

APLICAÇÃO DA MINERAÇÃO DE TEXTOS EM FÓRUNS DE DISCUSSÃO

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

Resumo: Este artigo apresenta um estudo realizado com a técnica de mineração de textos utilizando grafos. Esta pode ser utilizada para efetuar uma análise qualitativa das contribuições textuais registradas por alunos em fóruns de discussão. A técnica foi empregada para identificar quais discentes redigiram mensagens que contemplam conceitos relativos ao tema da discussão, e quais não o fizeram. Desta forma, é possível ter subsídios para investigar quais alunos necessitam de maior auxílio, e motivá-los para discutir os conceitos importantes que fazem parte do tema em debate.

Palavras-chave: mineração de textos, fórum de discussão, análise textual.

Abstract: This article presents a study of the technique of text mining using graphs. This technique can be used to make a qualitative analysis of textual contributions recorded by students in discussion forums. The technique was used to identify which students drafted messages that include concepts related to the topic of discussion, and what did not. Thus, it is possible to have subsidies to investigate which students need more help, and motivate them to discuss the important concepts that are part of the topic under discussion.

Keywords: text mining, discussion forum, textual analysis.

1. Introdução

O fórum de discussão é um recurso que permite aos sujeitos participantes discutir um determinado assunto e trocar ideias. 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 composto por quadros de diálogo, onde as mensagens redigidas podem ser classificadas tematicamente. Nestes espaços, os alunos podem realizar contribuições, refutar outras, esclarecer dúvidas, entre outros. A comunicação é efetuada de forma assíncrona e as mensagens digitadas permanecem à disposição dos participantes. De acordo com Palloff e Pratt (2004), as interações dos alunos nas discussões proporcionam um momento de reflexão sobre os conteúdos educacionais abordados.

O envolvimento em fóruns de discussão é uma parte importante das atividades dos alunos que estudam através da modalidade de educação a distância, permitindo que o professor possa diagnosticar informações sobre os discentes. No entanto, se o docente possuir uma grande quantidade de alunos, o tempo necessário para que ele consiga analisar as discussões será grande. Uma técnica que permita fazer uma análise dos conteúdos das mensagens pode auxiliar o professor. Isto permitirá a identificação dos alunos que estão (e aqueles que não estão) discutindo os conceitos do assunto 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.

Para auxiliar na análise da relevância dos textos produzidos pelos alunos em um fórum de discussão, este artigo apresenta um estudo com a técnica de mineração de textos utilizando grafos.

A próxima seção descreve algumas técnicas de mineração de textos. A seção 3 relata alguns trabalhos correlacionados. A seção 4 aborda a técnica de mineração de textos utilizando grafos. A seção 5 apresenta as experiências piloto realizadas. A seção 6 relata as considerações finais do estudo.

2. Mineração de textos

A mineração de textos trata do processo de extrair padrões interessantes e não-triviais de conhecimento a partir de textos (Tan, 1999). Ela 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 (Feldman e Sanger, 2007).

Os sistemas de mineração de textos baseiam-se em rotinas de pré-processamento, algoritmos para descoberta de padrões, e elementos para apresentação dos resultados. As etapas que compõe a arquitetura de um sistema para mineração de textos são: operações de pré-processamento, geração de documentos processados, mineração, apresentação dos resultados. O usuário do sistema interage com a etapa de pré-processamento, com o núcleo de mineração e com a apresentação dos resultados.

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 modelo intermediário (Feldman e Sanger, 2007; Tan, 1999).

As operações que fazem parte do núcleo de mineração, também chamadas de processos de destilação do conhecimento, representam o coração de um sistema de mineração de texto, e incluem: descoberta de padrões, análise de tendências, e algoritmos incrementais para descoberta de conhecimento. Os padrões mais utilizados para a descoberta de conhecimento são as distribuições e as proporções, conjunto de conceitos frequentes, e associações. Estas operações também podem estar relacionadas a comparações, e a identificação de níveis de interesse, com alguns destes padrões. Sistemas avançados, ou orientados a domínio, para mineração de texto, também podem melhorar a qualidade das operações a partir de consultas realizadas em bases de conhecimento (Feldman e Sanger, 2007).

Os componentes que fazem parte da apresentação dos resultados representam a interface do sistema, com funcionalidades para navegação, e acesso a linguagem utilizada para consultas (Feldman e Sanger, 2007; Puretskiy et al., 2010; Tan, 1999).

