

Análise temática das mensagens de discussões online

Breno Fabrício Terra Azevedo
Instituto Federal de Educação, Ciência
e Tecnologia Fluminense
Rua Dr. Siqueira, 273, Pq Dom Bosco
Campos dos Goytacazes/RJ, Brasil,
CEP 28030-130
(22)2726-2861

bterra@iff.edu.br

Patricia Alejandra Behar
Universidade Federal do Rio Grande
do Sul
Av. Paulo Gama, 110, prédio 12105
3º andar sala 332, Porto Alegre/RS,
Brasil, CEP 90040-060
(51)3308-3901

pbehar@terra.com.br

Eliseo Berni Reategui
Universidade Federal do Rio Grande
do Sul
Av. Paulo Gama, 110, prédio 12105
3º andar sala 332, Porto Alegre/RS,
Brasil, CEP 90040-060
(51)3308-3739

eliseoreategui@gmail.com

RESUMO

Este artigo apresenta um estudo para investigar a relevância temática das contribuições textuais redigidas em fóruns de discussão. Foram realizadas experiências para analisar as mensagens postadas pelos alunos. O artigo apresenta os resultados provenientes do trabalho realizado.

Categorias e Descritores de Assunto

K.3.1 [Computadores e Educação]: Utilização do Computador na Educação – *Educação a Distância*.

Termos Gerais

Experimentação.

Palavras-chaves

Fóruns de discussão, mineração de textos, relevância temática.

1. INTRODUÇÃO

De acordo com Sánchez [21], o fórum de discussão para fins educacionais em um ambiente online é um espaço de comunicação composto por quadros de diálogo, onde as mensagens postadas podem ser classificadas tematicamente. Nestes espaços, os alunos podem interagir entre si, com os professores e tutores. Os discentes podem trocar ideias, realizar contribuições, emitir opiniões, esclarecer dúvidas, entre outros.

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 [4].

Segundo Carlini [5], no fórum devem ser realizadas discussões envolvendo troca de saberes, perspectivas, experiências, além das opiniões. Para a autora, é importante que o professor utilize o fórum para motivar trocas entre os alunos, a comunicação entre eles, a reflexão, e um posicionamento sobre o tema em debate.

A discussão assíncrona deve ser incentivada pelo professor, pois é a melhor forma de estabelecer e manter a interação com os alunos. As participações dos alunos nas discussões proporcionam um momento de reflexão sobre os conteúdos educacionais abordados. O fórum de discussão é um espaço propício para realizar este tipo de ação. O fato dos alunos responderem às questões de um debate e às mensagens indica uma possível reflexão [18].

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.

Este artigo apresenta uma metodologia que pode auxiliar o professor a analisar a relevância temática das mensagens redigidas por alunos em fóruns de discussão. Isto permitirá que o docente identifique os discentes que estão (e aqueles que não estão) discutindo os conceitos do tema em debate. Desta forma, o professor poderá dedicar um tempo maior para descobrir o motivo pelo qual alguns discentes não discutiram os conceitos referentes ao tema. Caso o docente identifique alunos com dificuldade para compreender os conceitos importantes, ele poderá oferecer ajuda.

Na próxima seção será apresentada uma breve introdução sobre mineração de textos. A seção 3 descreve a metodologia empregada no estudo. A quarta seção apresenta as experiências realizadas. A seção 5 relata as considerações finais da pesquisa e a última cita as referências bibliográficas utilizadas neste artigo.

2. MINERAÇÃO DE TEXTOS

Tan [23] explica que a mineração de textos trata do processo de extrair padrões interessantes e não-triviais de conhecimento a partir de textos.

Feldman e Sanger [10] definem a mineração de textos 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 elementos úteis a partir de coleções de documentos. As informações 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. As etapas que compõem 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 [10], [23].

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 [10].

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 [10], [19], [23].

Algumas técnicas empregadas na mineração de textos 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 [9], [12].

