# Mineração de Textos em Fóruns Educacionais: uma revisão da literatura

Máverick A. D. Ferreira<sup>1</sup>, Anderson P. Cavalcanti<sup>1</sup>, Rafael Ferreira<sup>1</sup>, Péricles Miranda<sup>1</sup>, Sebastião Neto<sup>2</sup>, Augusto Oliveira<sup>1</sup>

<sup>1</sup>Departamento de Estatística e Informática – Universidade Federal Rural de Pernambuco (UFRPE) – Recife – PE – Brasil

<sup>2</sup>Universidade Federal de Alagoas (UFAL) — Maceió — AL - Brasil {amaverick70, anderson.pinheiro27, rafaelflmello, periclesmiranda, sebast.rogers, auguustocesarf}@gmail.com

Abstract. Educational Forums are tools used in the e-learning because they are propitious environments for collaborative learning. Therefore, in the last years, several researches have been conducted with the objective of supporting pedagogical mediation in the Forums, through the use of Text Mining techniques. For the projection of new studies in this area, a systemic view of what has already been done is necessary. Thus, this research presents a systematic review of the literature of the last 10 years, in the context of Text Mining in Educational Forums. Manual searches were conducted in 56 databases, national and international, resulting in the selection of 30 relevant articles.

Resumo. Os Fóruns Educacionais são ferramentas utilizadas na Educação a Distância por serem ambientes propícios para a aprendizagem colaborativa. Diante disso, nos últimos anos diversas pesquisas têm sido conduzidas com o objetivo apoiar a mediação pedagógica, nos Fóruns, por meio do uso de técnicas de Mineração de Textos. Para a projeção de novos estudos nesta área, é necessária uma visão sistêmica do que já foi feito. Com isso, esta pesquisa apresenta uma revisão sistemática da literatura, dos últimos 10 anos, no contexto de Mineração de Textos em Fóruns Educacionais. Foram conduzidas buscas manuais em 56 bases de dados, nacionais e internacionais, obtendo-se como resultado a seleção de 30 artigos relevantes.

# 1. Introdução

Com o crescimento da educação a distância (EAD), os fóruns de discussões educacionais se apresentam como ferramentas importantes para o processo de ensino-aprendizagem. Eles possibilitam interações assíncronas entre professores e estudantes localizados em lugares geograficamente distintos. Dentre os benefícios de promover discussões em fóruns podem ser citados: as argumentações e reflexões sobre os temas discutidos que direcionam para uma aprendizagem colaborativa; e a demonstração de sentimentos/emoções, base para a criação de laços afetivos entre os envolvidos no debate [Xia et al. 2013].

DOI: 10.5753/cbie.sbie.2017.21 21

Ao considerar os ganhos decorrentes do uso dos fóruns na EAD é possível perceber que, além do impacto positivo na aprendizagem dos estudantes, o seu uso pode contribuir para a redução da evasão estudantil devido à sensação de pertencimento ao grupo, provocada nos estudantes, a partir de demonstrações afetivas nas discussões [Carvalho e Lima 2015]. Em contrapartida, com o aprofundamento das discussões e o consequente aumento do número de postagens proferidas pelos participantes, passa a ser exigido demasiado esforço dos professores/tutores para acompanhar o andamento do fórum [Scheuer e McLaren 2008].

Desse modo, considerando a importância da atuação do mediador nos fóruns, surge a demanda por formas/meios de apoio para o acompanhamento de fóruns educacionais de modo a utilizar o potencial desta ferramenta sem sobrecarregar o mediador [Bartholo et al. 2009].

Nesse sentido, como a maioria das contribuições dos estudantes em fóruns são textuais, pesquisas têm utilizado técnicas de Mineração de Textos para extrair informações relevantes nas postagens com a intenção de apoiar o professor/tutor na mediação de fóruns. O crescente interesse de pesquisadores pelo uso de técnicas de Mineração de Textos em fóruns educacionais justifica a realização de uma Revisão da Sistemática da Literatura (RSL) para detalhar os artigos já publicados e as oportunidades de pesquisas neste campo.