As técnicas empregadas na mineração de textos incluem:

- **Extração de informação:** identifica termos e frases principais em um texto, bem como seus relacionamentos. A extração é realizada através da busca de sequências pré-definidas, sendo este um processo chamado de casamento de

padrões. A extração infere os relacionamentos entre termos, pessoas, lugares, datas, para fornecer informação significativa ao usuário (Fan et al., 2006; Mooney e Nahm, 2003).

- **Rastreamento de tópicos:** mantém armazenados os perfis dos usuários, e baseado nos documentos que o usuário visualiza, prevê outros documentos que possam interessar ao usuário. O rastreamento pode ser utilizado para descobrir as referências de uma área de pesquisa (Gupta e Lehal, 2009; Mahalakshmi e Sendhilkumar, 2009).
- **Links conceituais:** conectam documentos relacionados a partir da identificação de conceitos comumente compartilhados. Esta técnica ajuda os usuários a encontrar informações que talvez não fossem encontradas com métodos tradicionais de busca. Os links conceituais favorecem a navegação na informação ao invés de auxiliar a sua busca (Fan et al., 2006; Gupta e Lehal, 2009).
- **Visualização de informação:** apresenta grandes quantidades de documentos em uma hierarquia visual, ou em um mapa, fornecendo opções de navegação, além da realização de pesquisas (Feldman e Sanger, 2007; Puretskiy et al., 2010; Tan, 1999).
- **Análise de perguntas e respostas:** é proveniente da área de processamento de linguagem natural, que trata de como encontrar a melhor resposta para uma respectiva pergunta (Gupta e Lehal, 2009; Ramakrishnan e Bhattacharyya, 2009).
- **Produção de sumários (sumarização):** é útil para tentar descobrir se um documento extenso atende ou não as necessidades do usuário, para que o mesmo avalie se é interessante ler o texto para obter mais informações. O objetivo da sumarização é reduzir o tamanho e o nível de detalhamento de um documento, retendo os pontos principais e o significado geral. A grande dificuldade para o software ainda é fazer a análise semântica e a interpretação de significado. Uma das estratégias mais utilizadas pela sumarização é a extração de sentenças importantes do texto, a partir do peso estatístico das sentenças. Algumas heurísticas adicionais, tais como informações sobre posicionamento também são usadas para sumarização. Os métodos de sumarização podem ser classificados em dois grandes grupos: abordagens superficiais, que são restritas ao nível sintático da representação; e as abordagens profundas, que envolvem um nível de representação semântico do texto original e utilizam processamento linguístico em algum nível (Castellanos, 2004; Gupta e Lehal, 2009).
- **Categorização (classificação):** envolve a identificação dos principais temas de um documento, colocando o mesmo em um conjunto pré-definido de tópicos (Gupta e Lehal, 2009). Neste processo, o texto é tratado como um conjunto de palavras, ao invés de realizar o processamento da informação real, como na extração de informação. A categorização conta as palavras que aparecem no documento, e a partir desta contagem, identifica os principais tópicos do texto. Esta técnica geralmente utiliza um método de classificação dos documentos, para ordenar aqueles que possuem mais conteúdo sobre um tema específico.
- **Agrupamento de textos (*clustering*):** é uma técnica utilizada para agrupar documentos semelhantes. Ela é diferente da categorização, pois ao invés do uso

de tópicos pré-definidos, os documentos são agrupados em tempo real. Outra vantagem do agrupamento é que os documentos podem aparecer em vários subtópicos, e assim garantir que um documento útil não seja omitido nos resultados da pesquisa. Um algoritmo de agrupamento básico cria um vetor de tópicos para cada texto, e calcula os pesos para identificar em qual grupo um documento deve fazer parte (Gupta e Lehal, 2009; Kobayashi e Aono, 2004).

O sistema de mineração de texto desenvolvido nesta pesquisa, detalhado na seção 4, emprega a técnica de categorização.

3. Trabalhos correlacionados

Esta seção apresenta algumas pesquisas realizadas com a aplicação de técnicas de mineração de textos para analisar fóruns de discussão.

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.

O trabalho de Ravi e Kim (2007) apresenta uma abordagem para identificar automaticamente perfis de interações de alunos em fóruns de discussões. Foram desenvolvidos classificadores de “ato de discurso” que identificam os papéis das mensagens individuais, como: pergunta, resposta, elaboração, correção. Os classificadores foram utilizados na busca de mensagens que contêm perguntas ou respostas. Foi utilizado um conjunto de regras para análise dos tópicos para descobrir aqueles que poderiam ter perguntas sem resposta e necessitar da atenção do professor.

