

Análise das mensagens de fóruns de discussão através de um software para mineração de textos

Breno Fabrício Terra Azevedo¹, Patricia Alejandra Behar², Eliseo Berni Reategui²

¹Instituto Federal de Educação, Ciência e Tecnologia Fluminense - Rua Dr. Siqueira, 273 - Parque Dom Bosco - CEP 28030-130 - Campos dos Goytacazes - RJ - Brasil

²Universidade Federal do Rio Grande do Sul - Av. Paulo Gama, 110 - prédio 12105 - 3º andar sala 332 - CEP 90040-060 - Porto Alegre - RS - Brasil

bterra@iff.edu.br, pbehar@terra.com.br,
eliseoreategui@gmail.com

Abstract. *This paper presents the results of a qualitative analysis of textual contributions recorded by students in discussion forums. The study was performed with the software MineraFórum. It applies text mining techniques using graphs to identify which students wrote relevant messages, and which students did not. Thus, it is possible to have subsidies to assess which students need more help, and motivate them to discuss the important concepts related to the topic under discussion.*

Resumo. *Este artigo apresenta os resultados de uma análise qualitativa das contribuições textuais registradas por alunos em fóruns de discussão. O estudo foi efetuado com o software MineraFórum. O referido programa aplica técnicas de mineração de textos utilizando grafos para identificar quais alunos redigiram mensagens relevantes e quais não o fizeram. Desta forma, é possível ter subsídios para avaliar os discentes que necessitam de maior auxílio e motivá-los para discutir os conceitos importantes que fazem parte do tema em debate.*

1. Introdução

O envolvimento em fóruns de discussão é uma atividade importante para os alunos. Através da análise da interação dos discentes nos fóruns, o professor pode diagnosticar informações sobre os estudantes. No entanto, se o docente possuir uma grande quantidade de alunos, o tempo necessário para que ele consiga analisar as discussões será grande. Um recurso que permita fazer, de forma automatizada e rápida, uma análise das mensagens postadas em um fórum de discussão, pode auxiliar o professor. Isto possibilitará que o docente identifique os alunos que estão (e aqueles que não estão) discutindo os conceitos do tema em debate. Desta forma, o docente poderá dedicar um tempo maior para descobrir o motivo pelo qual alguns discentes não discutiram os conceitos referentes ao tema. Caso o professor identifique alunos com dificuldade para compreender os conceitos importantes, ele poderá oferecer ajuda.

De acordo com Palloff e Pratt (2004), o papel do professor nos fóruns de discussão é muito importante. Além de escrever mensagens de apoio e motivação para os alunos, responder às perguntas, ele deve observar o nível de participação de cada discente. Caso seja detectado que um aluno não está participando ou está desviando do tema da discussão, o docente deve tentar ajudá-lo a suplantando as dificuldades e a

solucionar os problemas.

Para auxiliar na análise qualitativa dos textos produzidos pelos alunos em um fórum de discussão, este artigo apresenta um estudo realizado com o software MineraFórum.

O MineraFórum utiliza técnicas de mineração de textos para analisar as mensagens de fóruns de discussão. A partir desta análise, é possível identificar se as contribuições textuais produzidas pelos discentes são relevantes (ou não) para o debate.

Na próxima seção será apresentada uma breve introdução sobre mineração de textos. A seção 3 relata alguns trabalhos que empregam esta técnica na análise de fóruns de discussão. A seção 4 explica o MineraFórum. A seção 5 descreve as experiências realizadas. A seção 6 relata as considerações finais do artigo.

2. Mineração de textos

De acordo com Tan (1999), a mineração de textos trata do processo de extrair padrões interessantes e não-triviais de conhecimento a partir de textos.

Segundo Feldman e Sanger (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. Estas informações são identificadas em padrões interessantes nos dados textuais não estruturados.

Na mineração de textos, uma técnica bastante comum utilizada para representar um documento é identificar as características do mesmo 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].

A técnica de mineração de textos utilizando grafos descobre as palavras com maior ocorrência no texto, e identifica se elas estão próximas. O grafo obtido a partir da mineração apresenta em seus nós as palavras que mais ocorreram. As associações entre os nós do grafo indicam a proximidade entre as palavras. A figura 1 apresenta o grafo gerado a partir do texto “Existem diversas técnicas de mineração de textos”.



