

Mineração de Texto em Moocs: Análise da Relevância Temática de Postagens em Fóruns de Discussão

Vanessa Faria de Souza – PPGIE/UFRGS – vanessa.souza@ibiruba.ifrs.edu.br
Gabriela Trindade Perry – PPGIE/UFRGS – gabriela.perry@ufrgs.br

Resumo. Os MOOCs estão em gradativa evolução, o que ocorre devido à grande disseminação dos ambientes virtuais de aprendizagem, os quais disponibilizam meios de interação aos participantes, um deles é o fórum de discussão, que possui muitas informações a respeito do engajamento dos alunos. Contudo, realizar a leitura de todas as postagens é uma tarefa difícil, pois os MOOCs costumam ter uma faixa muito alta de alunos matriculados. Neste sentido, a mineração de texto pode auxiliar professores a obter conhecimentos relevantes sobre as postagens dos alunos. Desta forma, neste estudo foi realizada uma mineração de textos utilizando grafos, das postagens dos alunos em fóruns de discussão de dois MOOCs da plataforma Lúmina, um no qual o professor ofertante interage com os alunos e outro no qual o professor não interage. Com o intuito de verificar a relevância de tais postagens e observar o comprometimento dos alunos com o curso. Para tanto, foi utilizada a ferramenta Sobek para geração dos grafos e foi calculado o coeficiente de relevância temática (CRT) para cada fórum de discussão. Os resultados apontam que mediações nos fóruns de discussão, por meio de interações de tutores e/ou professores causam impacto significativo na relevância das postagens dos alunos.

Palavras-chave: mineração de texto, grafos, fóruns, MOOCs, coeficiente de relevância temática.

TEXT MINING IN MOOCs: ANALYSIS OF THEMATIC RELEVANCE IN DISCUSSION FORUMS

Abstract. MOOCs are gradually evolving, which is due to the widespread dissemination of virtual learning environments, which provide participants with a means of interaction, one of which is the discussion forum, which has a lot of information about student engagement. However, reading all posts is a difficult task, as MOOCs often have a very high range of students enrolled. In this sense, text mining can help teachers gain relevant knowledge about student posts. Thus, in this study a text mining was performed using graphs, from the students' posts in discussion forums of two Lúmina platform MOOCs, one in which the offering teacher interacts with the students and another in which the teacher does not interact. In order to verify the relevance of such posts and observe the students' commitment to the course. For this, the Sobek graph generation tool was used and the thematic relevance coefficient (TRC) was calculated for each discussion forum. The results indicate that mediations in discussion forums, through interactions of tutors and teachers, have a significant impact on the relevance of students' posts.

Keywords: text mining, graphs, forums, MOOCs, thematic relevance coefficient.

1. Introdução

Uma evidência do impacto da globalização na educação tem sido a proliferação dos *Massive Open Online Courses* (MOOCs) (Sanchez-Gordon e Luján-Mora, 2014). As novas tecnologias no campo da educação facilitam a abertura de cursos para um número muito maior de alunos do que o oferecido no ambiente tradicional. No entanto, a

implementação e administração dessas plataformas geram desafios importantes que devem ser resolvidos para que os alunos possam tirar proveito de todos os recursos ao máximo (Alisson et al., 2012).

A variedade de recursos de aprendizagem oferecidos pelos MOOCs, juntamente com o surgimento das mídias sociais, contribuiu para a criação de espaços para a interação aluno/professor e aluno/aluno. Essa interação propicia uma grande quantidade de dados que evidenciam o comportamento de aprendizagem e deixa rastros do processo educacional, os mesmos são úteis para a avaliação da aprendizagem (Paltoglou e Thelwall, 2012). O fórum de discussão, neste contexto, é um recurso que permite aos sujeitos participantes discutir um determinado assunto e trocar ideias.