A extração de informação identifica as frases principais e os relacionamentos dentro de um texto. A ação é realizada através da busca de sequências pré-definidas no texto, sendo este um processo chamado de casamento de padrões. O procedimento infere os relacionamentos entre todas as pessoas identificadas, lugares, datas, para fornecer informação significativa ao usuário [9], [17].

O 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 processo pode ser utilizado para descobrir as referências de uma área de pesquisa [12], [16].

Os 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 [9], [12].

A 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 [10], [19], [23].

A 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 [12], [20].

A 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. Abordagens profundas, que envolvem um nível de representação semântico do texto original e utilizam processamento linguístico em algum nível [6], [12].

A categorização (classificação) envolve a identificação dos principais temas de um documento, colocando o mesmo em um conjunto pré-definido de tópicos [12]. 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 documento. Esta técnica geralmente utiliza um método de classificação dos documentos, para ordenar aqueles que possuem mais conteúdo sobre um tema específico.

O 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 [12], [13].

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. Neste processo, 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, esse procedimento descarta informações importantes, como por exemplo, a ordem nos quais os termos aparecem, onde os mesmos aparecem, e a proximidade entre eles [22].

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 interessante, e podem ser utilizados para a representação de textos [22].

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 destaca em seus nós as palavras que mais ocorreram. As associações entre os nós do grafo indicam a proximidade entre as palavras.

Considerando-se o grafo gerado a partir de uma mensagem postada no fórum, pode-se analisar se ela refere-se ao contexto, e avaliar a relevância temática da mesma. Ao investigar 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.”. A figura 2 exhibe o grafo gerado a partir do texto “Existem diversas técnicas de mineração de textos”.

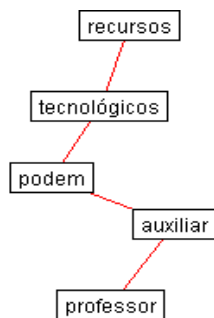


Figura 1: Grafo gerado a partir do texto “Os recursos tecnológicos podem auxiliar o professor.”.



Figura 2: Grafo gerado a partir do texto “Existem diversas técnicas de mineração de textos.”.

3. METODOLOGIA

Para efetuar a investigação das mensagens postadas em fóruns de discussão, este estudo definiu a seguinte metodologia:

- Escolha de uma técnica para analisar o conteúdo das mensagens redigidas.
- Seleção de um software para realizar a análise automática das mensagens.
- Definição dos fóruns de discussão nos quais as mensagens serão analisadas.
- Aplicação do software escolhido nas contribuições textuais redigidas nos fóruns estabelecidos.
- Análise do resultado apresentado pelo software.

No primeiro item descrito na metodologia, diversas técnicas podem ser utilizadas para analisar o conteúdo das mensagens redigidas. Algumas podem ser observadas em [3], [7], [8], [11], [14], [15]. Este estudo optou pela técnica de mineração de textos utilizando grafos, devido a resultados promissores apresentados no trabalho de Azevedo et al. [1].

A análise das mensagens postadas em fóruns pode ser efetuada manualmente. No entanto para auxiliar e acelerar o processo pode-se utilizar um software. Neste estudo, foi escolhido o MineraFórum [2], pois o mesmo apresenta características úteis ao estudo que foi realizado.

3.1 Análise temática de mensagens

Para efetuar a análise temática das mensagens postadas em um fórum foi utilizado o software MineraFórum. A figura 3 apresenta a interface principal do mesmo.

