

Aplicando Mineração de Textos na análise de artigos científicos sobre Sistemas de Recomendação com Mineração de Dados Educacionais e *Learning Analytics*

Fábio Josende Paz, PGIE - UFRGS e URCAMP, Fabiopaz@urcamp.edu.br
Silvio César Cazella, PGIE - UFRGS e UFCSPA, silvio.cazella@gmail.com

Resumo: A mineração de textos apresenta-se como uma alternativa para análise de grandes volumes de textos. Nesse contexto, este artigo tem como objetivo apresentar os resultados da extração de conhecimento obtido pela aplicação da mineração de textos em um *corpus* composto por artigos científicos relacionados a Sistemas de Recomendação com Mineração de dados Educacionais (MDE) e *Learning Analytics* (LA), buscando descrever os assuntos abordados e principais autores das áreas. Foi aplicado o processo de Descoberta de Conhecimento em Textos (DCT) no corpus analisado, sendo possível por fim evidenciar as relações extraídas pela mineração de textos.

Palavras chaves: Descoberta de Conhecimento em Textos, Extração de conhecimento, Mineração de textos.

Applying Text Mining in the analysis of scientific articles on Recommendation Systems with Educational Data Mining and Learning Analytics

Abstract: *The Text mining presents itself as an excellent alternative for the analysis of large volumes of texts. In this context, this article aims to present the results of the extraction of knowledge obtained through the application of text mining in a corpus composed of scientific articles related to Recommendation Systems with Educational Data Mining (MDE) and Learning Analytics (LA). describe the issues addressed and main authors of the areas. It was applied the process of Knowledge Discovery on Texts (DCT) in the analyzed corpus, being possible to finally highlight the relationships extracted by the mining of texts.*

Keywords: *Knowledge Discovery on Texts, Knowledge Extraction, Text Mining.*

1 Introdução

Analisar conteúdo de textos não estruturados ou semiestruturados constitui-se em uma tarefa árdua. Uma área de pesquisa da ciência da computação busca auxiliar nesse processo de análise, esta se denomina Mineração de textos (VIJAYARANI *et al.*, 2015). Esta constitui-se em uma etapa do Processo de Descoberta de Conhecimento em textos (DCT). Este processo contribui no processo de busca e extração de conhecimento útil em textos não estruturados ou semiestruturados. Segundo Romero e Ventura (2007), o aumento do uso da mineração de textos no campo educacional é devido às possibilidades que a mesma proporciona conforme o objetivo de quem está fazendo uso de suas práticas.

Estudos científicos publicados (BULEGON e MORO, 2010; PARREÑO *et al.*, 2016; KRASSMANN *et al.*, 2017) mostram que o processo de Descoberta de Conhecimento em Textos (DCT), juntamente com a etapa de Mineração de textos e suas técnicas de análise são produtivos para a extração automática de conhecimento de grandes coleções de textos semi ou não-estruturados. Ainda segundo Konchady (2006) a Mineração de Texto consiste em explorar e identificar termos relevantes e identificar padrões textuais, além de analisar a frequência destes termos no domínio, corroborando Vijayarani *et al.*

(2015) acrescentam que as técnicas utilizadas na etapa de Mineração de Texto podem ser utilizadas em diversos tipos de domínios de pesquisa, como processamento de linguagem natural, recuperação de informações, classificação e agrupamento de texto.

Nesse contexto, este artigo tem como objetivo apresentar os resultados da extração de conhecimento obtido pela aplicação da mineração de textos em um *corpus* composto por artigos científicos relacionados a Sistemas de Recomendação com Mineração de dados Educacionais (MDE) e *Learning Analytics* (LA), buscando descrever os assuntos abordados e principais autores das áreas. Para atingir o objetivo, o *corpus* foi analisado através de ferramentas de Mineração de Texto que integra o DCT. Na Figura 1, visualizam-se as etapas do processo de Descoberta de Conhecimento em Textos (DCT).



