO (TISE)**, XII, 2007, Santiago/Chile. Memorias del XII Taller Internacional. Santiago: LOM Ediciones S.A., 2007, v. 1, p. 321-338.

CHEN, Y.; CHENG, X.; HUANG, Y. A Wavelet-Based Model to Recognize High-Quality Topics on Web Forum. In: **IEEE/WIC/ACM INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE AND INTELLIGENT AGENT TECHNOLOGY**, 2008. Sydney. Proceedings. Sydney: IEEE/WIC/ACM, 2008, v.1, p.343-351.

DORNELLES, R. J. **A utilização de tecnologias de Internet na educação a distância: o caso de uma disciplina de graduação da Escola de Administração da Universidade Federal do Rio Grande do Sul**. Porto Alegre: UFRGS, 2001. Dissertação de Mestrado em Administração. Disponível em: < http://www.ea.ufrgs.br/professores/hfreitas/files/orientacao/mestrado/defesa/pdf/28_dissertacao_dornelles.pdf>. Acesso em: 28 ago. 2009.

DRINGUS, L. P.; ELLIS, T. Using data mining as a strategy for assessing asynchronous discussion forums. **Computers & Education**, Oxford, v. 1, n. 45, p.141-160, 2005.

FAN, W.; WALLACE, L.; RICH, S.; ZHANG, Z. Tapping the power of text mining. **Communications of ACM**, Nova York, v. 9, n. 49, p.76-82, 2006.

FELDMAN, R.; SANGER, J. **The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data**. Cambridge, MA: Cambridge University Press, 2007.

GEROSA, M. A.; PIMENTEL, M. G.; FUKS, H.; LUCENA, C. J. P. Coordenação de

² PGIE/UFRGS

V. 7 Nº 3, dezembro, 2009

- Fóruns Educacionais: Encadeamento e Categorização de Mensagens. In: **SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (SBIE)**, XIV, 2003, Rio de Janeiro. Anais. Rio de Janeiro: Sociedade Brasileira de Computação, 2003, p. 45-54.
- HEARST, M. Untangling Text Mining. In: **ANNUAL MEETING OF THE ASSOCIATION OF COMPUTATIONAL LINGUISTICS**, 37th, 1999. University of Maryland, MD, Association of Computational Linguistics, 1999, p.3-10.
- LI, Y; HUANG, R.; Analyzing Peer Interactions in Computer-Supported Collaborative Learning: Model, Method and Tool. **Lecture Notes in Computer Science (LNCS)**. Berlim: Springer-Verlag Berlim Heidelberg, v. 5169, p. 125-136, 2008.
- LIN, Fu-Ren; HSIEH, Lu-Shih; CHUANG, Fu-Tai. Discovering genres of online discussion threads via text mining. **Computers & Education**, Oxford, v. 2, n. 52, p.481-495, 2009.
- LORENZATTI, A. **SOBEK**: uma Ferramenta de Mineração de Textos. Caxias do Sul/RS: UCS, 2007. Trabalho de Conclusão de Curso.
- MACEDO, A. L. **Um recurso de apoio para acompanhamento e avaliação do estudante na construção de um texto coletivo**. Porto Alegre/RS: PPGIE/UFRGS, 2008. Proposta de Tese de Doutorado.
- MACEDO, A.; REATEGUI, E.; LORENZATTI, A.; BEHAR, P. Using Text-Mining to Support the Evaluation of Texts Produced Collaboratively. In: Tatnall, A., Jones, A.. (Org.). **Education and Technology for a Better World**: Selected papers of the 9th IFIP TC3 World Conference on Computers in Education, WCCE. Berlim: Springer, p. 368-377, 2009.
- MAZZOLINI, M. When to jump in: The role of the instructor in online discussion forums. **Computers & Education**, Oxford, v. 2, n. 49, p.193-213, 2007.
- PALLOFF, R. M.; PRATT, K. **O aluno virtual**: um guia para trabalhar com estudantes on-line. Porto Alegre: Artmed, 2004.
- SÁNCHEZ, Lourdes P. El foro virtual como espacio educativo: propuestas didácticas para su uso. **Verista Quaderns Digitals Net**, n.40, p.1-18, 2005. Disponível em: <http://www.quadernsdigitals.net/datos_web/hemeroteca/r_1/nr_662/a_8878/8878.html>. Acesso em: 27 ago. 2009.
- SCHENKER, Adam. **Graph-Theoretic Techniques for Web Content Mining**. Florida: University of South Florida, 2003. PhD Thesis.
- SPARK-JONES, K., WILLET, P. **Readings in Information Retrieval**. San Francisco: Morgan Kaufmann, 1997.
- TAN, A. Text Mining: The State of the Art and the Challenges. In: **Workshop on Knowledge Discovery from Advanced Databases (PKDAD'99)**, 1999, Beijing. Proceedings. Beijing, 1999, p.71-76.