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 mostra como os tópicos do debate foram modelados utilizando “atos de discurso”. Cada postagem foi classificada de acordo com categorias, como: pergunta, resposta, elaboração e correção. Ao classificar as contribuições da discussão, os autores foram capazes de identificar os papéis dos alunos e do professor nas discussões. 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.

Outra pesquisa para analisar fóruns de discussão com técnicas de mineração de textos é

apresentada por Lin et al. (2009). O trabalho propõe um sistema de classificação dos 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 Moodle¹ para realizar os experimentos. Para os pesquisadores, o professor pode participar de um debate para contribuir com o processo de aprendizagem dos alunos. 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.

Li e Huang (2008) apresentam 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. 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.

Um fórum de discussão com características tecnológicas avançadas é apresentado por Li et al.(2008). O projeto utiliza ontologia de domínio e técnicas de mineração de textos. No trabalho desenvolvido, as transcrições das discussões do fórum são automaticamente transformadas em uma modelagem estrutural em três etapas: o reconhecimento do tópico, a identificação do tipo de cada uma, e a associação semântica entre elas. A primeira etapa agrupa em um documento as mensagens de um mesmo conjunto de discussões. A segunda etapa identifica seis tipos de mensagens: pergunta, opinião, sugestão, recomendação, pedido, e citação. A terceira etapa organiza os textos com associação semântica. O fórum proposto oferece três funcionalidades aos discentes: busca de informações úteis às suas necessidades, navegação temática pelas mensagens, e recomendação de outros alunos que possuem interesse em realizar comunicação e colaboração. Um estudo experimental foi realizado para apresentar que a abordagem é eficaz para descobrir parceiros de aprendizagem com os mesmos interesses e a busca de mensagens com guia de navegação temático.

4. Mineração de textos utilizando grafos

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

¹ <http://moodle.org>

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).

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 partir do estudo realizado neste artigo, observou-se que a mineração utilizando grafos é uma opção viável para a análise das mensagens de um fórum 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.

Como exemplo de utilização dessa técnica, a figura 1 apresenta o grafo gerado a partir do texto “Os recursos tecnológicos podem auxiliar o professor.”.

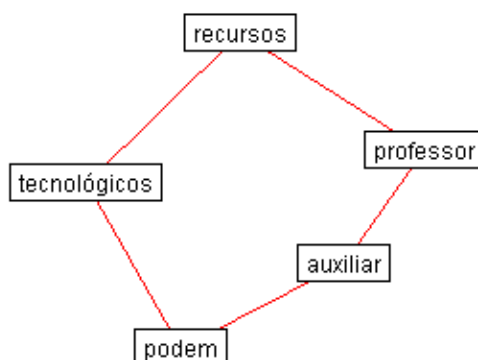


Figura 1 - Grafo gerado a partir do texto “Os recursos tecnológicos podem auxiliar o professor.”.

A partir dos resultados apresentados pela técnica, o professor poderá fazer um acompanhamento qualitativo das contribuições de cada aluno e de um grupo de alunos em um fórum de discussão. O professor poderá:

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

5. Experiências piloto

Para analisar as contribuições textuais produzidas pelos alunos nos fóruns, foi utilizado um software de mineração de textos que se encontra em desenvolvimento, chamado

MineraFórum² (Azevedo, 2010). 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) (Azevedo et al., 2009) de uma contribuição textual, cujo objetivo é analisar o quanto um texto é relevante a um determinado tema de discussão. O CRT indica o grau de relevância de uma mensagem de acordo com o tema do fórum.

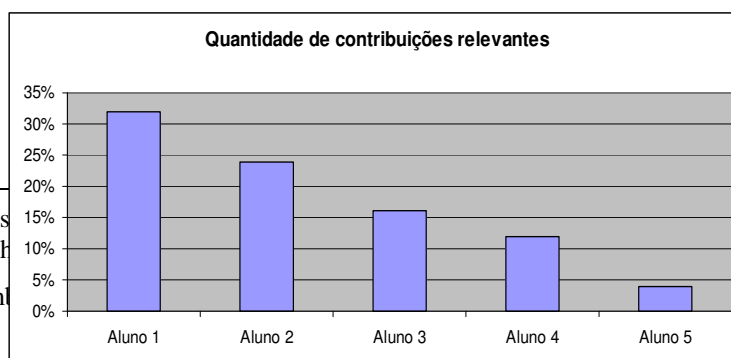
Foram efetuadas experiências piloto, com o objetivo de analisar/validar se a referida técnica de mineração é adequada para analisar as mensagens de fóruns de discussão.

