

## Mineração de Texto Aplicada à Identificação de Colaboração em Fóruns Educacionais

Máverick A. D. Ferreira<sup>1</sup>, Rafael Ferreira<sup>1</sup>, Anderson P. Cavalcanti<sup>1</sup>, Ruan Carvalho<sup>1</sup>, Sebastião Neto<sup>2</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,rafaelflmeillo, anderson.pinheiro27, ruanvbc, sebast.rogers}@gmail.com

**Abstract.** *Educational forum is an important resource for interaction between teachers and students in virtual learning environments. Such space that should be used to generate knowledge often comes down to student questions and teacher answers. To address this problem, this paper presents a approach for automatic identification of collaboration in educational forums. This approach aims to aid teachers in the follow-up of the post and encourage the students' participation with collaborative posts. To evaluate the proposal, a quantitative and qualitative experiment was carried out.*

**Resumo.** *O fórum educacional é um importante recurso para a interação entre professores e alunos em Ambientes Virtuais de Aprendizagem. Esse espaço que deveria ser usado para gerar conhecimento muitas vezes acaba se resumindo a perguntas de alunos e respostas do professor. Para lidar com esse problema, este trabalho apresenta uma abordagem para identificação automática de colaboração em fóruns educacionais. Essa abordagem visa principalmente auxiliar professores no acompanhamento das postagens e incentivar a participação de alunos com postagens colaborativas. Para avaliar a proposta foi realizado um experimento quantitativo e qualitativo, a partir da aplicação do fórum numa turma real.*

### 1. Introdução

O surgimento da modalidade de ensino a distância facilitou o acesso à educação a diferentes públicos, devido à flexibilidade em relação aos horários de estudo e acessibilidade. Por isso, a educação à distância (EAD) tem contribuído para a democratização do acesso à educação, principalmente, dos níveis técnico e superior. Por outro lado, o número reduzido de encontros presenciais pode desencadear a sensação de isolamento no estudante impactando negativamente na aprendizagem ou, até mesmo, levando à evasão [Daudt e Behar 2013].

Nesse contexto, os Ambientes Virtuais de Aprendizagem (AVA) são utilizados como forma de possibilitar interações entre estudantes e professores. Sendo o fórum educacional, uma das ferramentas integradas aos AVA, capaz oportunizar colaboração entre os envolvidos, por meio de discussões assíncronas, posicionando-se como um

aliado para evitar o isolamento de estudantes no contexto da EAD [Bastos e Silva 2009].

Porém, mesmo com as características colaborativas dos fóruns, as postagens inseridas por estudantes nas discussões são predominantemente direcionadas ao mediador [Ferreira et al. 2015]. Contudo, interações do tipo estudante-estudante são importantes para o processo de geração de conhecimento colaborativo. Também é relevante considerar que com o aprofundamento da discussão existe um aumento no número de postagens, tornando dispendioso para o mediador acompanhar o andamento do debate [Dringus e Ellis 2005].

Desse modo, a possibilidade de sobrecarga do mediador com o crescimento das discussões e a relevância da mediação pedagógica, para promoção de colaboração nos fóruns, implica no seguinte problema de pesquisa: como apoiar professores/tutores na atividade de monitoramento e incentivo à colaboração em fóruns educacionais? Diante deste cenário, este trabalho apresenta uma abordagem para a identificação automática de aspectos colaborativos em postagens de fóruns educacionais. Os principais objetivos desta abordagem são: (i) auxiliar professores no acompanhamento do fórum e; (ii) incentivar a colaboração entre estudantes.

A proposta foi avaliada a partir da condução de experimentos em uma base de dados com 600 postagens extraídas de fóruns educacionais e uma aplicação em um fórum real utilizado para atividades numa turma com 10 alunos. Os resultados mostram que a abordagem proposta é bastante promissora.

## 2. Trabalhos Relacionados

Dado o potencial dos fóruns de discussão, para promover a colaboração na EAD, pesquisas têm sido conduzidas ao longo dos últimos anos com o objetivo de fortalecer os aspectos colaborativos desta ferramenta. Nesse sentido, esta seção apresenta trabalhos já realizados em benefício da colaboração em fóruns educacionais.