Segundo Sánchez (2005), o fórum de discussão para fins educacionais em um ambiente online 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, dentre 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, por meio 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, quase que inviável para cursos MOOC. Desta forma, para que seja possível que o professor possa analisar todas as respostas dos alunos em fóruns, e demais ambientes de interação, o emprego de métodos computacionais pode ser de grande valia.

Neste contexto, a interação entre estudantes nos MOOCs, dada por diferentes mídias digitais, pode ser explorada usando várias técnicas de mineração de texto, como mineração de opinião (OM), análise de sentimentos (SA) e análise de subjetividade ou orientação emocional. Com o objetivo de melhorar o aprendizado, personalizar a interação do aluno no curso e/ou verificar o quanto o aluno está comprometido com a realização deste curso (Shatnawi, Medhat e Coceia, 2014), o que pode fornecer indícios de alunos pretensos a desistência, assim o professor pode realizar ações motivando a conclusão.

Desta forma, este trabalho teve como objetivo realizar a mineração de textos, produzidos pelos alunos em fóruns de discussão de MOOCs, da plataforma Lúmina pertencente a Universidade do Rio Grande do Sul (UFRGS). Pretende-se, com essa pesquisa, além de proporcionar o conhecimento de ferramentas de mineração para professores/tutores de curso MOOC, analisar a relevância dos textos dos fóruns de dois MOOCs disponíveis na plataforma.

Os dois MOOCs analisados diferem no formato de seus fóruns, em um deles os professores/tutores ofertantes interagem com os alunos, enquanto que no outro o professor não interage. Intenciona-se, deste modo investigar se a interação com o professor leva os alunos a postarem comentários mais relevantes a partir do tema proposto. Cabe evidenciar, que em cursos MOOCs não há o processo de tutoria, contudo há cursos em que o professor se propõe a interagir com os alunos matriculados com o intuito melhorar o processo de aprendizagem.

Este artigo apresenta um estudo com a técnica de mineração de textos utilizando grafos que foi realizada, por meio da ferramenta Sobek (LORENZATTI, 2007) e para direcionar o processo de análise das postagens foi aplicado o coeficiente de relevância temática (CRT) (AZEVEDO et al., 2009) de uma contribuição textual.

2. Trabalhos Relacionados

Esta seção apresenta algumas pesquisas realizadas com a aplicação de técnicas de mineração de textos para analisar fóruns de discussão e demais ambientes de interação textual em MOOCs.

No trabalho de Hsu et al. (2017) foi realizado um estudo sobre mineração de texto para detecção de erros em fóruns de MOOCs baseado em processamento de linguagem natural. Eles salientam que o fórum é o caminho direto para os alunos fazerem perguntas e esclarecerem suas dúvidas e erros. Desta forma, os autores utilizaram os textos sem formatação de postagens e respostas no fórum de MOOCs para extrair palavras-chave de conceitos errôneos para professores/tutores, com base na técnica de processamento de linguagem natural. Neste estudo foram pesquisados tais conceitos errôneos em mais de 120 mil palavras chinesas e 15 mil palavras em inglês em um curso. Os autores puderam observar que as nuvens de palavras são válidas na identificação de conceitos errôneos e suas correlações.

O trabalho desenvolvido por Yu e Zheng (2017) também aplicou o processamento de linguagem natural para extrair informações relevantes de interações textuais em MOOCs. No artigo foi implementada uma combinação da tecnologia de mineração de dados e processamento de linguagem natural e analisou-se o comportamento e o conteúdo de aprendizagem dos alunos nos itens interativos dos MOOCs, com vistas a identificar interesses, dificuldades, tendências e avaliar seu efeito nas atividades concluídas, por meio da interação entre professores/alunos e alunos/alunos. Os autores salientam que em posse dos resultados os professores podem otimizar a estrutura do curso, estabelecer mecanismos de incentivo para estimular o desejo de aprendizado dos alunos e assim melhorar a taxa de conclusão.