Diante deste cenário, este trabalho propõe uma RSL para identificar qual o atual cenário de pesquisas sobre o uso de técnicas de Mineração de Textos em Fóruns Educacionais. Para tal, foram realizadas buscas manuais em 56 bases de dados (conferências e periódicos), nacionais e internacionais, nas áreas de Computação em Educação, Inteligência Artificial e Processamento de Linguagem Natural (PLN).

## 2. Fórum Educacional

No modelo de educação presencial as interações do tipo estudante-estudante e estudante-professor ocorrem face a face. Por outro lado, no ensino a distância as interações costumam acontecer por meio de ferramentas integradas em AVA, tais como: *chats* e fóruns [Xia et al. 2013].

A ferramenta fórum é um meio de interação estruturado em árvore, onde a partir da inserção de uma questão/tópico de discussão, usuários podem responder/comentar a questão colocada ou responder/comentar postagens já inseridas, por outros participantes. Outra característica importante do fórum é a flexibilidade com relação à participação dos envolvidos. Em outras palavras, não é necessária a inserção de postagens ao mesmo tempo, o perfeito andamento da discussão pode ocorrer com a inserção de postagens com diferenças de horas ou até de dias.

Diante disso, os fóruns são amplamente utilizados no âmbito educacional para a promoção de discussões sobre conteúdos curriculares ou questões geralmente definidas pelo professor numa perspectiva estruturada e assíncrona. Nesse aspecto, de acordo com Bartholo *et al.* (2009), a atuação do mediador nos fóruns consiste em atividades como: esclarecer dúvidas, dar *feedbacks* para os estudantes e propor perguntas para um maior aprofundamento das discussões.

## 3. Revisão Sistemática da Literatura

Uma Revisão Sistemática da Literatura (RSL) permite a identificação e posterior análise das pesquisas mais relevantes, já realizadas, em um determinado campo de pesquisa [Babar 2009]. Nessa perspectiva, com o objetivo de realizar uma RSL reaplicável e provida de imparcialidade na seleção dos estudos, esta pesquisa foi organizada seguindo o percurso metodológico, proposto por Kitchenham (2004), que divide a RSL em planejamento, execução e síntese dos resultados, como mostra as subseções a seguir.

### 3.1. Planejamento

Nesta etapa, foram definidas as estratégias a serem adotadas durante a realização da RSL, cujo objetivo é responder a seguinte questão de pesquisa: "Qual o cenário atual de pesquisas sobre a aplicação de técnicas de Mineração de Textos em Fóruns Educacionais?". Como primeira estratégia, para responder a questão de pesquisa definida, foram delineadas quatro questões específicas para serem respondidas a partir da análise das publicações existentes, no campo de Mineração de Textos em Fóruns Educacionais, são elas:

- Quais as técnicas de PLN estão sendo utilizadas?
- Quais as ferramentas, com direcionamento para Processamento de Textos ou Aprendizagem de Máquina, estão sendo utilizadas?
- Quais os Ambientes Virtuais de Aprendizagem (AVA) estão sendo utilizados?
- Quais os objetivos educacionais norteiam as pesquisas publicadas?

Foram realizadas buscas nas páginas Google Scholar, ACM Digital Library e IEEExplore com os termos de busca "Fórum" + "Mineração de texto" e "Forum" + "Text mining". Após o resultado inicial foram levantados os principais periódicos e conferências das áreas de "Computação e Educação", "Inteligência Artificial" e "Processamento de Linguagem Natural" que possibilitem responder as questões levantadas. A partir disso, foi realizada uma busca manual, em 56 bases de dados (conferências e periódicos), nacionais e internacionais, nas áreas citadas. A Tabela 1 apresenta as 10 bases de dados com mais publicações em cada área.