Em Guimarães *et al.* (2013), foi utilizado um modelo de reconhecimento de colaboração, proposto por Murphy (2004), com o objetivo de identificar o nível de colaboração em um fórum de discussão *online* de uma comunidade voltada para a aprendizagem de técnicas artesanais. As características de colaboração foram identificadas de forma manual. Os resultados obtidos evidenciaram as ações colaborativas dos estudantes no decorrer das discussões mostrando o potencial do modelo utilizado.

Ferreira *et al.* (2015) apresentaram um modelo de fórum baseado em mapas conceituais. A proposta além de diminuir o papel hierárquico do professor/tutor favorece interações do tipo estudante-estudante impactando, assim, positivamente para a colaboração em fóruns. Como forma de avaliar o modelo, os autores conduziram simulações utilizando o método formal rede de Petri Colorida.

Com um olhar para os altos índices de evasão na EAD e defendendo a importância de estimular nos estudantes a sensação de pertencimento ao grupo, em Silva *et al.* (2012), é proposta uma abordagem destinada à identificação automática de presença social em fóruns e chats. Durante o estudo os autores se propuseram a criar regras, para a identificação de presença social, específicas para as bases de dados

utilizadas e também regras genéricas (podem ser aplicadas em outros fóruns). Para testá-las, foram conduzidos dois experimentos em cada base de dados. A abordagem em questão chegou a atingir com as regras específicas 93,95% (base 1) e 88,80% (base 2) de taxa de acerto e com as regras genéricas 58,90% (base 1) e 5,00% (base 2). A maior dificuldade dessa abordagem é a forte dependência de um anotador inicial para os termos de colaboração.

Gomes (2012), em sua dissertação de mestrado, defende um processo de ensino-aprendizagem centrado nas interações entre estudantes. Para tanto, apresenta um sistema baseado em PLN direcionado para a análise automática das interações em fóruns para, em seguida, gerar indicadores que permitam ao aluno avaliar seu próprio desempenho e ao tutor/professor mensurar o desempenho dos participantes da discussão.

Com a mesma perspectiva dos trabalhos listados acima, esta pesquisa busca estimular a utilização dos aspectos colaborativos presentes nos fóruns de discussão. No entanto, visualiza-se como principal diferencial deste estudo, frente aos demais, o posicionamento de uma abordagem para identificação automática de colaboração, para fóruns, baseada no modelo proposto por Murphy (2004).

### 3. Identificação Automática de Colaboração

A abordagem proposta tem como objetivo identificar, a partir da análise de postagens inseridas em fóruns educacionais, aspectos que indiquem colaboração entre estudantes em discussões assíncronas. Para isso, foi utilizado o modelo proposto por Murphy (2004), sendo consideradas as seguintes dimensões: **Presença Social:** reconhecer a presença do grupo e elogiar/expressar apreciação pelos outros participantes; **Acolhimento ou reflexão das perspectivas do outro:** introduzir novas perspectivas e; **Co-construção de perspectivas e significados partilhados:** partilhar informação e recursos.

De acordo com a Figura 1, a proposta foi estruturada em dois módulos os quais utilizam técnicas de PLN para estruturar regras direcionadas à identificação de aspectos colaborativos nas postagens.

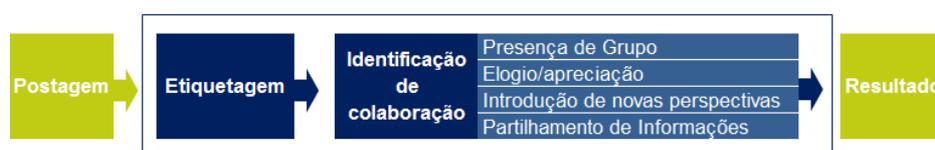
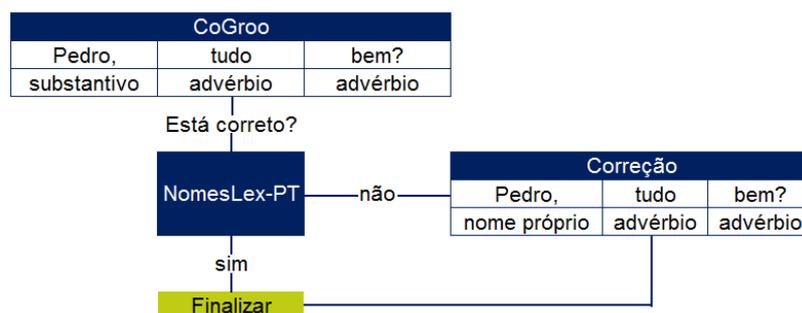


Figura 1. Estrutura da abordagem proposta

Conforme sintetizado, na Figura 1, os módulos integrantes da abordagem são: *etiquetagem* e *identificação de colaboração*. As seções a seguir detalham o funcionamento de cada um dos módulos.