Figura 1. Grafo gerado a partir de um texto.

3. Análise de fóruns de discussão com mineração de textos

Dringus e Ellis (2005) conduziram um estudo para avaliar a utilização de mineração de dados e de textos para analisar fóruns de discussão. No artigo, os autores discutem os problemas que contribuem para dificultar a avaliação destes ambientes. Eles sugerem indicadores de participação que o professor pode extrair das discussões e utilizar para avaliar o progresso e o desempenho dos alunos. Os autores recomendam que a mineração de dados seja utilizada para descobrir a distribuição e frequência das contribuições do participante ao longo do tempo de duração do fórum. Ela pode ser empregada também para identificar iniciativas ou respostas, e descobrir quanto tempo após a postagem inicial, a resposta foi emitida. A pesquisa indica que a mineração de textos seja utilizada para avaliar se uma contribuição está relacionada ao tópico do fórum, e se os alunos compartilham recursos, como referências da Web ou citações da literatura. De acordo com o estudo, as técnicas de mineração de textos e dados podem ser parte da solução para auxiliar o professor a analisar fóruns de discussão e obter *feedback* dos alunos.

Kim et al. (2007) apresentaram em seu artigo um agente inteligente, que foi implementado dentro de um fórum de discussão, para fornecer automaticamente respostas às perguntas dos alunos. O trabalho apresenta como os tópicos de discussão foram modelados utilizando “atos de discurso”. Cada postagem foi classificada de acordo com categorias de “atos de discurso”, como: pergunta, resposta, elaboração e correção. Os autores desenvolveram um conjunto de padrões para analisar as interações dos discentes nas discussões. Alguns destes padrões foram utilizados para descobrir mensagens onde os alunos poderiam ter perguntas não respondidas. O agente inteligente utiliza técnicas de mineração de texto para extrair palavras e suas frequências da pergunta do aluno, dos documentos do curso, e das discussões anteriores. A medida de similaridade cosseno¹ é utilizada para descobrir quais destes documentos e/ou discussões passadas possuem semelhança com o texto envolvido na pergunta do aluno.

Li e Huang (2008) apresentaram uma pesquisa para fornecer um retrato mais completo das interações entre os sujeitos, na aprendizagem colaborativa suportada por computador. Os autores propõem um modelo de análise multidimensional para investigar as interações, baseado em técnicas de análise de conteúdo, mineração de texto, e análise de redes sociais. A análise de conteúdo é empregada para pesquisar como os alunos interagem, descobrindo os possíveis padrões de processo (a intenção do discurso) dentro da conversação. A mineração de texto é utilizada para descobrir os tópicos que apareceram nos debates. As mensagens de um conjunto de discussões são agrupadas em um documento, que é representado por um vetor de termos ponderados. O método do cosseno é utilizado para calcular a similaridade entre o vetor do documento e o vetor do tema, definido pelo professor. O estudo também descreve o projeto e implementação de uma ferramenta para análise inteligente de conteúdo, denominada VINCA (Visual Intelligent Content Analyzer). Foi realizado um experimento com a ferramenta para analisar um conjunto de discussões, com a finalidade de desvendar a interação dos alunos em termos de padrão de processo, espaço de tópicos, e rede social.

Outra pesquisa para analisar fóruns de discussão é apresentada por Lin et al. (2009). O trabalho utiliza mineração de textos e propõe um sistema de classificação dos

¹ A medida de similaridade cosseno é um método para calcular a distância entre dois vetores que representam documentos [Feldman e Sanger 2007].

gêneros das contribuições textuais, 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. Foram coletados dados de um fórum de discussão do Moodle para realizar os experimentos. A maior questão de pesquisa no estudo foi validar a coerência dos resultados codificados por um sistema automático de classificação de gêneros e a análise realizada por juízes humanos. O artigo conclui que o modelo em cascata, embutido no sistema desenvolvido, pode facilitar o processo de codificação da análise de conteúdo de fóruns.