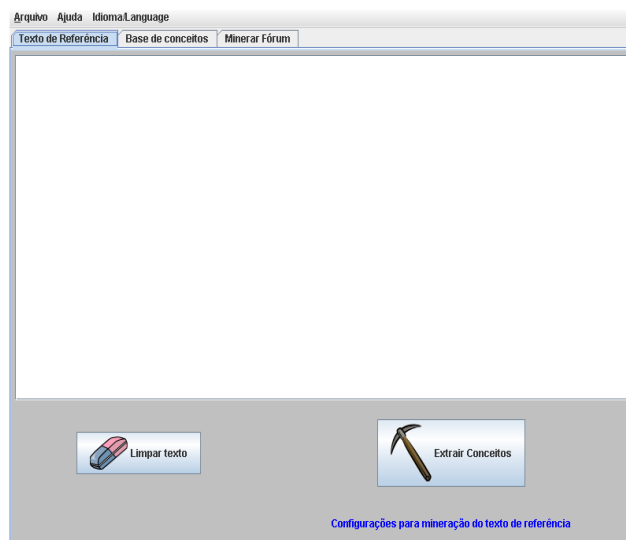


Figura 3: Interface principal do MineraFórum.

Alguns recursos oferecidos pelo MineraFórum são relacionados abaixo:

- Permite que o usuário informe um texto de referência sobre o tema em debate. A figura 3 apresenta a interface na qual o usuário pode indicar o texto.
- Possibilita que o usuário digite conceitos que são relevantes ao tópico da discussão e realize associações entre eles. A figura 4 apresenta a interface referente a este recurso.
- Utiliza um dicionário de sinônimos no processo de mineração. Este dicionário pode ser informado pelo usuário, ou pode ser utilizado o que foi previamente definido dentro do software.
- Permite que o usuário informe palavras que possuem equivalência semântica.
- Emprega a técnica de mineração de textos utilizando grafos. O processo é realizado com o texto de referência indicado pelo usuário e com cada mensagem postada no fórum.
- Calcula a relevância temática de cada mensagem postada. O valor mínimo da relevância a ser considerada na análise pode ser informado pelo usuário.
- Exibe um relatório com informações sobre a análise das mensagens redigidas no fórum. Estas informações são: o total de mensagens postadas por cada autor, a quantidade de contribuições relevantes e não relevantes ao tema feitas por cada autor, os conceitos importantes que foram citados nas contribuições relevantes ao tópico da discussão.

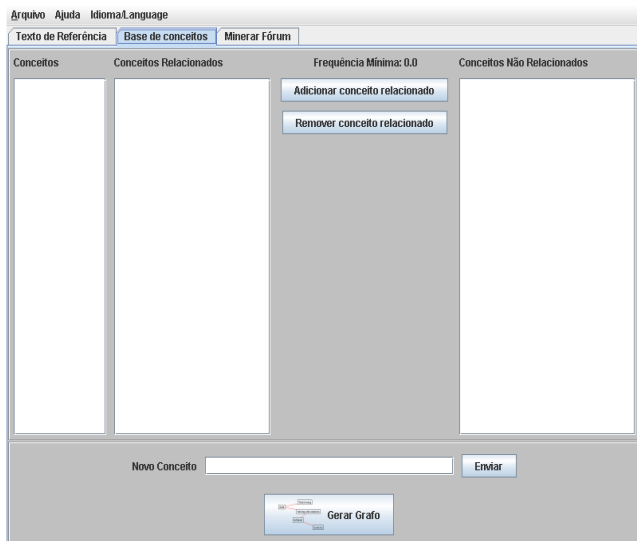


Figura 4: Interface para digitação de conceitos e suas associações.

O MineraFórum é um recurso que está em fase final de desenvolvimento, para ser incorporado ao Ambiente Virtual de Aprendizagem ROODA¹. Para utilizar o MineraFórum, o usuário deve seguir os seguintes passos:

1. Escolher o fórum de discussão a ser analisado.
2. Informar um texto de referência que contenha conceitos importantes referentes ao assunto em debate no fórum. Ao invés de indicar um texto de referência, o usuário pode digitar conceitos relevantes ao tema e seus relacionamentos.
3. O usuário possui a opção de informar palavras que possuem equivalência semântica.
4. Definir o valor mínimo da relevância temática a ser considerada na análise das mensagens.
5. Solicitar a mineração das mensagens.