#### 3.1. Etiquetagem

O módulo de *etiquetagem* é designado para a realização de análise morfológica das postagens e para etiquetar cada palavra com sua respectiva classe gramatical. Para essa etapa foi utilizado o componente gramatical CoGroo [Silva e Finger 2013].



**Figura 2. Exemplo de utilização do NomesLex-PT como supervisor do CoGroo**

A partir de uma análise de resultados preliminares, constatou-se uma dificuldade por parte do CoGroo para etiquetar corretamente nomes próprios. Por isso, optou-se pelo uso de um léxico com 2027 nomes em Português, NomesLex-PT<sup>1</sup>, para supervisionar o CoGroo no processo de etiquetagem, conforme mostrado na Figura 2.

### 3.2. Identificação de colaboração

O módulo de *Identificação de Colaboração* avalia a postagem etiquetada buscando identificar os quatro aspectos considerados no modelo adotado. Para tal, foi estruturado um sistema de regras em conjunto com uma tutora, da Universidade Aberta do Brasil (UAB), com formação básica em Licenciatura em Letras.

#### 3.2.1. Presença de Grupo

A estrutura de classificação destinada para o reconhecimento de Presença de Grupo foi organizada com quatro regras. Na regra um, *Interjeições no início da postagem*, é verificada a existência de palavras etiquetadas como interjeição nas três primeiras posições da postagem analisada. Esta estratégia foi adotada por entender-se que interjeições como “oi!” e “olá!” são costumeiramente introduzidas no início da fala. A regra dois, *Cumprimentos no início da frase*, procura cumprimentos no início da postagem, sendo consideradas expressões como “bom dia”, “boa tarde” e “boa noite”.

O envolvimento de pessoas em um discurso pode dar-se em primeira pessoa (quem fala), segunda pessoa (com quem fala) ou terceira pessoa (de quem fala). Com isso, ao entender como um forte indício de reconhecimento de Presença de Grupo o ato de um estudante (quem fala) dirigir-se a outros estudantes (com quem fala) por meio de uma postagem, a terceira regra busca localizar na postagem em análise: pronomes pessoais na primeira e segunda pessoa do plural. Mais especificamente, pronomes pessoais caso reto (ex. nós, vós), pronomes pessoais oblíquos (ex. conosco, convosco) e pronomes pessoais de tratamento (ex. vocês, senhores).

Para estruturação da quarta regra, montou-se um conjunto *A* com os seguintes termos: “{colegas, amigos, companheiros, galera}”. Logo, a regra quatro consiste em aferir a intersecção entre o conjunto *B* “{substantivos presentes na postagem em análise}” e o conjunto *A*. Ao término desta operação, se a intersecção entre *A* e *B* resultar em um conjunto com um ou mais elementos, a postagem analisada possui palavras que representam o coletivo de estudantes.

1 <http://xldb.fc.ul.pt/wiki/NomesLex-PT01>

Caso a intersecção entre  $A$  e  $B$  seja vazia, isto é, não existem substantivos associados ao coletivo de estudantes na postagem analisada, o passo seguinte é montar um conjunto  $C$  contendo sinônimos de cada substantivo presente na postagem em análise. A preparação do conjunto de sinônimos conta com um dicionário *léxico* para Português Brasil, o WordNet.PT<sup>2</sup>. Em seguida, verifica-se a intersecção entre os conjuntos  $C$  e  $A$  e caso obtenha-se como resultante um conjunto com um ou mais elementos, a postagem analisada possui palavras que representam o coletivo de estudantes.

Após a execução dos passos mencionados acima e tendo sido encontrado, na postagem analisada, pelo menos um dos substantivos presentes no conjunto  $A$ , passa a ser verificada a posição do mesmo na postagem. O objetivo de identificar a sua posição na postagem se resume à importância de confirmar a não existência de pronomes em terceira pessoa na posição anterior.

### 3.2.1. Elogio/Apreciação por outros participantes

Para a identificação de *Elogio/Apreciação por outros participantes*, foram estruturadas um total de 15 regras utilizando os termos da postagem etiquetados em: substantivos, pronomes pessoais segunda pessoa, pronomes possessivos segunda e terceira pessoa, adjetivos positivos, advérbios positivos, advérbios negativos e nomes próprios.