4. MineraFórum

O MineraFórum é um software que realiza a análise qualitativa das mensagens redigidas em fóruns de discussão. Este foi desenvolvido no NUTED²-PGIE/UFRGS. Ele é capaz de calcular a relevância de cada postagem dentro do debate. Para analisar o conteúdo das contribuições textuais, o programa emprega a técnica de mineração de textos utilizando grafos.

Devido aos resultados promissores apresentados no trabalho de Azevedo et al. (2010), verificou-se que a mineração utilizando grafos é uma opção viável para a análise das mensagens de fóruns de discussão. Considerando-se o grafo gerado a partir de uma mensagem, pode-se avaliar se ela refere-se ao contexto, e avaliar a relevância temática da mesma. Ao analisar uma mensagem, quanto mais palavras relativas ao tema forem citadas, e quanto mais próximas elas estiverem umas das outras, maior será a relevância temática. Desta forma, esta técnica é utilizada pelo MineraFórum.

Alguns recursos oferecidos pelo MineraFórum são relacionados abaixo:

- Permite que o usuário informe um texto de referência sobre o tema em debate.
- Caso o usuário deseje, ao invés de informar um texto de referência, é possível digitar os conceitos que são relevantes ao tópico da discussão e realizar associações entre eles.
- Utiliza um dicionário de sinônimos no processo de mineração. Este dicionário foi previamente definido dentro do software. No entanto, caso o usuário necessite, outro dicionário pode ser informado. Os sinônimos são importantes quando o MineraFórum compara as palavras digitadas nas mensagens com os conceitos relevantes identificados no texto de referência.
- Além do dicionário de sinônimos, o usuário pode informar palavras que possuam equivalência semântica.
- Calcula a relevância de cada postagem.
- Exibe um gráfico com a média dos valores de relevância das mensagens postadas por cada autor.
- Identifica se existem mensagens similares redigidas no fórum de discussão.
- Possibilita que os resultados do processo de mineração sejam armazenados em arquivo (no formato “html”).

² Núcleo de Tecnologia Digital aplicada à Educação / Pós-graduação em Informática na Educação

- Exibe um relatório com informações sobre a análise das postagens: o total de mensagens redigidas por cada aluno, a quantidade de contribuições relevantes feitas por cada discente, os conceitos que foram utilizados nas postagens relevantes, a relevância de cada mensagem, se a mensagem é similar (ou não) a alguma outra do fórum, a média das relevâncias das mensagens de cada aluno, a quantidade de vezes que cada mensagem foi citada dentro do debate.

A figura 2 exibe a interface do MineraFórum após o usuário selecionar o botão “Minerar fórum”. As informações sobre a mineração das postagens são apresentadas. Nesta tela, o usuário também pode escolher outras funcionalidades.



Figura 2. Interface do MineraFórum com a aba “Minerar fórum” selecionada.

No processo de análise das mensagens postadas, o MineraFórum organiza e agrupa as mensagens de cada aluno. O software calcula o valor da relevância de cada mensagem. Para realizar este procedimento, três critérios são considerados: a relevância temática da mensagem (RT), a relevância de citações da mensagem (RM), e a similaridade da mensagem (SM).

Para calcular a relevância temática das mensagens (RT), o MineraFórum executa as seguintes ações:

- a) Constrói um grafo a partir do texto de referência sobre o tema do debate, indicado pelo usuário. Neste processo, as *stopwords* (palavras que podem ser descartadas no processo de mineração, como: advérbios, artigos, preposições, entre outras) são eliminadas e as palavras com maior ocorrência dentro do texto são identificadas. As palavras com maior incidência representam os conceitos mais relevantes do texto minerado e correspondem

aos vértices do grafo. As arestas entre os vértices são criadas de acordo com a proximidade entre as palavras dentro do texto. Caso o usuário decida por digitar os conceitos importantes relacionados ao tópico da discussão, ao invés de indicar o texto de referência, o MineraFórum constrói o grafo a partir destes conceitos.