De forma semelhante, Buenaño-Fernández, Luján-Mora e Villegas-Ch (2017), propuseram uma melhoria em MOOCs por meio de mineração de texto de e-mails dos alunos. O objetivo deste estudo foi o de analisar as opiniões dos alunos sobre seus cursos, instrutores e as principais ferramentas usadas. A pesquisa enfoca o cálculo e a análise da frequência dos termos, a análise de concordâncias e agrupamentos. O estudo de caso usado neste artigo é um MOOC sobre o tema de “Desenvolvimento Web” com mais de 40.000 estudantes matriculados. Foi explorado pelos autores a opinião desses alunos que forneceram informações úteis para um feedback quanto ao curso realizado e que pode ser posteriormente generalizado.

Os mesmos autores também desenvolveram uma pesquisa (Buenaño-Fernández, Villegas-Ch, e Luján-Mora, 2018) empregando mineração de texto para avaliar a interação de alunos em Ambientes Virtuais de Aprendizagem (AVAs). Neste trabalho eles exploraram as interações dos alunos em diferentes ambientes em que postam textos em MOOCs. A pesquisa enfoca o cálculo e a análise da frequência dos termos, a análise de concordâncias e agrupamentos. O presente estudo concentrou-se em duas áreas: Primeiramente, a análise da frequência dos termos, que forneceu alguns resultados interessantes como ponto de partida para melhorar a aplicação de técnicas de análise linguística e semântica em mensagens de e-mail. Em segundo lugar, a proposta de um modelo preditivo da polaridade de sentimentos para os documentos compilados.

Guo et al. (2019) implementaram em seu trabalho uma rede neural híbrida com caracteres de palavras baseadas em atenção, com informações semânticas e estruturais para identificação de postagens urgentes nos fóruns de discussão de um MOOC. Nesta pesquisa os autores, apresentam uma nova rede neural híbrida para identificar postagens que requerem atenção imediata de professores/tutores, os autores definem como “urgentes” as mensagens que podem determinar se este aluno irá permanecer ou não no curso. Este trabalho pode ajudar os professores a priorizar suas respostas e gerenciar melhor várias postagens. Por consequência, os professores podem responder às perguntas dos alunos em tempo hábil e ajudar a reduzir as taxas de evasão.

3. Mineração de Texto Utilizando Grafos

Antes de definir a mineração de texto utilizando grafos, cabe explicar o processo como um todo. A mineração de textos busca 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.

Neste cenário, uma técnica bastante comum utilizada para representar um documento e identificar as características do mesmo, por meio de um modelo de espaço vetorial, é a mineração de texto utilizando grafos. Nesta técnica, cada termo do documento torna-se uma característica dimensional. O valor de cada dimensão pode indicar o número de vezes que o termo aparece no texto, ou pode indicar o peso do termo a ser considerado, como por exemplo, a quantidade de documentos nos quais o termo aparece. No entanto, essa técnica descarta informações importantes, como por exemplo, a ordem nos quais os termos aparecem, onde os mesmos aparecem, e a proximidade entre eles (Schenker, 2003).

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 em MOOCs. Considerando-se o grafo gerado a partir de uma mensagem, pode-se avaliar se ela se refere 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.

4. Procedimentos Metodológicos

Essa pesquisa consistiu em um estudo de dois MOOCs ofertados na plataforma Lúmina, que pertence à UFRGS, os cursos disponibilizam vários recursos de aprendizagem para os alunos, como textos, vídeo aulas, sugestões de vídeos, referências (textos, livros, artigos), questionários, estudos dirigidos e o fórum de discussão, na Tabela 01 pode ser observada uma breve descrição de cada curso.