Como o módulo de etiquetagem (Subseção 3.1) atribui apenas as classes gramaticais para cada palavra, para classificar os adjetivos/advérbios em positivos e negativos foi utilizado o *léxico* para sentimentos denominado de SentiLex-PT [Carvalho e Silva 2015]. Este *léxico* possibilitou a análise dos adjetivos/advérbios de cada postagem, tendo como retorno -1 (negativa), 0 (neutra) ou 1 (positiva).

Do mesmo modo, com o objetivo de identificar substantivos com menção a estudantes, montou-se um conjunto  $A$  com os seguintes termos: “{colega(s), amigo(s), companheiro(s)}”. E como complemento, um conjunto  $B$  “{substantivos presentes na postagem em análise}” para, a partir da intersecção entre  $A$  e  $B$ , identificar na postagem se existe e quais os substantivos se referem a estudantes.

**Tabela 1a. Regras para identificação de Elogio/Apreciação**

Regras	Exemplos
nome próprio + verbo de ligação + adjetivo	João é legal.
nome próprio + verbo de ligação + advérbio + adjetivo	João é bem inteligente.
substantivo + verbo de ligação + adjetivo	Pelas discussões, a turma é legal.
nome próprio + verbo + adjetivo	João colocou de forma inteligente.
nome próprio + verbo + advérbio	Pedro estudou bem o assunto
substantivo + verbo + adjetivo	Θ colega colocou de maneira interessante.
substantivo + verbo + advérbio	Θ colega pontuou devidamente.

A partir das Tabelas 1a e 1b é possível perceber que as regras elaboradas buscam identificar elogios/apreciação de duas formas. A primeira regra, por exemplo, aponta o elogio “legal” como característica do sujeito “João” por meio de um verbo de ligação. Em contrapartida, a quarta direciona um Elogio/Apreciação “inteligente” para uma ação “colocou” realizada pelo sujeito “João”.

<sup>2</sup> <https://github.com/own-pt/openWordnet-PT>

**Tabela 1b. Regras para identificação de Elogio/Apreciação**

pronome pessoal (2ª pessoa) + verbo + advérbio	Você colocou bem.
adjetivo + nome próprio	Interessante ponto de vista, João.
advérbio + verbo + nome próprio	Perfeitamente discutido, Maria.
advérbio + verbo + substantivo	Bem colocado, colega.
verbo + advérbio + nome próprio	Discutiu bem, Pedro.
substantivo + adjetivo	Pense numa turma boa.
adjetivo (no início da postagem) + substantivo	Boa, galera.
adjetivo + pronome possessivo (2ª pessoa e 3ª do Singular)	<del>Não sei por que,</del> mas achei legal sua colocação.

É importante destacar que as regras listadas nas Tabelas 1a e 1b consideram apenas substantivos com menção a estudantes. Além disso, como o objetivo desta etapa é identificar indícios de Elogio/Apreciação nas postagens, os adjetivos/advérbios explícitos nas regras (Tabelas 1a e 1b) são positivos. Ainda, como um advérbio pode modificar um verbo, outro advérbio ou um adjetivo, após a validação de qualquer uma das regras adotadas, é examinada a presença de advérbios negativos junto a alguma palavra (verbo, advérbio ou adjetivo) ligada ao adjetivo/advérbio positivo presente na regra em questão. Isto acontece porque em uma frase “João não discutiu bem.”, onde um advérbio positivo “bem” modifica o verbo “discutiu” e um advérbio negativo “não” também modifica o mesmo verbo “discutiu”, o sentido negativo prevalece.

### 3.2.2. Introdução de Novas Perspectivas

Para construir regras capazes de detectar a *Introdução de Novas Perspectivas* na postagem, inicialmente foi conceituada a expressão *novas perspectivas* no contexto desta abordagem. Nesse sentido, ficou definido como *novas perspectivas* os seguintes cenários: a inserção de ideias complementares e a inserção de ideias de contraposição. A regra um, associada ao primeiro cenário, busca identificar na postagem conjunções aditivas, isto é, expressões direcionadas para a complementação de ideias, tais como: “como também”, “mas ainda” e “não só”.