Tabela 1. Lista de Periódicos e Conferências considerados na revisão

Computação e Educação	Inteligência Artificial	Processamento de Linguagem Natural
<ol> <li>IEEE International Conference on Advanced Learning Technologies;</li> <li>Journal of Distance Education Technologies;</li> <li>Frontiers in Education Conference;</li> <li>CBIE</li> <li>RENOTE;</li> <li>Computers &amp; Education Journal;</li> <li>Conference on Educational Data Mining;</li> <li>Artificial Intelligence in Education;</li> <li>International Conference on Computers and Advanced Technology in Education;</li> <li>Journal of Educational Computing Research.</li> </ol>	<ol> <li>Computers in Human Behavior;</li> <li>Expert Systems with Application;</li> <li>Artificial Intelligence Review;</li> <li>International Conference on Machine Learning Applications;</li> <li>BRACIS;</li> <li>Artificial Intelligence;</li> <li>Knowledge Discovery and Data Mining;</li> <li>Journal of Artificial Intelligence Research;</li> <li>Conference on Artificial Intelligence;</li> <li>International Conference on Web Information Systems Engineering.</li> </ol>	<ol> <li>NAACL;</li> <li>COLING;</li> <li>ACL;</li> <li>Interspeech;</li> <li>Computer Speech and Language;</li> <li>CoNLL;</li> <li>SigDial;</li> <li>EMNLP;</li> <li>IJCAI;</li> <li>NAACL.</li> </ol>

Com o objetivo de selecionar em cada base de dados (Tabela 1) os artigos relevantes ao contexto da pesquisa, assumiram-se alguns critérios de inclusão e exclusão conforme exibe a Tabela 2.

Tabela 2. Critérios de inclusão e exclusão

Critérios de Inclusão	Critérios de Exclusão	
<ul> <li>[I1] Apresentam técnicas/modelos de mineração de textos para fóruns educacionais;</li> <li>[I2] Avaliam técnicas/modelos de mineração de textos para fóruns educacionais;</li> </ul>	[E1] Trabalhos dedicados a apresentar/experimentar técnicas/modelos de mineração de textos, para fóruns, sem objetivos educacionais;	
[I3] Apresentam ferramentas baseadas em	[E2] Revisões/mapeamentos da literatura; [E3] Trabalhos publicados fora do período de 2007 - 2016.	

## 3.2. Fases da Pesquisa

Alinhado aos critérios de inclusão e exclusão, definidos (Subseção 3.1), foram definidas duas fases para seleção das pesquisas (Figura 1). A fase 1 consiste da leitura do título e/ou resumo de cada artigo e posterior aplicação dos critérios de inclusão/exclusão adotados. A fase 2, por sua vez, compreende uma leitura de todo o texto dos artigos selecionados, na fase 1, além da aplicação dos critérios de inclusão/exclusão.

Fase 1	Leitura do título e/ou resumo de cada artigo e aplicação dos critérios de inclusão e exclusão.
Fase 2	Leitura de todo o texto de cada selecionado na etapa 1 e aplicação dos critérios de inclusão e exclusão.

Figura 1. Fases adotadas para seleção dos artigos

Como forma de reduzir as chances de trabalhos relevantes serem excluídos durante o processo, organizou-se a inclusão/exclusão de cada pesquisa por pares. A execução da primeira fase resultou na seleção de 55 artigos, conforme exibe a Tabela 3. Logo após, ao executar a fase 2, dos 55 artigos selecionados na fase 1, apenas 30 artigos foram enquadrados como relevantes em relação ao contexto desta pesquisa.

Tabela 3. Número de artigos selecionados por fase

Fase da seleção	Número de artigos selecionados
Fase 1	55
Fase 2	30

Dessa forma, o passo seguinte foi extrair de cada artigo selecionado as informações necessárias para posterior análise, como segue: a referência completa; técnicas de PLN utilizadas; AVA utilizados; ferramentas de Processamento de Textos ou Aprendizagem de Máquina utilizadas e; o objetivo da pesquisa.

#### 4. Resultados

Nesta seção são apresentados os resultados obtidos. Como já foi dito, 30 artigos foram selecionados como relevantes para esta pesquisa. A Figura 2 apresenta a distribuição anual e as instituições com mais publicações selecionadas.

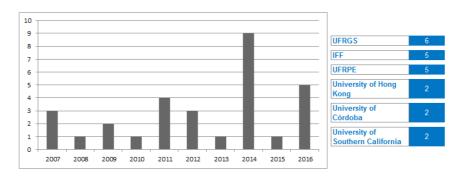


Figura 2. Distribuição anual e instituições com mais publicações selecionadas

É possível perceber a incidência de publicações, no contexto de Mineração de Textos em Fóruns Educacionais, pois nos últimos 10 anos pelo menos uma pesquisa foi publicada anualmente. Atrelado a isso, apenas 6 instituições possuem duas ou mais publicações no contexto pesquisado e 16 apenas uma pesquisa cada. Sendo importante destacar a ocorrência de pesquisas realizadas envolvendo mais de uma instituição.