Tabela 01 – Descrição dos Cursos

No	CURSO	DESCRIÇÃO	RECURSOS
1	Alimentação Saudável na Escola	O curso tem como objetivo proporcionar conhecimento e atualização sobre a promoção da alimentação saudável no ambiente escolar e sensibilizar profissionais e estudantes de diferentes regiões brasileiras, tanto na área de Nutrição e de Educação, demais interessados dentro e fora da comunidade escolar.	Possui três módulos, em cada um destes há uma videoaula gravada pelo professor, vídeos de plataformas web que complementam os estudos, textos em PDF e referências para outros textos importantes, e um fórum de discussão. Possui um questionário para traçar o perfil dos estudantes, materiais complementares, e um questionário final.
2	Esportes e Atividades ao Ar Livre	Pretende-se, com a realização deste curso, sensibilizar os participantes em relação à abrangência dos esportes e atividades ao ar livre; relevância da educação ambiental ao ar livre; importância da gestão de riscos para a prática de esportes e atividades ao ar livre e às possibilidades de práticas, inclusive de docência.	Possui quatro módulos, em cada um destes há uma videoaula gravada pelo professor, vídeos de plataformas web que complementam os estudos, textos em PDF e referências para outros textos importantes, e um fórum de discussão. Exceto no último módulo que possui apenas as referências e o fórum. Possui um questionário para traçar o perfil dos estudantes e um questionário final.

Esses dois cursos disponibilizam fóruns de interação entre os estudantes, todavia no primeiro o professor não interage com os alunos, enquanto que no segundo os professores ofertantes do curso e tutores interagem em todos os fóruns de discussão. Diante dessa informação formulou-se a questão de pesquisa que norteará este estudo: **Questão de Pesquisa:** A interação do professor em um fórum de discussão influencia na relevância das postagens dos alunos em cursos MOOC?

Para responder essa questão foram empregados alguns procedimentos que configuram a metodologia adotada nesta pesquisa, que em termos gerais resume-se em:

- Foram minerados os textos de um fórum de cada curso, com o objetivo de analisar a relevância das postagens dos alunos;
- Para analisar as contribuições textuais produzidas nos fóruns, foi utilizado um software de mineração de textos, chamado Sobek (LORENZATTI, 2007);
- A partir dos grafos gerados pelo software, foi realizado manualmente o cálculo do coeficiente de relevância temática (CRT) (Azevedo et al., 2009) de uma contribuição textual, cujo objetivo é analisar o quanto um texto é relevante para um determinado tema de discussão.

Para maior detalhamento da realização deste estudo, os procedimentos realizados foram divididos em etapas que podem ser visualizadas na Tabela 02.

Tabela 02 – Etapas da Realização do Estudo

ETAPA	DESCRIÇÃO
1	Escolha do fórum de discussão: de cada curso foi escolhido um fórum envolvendo a temática central do mesmo
2	Coleta das contribuições textuais produzidas pelos alunos no fórum e limpeza do texto - exclusão de itens que não colaboram na mineração
3	Definição dos conceitos importantes relativos ao tema do fórum, bem como a definição da associação entre estes conceitos
4	Definição de possíveis termos que podem ser considerados como equivalentes semânticos aos conceitos envolvidos
5	Execução do processo de mineração, para cada fórum e geração do grafo com os conceitos envolvidos
6	Cálculo do coeficiente de relevância temática (CRT) de cada um dos fóruns, a partir dos grafos
7	Definição do fórum com as postagens mais relevantes para o contexto da discussão

Desta forma, após a formulação da metodologia do estudo, levantaram-se duas hipóteses que foram validadas: **Hipótese nula, ou H₀**: A relevância das postagens no fórum do curso “Esportes e Atividades ao Ar Livre” é menor ou igual a das postagens do curso “Alimentação Saudável na Escola”. **Hipótese alternativa, ou H₁**: A relevância das postagens no fórum do curso “Esportes e Atividades ao Ar Livre” é maior do que a do curso “Alimentação Saudável na Escola”.