No mesmo sentido, a regra dois concentra-se na descoberta de conjunções adversativas na postagem analisada. Isto acontece porque tais conjunções ligam orações por meio de oposição de ideias. Alguns exemplos de expressões adversativas são “porém”, “contudo” e “entretanto”.

### 3.2.3. Partilhamento de Informações

A identificação de *Partilhamento de Informações* conta apenas com uma regra. Esta tem como objetivo identificar o compartilhamento de links para materiais relacionados aos assuntos discutidos no fórum.

Inicialmente é realizada uma procura, na postagem em análise, pelos termos “http(s)” e “www” por serem associados a *urls*. Como no módulo de *etiquetagem* uma *url* é tratada como uma única palavra, quando encontradas palavras com termos associados a *urls*, estas são processadas em um motor de busca. Após o processamento, caso seja retornado um conteúdo html, são extraídos os conteúdos textuais presentes nas *tags* <p>.

Em seguida é calculado o TF-ISF, medida proposta por Larocca *et al.* (2000), que tem como objetivo extrair os principais conceitos de um texto por meio do

ranqueamento de suas palavras. O resultado obtido com TF-ISF, isto é, as palavras representativas do conteúdo compartilhado, é utilizado para projetar um conjunto *A*. Em paralelo, é criado um conjunto *B* com os substantivos contidos na problemática do fórum no qual a postagem analisada foi inserida. Com isso, se o conjunto resultante da intersecção dos conjuntos *A* e *B* não for vazio, é identificado o *Partilhamento de Informações* relacionadas ao conteúdo proposto para discussão no fórum.

#### 4. Avaliação

Para o processo de avaliação foi construída uma base de dados com 600 postagens extraídas das seguintes disciplinas do curso de Ciência da Computação: Análise de dados, Sistemas de Informação, Internet e Web e Programação Orientada a Objetos. Cada postagem da base foi avaliada manualmente por dois avaliadores externos, sendo solicitado um terceiro avaliador em caso de divergência de avaliação.

As postagens foram categorizadas seguindo as dimensões consideradas do modelo de Murphy (2004): Presença de Grupo, Elogio/Apreciação por outros participantes, Introdução de Novas Perspectivas e Partilhamento de Informações. A Tabela 2 mostra a distribuição das postagens em cada dimensão, onde *identificado* significa que a postagem contém a dimensão avaliada e *não identificado* a postagem não contém.

**Tabela 2. Distribuição das postagens**

Dimensões avaliadas	Identificado	não-identificado
Presença de Grupo	19	581
Elogio/Apreciação	78	522
Introdução de Novas Perspectivas	92	508
Partilhamento de Informações	17	583

Após a distribuição das postagens, foram realizados dois experimentos utilizando a base de dados preparada, um com a abordagem proposta e outro utilizando um classificador *Support Vector Machine* (SVM) utilizando todas as palavras como atributos. Este algoritmo foi escolhido devido aos bons resultados alcançados no domínio de classificação de textos [Haddoud *et al.* 2016]. O resultado de cada experimento foi comparado com o da avaliação manual para, em seguida, verificar qual o classificador obteve melhor desempenho. Para fins de comparação adotaram-se as medidas precisão, cobertura e *f-measure*.

A Tabela 3 mostra os resultados obtidos com a abordagem proposta. Todas as dimensões obtiveram resultados acima de 50% de *f-measure*, chegando até 82,8% para *Partilhamento de Informação* e 78,5% para *Presença de Grupo*.

É importante destacar que o banco de dados é bastante desbalanceado, devido à natureza do fórum, no qual a maioria das postagens são perguntas ou respostas. Isso dificultou a identificação das dimensões apresentadas. Em relação à aplicação do algoritmo SVM na base de dados original, todas as postagens foram classificadas como *não identificado*. Em outras palavras, o classificador não conseguiu definir bem a separação da classe. Para realizar uma comparação com a proposta, foi realizado um balanceamento do banco de dados e nova execução do algoritmo SVM. A Tabela 4 mostra os resultados alcançados, onde apenas a dimensão *Elogio/Apreciação* obteve um

resultado consideravelmente maior que o da proposta. Enquanto, na *Presença de Grupo* e *Partilhamento de Informações* a abordagem proposta atingiu resultados melhores.