Figura 1: Etapas do processo DCT

Fonte: Adaptado de Tan (1999) e Vijayarani *et al.* (2015)

A etapa de Coleta, consiste na formação da base de documentos ou Corpus para o estudo, o pré-processamento ou limpeza dos dados tem o intuito de remover ruídos e preparar os dados para a mineração, ou seja, retirar caracteres indesejados, correção ortográfica e morfológica, também a realização da análise semântica e normalização dos vocabulários, na etapa da Mineração são aplicados métodos ou tarefas de mineração através de algoritmos específicos e por fim a interpretação dos resultados obtidos (TAN, 1999; SCHIESSL E BRASCHER, 2011).

Algumas pesquisas que aplicaram a Mineração de Textos têm apresentado resultados interessantes, como por exemplo, o trabalho de Krassmann *et al.* (2017) no qual a aplicação de técnicas de Mineração de Textos demonstrou ser eficiente para analisar um *corpus* de textos acadêmicos. No estudo de Parreño *et al.* (2016) foi realizada uma análise bibliométrica e posteriormente aplicadas técnicas de Mineração de Textos a fim de buscar informações atualizadas sobre o estado da arte no uso de jogos na educação, este estudo permitiu a identificação rápida dos principais autores e instituições relevantes, além de temas e tendências importantes no desenvolvimento do conhecimento deste domínio.

Este artigo está estruturado em seis seções, incluídas a introdução e a conclusão. Na seção 2 é apresentado uma breve descrição sobre Sistemas de Recomendação educacionais. Na seção 3, os materiais e métodos são apresentados, na sequência na seção 4, é aplicado o processo de Descoberta de Conhecimento em Textos (DCT), e a seção 5 apresenta a análise dos resultados e discussão. Por fim apresentam-se as referências deste estudo.

2 Sistemas de Recomendação Educacionais

Identificar e recomendar materiais personalizados para cada estudante, baseando-se em suas necessidades, interesses e competência a serem desenvolvidas, é um grande desafio (CAZELLA *et al.*, 2009). Nesse sentido, os Sistemas de Recomendação são considerados ferramentas de grande utilidade, principalmente para usuários de sistemas online, pois os mesmos conseguem lidar adequadamente com grandes bases de

informações. (RICCI et al., 2011), já os Sistemas de Recomendação Educacionais (SRE) têm a finalidade de produzir recomendações para seus usuários geralmente alunos e professores (LIU et al., 2013; IMIRAN et al., 2014; FULANTELLI et al., 2015). Ainda, podem ser usados provendo materiais didáticos e informações úteis que considerem as preferências dos estudantes e ajudando-os a alcançar os objetivos de aprendizagem do curso (LIU et al., 2013; COSTA et al., 2013).

Nessa perspectiva, o processo de ensino e aprendizagem é o objeto central do sistema, pois, a partir de uma modelagem, o sistema é capaz de filtrar recursos educacionais, fazer sugestões, personalizar conteúdo, com o objetivo de proporcionar ao aluno uma experiência melhor de aprendizado, assim, o processo educativo torna-se mais interativo, dinâmico e inovador. (COSTA, AGUIAR, MAGALHÃES, 2013)

Diversas pesquisas têm sido realizadas sobre a temática, investigando técnicas de recomendação que melhor se adaptam às necessidades de estudantes e professores nos processos de ensino e aprendizagem (COSTA et al., 2013), algumas delas são as técnicas de *Learning Analytics* (LA) e de Mineração de Dados Educacionais (MDE).

A MDE tem como objetivo desenvolver métodos para explorar conjuntos de dados coletados em ambientes educacionais e possui um grande potencial para melhorar a qualidade de ensino (BAKER et al., 2011), pode ser utilizado para recomendações pedagógicas (Paiva et al., 2013; França e Amaral, 2013). *Learning Analytics* (LA) é a medição, coleta, análise e interpretação dos dados produzidos nos ambientes educacionais, permitindo avaliar o progresso acadêmico, prever o futuro e identificar possíveis problemas. (SIEMENS E LONG, 2011; JOHNSON et al. (2011). Ainda Rigo et al. (2014) afirmam que *Learning Analytics* (LA) pode auxiliar no acompanhamento e visualização de diversos aspectos fundamentais no processo de ensino e aprendizagem, nesse contexto, alguns estudos utilizaram a análise automática dos dados denominada *Learning Analytics* (Arnold, 2010; Rigo et al., 2014).