5. Resultados

A realização do cálculo da CRT (Azevedo et al., 2009) envolve a mineração do texto correspondente a cada um dos fóruns – os quais tem suas características apresentadas na Tabela 03 – e a geração do grafo equivalente. Para tanto foi necessário efetuar a coleta das contribuições textuais produzidas pelos alunos no fórum e a limpeza do texto, que tem o objetivo de deixar o texto apenas com os itens que devem ser minerados. Tal tarefa foi elaborada da seguinte forma, todas as postagens dos fóruns foram copiadas para um editor de texto, o qual possibilitou a retirada dos seguintes termos:

1. Saudações, como: boa tarde, bom dia, boa noite, olá, dentre outras;
2. Agradecimentos, e demais variações;
3. Postagens de apenas uma palavra como “concordo”, “boa”, “ótimo”, dentre outras;
4. Nomes dos alunos;
5. Todas as postagens dos professores e tutores do curso Esportes e Atividades ao Ar Livre, permanecendo apenas as postagens dos alunos.

Tabela 03 – Descrição dos Fóruns

CURSO	FÓRUM	TEMA	Nº DE POSTAGENS
Alimentação Saudável na Escola	Fórum 3	Comente de que maneira você pode (ou poderia) contribuir com a educação alimentar e nutricional no ambiente escolar?	1.573
Esportes e Atividades ao Ar Livre	Fórum 1	Que características dos esportes e atividades ao ar livre podem estar relacionadas ao crescente interesse da população nessas práticas? Quais os benefícios que podem ser gerados pela prática desses esportes e atividades ao ar livre? Quais aspectos devem ser levados em consideração para a promoção de iniciativas focadas em esportes e atividades ao ar livre?	753

Após essa etapa, procedeu-se a delimitação dos conceitos mais importantes relativos ao tema do fórum, bem como a definição da associação entre estes conceitos. Para efetuar essa etapa recorreu-se a pesquisa apurada nos materiais disponibilizados para os alunos realizarem cada um dos fóruns, foi então implementada a extração dos principais termos desses materiais utilizando a ferramenta Sobek (LORENZATTI, 2007), desta forma foi possível chegar aos seguintes conceitos e associações, observados na Tabela 04.

Tabela 04 – Conceitos e Associações

CURSO	CONCEITOS IMPORTANTES/ASSOCIAÇÕES
Alimentação Saudável na Escola:	“Educação”, “Alunos”, “Alimentar” “Alimentação”, “Alimentos”, “Saudáveis”, “Saúde”, “Hortas”, “Conscientização”, “Frutas” e “Verduras”. As associações relevantes entre os conceitos ficaram estabelecidas da seguinte maneira: O conceito “Conscientização” está relacionado com “Alunos”, “Educação” e “Alimentar”, “Alimentos” e “Saudáveis”, “Alimentação” e “Saudáveis” também estão relacionados, assim como com “Alunos”, “Hortas”, “Frutas” e “Verduras”.
Esportes e Atividades ao Ar Livre	“Saúde”, “Mental”, “Contato”, “Natureza”, “Bem”, “Estar”, “Ar Livre”, “Preservação”, “Prática”, “Esportes”. As associações relevantes entre os conceitos ficaram estabelecidas da seguinte maneira: Os conceitos “Natureza”, “Ar Livre”, “Bem” e “Estar” estão relacionados, e também “Prática”, “Esportes”, “Saúde” e “Mental”. Os termos mental e psicológico foram definidos como equivalentes semânticos.

Depois da limpeza do texto foi possível realizar a geração dos grafos, esse procedimento incluiu importar o texto em PDF, já limpo, para o software Sobek (LORENZATTI, 2007), o processamento durou poucos minutos, destaca-se que todas as postagens dos alunos estavam contidas em um único documento gerado para cada fórum de discussão – Alimentação Saudável na Escola¹ e Esportes ao Ar Livre². Na Figura 01 pode ser observado o grafo correspondente ao texto do fórum Alimentação Saudável na Escola.