**Tabela 3. Resultados obtidos com a abordagem proposta**

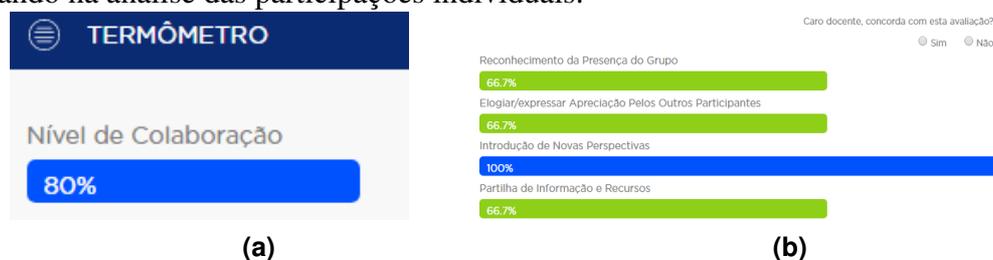
Dimensões avaliadas	Precisão	Cobertura	<i>F-mesurare</i>
Presença de Grupo	0,807	0,766	0,785
Elogio/Apreciação	0,687	0,598	0,639
Introdução de Novas Perspectivas	0,605	0,559	0,581
Partilhamento de Informações	0,846	0,811	0,828

**Tabela 4. Resultados obtidos com o *support vector machine* (base balanceada)**

Dimensões avaliadas	Precisão	Cobertura	<i>F-mesurare</i>
Presença de Grupo	0,711	0,710	0,709
Elogio/Apreciação	0,719	0,717	0,716
Introdução de Novas Perspectivas	0,586	0,586	0,586
Partilhamento de Informações	0,429	0,441	0,416

## 5. Aplicação da Proposta

Com o objetivo de realizar uma avaliação da abordagem proposta em um contexto real, foi desenvolvido um fórum com alguns diferenciais conforme ilustrado na Figura 3a - uma barra mostrando o **nível de colaboração** dos alunos no fórum e; na Figura 3b - detalhes sobre o **nível de participação de cada aluno**. Vale ressaltar que o **nível de colaboração** se refere à participação de todos os estudantes no fórum, sendo este fixado na tela de discussão com o objetivo de motivar os estudantes a demonstrarem aspectos colaborativos e, assim, manterem o termômetro elevado. Já o **nível de participação de cada aluno** é exibido apenas para o professor (administrador do fórum criado), auxiliando na análise das participações individuais.



**Figura 3. Estatísticas exibidas para estudantes e professores**

Além disso, foram definidas algumas mensagens para *feedbacks* a serem exibidas individualmente para cada estudante após o mesmo inserir uma postagem no fórum (Tabela 5).

Desse modo, a avaliação da abordagem proposta se deu em dois momentos. No primeiro, realizou-se um experimento em uma turma de tópicos Avançados em Inteligência Artificial com 10 alunos. Neste experimento foi utilizada a ferramenta fórum do AVA Moodle, para discussão de artigos científicos relacionados ao tema da disciplina em questão, tendo os alunos uma semana para realizar as postagens. Ao fim da semana, apenas 10 postagens foram registradas.

**Tabela 5. Exemplos de *feedbacks* para cada dimensão identificada**

<b>Dimensões identificadas</b>	<b>Exemplos de <i>feedbacks</i></b>
Presença de Grupo	“De fato, não tem como debater sozinho. Por isso, é sempre bom reconhecer a presença dos colegas! Vamos em frente!”
Elogio/Apreciação	“Parabéns! Ao valorizar as contribuições dos seus colegas, você está contribuindo para o aprofundamento do debate.”
Introdução de Novas Perspectivas	“Parabéns pela postagem! Criticidade é algo muito importante para o andamento da discussão!”
Partilhamento de Informações	“É muito interessante trazer conteúdos externos para enriquecer a discussão! Mas cuidado, é bom sempre analisar se o conteúdo externo pode contribuir para a discussão.”

No segundo momento, foi passado outro artigo para a mesma turma, porém fazendo uso da ferramenta desenvolvida como experimentação à abordagem proposta. Como consequência, o número de postagens subiu para 30. Além do acréscimo do número de postagens, na primeira semana a maioria das postagens inseridas foram simplesmente um resumo do artigo proposto para discussão e nem todos os alunos participaram da atividade.

Por outro lado, na segunda semana, todos os alunos matriculados na disciplina participaram do fórum e houve um número maior de interações entre os alunos contando não só com postagens que resumiam o artigo, como também textos que contemplavam todas as quatro categorias de colaboração identificadas na abordagem proposta.