O MineraFórum identifica os sinônimos das palavras do texto baseando-se num *thesaurus*². Os sinônimos são importantes quando o MineraFórum compara as palavras digitadas nas mensagens com os conceitos previamente informados no passo 2, citado anteriormente. Caso exista na mensagem uma palavra que seja sinônima de um conceito cadastrado, ela será considerada como relevante ao tema. Se o usuário digitou palavras que possuem equivalência semântica, conforme citado no passo 3, elas serão consideradas como sinônimas.

No processo de análise das mensagens postadas, o MineraFórum organiza e agrupa as mensagens de cada autor, calcula a quantidade total de mensagens de cada um, a quantidade de mensagens relevantes e não relevantes ao tema em debate. Uma mensagem é considerada como relevante, se o resultado do cálculo de sua relevância temática for maior que o mínimo estipulado no passo 4, citado anteriormente.

Para calcular a relevância temática das mensagens, o MineraFórum realiza as seguintes ações:

- a) Construção de um grafo a partir do texto de referência sobre o tema do debate, indicado pelo usuário. Neste processo, as *stopwords*³ 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. A fórmula efetua o somatório dos três fatores citados. O valor deste resultado corresponde à relevância temática da mensagem.

Com o intuito de verificar a metodologia proposta, cinco fóruns de discussão foram selecionados para as experiências. As contribuições textuais redigidas nestes fóruns foram analisadas pelo MineraFórum.

A próxima seção apresenta a descrição dos fóruns escolhidos. O resultado da análise da relevância temática das mensagens postadas em cada fórum é exibido na forma de gráficos. Cada gráfico apresenta a quantidade de contribuições relevantes ao tema em debate, redigidas por cada aluno.

4. EXPERIÊNCIAS REALIZADAS

Foram realizados cinco experimentos, com fóruns de discussão distintos, para analisar a relevância temática das mensagens redigidas pelos alunos.

Em todos os experimentos, foi definido o valor mínimo 01 para a relevância temática (RT) a ser considerada na análise dos textos. Isto significa que após o cálculo da relevância temática de cada mensagem, aquelas cuja RT foi maior ou igual a 01 foram consideradas como relevantes ao tema do debate.

Os três primeiros experimentos foram realizados em fóruns 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 primeiro fórum foi “Mapas conceituais”, que apresentou 16 contribuições textuais. No segundo fórum, o tema foi “TICs e Educação”, com 13 contribuições textuais. O tema do terceiro fórum foi “Comunidades virtuais”, que apresentou 26 contribuições textuais.

O quarto e quinto experimentos foram realizados em fóruns 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 postaram 42 mensagens. O segundo fórum contou com a participação de 28 alunos, que realizaram 67 contribuições textuais.

Nos três primeiros experimentos, o professor informou ao MineraFórum um texto de referência sobre o assunto “Mapas conceituais”. Na quarta e quinta experiência, o texto indicado pelo docente referiu-se ao tema “Aprender com os outros”.

Os resultados das experiências são apresentados nas figuras 5 a 9. Os gráficos indicam o valor da relevância temática das mensagens postadas por cada aluno.

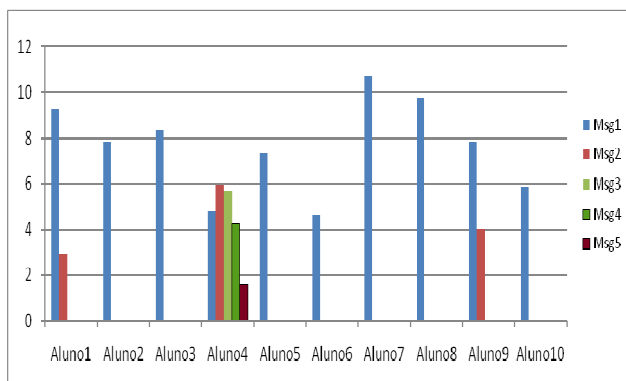


Figura 5: Relevância temática das mensagens postadas pelos alunos no primeiro fórum analisado.