Diante desse contexto, Baker e Yacef (2009) afirmam que a MDE está relacionada com o desenvolvimento de métodos para análise de dados de aprendizagem, enquanto a LA está relacionada com o uso desses dados para a melhoria do processo de aprendizagem.

3 Materiais e Métodos

Esta pesquisa caracteriza-se por ser uma pesquisa exploratória de acordo com seu objetivo, de natureza aplicada, com abordagem quantitativa uma vez que métricas são assumidas para avaliar os modelos obtidos. Foram aplicadas técnicas de Mineração de Textos em um corpus, ou seja, em uma coleção de textos. Para obter o corpus desta pesquisa, levou-se em consideração o foco de pesquisa do autor, foram utilizados três termos na composição de uma string de busca: “*Learning Analytics and Educational Data Mining or Recommendation Systems*”, esses termos deveriam aparecer no texto do artigo, o período de 2007 até 2017 foi usado como critério de inclusão, assim como a necessidade de serem artigos revisados por pares e na língua inglesa. A busca por artigos foi realizada no portal de Periódicos da Capes e como resultado da aplicação da string de busca e os critérios, obteve-se 251 artigos. Os critérios de exclusão utilizados foram: Artigos de revisão; artigos duplicados; artigos somente MDE ou LA, artigos fora do contexto da pesquisa. Foram lidos os resumos de todos os artigos para a aplicação destas regras de exclusão, ao final resultaram 10 artigos para a construção do corpus e

aplicação do processo de mineração de textos, os quais estão descritos na Tabela 1.

| ID | Título Artigo | Ano | País de Origem | Periódico | Autores | Número de palavras |
|----|--|------|----------------|--|---|--------------------|
| 1 | A framework to support educational decision making in mobile learning | 2014 | Italia | Computers in Human Behavior | Giovanni Fulantelli, Davide Taibi, Marco Arrigo | 7431 (10,17%) |
| 2 | Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory | 2014 | USA | Computers in Human Behavior | Xing Wanli, Guo Rui, Petakovic Eva, Goggins Sean | 9639 (13,19%) |
| 3 | Design and Implementation of a Learning Analytics Toolkit for Teachers | 2012 | Alemanha | Educational Technology & Society | Anna Lea Dyckhoff, Dennis Zielke, Mareike Bultmann, Mohamed Amine Chatti, Ulrik Schroeder | 10104 (13,82%) |
| 4 | Educational Data Mining and Learning Analytics: Applications to Constructionist Research | 2014 | USA | Technology, Knowledge and Learning | Matthew Berland, Ryan S. Baker, Paulo Blikstein | 7001 (9,58%) |
| 5 | Learning Analytics and Digital Badges: Potential Impact on Student Retention in Higher Education | 2016 | Alemanha | Technology, Knowledge and Learning | Dana-Kristin Mah | 7214 (9,87%) |
| 6 | Student Facing Dashboards: One Size Fits All? | 2017 | USA | Technology, Knowledge and Learning | Stephanie D. Teasley | 2266 (3,10%) |
| 7 | AMOEBAs: Designing for collaboration in computer science classrooms through live learning analytics | 2015 | USA | International Journal of Computer-Supported Collaborative Learning | Matthew Berland, Don Davis, Carmen Petrick Smith | 5778 (7,90%) |
| 8 | Requirements for Semantic Educational Recommender Systems in Formal E-Learning Scenarios | 2011 | Espanha | Algorithms | Olga C. Santos, Jesus G. Boticario | 8670 (11,86%) |
| 9 | Identifying recommendation opportunities for computersupported collaborative environments | 2016 | Espanha | Expert Systems | Jesus L. Lobo, Olga C. Santos, Jesus G. Boticario, Javier Del Ser | 8096 (11,08%) |
| 10 | The effect of recommendation systems on Internet-based learning for different learners: A data mining analysis | 2013 | China / Taiwan | British Journal of Educational Technology | Chen-Chung Liu, Chia-Jung Chang, Jui-Min Tseng | 6895 (9,43%) |