Por fim, também foi relevante para o professor o fato de poder contar com as estatísticas relacionadas às postagens. Esse fator foi levado em consideração para a atribuição final da nota dos alunos nos fóruns.

## **6. Considerações Finais**

O trabalho apresentou uma abordagem para identificação automática de postagens colaborativas em fóruns educacionais. Os objetivos do trabalho são: (i) incentivar a participação de estudantes nos fóruns, principalmente por meio de interações estudante-estudante; (ii) auxiliar professores e tutores na mediação do fórum.

Os resultados obtidos demonstram o potencial da proposta para identificar automaticamente interações entre discentes em fóruns. Além disto, a abordagem foi integrada a um fórum educacional real, facilitando a análise do professor nos fóruns testados, ajudando o docente a realizar avaliação formativa de cada estudante. Em relação aos alunos a proposta impulsionou a quantidade de interações do tipo estudante-estudante.

Como trabalhos futuros, busca-se: (i) criar uma abordagem híbrida entre regras e aprendizagem de máquina para melhorar os resultados obtidos; (ii) ampliação da base de dados, utilizando fóruns de diferentes cursos; (iii) realização de novos experimentos contando com um número maior de estudantes; (iv) identificação de colaboração utilizando outras dimensões.

## Referências

- Bastos, H. P. P., Silva, J. M. S. (2009). Fatores de evasão em curso a distância: Relato de Pesquisa sobre Evadidos do Curso "Leitura Instrumental em Inglês a Distância" no IFF, RJ. In: RENAME, v.7, n.3.
- Carvalho, P., Silva, M. (2015). SENTILEX-PT: Principais Características e potencialidades. In: *Oslo Studies in Language*. v.7, n.1, pág. 425-438.
- Daudt, S. I. D., Behar, P. A. (2013). A gestão de cursos de graduação a distância e o fenômeno da evasão. In: *Revista Educação*. v. 36, n. 3, pág. 412-421.
- Dringus, L. P., Ellis, T. (2005). Using data mining as a strategy for assessing asynchronous discussion forums. In: *Computers & Education*, v.45, n.1, pág. 141 – 160.
- Ferreira, F. J. M.; Miranda, S. K. O.; Costa, E. B.; Costa, F. P. D.; Rocha, H. J. B. (2015) Um Modelo de Fórum de Discussão com Suporte às Interações entre Aprendizes utilizando Mapas Conceituais. In: XXIV Simpósio Brasileiro de Informática na Educação (SBIE), pág. 416-425.
- Guimarães, I. C., Cação, O., Coutinho, V. (2013). Da interação à colaboração em comunidades e fóruns de discussão. In: *Internet Latent Corpus Journal*. v.3, n.1.
- Gomes, A. F. (2012). Eu-Tu: O emprego da classificação automática de mensagens em fóruns eletrônicos de discussões para análise do processo de ensino e aprendizagem centrado em interações. Dissertação de Mestrado - Universidade Federal do Rio de Janeiro.
- Haddoud, M., Mokhtari, A., Lecroq, T., & Abdeddaïm, S. (2016). Combining supervised term-weighting metrics for SVM text classification with extended term representation. In: *Knowledge and Information Systems*, 49(3), 909-931.
- Larocca N. J., Santos, A. D., Kaestner, A. A. C, Freitas, A. A. (2000). Generating Text Summaries through the Relative Importance of Topics. In: *the Proceedings of the International Joint Conference IBERAMIA/SBIA*, Atibaia, SP.
- Murphy, E. (2004). Recognising and promoting collaboration in an online asynchronous discussion. In: *British Journal of Educational Technology*, v.35, n.4, p. 421-431.
- Silva, J. K., Bastos, H. P. P., Bercht, M., Wives, L. K. (2012). Automatização do Processo de Identificação de Presença Social em Fóruns e Chats. In: Simpósio Brasileiro de Informática na Educação.
- Silva, W. D. C. M., Finger, M. (2013). Improving CoGrOO: the Brazilian Portuguese Grammar Checker. In: *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, pág. 21-29.
- Tortoreli, A. C., Gasparin, J. L. (2012). A interação do professor e alunos no ambiente virtual de aprendizagem: a ferramenta assíncrona fórum. In: XVI Encontro Nacional de Didática e Práticas de Ensino (ENDIPE), p. 23-34.