De acordo com a Figura 3, dos artigos selecionados, 4 (13%) foram extraídos de periódicos e 26 (87%) de conferências. Nesse contexto, chama a atenção o número representativo de publicações selecionadas (8 publicações) do Congresso Brasileiro de Informática na Educação (CBIE).

	Períodos/conferências	Nº de estudos
Conferências Periódicos	RENOTE	2
	Computers & Education	1
	Artificial Intelligence Review	1
13%	Artificial Intelligence in Education	2
	IEEE International Conference on Advanced Learning Technologies	4
	CBIE	8
	IEEE Frontiers in Education Conference	2
	BRACIS	3
	International Conference on Interactive Collaborative Learning	1
	International Conference on Educational Data Mining	2
87%	IEEE Global Engineering Education Conference	1
	ASEE Annual Conference and Exposition	1
	International Conference on Advances in Web based Learning	1
	IEEE International Conference on Data Engineering	1
	Total	30

Figura 3. Quantidade de artigos selecionados por congresso/periódico

Outro ponto interessante é que pesquisas com as características buscadas têm sido publicadas em periódicos ou conferências nas áreas de Computação e Educação e Inteligência Artificial. Esta afirmação é baseada no fato de não terem sido extraídos artigos dos periódicos e das conferências de PLN consideradas.

Em resposta à primeira questão específica de pesquisa, "Quais as técnicas de PLN estão sendo utilizadas?", a Figura 4 exibe um detalhamento das técnicas utilizadas nas pesquisas selecionadas. Com isso, percebe-se que seis pesquisas (21%) não utilizaram/especificaram técnicas de PLN no estudo enquanto, em contrapartida, vinte e três (79%) relataram o uso de uma ou mais técnicas.

Ainda, na Figura 4, é exibido um detalhamento das técnicas utilizadas, na seguinte ordem: remoção de *Stopwords* – remove palavras sem importância semântica para o texto [Azevedo et al. 2011; Lui et al. 2007; Júnior e Esmin 2012], TF-IDF – pondera a importância de cada palavra para um conjunto de documentos [Ferreira et al. 2016], *Stemming* – reduz palavras ao radical [Rolim et al. 2016], LSA – analisa as relações entre documentos de texto [Yoo e Kim 2014], *N-grams* – monta grupos de

palavras de modo a possibilitar a verificação de possíveis dependências [Ravi e Kim 2007], *POS Tagger* – etiqueta palavras com suas respectivas classes gramaticais [Lau 2007], *Tokenization* – remove caracteres especiais e divide o texto em *tokens* a partir do caractere espaço [Silva et al. 2015], *Segmentation* – divide o texto seguindo sua estrutura semântica, por exemplo, palavras e orações [Lin et al. 2009] e *Lemmatization* – transfere as palavras para sua forma de dicionário [Lau 2007].



Figura 4. Técnicas de PLN utilizadas nas pesquisas selecionadas

Em relação à segunda pergunta de pesquisa, "Quais as ferramentas, com direcionamento para Processamento de Textos ou aprendizagem de máquina, estão sendo utilizadas?" de acordo com a Figura 5, apenas dezessete pesquisas (59%) utilizaram ferramentas direcionadas para Processamento de Texto ou Aprendizagem de Máquina. Dentre as ferramentas usadas, podem ser citadas: o WEKA – biblioteca composta por diversos classificadores de Aprendizagem de Máquina [Lin et al. 2009] e; o dicionário para análise de sentimentos de textos em Inglês LIWC [Yoo e Kim (2014)].