Tabela 1: Artigos selecionados para composição do Corpus

Na Tabela 1, ao analisarmos os metadados dos artigos percebe-se que 80% dos documentos do corpus são oriundos de pesquisas realizadas em países como Estados Unidos, Alemanha e Espanha, o que evidencia que o assunto abordado nesta pesquisa é relevante nestes países. Observa-se também que dois periódicos se destacam com publicações neste tema, são eles *Technology, Knowledge and Learning* e *Computers in Human Behavior*, com 50% dos artigos que compõem o corpus. As informações descritas podem ser relevantes para pesquisadores da área que desejam procurar eventos e periódicos para leitura ou mesmo publicação de suas pesquisas. A distribuição dos artigos no corpus estudado também é apresentada o que demonstra que nenhum artigo representou mais do que 18% de todo o corpus, porém o ID 9 representa apenas 3 % o que não nos permite generalizar os resultados a serem apresentados, na Figura 2 apresenta-se a divisão do corpus de forma visual, adquirida no componente visual *Prospect Viewer* da ferramenta Voyant Tools, identificando a fatia correspondente a cada publicação do corpus.

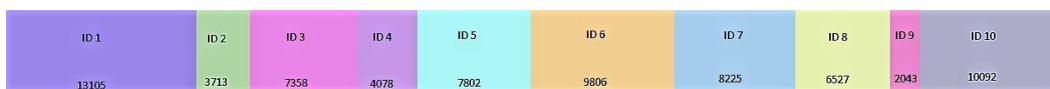


Figura 2: Distribuição dos artigos no corpus

Na Tabela 2, apresentam-se as ferramentas gratuitas utilizadas na fase de processamento, que também indicam o propósito adotado para cada um e onde podem ser encontrados seus resultados neste estudo.

| Ferramenta | Descrição da ferramenta | Objetivo |
|---|---|---|
| Word Counter | Online e gratuito que calcula o número de palavras, frases e termos compostos, entre outros. | Descobrir termos com maior incidência. Tabelas: 3 4 e 5. |
| Voyant Tools (ROCKWELL, SINCLAIR, 2016) | Online e gratuito que permite gerar diversos formatos de gráficos para análise. | Verificar a variação de incidência dos termos, gerar a nuvem de palavras, além da distribuição do corpus. Figuras: 2 3 4 e 5. |
| Sobek (REATEGUI, 2011) | Online e Desktop gratuito, com foco educacional, que gera gráficos com o máximo de palavras relevantes. | Representação gráfica dos termos com maior incidência e suas relações. Figura: 5. |

Tabela 2: Ferramentas utilizadas para a Mineração de textos.

Ainda para analisar os metadados dos artigos, utilizou-se a ferramenta Word Counter para identificar os principais autores citados nesse corpus foi utilizado um arquivo

contendo apenas as referências bibliográficas dos textos, obtendo a relação da quantidade de artigos citados, lembrando que o mesmo artigo pode ter sido citado em mais de um texto, na Tabela 3 apresenta-se os principais autores.

| Ordem | Autor | Quantidade de artigos citados | Domínio de publicação |
|-------|-----------------|-------------------------------|---|
| 1 | Santos O. C. | 19 | Sistemas de Recomendação para a educação |
| 2 | Ventura S. | 18 | DM, MDE, LA Sistemas de recomendação, predição |
| 3 | Boticario J. G. | 17 | Sistemas de recomendação, frameworks, modelos, para a educação. |
| 4 | Romero C. | 15 | DM, MDE, LA Sistemas de recomendação, predição |
| 5 | Baker R. | 15 | MDE, LA, recomendação, predição para educação. |
| 6 | Blikstein P. | 14 | LA, Educação |
| 7 | Siemens G. | 13 | LA, MDE |
| 8 | Berland M. | 11 | LA |
| 9 | Drachslar H. | 11 | LA e sistemas de recomendação |
| 10 | Johnson L. | 11 | Análises na Educação Universitária |
| 11 | Liu C. C. | 9 | Análises na educação |
| 12 | Goggins S. P. | 8 | Interatividade social |
| 13 | Ferguson R. | 8 | LA |
| 14 | Arnold K. E. | 8 | LA |
| 15 | Martin T. | 7 | LA |

Tabela 3: Principais autores do Corpus

A Tabela 3 apresenta os autores mais citados, entre os quais se pode destacar Santos, Boticario, Bliksten e Berland apesar de possuírem uma quantidade significativa de citações, os mesmos concentram-se em apenas 4 artigos do corpus. No entanto, os autores Romero, Ventura e Baker estão presentes em 90% dos artigos o que nos permite concluir que os mesmos são importantes em seus domínios. Ainda apresentam-se os domínios de pesquisa de cada autor o que pode auxiliar para novas pesquisas

4 Processo de Descoberta de Conhecimento em Textos

A seguir será descrita a aplicação detalhada do processo de Mineração de Textos no corpus objeto desta pesquisa.