A metodologia empregada nas experiências piloto envolveu:

- a) A escolha do 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 escolha do valor mínimo do coeficiente de relevância temática considerado para análise.
- e) A coleta das contribuições textuais produzidas pelos alunos no fórum.
- f) Para cada contribuição textual, foi executado o processo de mineração e gerado o grafo com os conceitos envolvidos.
- g) A partir de cada grafo gerado, foi calculado o coeficiente de relevância temática (CRT) 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. Foi definido o valor mínimo 01 para o CRT a ser considerado na análise das mensagens. Isto significa que após o cálculo do coeficiente para cada contribuição textual, aquelas cujo CRT foi maior ou igual a 01 foram consideradas como relevantes ao tema do debate.

O primeiro experimento foi realizado em um fórum de discussão do Ambiente ROODA³ (Behar et al., 2007), 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. A figura 2 apresenta a quantidade de contribuições relevantes identificadas neste experimento.



² Este software es

³ Disponível em h

Figura 2 - Quantidade de contribuições relevantes do primeiro experimento

O segundo e terceiro experimentos foram realizados em um fórum de discussão do Ambiente ROODA (Behar et al., 2007), 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. As figuras 3 e 4 apresentam a quantidade de contribuições relevantes identificadas.

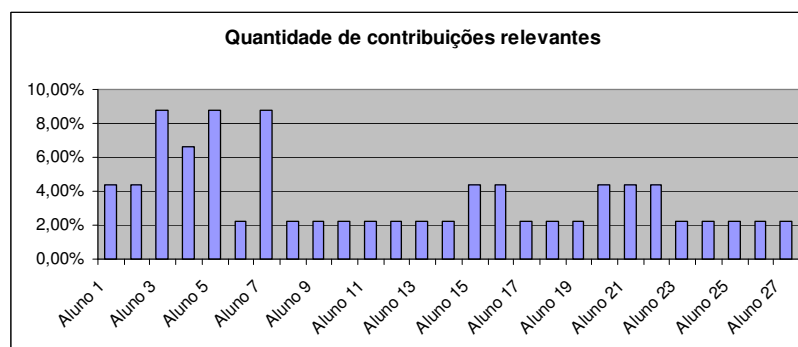


Figura 3 - Quantidade de contribuições relevantes do primeiro grupo

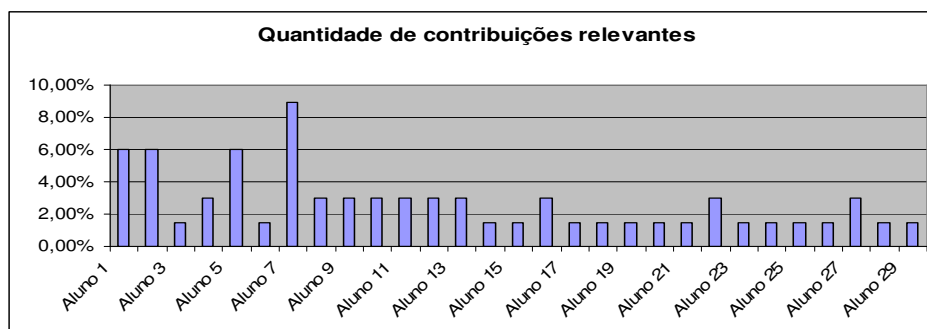


Figura 4 - Quantidade de contribuições relevantes do segundo grupo

A partir dos resultados apresentados nas figuras 2, 3, 4, percebe-se que a técnica de mineração de textos utilizando grafos tem um grande potencial para apoiar o professor na análise qualitativa das mensagens de um fórum de discussão. Isto é possível na medida em que a técnica verifica se conceitos importantes de serem tratados estão efetivamente sendo considerados nas contribuições dos alunos.

6. Considerações finais

A principal contribuição deste artigo é apresentar um estudo realizado para investigar fóruns de discussão, demonstrando como uma técnica de mineração de texto pode

contribuir com o trabalho do professor na análise qualitativa das postagens dos alunos. Foram realizadas experiências piloto com a mineração de textos utilizando grafos. Através desta técnica é possível realizar uma análise qualitativa das mensagens redigidas em fóruns.

Considerando-se o grafo gerado a partir de uma mensagem, pode-se analisar se ela se refere ao contexto do debate e avaliar a relevância temática da mesma. A partir desta análise, o professor obtém informações para auxiliá-lo a acompanhar as mensagens redigidas pelos alunos.