<ul><li>Utilizaram ferramentas</li></ul>	Ferramentas	Nº de estudos	% de estudos
43%	WEKA	5	17%
	MineraFórum	2	7%
	Dbpedia Spotlight tool	1	3%
	Cluto-agglomerative clustering	1	3%
	LIWC	1	3%
	CKIP	1	3%
	RapidMiner	1	3%
	SO-CAL	1	3%
	HowNet	1	3%
	UCINET	1	3%
	SOBEK	1	3%
	WordNet	1	3%
	SMART Retrieval System	1	3%

Figura 5. Cenário das ferramentas de PLN e Aprendizagem de Máquina

A Figura 6, por sua vez, expõe o percentual de pesquisas que fizeram uso de Ambientes Virtuais de Aprendizagem. A propósito, é considerado como uso de AVA, nas pesquisas, os seguintes cenários: cenário 1 - aplicação de sistema/modelo de Mineração Textos em um dado AVA ou; cenário 2 - extração de postagens, geradas em um dado AVA, para testar sistemas de Mineração de Textos.

Dessa forma, em resposta à terceira questão de pesquisa: "Quais os Ambientes Virtuais de Aprendizagem (AVA) estão sendo utilizados?", os AVA utilizados nas pesquisas selecionadas são: o Moodle utilizado em 12 estudos [Rubio e Villalon 2016; Lopez et al. 212], a plataforma de Ensino a Distância da UFRGS ROODA em 3 estudos [Azevedo et al. 2011; Silva et al. 2015], o Fórum.Edu no estudo de Ferreira et al. (2016) e o Coursera em [Wen et al. (2014)].

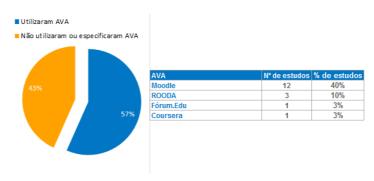


Figura 6. Ambientes Virtuais de Aprendizagem utilizados

Por fim, para responder a quarta questão de pesquisa: "Quais os objetivos educacionais norteiam as pesquisas publicadas?", na Figura 6 são exibidos os objetivos educacionais identificados nas pesquisas selecionadas. Como pode ser visualizado, as 30 pesquisas selecionadas foram agrupadas em 9 objetivos macro.

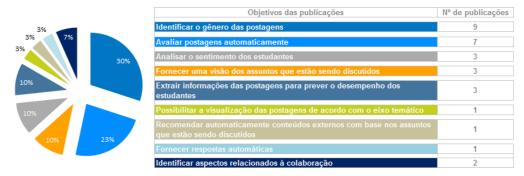


Figura 6. Objetivos educacionais das pesquisas selecionadas

Em cada objetivo listado, na Figura 6, estão agrupadas pesquisas relacionadas com o objetivo em questão. Por exemplo, o objetivo de pesquisa "Identificar o gênero das postagens" agrupa 9 estudos. A seguir são exemplificados alguns dos estudos selecionados para cada objetivo:

**Identificar o gênero das postagens:** constam, neste objetivo, pesquisas direcionadas para a identificação de perguntas não respondidas em uma discussão *online* [Ravi e Kim 2007]; dedicadas para a identificação de postagens sobre dúvidas [Rolim et al. 2016].

**Avaliar postagens automaticamente:** pesquisas com foco em avaliar a relevância das contribuições do estudante em relação à discussão corrente [Rubio e Villalon 2016].

**Analisar o sentimento dos estudantes:** pesquisas direcionadas para a análise dos sentimentos dos estudantes com relação ao curso [Wen et al. 2014].

Fornecer uma visão dos assuntos que estão sendo discutidos: pesquisas que propõem a extração de tópicos, relevantes, para serem discutidos no decorrer do fórum [Nunes et al. 2014]; sugerem a criação de mapas conceituas com o objetivo de sintetizar o andamento da discussão [Lau 2007].

Extrair informações das postagens para prever o desempenho dos estudantes: pesquisas voltadas para prever as notas finais dos estudantes, ou o desempenho em projetos específicos, a partir das postagens proferidas em fóruns de discussão [López et al. 2012; Yoo e Kim 2014];

**Possibilitar a visualização das postagens de acordo com o eixo temático:** Estudos sobre a possibilidade de navegação temática e recomendação das postagens [Li et al. 2008].

Recomendar automaticamente conteúdos externos com base nos assuntos que estão sendo discutidos: pesquisas com viés para a recomendação de links relacionados ao assunto discutido [Costa et al. 2014].

**Fornecer respostas automáticas:** pesquisas sobre o uso de informações textuais, extraídas dos fóruns, para orientar agentes pedagógicos na resposta a perguntas dos estudantes [Kim e Shaw 2014].