4.1 Coleta

A primeira etapa do processo DCT é a coleta, neste estudo foram utilizados 10 artigos científicos para compor o corpus da pesquisa, os quais já foram apresentados na Tabela 1, na seção materiais e métodos.

4.2 Pré-Processamento

Nesta etapa para melhorar a técnica de Mineração do Textos, ocorreu à limpeza dos dados contida no corpus com o intuito de preparar os textos para a mineração de textos. Foram retirados dos textos as referências bibliográficas, rodapés, títulos, tabelas, figuras, autores e identificações tanto de autores quanto dos periódicos, uma vez que este conteúdo não agregaria valor na extração de conhecimento dos corpos. Também foram salvos textos apenas com as referências bibliográficas para verificar um dos objetivos deste trabalho que é descobrir os autores mais citados. Os textos foram salvos no formato "txt" aceito nas ferramentas de mineração de textos utilizadas neste estudo.

4.3 Mineração de Texto

Sobre o corpus pré-processado foram aplicados algoritmos de mineração de textos para extrair conhecimentos relevantes a serem analisados, os quais são apresentados na seção de análise dos resultados.

4.4 Pós-Processamento

Nessa fase foram organizados os conhecimentos adquiridos na fase de mineração de texto para serem apresentados apenas os resultados relevantes a esse estudo, os quais são apresentados na próxima seção.

5 Análise de Resultados

Para melhor apresentação e entendimento dos resultados optou-se por apresentá-los de acordo com os objetivos apresentados no início deste artigo. A primeira análise dos trabalhos selecionados objetivou identificar a quantidade e a porcentagem de incidência dos termos no corpus, para isso utilizou-se a ferramenta Word Counter, onde todos os textos foram colocados em um único arquivo para análise, a Tabela 4 apresenta esses resultados, nos quais foram encontrados 10 termos únicos, compostos duplos e triplos.

| Ordem | Termos com maior incidência | | | |
|-------|-----------------------------|-------|---|--|
| | Únicos | Perc. | Compostos (duplos) | Compostos (triplos) |
| 1° | Students (936) | 5% | Learning Analytics (164) | Student performance prediction (15) |
| 2° | Learning (790) | 4% | Digital Badges (85) | Educational data mining (13) |
| 3° | Data (535) | 3% | Higher Education (60) | task interaction framework (11) |
| 4° | Recomendation (396) | 2% | Data Mining (53) | Learning analytics tools (11) |
| 5° | Information (332) | 2% | Mobile learning (46) | Learning analytics digital (10) |
| 6° | Analytics (211) | 1% | Learning environments (39) | Constructionist learning environments (10) |
| 7° | Model (211) | 1% | Student retention (34) | Formal learning scenarios (10) |
| 8° | Based (201) | 1% | Generic Skills (32) | Graphical user interface (10) |
| 9° | Support (174) | 1% | Real time (32) | Higher education institutions (9) |
| 10° | Participants (172) | 1% | Prediction model (31) Recomender System (31) | Analytics Digital Badges (9) |

Tabela 4: Principais termos do corpus

Na Tabela 4, os termos com maior incidência “*Students*”, “*Learning*”, “*Data*” e “*Recommendation*” nos sinalizam que os textos apresentam o domínio pretendido no início da pesquisa, porém ainda podemos citar o termo “*Model*” na 7ª posição, o que nos evidencia que modelos foram criados ou utilizados nos textos. Complementando a análise com os termos compostos percebemos que os textos apresentam prioritariamente técnicas de *Learning Analytics* e Mineração de Dados Educacionais para prever a performance de estudantes em tempo real, ainda podem ser através de frameworks com foco em ambientes de aprendizagem, ou seja, sistemas de recomendação que evidencia também a retenção de estudantes.