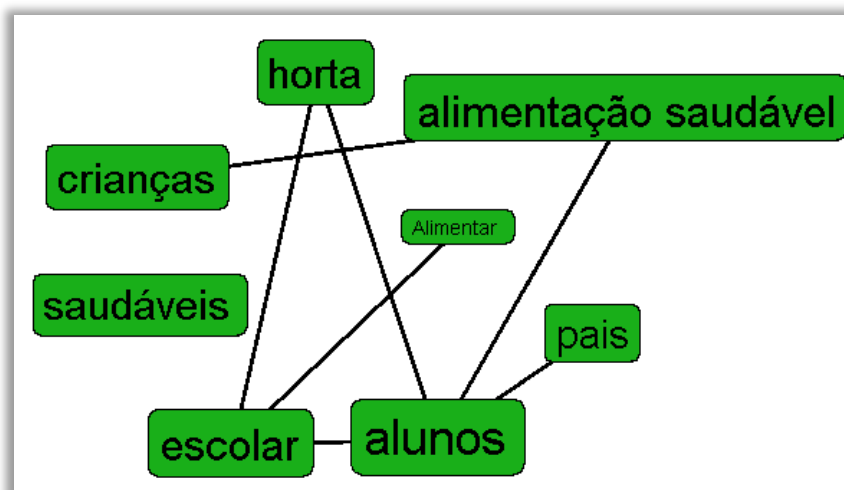


Figura 01 – Grafo alimentação saudável na escola

Na figura 02 está representado o grafo dos termos minerados do curso Esportes e Atividades ao Ar Livre.

¹ https://docs.google.com/document/d/1wl62xzEhJ7BFfcyqyalDF_Y8Ww1lMbj3t_csx7mefXI/edit

² <https://docs.google.com/document/d/1w0J3jEiuyc1Ohj2LeNmNUyGLnGC-kUiK9xjNndXuurM/edit>

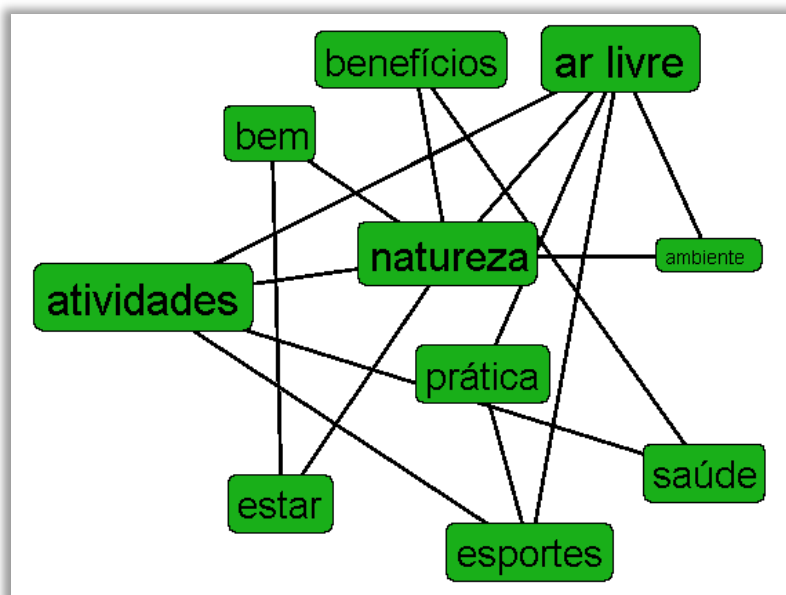


Figura 02 – Grafo esportes e atividades ao ar livre

Por fim, prossegue-se ao cálculo do coeficiente de relevância temática (CRT) de uma contribuição textual, definido por Azevedo et al. (2009), cujo objetivo é analisar o quanto um texto é relevante a um determinado tema de discussão. O coeficiente de relevância temática indica o grau de relevância da contribuição textual de acordo com o tema do fórum. O coeficiente de relevância pode ser calculado a partir da seguinte fórmula:

$$\text{CRT} = \text{NC} + \text{NA}$$

- NC: número de conceitos relevantes utilizados no texto
- NA: número de associações entre os conceitos relevantes utilizados no texto

Para o cálculo do NC, os termos considerados como equivalentes semânticos também são considerados como conceitos relevantes. Os prefixos, sufixos, e plurais dos conceitos relevantes citados no texto são convertidos, para também serem considerados como conceitos relevantes. Com o CRT, os grafos que possuem mais conceitos importantes relativos ao tema, com mais associações entre os conceitos, possuem maior coeficiente de relevância temática. O cálculo do CRT de cada um dos fóruns está expresso na Tabela 05.