**Identificar aspectos relacionados à colaboração:** pesquisas com foco na identificação de Presença Social e sensação de Pertencimento em Fóruns Educacionais [Bastos et al. 2011].

#### 3. Discussão dos Resultados

Os resultados obtidos evidenciam o interesse da comunidade de Informática na Educação, nos últimos 10 anos, por pesquisas direcionadas para a aplicação de técnicas de Mineração de Textos em Fóruns Educacionais. De acordo com resultados, as técnicas de PLN mais utilizadas para potencializar a mineração textual é a remoção de *stopwords* e o TF-IDF. Destaca-se também o uso de técnicas/ferramentas de Aprendizagem de Máquina nos estudos selecionados, tais como o WEKA.

Como ambiente para a aplicação das técnicas propostas nos estudos, o Moodle tem sido amplamente escolhido. Além disso, é possível perceber basicamente duas correntes de pesquisas na área de Mineração de Textos em Fóruns Educacionais. Em uma o foco é automatizar processos, de modo a facilitar a mediação pedagógica nos fóruns, é o caso de pesquisas focadas em avaliar qualitativamente postagens, extrair tópicos relevantes para serem discutidos ao longo do debate e organizar as postagens de acordo com o tema abordado. Na segunda corrente, são conduzidas pesquisas sobre como a aplicação de técnicas de Mineração de Textos pode auxiliar os professores/tutores em ações preventivas contra baixos desempenhos de aprendizagem e evasão, podem ser citados como exemplos os trabalhos dedicados para a previsão do desempenho e a análise de sentimentos dos estudantes.

Apesar do visível avanço neste campo de pesquisa, uma análise do atual cenário considerando as características colaborativas do fórum, mostra algumas oportunidades de estudos. Por exemplo, práticas não colaborativas como a predominância de postagens direcionadas do estudante exclusivamente para o mediador, são comuns no atual contexto da EAD, abrindo margem para a condução de investigações focadas em utilizar técnicas de Mineração de Textos para fornecer indicadores acerca dos aspectos colaborativos das discussões. Por sinal, das pesquisas selecionadas duas se posicionam nesse aspecto, porém focadas apenas na identificação de um dos aspectos colaborativos, a presença social.

Também, saber o perfil de aprendizagem dos estudantes ou o interesse por determinados conteúdos, identificar os níveis de plágio nos fóruns são algumas das oportunidades visualizadas para o emprego de técnicas de Mineração de Textos em

Fóruns Educacionais. Tais pontos, uma vez integrados no ambiente educacional, podem subsidiar a montagem de pares de discussões ou ações corretivas/preventivas dos professores/tutores.

# 4. Considerações Finais

Esta pesquisa se propôs, por meio de uma RSL, identificar qual o atual cenário de pesquisas sobre o uso de técnicas de Mineração de Textos em Fóruns Educacionais. Para tal, foram realizadas buscas em 56 bases de dados, nacionais e internacionais, com o objetivo de ter-se uma visão sistêmica das pesquisas já realizadas.

Com isso, além das técnicas/ferramentas utilizadas e dos objetivos educacionais presentes nos estudos selecionados, esta pesquisa mostrou os avanços no que se refere ao uso de técnicas de Mineração de Textos. Mesmo com o visível avanço do campo de pesquisa estudado, apenas duas pesquisas foram identificadas com o objetivo de fortalecer, a principal característica dos fóruns, a colaboração.

Espera-se a partir desta revisão estimular a condução de pesquisas voltadas para a utilização de técnicas de Mineração de Textos em Fóruns Educacionais de modo a potencializar o uso desta ferramenta sem sobrecarregar o mediador.