- b) Criação de um grafo a partir de cada mensagem redigida no fórum.
- c) Para calcular a relevância temática de uma mensagem, o MineraFórum deve analisar a correspondência entre o grafo gerado a partir do texto de referência e o grafo construído a partir da mensagem. Na primeira etapa desta análise são identificados quais vértices do primeiro grafo são equivalentes a outros no segundo grafo. No contexto desta pesquisa, dois vértices são equivalentes se eles possuem conteúdo semelhante, isto é, se possuem as mesmas palavras, ou se as palavras podem ser reduzidas ao mesmo radical, ou se possuem sinônimos, ou se possuem palavras com equivalência semântica. Na segunda etapa da análise, é utilizada uma fórmula que considera três aspectos dos vértices equivalentes: a quantidade destes vértices nos dois grafos, a distância entre eles dentro do seu respectivo grafo e o peso deles no seu grafo. O valor deste resultado corresponde à relevância temática da mensagem (RT).

Para calcular o valor de RM de uma mensagem, o software divide a quantidade de citações da mensagem pelo total de postagens no fórum. Para identificar se uma mensagem é similar a alguma outra do fórum (SM), o MineraFórum analisa a correspondência entre os grafos de todas as mensagens. Caso a mensagem seja semelhante a outra do fórum, o valor de SM será igual a RT, com sinal negativo.

O valor da relevância de uma mensagem (RF) será obtido com a média ponderada entre RT e RM. Caso a mensagem seja similar a outra do fórum, o valor de SM será subtraído de RF. Se a contribuição textual não for semelhante a outra do fórum, então o valor calculado para RF será mantido.

O valor final de RF é convertido em um valor inteiro, em uma escala de 0 (zero) a 5 (cinco). O valor zero significa que a mensagem não possui relevância dentro do debate. O valor cinco indica que a postagem possui relevância máxima.

O MineraFórum segue as sugestões indicadas no estudo de Dringus e Ellis (2005), pois utiliza mineração de textos para avaliar se uma mensagem está relacionada ao tema em debate, e auxilia o professor a obter informações sobre os discentes.

A tabela 1 apresenta uma comparação entre o MineraFórum e os trabalhos correlatos apresentados na seção 3.

Tabela 1. Comparação entre o MineraFórum e trabalhos correlatos.

Autores	Semelhanças entre o MineraFórum e trabalhos correlatos	Diferenças entre o MineraFórum e trabalhos correlatos
Kim et al. (2007)	Agente inteligente que analisa o conteúdo de mensagens de fóruns.	As técnicas de mineração de textos utilizadas são distintas do MineraFórum. O objetivo do agente é

		responder automaticamente as perguntas dos alunos, e o intuito do MineraFórum é apresentar indicadores para auxiliar o docente na análise de fóruns.
Li e Huang (2008)	Ferramenta que utiliza técnicas de mineração de textos para analisar a relevância das mensagens em relação ao tema em debate.	O trabalho representa os textos com o modelo de espaço vetorial e realiza a comparação dos mesmos com a medida de similaridade cosseno. A ferramenta desenvolvida não está integrada a um AVA.
Lin et al. (2009)	Sistema de classificação de gêneros para investigar o conteúdo das mensagens de fóruns. A ferramenta auxilia o professor.	As técnicas utilizadas para analisar o conteúdo das mensagens e os resultados fornecidos são distintos do MineraFórum. Não há integração com AVA.

O MineraFórum é um recurso que pode ser utilizado pelo docente dentro do Ambiente Virtual de Aprendizagem (AVA). Ele está integrado ao ROODA³ [Behar 2007], ETC⁴ [Macedo et al. 2010] e MOODLE⁵. O docente poderá escolher os fóruns de discussão para que as mensagens sejam mineradas. A figura 3 apresenta a interface do ETC, na qual o professor pode selecionar o botão do MineraFórum.

Figura 3. Interface do ETC com o botão do “MineraFórum”.

³ O ROODA é uma das plataformas de Educação a Distância utilizada pela UFRGS. O ROODA está disponível em <https://www.ead.ufrgs.br/rooda/>

⁴ O ETC é um editor de texto coletivo desenvolvido pelo NUTED, utilizado na UFRGS e também em cursos de extensão, disponível em <http://www.nuted.ufrgs.br/etc2/>

⁵ O MOODLE, com o MineraFórum integrado, está disponível em <http://www.nie.iff.edu.br/moodle/>

5. Experiências realizadas

Para validar os resultados do MineraFórum, foram realizados dez experimentos. Nas experiências, cinco professores analisaram 403 mensagens de dez fóruns de discussão. Os temas dos fóruns foram distintos, bem como, o nível de ensino e a modalidade dos cursos dos alunos. Cabe ressaltar que os fóruns foram realizados nos Ambientes Virtuais de Aprendizagem ROODA, ETC e MOODLE.