Uma nova análise com a ferramenta Word Counter foi realizada para identificar os três termos que mais apareceram em cada um dos artigos, juntamente com o seu percentual em relação a todo o artigo. Na Tabela 5, destacamos com * a palavra “*Student*” que aparece em 7 dos 10 artigos, o que nos permite afirmar que os estudos tiveram estudantes como foco.

| ID | 1° Termo | 2° Termo | 3° Termo |
|----|-------------------|-------------------|---------------------|
| 1 | Learning (227) 9% | *Student (111) 4% | Mobile (87) 3% |
| 2 | *Student (164) 5% | Model (115) 4% | Prediction (97) 3% |
| 3 | Data (178) 6% | Learning (120) 4% | *Student (112) 4% |
| 4 | *Student (105) 5% | Data (94) 4% | Learning (92) 4% |
| 5 | *Student (154) 6% | Learning (130) 5% | Digital (115) 4% |
| 6 | *Student (71) 8% | Feedback (26) 3% | Performance (25) 3% |

| | | | |
|----|-------------------------|-----------------------|---------------------|
| 7 | *Student (119) 7% | Programming (45) 3% | Real (30) 2% |
| 8 | Recommendation (208) 7% | Learning (85) 3% | Information (71) 2% |
| 9 | Participants (109) 4% | Collaborative (63) 2% | Indicators (45) 2% |
| 10 | Recommendation (135) 5% | Information (109) 4% | Searching (92) 3% |

Tabela 5: Principais termos por artigo

Também nessa análise buscamos identificar empiricamente palavras que podem ter sido usadas como sinônimo ou técnica, que pode ser associada aos termos da *string* de pesquisa, os quais destacam-se alguns resultados esperados de Sistemas de Recomendação: *feedback*, *information*, *prediction*, *performance*, *real* e *indicators*.

Ainda, na Figura 3, gerada através da ferramenta *Trends* do Software *Voyant Tools*, pode-se visualizar no eixo Y a variação da incidência de cinco termos (“*Student*”, “*Learning*”, “*Data*”, “*Recommendation*”, “*Information*”, “*Model*”) nos 10 textos que compõem o corpus (eixo X).

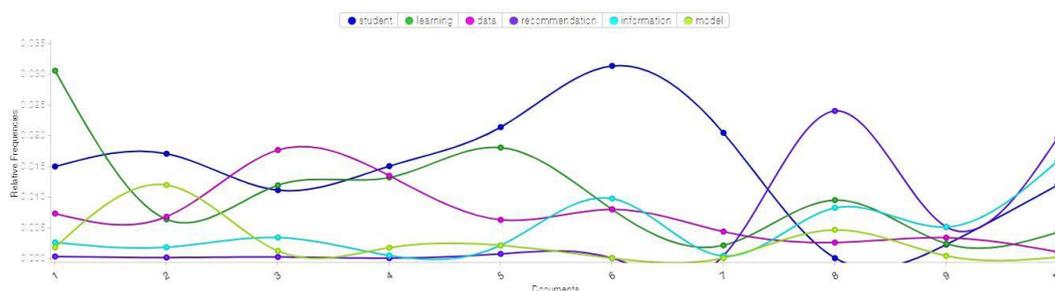


Figura 3: Variação na incidência de termos prevaletentes no corpus.

Na figura 3, é possível verificar que os termos *learning* e *student* possuem uma alta frequência em praticamente todos os textos o que corrobora análises já realizadas nas Tabelas 2 e 3, porém outras análises são interessantes de serem realizadas: a) os termos *Information*, *model* e *recommendation* estão presentes em todos os textos, o que nos sugere que todos apresentam sistemas de recomendação, porém destacam-se nos Id's 7,8,9 e 10 com uma maior incidência; b) O termo *model* está com alta incidência no texto de Id 2; c) O texto de Id 8 possui alta incidência dos termos *recommendations* e *model* e os Id's 6 e 7 são os que possuem menor incidência destes termos; d) Podemos destacar que os textos possuem os principais termos com uma relativa frequência o que demonstra que os mesmos podem e devem ser utilizados para pesquisas na área.

Na sequência utilizou-se a ferramenta Sobek visando extrair o grafo referente a relação dos termos mais relevantes dos textos.