#### Referências

- Azevedo, B. F. T., Behar, P. A., Reategui, E. B. (2011). Análise das mensagens de fóruns de discussão através de um software para mineração de textos. In: Simpósio Brasileiro de Informática na Educação, pág. 20-29.
- Babar, Ali., Muhammad. (2009). Systematic literature reviews in software engineering: Preliminary results from interviews with researchers. *In: IEEE Third International Symposium on Empirical Software Engineering and Measurement*, pág. 346-355.
- Bastos, H. P. P. et al. (2011). Presença Social e Pertencimento em Fóruns Educacionais: Manifestação e Percepção de Afetividade. In: Simpósio Brasileiro de Informática na Educação, pág. 1047-1056.
- Bartholo, V. F., Amaral, M. A., Cagnin, M. I. (2009). Uma Contribuição para a Adaptabilidade de Ambientes Virtuais de Aprendizagem para Dispositivos Móveis. *In: Brazilian Journal of Computers in Education*, v.17, n.2, pág. 36-47.
- Carvalho, M. R., Lima, R. L. (2015). A importância da afetividade na EaD: uma perspectiva de Wallon. In: Revista EDAPECI, v.15, n.1, pág. 192-205.
- Costa, L. C. et al. (2014). Sistema de Recomendação de Links para o fomento de discussões em fóruns de um Ambiente Virtual de Aprendizagem. In: RENOTE.
- Ferreira, M. A. D. et al. (2016). Análise de Classificadores para Avaliação automática em Fóruns Educacionais. In: ENIAC.
- Júnior, R. L. O., Esmin, A. A. A. (2012). Monitoramento Automático de Mensagens de Fóruns de Discussão Usando Técnica de Classificação de Texto Semi-Supervisionado. In: Simpósio Brasileiro de Informática na Educação.

- Kim, J., Shaw, E. (2014). Scaffolding student online discussions using past discussions: PedaBot studies. *In: Artificial Intelligence Review*.
- Kitchenham, B. (2004). Procedures for performing systematic reviews. Technical report, Keele University and NICTA.
- Lau, R. Y. K. (2007). Towards Fuzzy Domain Ontology Based Concept Map Generation for E-Learning. *In: Advances in Web Based Learning*, pág. 90-101.
- Li, Y., Dong, M., Huang, R. (2008). Semantic Organization of Online Discussion Transcripts for Active Collaborative Learning. *In: IEEE International Conference on Advanced Learning Technologies*.
- Lin, F., Hsieh, L., Chuang, F. (2009). Discovering Genres of Online Discussion Threads via Text Mining. *In: Computers & Education*, pág. 481-495.
- Lui, A. K., Li, S. C., Choy, S. O. (2007). An Evaluation of Automatic Text Categorization in Online Discussion Analysis. In: *IEEE International Conference on Advanced Learning Technologies*.
- López, M. I. et al. (2012) Classification via clustering for predicting final marks based on student participation in forums. *In: International Conference on Educational Data Mining*.
- Nunes, B. P., et al. (2014). A Topic Extraction Process for Online Forums. In: *IEEE* 14th International Conference on Advanced Learning Technologies, pág. 541-543.
- Rolim, V. B., Ferreira, R., Costa, E. (2016). Identificação Automática de Dúvidas em Fóruns Educacionais. In: Simpósio Brasileiro de Informática na Educação, pág.
- Rubio, D., Villalon, J. (2016). A Latent Semantic Analysis method to measure participation quality online forums. *In: IEEE 16th International Conference on Advanced Learning Technologies*, pág.18-19.
- Ravi, S., Kim, J. (2007). Profiling Student Interactions in Threaded Discussions with Speech Act Classifiers. *In: Artificial Intelligence in Education*.
- Scheuer, O.; McLaren, B. M. (2008). Helping Teachers Handle the Flood of Data in Online Student Discussions. *In: Intelligent Tutoring Systems*, pág. 323-332.
- Silva, L. A. et al. (2015). Mineração de Dados em publicações de Fóruns de Discussões do Moodle como geração de Indicadores para aprimoramento da Gestão Educacional. In: Simpósio Brasileiro de Informática na Educação, pág. 1084-1093.
- Wen, M. et al. (2014). Sentiment Analysis in MOOC Discussion Forums: What does it tell us?. *In: International Conference on Educational Data Mining*.
- Xia, J. C., Fielder, J., Siragusa, L. (2013). Achieving better peer interaction in online discussion forums: A reflective practitioner case study. *In: Issues in Educational Research*, v.23, n.1, pág. 97-113.
- Yoo, J., Kim, J. (2014). Can Online Discussion Participation Predict Group Project Performance? Investigating the Roles of Linguistic Features and Participation Patterns. *In: Artificial Intelligence in Education*, pág. 8-32.