O objetivo das experiências foi comparar a média das relevâncias das mensagens, calculada pelo MineraFórum, com a média das avaliações das postagens feitas pelos professores. É importante lembrar que o MineraFórum calcula um valor de relevância entre zero e cinco, para cada contribuição textual. Desta forma, nos experimentos realizados, foi solicitado aos docentes para que na avaliação de cada mensagem fosse atribuído um valor entre zero e cinco. Na avaliação das mensagens participaram dois professores que possuem Doutorado e três com Mestrado. Dois docentes possuem grande experiência em EAD, e três possuem pouca.

A tabela 2 descreve as características dos fóruns analisados: o AVA no qual o fórum foi realizado, o tema, o curso, o nível de ensino e a modalidade de cada curso. A tabela 3 apresenta a tabulação dos valores obtidos com as análises feitas pelo MineraFórum e pelos docentes. As informações são: o professor que avaliou as mensagens, a quantidade de alunos que participou de cada fórum, a quantidade de postagens, a média das análises do software, a média das avaliações do docente, o grau de similaridade entre as médias, a média da diferença entre as análises, o desvio padrão de cada análise.

A tabela 3 apresenta o grau de similaridade entre a média da análise do MineraFórum e a do professor. Este dado revela que a média das análises das mensagens calculada pelo software é semelhante à média das avaliações feitas pelos docentes. O fórum nove apresentou o menor valor para este dado, 70,30%. No fórum sete, obteve-se o maior valor, 96,43%.

Tabela 2. Características dos fóruns analisados.

Fórum	AVA	Tema do fórum	Curso	Nível de ensino	Modalidade
1	Rooda	Mapas conceituais	Informática na Educação	Doutorado	Presencial
2	Rooda	Aprender com os outros	Pedagogia	Graduação	EAD
3	Rooda	Aprender como transformação	Pedagogia (grupo 1)	Graduação	EAD
4	Rooda	Aprender como transformação	Pedagogia (grupo 2)	Graduação	EAD
5	ETC	Trabalho em equipe	Extensão	Extensão	Presencial
6	ETC	Desenvolvimento de competências	Extensão	Extensão	Presencial
7	Moodle	Derivadas - Origens Históricas	Sistemas de Informação	Graduação	Presencial
8	Moodle	Certificação Digital	Sistemas de Informação	Graduação	Presencial
9	Moodle	Educação a distância	Técnico em Segurança do Trabalho	Ensino Médio	EAD
10	Moodle	EAD e educação aberta	Técnico em Segurança do Trabalho	Ensino Médio	EAD

Tabela 3. Resultados das análises dos fóruns.

Fórum	1	2	3	4	5	6	7	8	9	10
Professor	A	A	A	A	B	B	C	D	E	E
Quantidade de alunos	10	28	28	31	18	11	20	12	23	24
Quantidade de mensagens	16	67	48	73	76	42	22	12	23	24
Média da análise do MineraFórum	3,19	3,00	2,92	2,88	3,00	3,21	3,68	3,42	3,09	2,88
Média da análise do professor	3,63	2,25	2,79	2,32	3,61	3,90	3,82	2,67	4,39	4,08
Grau de similaridade entre a média da análise do MineraFórum e a do professor	87,93 %	75,12 %	95,71 %	80,48 %	83,21 %	82,32 %	96,43 %	78,05 %	70,30 %	70,41 %
Média da diferença entre as análises	1,06	1,37	1,04	1,33	1,29	1,50	1,23	1,25	1,83	1,38
Desvio padrão da análise do MineraFórum	1,11	1,36	1,11	1,19	1,63	1,39	0,72	0,67	0,95	0,34
Desvio padrão da análise do professor	1,41	1,11	0,80	1,22	1,59	1,32	1,10	1,44	0,78	1,06
Desvio padrão da diferença entre as análises	0,85	1,27	1,05	1,34	0,99	1,02	0,75	1,54	0,78	0,97

Destaca-se que fóruns, com a mesma quantidade de alunos, podem ter valores distintos no grau de similaridade entre as médias. Nos fóruns 2 e 3 houve a participação de 28 alunos. No fórum 2, o grau de similaridade foi 75,12% e no fórum 3 foi 95,71%.