Com os resultados apresentados pela técnica de mineração, o docente pode analisar quais contribuições textuais necessitam de intervenção. É possível identificar os alunos que colocaram apenas algumas mensagens relevantes no fórum para oferecer-lhes maior auxílio. O professor também poderá, a partir destas informações, motivar os alunos que colocaram muitas contribuições relevantes para interagir com aqueles que escreveram poucas.

Referências bibliográficas

AZEVEDO, B. F. T.; BEHAR, P. A.; REATEGUI, E. B. Estudo de análise qualitativa em fórum de discussão. **Revista Novas Tecnologias na Educação (RENOTE)**. Porto Alegre, v. 7, n. 2, 2009.

AZEVEDO, B. F. T. **MineraFórum: um recurso de apoio para análise qualitativa em fóruns de discussão**. Porto Alegre: PPGIE/UFRGS, 2010. Proposta de Tese de Doutorado.

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. Memórias del XII Taller Internacional. Santiago: LOM Ediciones S.A., 2007, v. 1, p. 321-338.

CASTELLANOS, M. HotMiner: Discovering Hot Topics from Dirty Text. In: BERRY, M. W. (ed.). **Survey of text mining: clustering, classification, and retrieval**. New York: Springer-Verlag, 2004. p. 123-157.

DRINGUS, L. P.; ELLIS, T. Using data mining as a strategy for assessing asynchronous discussion forums. **Computers & Education**, n. 45(1), 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.

GUPTA, V.; LEHAL, G. S. A Survey of Text Mining Techniques and Applications. **Journal of Emerging Technologies in Web Intelligence**, v. 1, n. 1, 2009.

KIM, J., et al. An Intelligent Discussion-Bot for Guiding Student Interactions in Threaded Discussions. In: **AAAI SPRING SYMPOSIUM ON INTERACTION CHALLENGES FOR INTELLIGENT ASSISTANTS**, 2007. Palo Alto. Proceedings. Palo Alto: AAAI Press/Stanford University, 2007.

- KOBAYASHI, M.; AONO, M. Vector Space Models for Search and Cluster Mining. In: BERRY, M. W. (ed.). **Survey of text mining: clustering, classification, and retrieval**. New York: Springer-Verlag, 2004. p. 103-122.
- LI, Y.; HUANG, R.; Analyzing Peer Interactions in Computer-Supported Collaborative Learning: Model, Method and Tool. **Lecture Notes in Computer Science (LNCS)**. Berlin: Springer-Verlag Berlin Heidelberg, v. 5169, p. 125-136, 2008.
- LI, Y; DONG, M.; HUANG, R. Semantic Organization of Online Discussion Transcripts for Active Collaborative Learning. In: **IEEE INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES**, 2008. Santander. Proceedings. Santander: IEEE Computer Society Press, 2008, p. 756-760.
- LIN, F.; HSIEH, L.; CHUANG, F. Discovering genres of online discussion threads via text mining. **Computers & Education**, Oxford, v. 2, n. 52, p.481-495, 2009.
- MAHALAKSHMI, G. S.; SENDHILKUMAR, S. Automatic Reference Tracking. In: SONG, M.; WU, Y. (Org.). **Handbook of research on text and web mining technologies**. Hershey: Information Science Reference, 2009. p. 483-499.
- MOONEY, R. J.; NAHM, U. Y. Text Mining with Information Extraction. In: **INTERNATIONAL MIDP COLLOQUIUM**, 2003. Bloemfontein. Proceedings. Bloemfontein: 2003, p. 141-160.
- PALLOFF, R. M.; PRATT, K. **O aluno virtual: um guia para trabalhar com estudantes on-line**. Porto Alegre: Artmed, 2004.
- PURETSKIY, A. A.; SHUTT, G. L.; BERRY, M. W. Survey of text visualization techniques. In: BERRY, M. W.; KOGAN, J. **Text mining: applications and theory**. John Wiley & Sons Ltd, 2010, p. 107-127.
- RAMAKRISHNAN, G.; BHATTACHARYYA, P. Question Answering Using Word Associations. In: SONG, M.; WU, Y. (Org.). **Handbook of research on text and web mining technologies**. Hershey: Information Science Reference, 2009. p. 571-603.
- RAVI, S.; KIM, J. Profiling Student Interactions in Threaded Discussions with Speech Act Classifiers. In: **AI IN EDUCATION CONFERENCE (AIED)**, 2007. Los Angeles. Proceedings. Los Angeles: 2007.
- SÁNCHEZ, L. 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, A. **Graph-Theoretic Techniques for Web Content Mining**. Florida: University of South Florida, 2003. PhD Thesis.
- 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.