Pode-se observar na figura 5 que o Aluno7 redigiu apenas uma mensagem, porém com um alto valor de relevância temática. Verifica-se que o Aluno4 redigiu cinco mensagens. No entanto, a última mensagem possui um baixo valor de relevância.

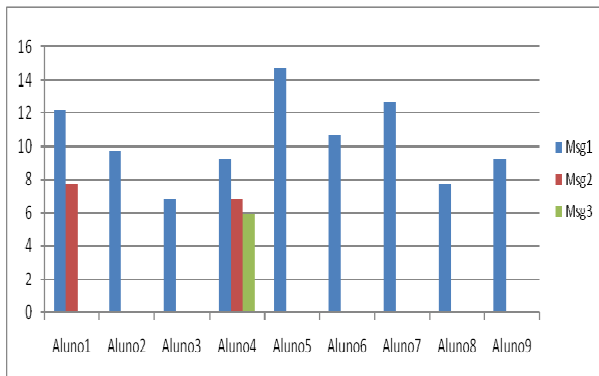


Figura 6: Relevância temática das mensagens postadas pelos alunos no segundo fórum analisado.

Na figura 6, observa-se que o Aluno1, Aluno5 e Aluno7 foram os que redigiram mensagens com maior relevância temática. Neste fórum, verifica-se que todas as mensagens redigidas pelos alunos foram relevantes ao tema do debate.

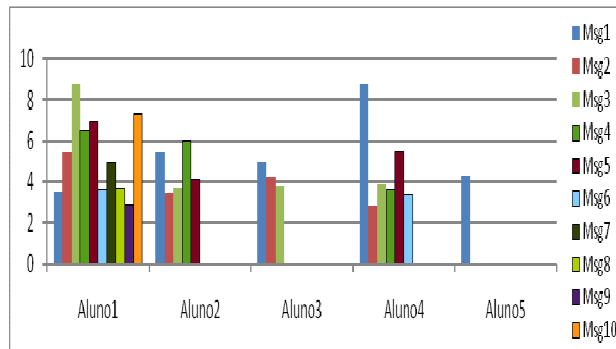


Figura 7: Relevância temática das mensagens postadas pelos alunos no terceiro fórum analisado.

Na figura 7 observa-se que o Aluno1 participou bastante do fórum, redigindo várias mensagens relevantes ao tema do debate. No entanto, o Aluno5 escreveu apenas uma mensagem.

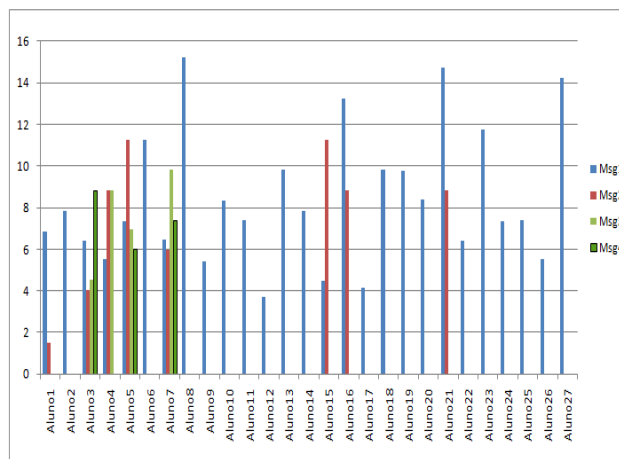


Figura 8: Relevância temática das mensagens postadas pelos alunos no quarto fórum analisado.

Na figura 8, observa-se que o Aluno3, Aluno4, Aluno5 e Aluno7 foram os que mais participaram do fórum, redigindo mensagens relevantes. Verifica-se que Aluno9, Aluno12, Aluno17, Aluno26 e Aluno27 postaram apenas uma mensagem com valor de relevância entre 3.0 e 5.0.