É importante citar que o software utiliza três critérios para analisar as mensagens. A relevância temática da contribuição textual, a relevância de citações da mensagem e a similaridade da mensagem. Ao avaliar as postagens, cada professor utilizou seus próprios critérios pessoais.

Cabe ressaltar que as mensagens trocadas entre os participantes do fórum, que expressam apoio e motivação são muito importantes para promover a sensação de comunidade entre os sujeitos [Palloff e Pratt 2002]. Esta sensação fortalece o processo de aprendizagem. O MineraFórum somente considerará estas mensagens como relevantes se elas envolverem conceitos relacionados ao tema do debate.

6. Considerações finais

O MineraFórum é um recurso para auxiliar o docente na análise qualitativa das mensagens de fóruns. O software permite executar, de forma automatizada, o referido processo. Cabe ressaltar que o papel do docente nos debates é fundamental.

A partir dos resultados apresentados na seção 5, pode-se dizer que o objetivo dos experimentos foi alcançado. Verificou-se que a média das análises das mensagens,

calculada pelo MineraFórum, é semelhante à média das avaliações dos professores.

O MineraFórum é capaz de apresentar ao professor uma visão sobre as contribuições escritas pelos discentes, organizando e agrupando as postagens de cada aluno. O software disponibiliza informações que podem auxiliar o docente.

Entende-se que, a partir das informações apresentadas pelo MineraFórum, o professor poderá direcionar seu apoio aos alunos que colocaram poucas contribuições relevantes no fórum. O docente também poderá motivar a interação entre os discentes que mais registraram textos relevantes com os que redigiram poucos.

Referências

- Azevedo, B. F. T., Reategui, E. B., Behar, P. A. (2010) “Qualitative Analysis of Discussion Forums”, In: IADIS International Conference on e-Learning, Freiburg, Alemanha, p. 251-258.
- Behar, P. A., Bernardi, M., Souza, A. P. F. de Castro e Kellen, K. (2007) “ROODA: desenvolvimento, implementação e validação de um AVA para UFRGS”, In: XII Taller Internacional de Software Educativo, Santiago, Chile, LOM Ediciones S.A., v. 1, p. 321-338.
- Dringus, L. P., Ellis, T. (2005) “Using data mining as a strategy for assessing asynchronous discussion forums”, *Computers & Education*, n. 45(1), p. 141-160.
- Feldman, R. e Sanger, J. (2007), *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press.
- Kim, J., et al. (2007) “An Intelligent Discussion-Bot for Guiding Student Interactions in Threaded Discussions”, In: *Proceedings of the AAAI Spring Symposium on Interaction Challenges for Intelligent Assistants*, Stanford University.
- Li, Y. e Huang, R. (2008) “Analyzing Peer Interactions in Computer-Supported Collaborative Learning: Model, Method and Tool”, In: *LNCS*, n. 5169, p. 125-136.
- Lin, F., Hsieh, L., Chuang, F. (2009) “Discovering genres of online discussion threads via text mining”, *Computers & Education*, n. 52(2), p. 481-495.
- Macedo, A. L. et al. (2010) “ETC: o que mudou e por quê?”, *Revista Novas Tecnologias na Educação*, v. 8, p. 2-12.
- Palloff, R. M. e Pratt, K. (2002), *Construindo comunidades de aprendizagem no ciberespaço*, Porto Alegre, Artmed.
- Palloff, R. M. e Pratt, K. (2004), *O aluno virtual: um guia para trabalhar com estudantes on-line*, Porto Alegre, Artmed.
- Schenker, A. (2003), *Graph-Theoretic Techniques for Web Content Mining*, Tese de Doutorado, University of South Florida, Florida.
- Tan, A. (1999) “Text Mining: The State of the Art and the Challenges”, In: *Proceedings of the PAKDD’99 Workshop on Knowledge Discovery from Advanced Databases*, Beijing, p.71-76.