Tabela 05 – Cálculo do CRT

CURSO	NC	NA	CRT
Alimentação Saudável na Escola	Horta Alimentação Saudável Saudáveis	Alimentos Saudáveis com Alunos	4
	Total = 3	Total = 1	
Esportes e Atividades ao Ar Livre	Saúde Natureza Bem-Estar Ar Livre Prática Esportes	Natureza com Ar livre / e com Bem-Estar Prática com Esportes	9
	Total = 6	Total = 3	

Como pode ser observado nos resultados apresentados do CRT a Hipótese alternativa (H) foi validada como sendo verdadeira, o que indica que mediações em fóruns de discussão, por meio de interações de tutores e/ou professores, causam impacto positivo na relevância das postagens dos alunos.

6. Considerações Finais

Este trabalho realizou um processo de mineração de texto, utilizando grafos, de dois fóruns de MOOCs disponíveis na plataforma Lúmina da UFRGS, os cursos tinham assuntos diversos, um voltado alimentação no ambiente escolar e outro a esportes ao ar livre. Todavia, os aspectos que chamaram a atenção para suas escolhas foram: no primeiro os fóruns eram realizados sem mediação, ou seja, apenas os alunos interagem entre si. Enquanto no segundo havia sempre um tutor intermediando os comentários dos alunos e também o professor do curso fazia postagens seguidamente, corroborando com as discussões.

Desta forma, a partir dos grafos gerados com a mineração do texto de cada fórum, efetuou-se cálculo do coeficiente de relevância temática, que indica o grau de relevância da contribuição textual, de acordo com o tema proposto. Buscou-se com a obtenção desse coeficiente, verificar se a interação com tutores e professores ofertantes do curso aumenta o envolvimento dos alunos sobre a temática em discussão.

Para tal, foram levantadas duas hipóteses: **Hipótese nula, ou H₀** – A relevância das postagens no fórum do curso “Esportes e Atividades ao Ar Livre” é menor ou igual a das postagens do curso “Alimentação Saudável na Escola” e **Hipótese alternativa, ou H** – A relevância das postagens no fórum do curso “Esportes e Atividades ao Ar Livre” é maior do que a do curso “Alimentação Saudável na Escola”. Por meio, do cálculo do CRT, foi possível observar que a Hipótese alternativa é verdadeira, pois o coeficiente de Esportes e Atividades ao Ar Livre foi maior. Isso também foi observado no decorrer do processo de limpeza das postagens dos fóruns.

No curso onde ocorreu mediação, o tutor e professor estavam em constante interação com os alunos, complementando suas respostas e fomentando maiores discussões, para não haver perda de foco, estes também indicavam materiais de referência para que os alunos pesquisassem e melhorassem suas postagens. Contudo, no curso em que não ocorreu intermediação, foi possível constatar que muitas postagens eram apenas com palavras do tipo, “concordo”, “sim”, “gostei”, que não ampliavam ou aprofundavam as discussões, e por muitas vezes os alunos fugiam do foco das discussões, colocando argumentos que não faziam parte do tema proposto.

Essas afirmações foram confirmadas, pois para assegurar o resultado do CRT total obtido, foram também realizados os cálculos de forma individual para dez alunos escolhidos aleatoriamente que participaram de cada fórum, e nesses cálculos também o número do coeficiente ficou maior para aqueles que participaram do curso Esportes e Atividade ao Ar Livre. Mesmo, na observação dos grafos gerados, pode-se perceber uma maior quantidade de termos e complexidade de relacionamentos nos conceitos minerados neste MOOC.