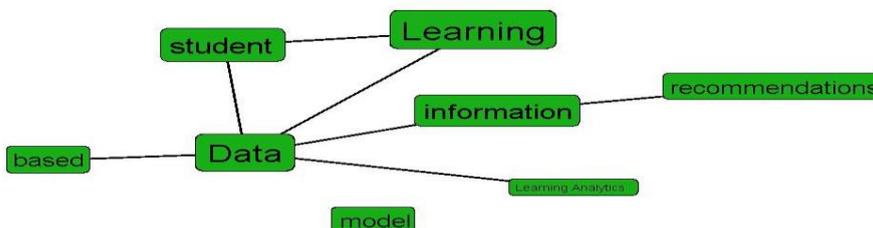


Figura 4: Grafo da relação entre os termos de maior frequência

A Figura 4 apresenta as conexões entre os termos de maior frequência no corpus, identificando uma associação direta de “*recommendation*” com “*information*” através

A Figura 6 foi obtida com o Voyant Tools na ferramenta *Cirrus* e reforça graficamente algumas descobertas encontradas neste estudo. Alguns termos merecem destaque porque estão diretamente associados com Sistemas de Recomendação de estudantes: *indicators, feedback, performance, models, information, support* e *prediction*. Nesse sentido, percebemos que foi eficiente para extrair conhecimento que relacione os termos estudados e permitiu a visualização de suas relações com base no corpus.

6 Conclusões

O processo de descoberta de conhecimento em textos permitiu alcançar o objetivo do estudo e mostrou-se eficiente para extrair conhecimento e apresentar uma perspectiva sobre Sistemas de Recomendação com Mineração de Dados Educacionais e *Learning analytics*. Algumas técnicas e ferramentas de Mineração de Texto foram apresentadas e aplicadas, permitindo análises que permitem uma visão aprofundada do corpus estudado.

Nesse sentido, foi possível apresentar os principais autores dos domínios estudados, observa-se que nos trabalhos selecionados foram abordados aspectos ligados a Sistemas de recomendação como: *indicators, feedback, performance, models, information, support* e *prediction*. Estes podem servir de base para novas *Strings* de busca, resultando em novas descobertas.

A Mineração de Texto mostrou-se um recurso eficiente para analisar dados textuais desestruturados e que a mesma contribui com a comunidade acadêmica permitindo novos conhecimentos e descoberta de padrões. Como limitação deste trabalho, apenas 10 artigos compuseram o corpus o que não nos permite generalizar os resultados e como trabalhos futuros pretende-se aplicar a Mineração de Textos em uma revisão sistemática de literatura e analisar suas contribuições.

Referências

- ARNOLD, K. E. Signals: Applying Academic Analytics. *EDUCAUSE Quarterly*. 33 (1). Disponível em <http://er.educause.edu/articles/2010/3/signals-applying-academic-analytics>. Acesso dia 25 de agosto de 2017, 2010.
- BAKER, R. and YACEF K. “The State of Educational Data Mining in 2009: A Review and Future Visions.” Pages 3-17. **JEDM -Journal of Educational Data Mining**, Volume 1, Issue 1, October, 2009.
- BAKER, R., ISOTANI, S., CARVALHO, A., Mineração de Dados Educacionais: Oportunidades para o Brasil. **Revista Brasileira de Informática na Educação**, 19(2), p. 3-13. 2011.
- BULEGON, Hugo; MORO, Claudia Maria Cabral. Mineração de texto e o processamento de linguagem natural em sumários de alta hospitalar. **Journal of Health Informatics**. Abril-Jun; 2(2), pag. 51-56. 2010.
- CAZELLA, S. C., REATEGUI, E. B., MACHADO, M., & BARBOSA, J. L. V. Recomendação de objetos de aprendizagem empregando filtragem colaborativa e competências. In **Simpósio Brasileiro de Informática na Educação (SBIE)**, Florianópolis, SC, 2009.
- COSTA, E., AGUIAR, J., & MAGALHÃES, J. Sistemas de Recomendação de Recursos Educacionais: conceitos, técnicas e aplicações. In **Jornada de Atualização em Informática na Educação (JAIE)**, Campinas, SP, 2013.