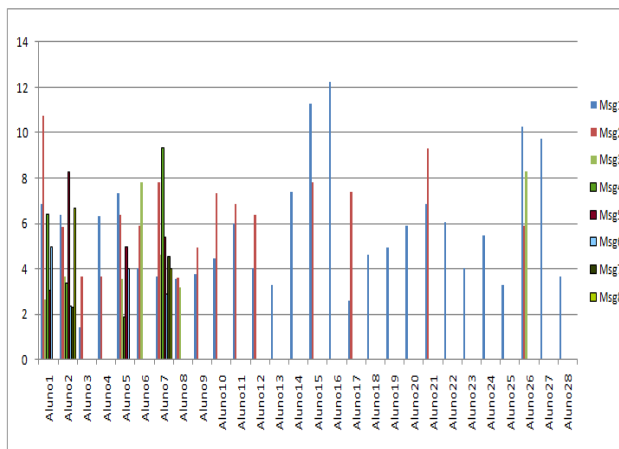


Figura 9: Relevância temática das mensagens postadas pelos alunos no quinto fórum analisado.

Na figura 9, observa-se que o Aluno1, Aluno2, Aluno5 e Aluno7 foram os que mais participaram do fórum, redigindo mensagens relevantes. Verifica-se que o Aluno13, Aluno23, Aluno25, Aluno28 postaram apenas uma mensagem com valor de relevância entre 3.0 e 4.0.

A partir dos resultados apresentados nesta seção, percebe-se que a metodologia proposta neste artigo proporcionou a análise da relevância temática das mensagens dos fóruns de discussão. Foi possível observar quais alunos redigiram mais contribuições textuais relativas ao tópico da discussão, e quais escreveram poucas mensagens relevantes. Além disso, verificaram-se quais foram os discentes que postaram as mensagens com maior relevância ao tema em debate.

5. CONSIDERAÇÕES FINAIS

A principal contribuição deste artigo é apresentar um estudo realizado para analisar as mensagens redigidas em fóruns de discussão. A pesquisa demonstra que a metodologia proposta na seção três pode ser aplicada na análise qualitativa das mensagens postadas. Foram realizadas experiências para observar os resultados obtidos com a aplicação da metodologia.

O software utilizado para analisar as contribuições textuais redigidas nos fóruns é capaz de calcular a relevância temática das mensagens. A partir do valor desta relevância, pode-se identificar se os textos referem-se, ou não, ao tópico da discussão.

Destaca-se que o uso de um software auxilia e acelera a análise temática das mensagens de um fórum. No entanto, cabe ressaltar que o tempo necessário para que o software execute o processo depende da quantidade de mensagens postadas. Nos experimentos realizados, cujo total de mensagens de cada fórum, foi respectivamente, 16, 13, 26, 42, 67, o MineraFórum efetuou a análise em poucos segundos.

Com a utilização de um software que permita analisar se as mensagens escritas pelos alunos referem-se ao contexto pretendido, o docente pode direcionar seu apoio aos discentes. A partir da análise da relevância temática dos textos, o professor obtém informações para auxiliá-lo a acompanhar as contribuições redigidas pelos alunos.

A análise da relevância temática das mensagens auxilia o professor a realizar um diagnóstico sobre os alunos. Nos experimentos realizados foi possível observar que alguns discentes escreveram várias mensagens que possuem relevância temática mediana. Por outro lado, alguns alunos redigiram somente uma ou duas mensagens com alta relevância.

A partir dos resultados apresentados neste estudo, o docente pode analisar quais contribuições textuais necessitam de intervenção. É possível identificar os discentes 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.