Como trabalho futuro, pretende-se verificar se a interação de tutores e professores nos fóruns causam maiores conclusões de alunos nos cursos, desta forma analisar se esse tipo de ação tem relação com a desistência e permanência de alunos em MOOCs.

Referências

ALLISON, C., MILLER, A., OLIVER, I., MICHAELSON, R., THANASSIS, T. The web in education. **Computer Networks**, vol. 56, no. 18, pp. 3811–3824, 2012.

AZEVEDO, B. F. T.; BEHAR, P. A.; REATEGUI, E. B. Estudo de análise qualitativa em fórum de discussão. **Renote - Revista Novas Tecnologias na Educação**. Porto Alegre, v. 7, n. 3, p. 1-10, 2009.

V. 17 N° 3, dezembro, 2019

RENOTE

DOI:

BUENAÑO-FERNÁNDEZ, D., LUJÁN-MORA, S., VILLEGAS-CH, W. Improvement of massive open online courses by text mining of students' emails: a case study. In: **Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality**, Cádiz, Spain, 2017, p. 1-7.

BUENAÑO-FERNÁNDEZ, D., VILLEGAS-CH, W., LUJÁN-MORA, S. Using text mining to evaluate student interaction in virtual learning environments. In: **World Engineering Education Conference (EDUNINE)**, Buenos Aires, Argentina, 2018, p. 1-6.

FELDMAN, R.; SANGER, J. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge, MA: **Cambridge University Press**, 2007.

GUO, S. X., SUN, X., WANG S. X., GAO, Y., FENG, J. Attention-Based Character-Word Hybrid Neural Networks With Semantic and Structural Information for Identifying of Urgent Posts in MOOC Discussion Forums. In: **IEEE Access**, v. 7, p. 120522 – 120532, 2019.

HSU, H. H., HUANG, N. F., CHEN, S. C., LEE, C. A., TZENG, J. W. Misconceptions mining and visualizations for Chinese-based MOOCs forum based on NLP. In: **2nd International Conference on Big Data Analysis (ICBDA)**, Beijing, China, 2017, p. 634-639.

LORENZATTI, A. **SOBEK: uma Ferramenta de Mineração de Textos**. 2007. Trabalho de Conclusão – Departamento de Informática, Universidade de Caxias do Sul, 2007, Caxias do Sul, BR-RS.

PALLOFF, R. M.; PRATT, K. *O aluno virtual: um guia para trabalhar com estudantes on-line*. Porto Alegre: **Artmed**, 2004.

PALTOGLOU G., THELWALL, M. Twitter, MySpace, Digg: Unsupervised Sentiment Analysis. **Social Media, in ACM Transactions on Intelligent Systems and Technology**, vol. 3, no. 4, 2012.

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 set. 2019.

SANCHEZ-GORDON, S., LUJÁN-MORA, S. MOOCs gone wild. In: **Proceedings of the 8th International Technology, Education and Development Conference (INTED 2014)**, 2014, p. 1449-1458.

SCHENKER, A. **Graph-Theoretic Techniques for Web Content Mining**. Florida: University of South Florida, 2003. PhD Thesis.

SHATNAWI,S., MEDHAT, M., COCEA, M. Text stream mining for Massive Open Online Courses: review and perspectives. **Systems Science & Control Engineering**, vol. 2, no. 1, pp. 664-676, 2014.

TAN, A. Text Mining: The State of the Art and the Challenges. In: **Workshop on Knowledge Discovery from Advanced Databases (PKDAD'99)**, Beijing. Proceedings. Beijing: 1999, p.71-76.

YU, F., ZHENG, D. Education data mining: How to mine interactive text in MOOCs using natural language process. In: **12th International Conference on Computer Science and Education (ICCSE)**, Houston, TX, USA, 2017, p. 694-699.