- FRANÇA, R. S; AMARAL, H. J. do. Mineração de Dados na Identificação de Grupos de Estudantes com Dificuldades de Aprendizagem no Ensino de Programação. **Novas Tecnologias na Educação** V. 11 N° 1, julho, 2013.
- FULANTELLI, G.; TAIBI, D.; ARRIGO, M. A framework to support educational decision making in mobile learning. **Computers in Human Behavior**, N. 47, p. 50–59, 2015.
- IMRAN, H.; ZADEH, M. B.; CHANG, T.W.; KINSHUK; GRAF, S. A Framework to Provide Personalization in Learning Management Systems through a Recommender System Approach. **ACIIDS 2014**, p. 271-280, 2014.
- JOHNSON, L., R. SMITH, H. WILLIS, A. Levine, HAYWOOD, K., Learning Analytics. *The 2011 Horizon Report*. Austin, **Texas: The New Media Consortium**. Disponível em: <http://net.educause.edu/ir/library/pdf/HR2011.pdf>. 2011.
- KRASSMANN, A. L., HERPICH, F., BERCHT, M., CAZELLA, S. C. Analyzing trends in academic papers about ubiquitous virtual worlds in education using text mining. **International Journal for Innovation Education and Research**. Vol:-5, No-04, p. 167-180, 2017.
- KONCHADY M. Text mining application programming. Boston: **Charles River Media**; 2006.
- LIU, C.; CHANG, C. TSENG, J. The effect of recommendation systems on Internet-based learning for different learners: A data mining analysis. **British Journal of Educational Technology**, Vol. 44, N. 5, p. 758–773, 2013.
- PARREÑO, J. M., IBÁÑEZ, E. M., ARROYO, A. The use of gamification in education: a bibliometric and text mining analysis. **Journal of Computer Assisted Learning**, v. 32, p. 663–676, 2016.
- PAIVA, R.; BITTENCOURT, Ig I.; SILVA, A. Uma Ferramenta para Recomendação Pedagógica Baseada em Mineração de Dados Educacionais. **II Congresso Brasileiro de Informática na Educação - CBIE 2013**. p. 111-120, 2013.
- REATEGUI, E; EPSTEIN, D; LORENZATTI, A.; KLEMANN, M.. Sobek: a Text Mining Tool for Educational Applications. In: **International Conference on Data Mining**, 2011, Las Vegas, Estados Unidos. Anais do DMIN '11, p. 59-64, 2011.
- RIGO, S. J.; CAMBRUZZI, W.; BARBOSA, J. L. V.; CAZELLA, S. C., Aplicações de Mineração de Dados Educacionais e *Learning Analytics* com foco na evasão escolar: oportunidades e desafios. **Revista Brasileira de Informática na Educação**, Volume 22, Número 1, p. 132-146, 2014.
- RICCI, F.; Rokach, L.; Shapira. B. Introduction to recommender systems handbook. Kantor, editors, **Recommender Systems Handbook**, pages 1–35. Springer, 2011.
- ROCKWELL, G.; SINCLAIR, S. **Hermeneutica: Computer-Assisted Interpretation in the Humanities**. Cambridge: MIT Press, 2016.
- ROMERO, C; VENTURA, S. Educacional data mining: a survey from 1995 to 2005. **Expert systems with applications**. v. 33, n. 11, p. 135-146, 2007.
- SCHIESSI, M.; BRASCHER, M. Descoberta de Conhecimento em Texto aplicada a um Sistema de Atendimento ao Consumidor. **RICI: Revista Ibero-americana de Ciência da Informação**, v. 4, n. 2, p. 94-110, ago./dez, 2011.
- SIEMENS, G., LONG, P. Penetrating the Fog: Analytics in Learning and Education. **EDUCAUSE review**, 46 (5), 2011.
- TAN, A. H. Text Mining: The state of the art and the challenges || , Proceedings of the



Pacific-Asia. **Conference on Knowledge Discovery and Data Mining (PAKDD)**, Beijing - China, 26-28 April, 1999.

VIJAYARANI, S., ILAMATHI, J., NITHYA. **Preprocessing Techniques for Text Mining - An Overview**. International Journal of Computer Science & Communication Networks, Vol 5(1), 7-16, 2015.