6. REFERÊNCIAS

- [1] 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. Anais do EL 2010, 251-258.
- [2] Azevedo, B. F. T. 2010. *MineraFórum: um recurso de apoio para análise qualitativa em fóruns de discussão*. Proposta de Tese de Doutorado. PPGIE/UFRGS.
- [3] Bassani, P. S., Behar, P. A. 2006. Análise das interações em ambientes virtuais de aprendizagem: uma possibilidade para avaliação da aprendizagem em EAD. *Revista Novas Tecnologias na Educação*, 4, 1.
- [4] Behar, P. A., et al. 2007. ROODA: desenvolvimento, implementação e validação de um AVA para UFRGS. In *XII Taller Internacional de Software Educativo (TISE)*, Santiago, Chile, 321-338.
- [5] Carlini, A. 2010. *20% a distância: e agora?: orientações práticas para o uso de tecnologia de educação a distância*. Pearson Education do Brasil, São Paulo.
- [6] Castellanos, M. 2004. HotMiner: Discovering Hot Topics from Dirty Text. In *Survey of text mining: clustering, classification, and retrieval*. Springer-Verlag, New York, 123-157.
- [7] Chen, Y., et al. 2008. A Wavelet-Based Model to Recognize High-Quality Topics on Web Forum. In *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. Sydney, Australia, 1, 343-351.
- [8] Dringus, L. P., Ellis, T. 2005. Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education*, 1, 45, 141-160.
- [9] Fan, W., et al. 2006. Tapping the power of text mining. *Communications of ACM*, 9, 49, 76-82.
- [10] Feldman, R., Sanger, J. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, Cambridge, MA.
- [11] Gerosa, M. A., et al. 2003. Coordenação de Fóruns Educacionais: Encadeamento e Categorização de Mensagens. In *Anais do XIV Simpósio Brasileiro de Informática na Educação*, Rio de Janeiro, Brasil, 45-54.
- [12] Gupta, V., Lehal, G. S. 2009. A Survey of Text Mining Techniques and Applications. *Journal of Emerging Technologies in Web Intelligence*, 1, 1.
- [13] Kobayashi, M., Aono, M. 2004. Vector Space Models for Search and Cluster Mining. In *Survey of text mining: clustering, classification, and retrieval*. Springer-Verlag, New York, 103-122.
- [14] Li, Y., Huang, R. 2008. Analyzing Peer Interactions in Computer-Supported Collaborative Learning: Model, Method and Tool. *Lecture Notes in Computer Science (LNCS)*. Springer-Verlag, Berlin, 5169, 125-136.
- [15] Lin, F., et al. 2009. Discovering genres of online discussion threads via text mining. *Computers & Education*, 2, 52, 481-495.
- [16] Mahalakshmi, G. S., Senthilkumar, S. 2009. Automatic Reference Tracking. In *Handbook of research on text and web mining technologies*. Hershey: Information Science Reference, 483-499.
- [17] Mooney, R. J., Nahm, U. Y. 2003. Text Mining with Information Extraction. In *Proceedings of International Midp Colloquium*. Bloemfontein, 141-160.

- [18] Palloff, R. M., Pratt, K. 2004. *O aluno virtual: um guia para trabalhar com estudantes on-line*. Artmed, Porto Alegre.
- [19] Puretskiy, A. A., Shutt, G. L., Berry, M. W. 2010. Survey of text visualization techniques. In *Text mining: applications and theory*. John Wiley & Sons Ltd, 107-127.
- [20] Ramakrishnan, G., Bhattacharyya, P. 2009. Question Answering Using Word Associations. In *Handbook of research on text and web mining technologies*. Hershey: Information Science Reference, 571-603.
- [21] Sánchez, L. P. 2005. El foro virtual como espacio educativo: propuestas didácticas para su uso. *Verista Quaderns Digitals Net*, 40, 1-18.
- [22] Schenker, A. 2003. *Graph-Theoretic Techniques for Web Content Mining*. Tese de Doutorado. University of South Florida, Florida.
- [23] 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, 71-76.

¹ O ROODA [4] é um ambiente virtual de aprendizagem e uma das plataformas de Educação a Distância utilizada pela Universidade Federal do Rio Grande do Sul (UFRGS), RS, Brasil e se encontra disponível em <https://www.ead.ufrgs.br/rooda/>

² Um *thesaurus* (tesauro) é um conjunto de palavras relacionadas semanticamente que abrange um domínio de conhecimento.

³ *Stopwords* são palavras que podem ser descartadas no processo de mineração, como: advérbios, artigos, preposições, entre outras.