

Entropia em redes semânticas de títulos e o futuro do petróleo em publicações científicas

Marcelo do Vale Cunha (Doutorando - MCTI), celaocunha@gmail.com;

Hernane Borges de Barros Pereira (Orientador - MCTI), hbbpereira@gmail.com;

Faculdade SENAI CIMATEC

Palavras Chave: *Redes Semânticas, Redes temporais; Comunicação científica; Entropia da informação, Petróleo.*

Introdução

O conjunto de trabalhos que compõe uma revista de publicação científica faz parte de um sistema formal de comunicação. Este sistema expressa em palavras, diagramas, imagens e equações o conhecimento de atividades de pesquisa, não só para contribuir com o avanço científico da humanidade, mas também para reforçar os laços de comunicação entre cientistas e da ciência com a sociedade em geral (Cunha, 2013).

As redes semânticas de títulos (RST) de artigos científicos propostas recentemente trouxeram possibilidades viáveis de estudo para responder questões relevantes sobre a difusão do conhecimento humano, em periódicos científicos (Pereira et al 2011, Fadigas e Pereira 2009, Cunha et al 2013).

A indústria do petróleo e suas ramificações na sociedade são fortemente influenciadas pelo conhecimento contido em publicações técnico-científicas. Temáticas de interesse para indústria, como por exemplo, "Exploração e produção", "petróleo", "matriz energética", "biocombustíveis" e "gás natural", são difundidas em revistas científicas de alto impacto. A maneira como são difundidos e, por extensão, como atingem a sociedade, depende de como são abordados e contextualizados na publicação e revista a qual pertencem.

O conhecimento e seu contexto dentro da publicação muda ao longo do tempo. Daí surge a necessidade da estudo da difusão destes conceitos, utilizando métodos de redes variáveis no tempo, como os descritos em Holme and J. Saramäki (2013) e de métodos de análise de correlação temporal de propriedades de redes que mudam ou não com o passar do tempo, como os descritos em Zebende (2011).

Estudos sobre a difusão do conhecimento associado à temática de petróleo, bem como o impacto social e econômico das publicações são escassos na literatura.

Seminário Anual de Pesquisa - 2016

Linguagem e comunicação vem sendo cada vez mais estudados pela física estatística e redes complexas (Caldeira et al, 2006; Teixeira et al 2010).

Além disso, a teoria da informação, proposta por Shannon (1948), revela um grande potencial para mensurar e diferenciar sistemas de representação do conhecimento.

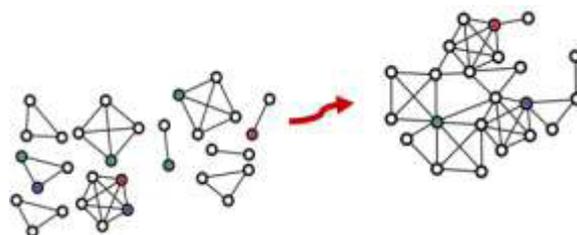
Então, faz-se necessário propor um método que seja capaz de realizar este estudo utilizando a entropia da informação de Shannon em redes de títulos, sob a óptica de um sistema complexo.

Sendo assim, o objetivo deste trabalho é investigar a evolução dos temas associados a petróleo em periódicos científicos de alto impacto, a partir da análise da entropia da informação em redes semânticas de títulos variáveis no tempo.

Resultados e Discussão

Encontra-se em desenvolvimento um banco de dados sobre a produção intelectual sobre exploração de petróleo, gás natural e biocombustíveis. Os canais de comunicação científica escolhidos são periódicos de alto impacto (eg. Journal of Petroleum Science and Engineering) e relatórios importantes para a indústria (eg. Relatórios da ONU). As publicações estão sendo usadas para a construção de redes de títulos, a partir das palavras contidas nos títulos, onde palavras são os vértices da rede e uma aresta liga duas palavras que pertencem ao mesmo título. A Figura 1 ilustra o processo de construção de uma rede de títulos.

Figura 1. Processo de formação de uma RST.

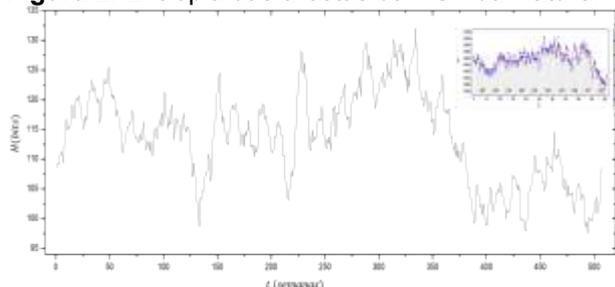


Foi construído para esta pesquisa um conjunto de parâmetros metodológicos capazes de estudar as redes de títulos. Esta metodologia utiliza a aplicação da entropia da informação na frequência de aparição das principais temáticas de interesse para indústria, nas publicações ao longo do tempo.

Além disso, o método matemático desenvolvido analisa tendências em séries temporais dos principais índices de redes variáveis no tempo.

A Figura 2 mostra uma série temporal dos valores das entropias de Shannon, associada a frequência de aparição dos pares de palavras em um TVG da rede semântica de títulos da Nature, de 1998 a 2008. Cada ponto do gráfico representa o valor da entropia das arestas para uma janela bimensal de publicações.

Figura 2. Entropia das arestas da RST da Nature.



Conclusões

Os resultados parciais validam bem o método que está sendo desenvolvido. A partir do banco de dados que está sendo gerado, novas investigações poderão ser feitas, na busca de diferenciar meios de comunicação e épocas onde são abordados temas relacionados a petróleo.

Os métodos aqui desenvolvidos poderão contribuir para uma compreensão mais aprofundada sobre a comunicação científica, inclusive com um dado foco específico, aqui usa-se temáticas associadas a petróleo.

O modelo computacional que é o produto final desta pesquisa, será de grande utilidade para a Indústria do Petróleo, pois é capaz de dispor de um banco de dados sobre a produção intelectual de interesse, organizado de forma a facilitar a construção de redes de palavras (eg. rede de títulos e redes de palavras chaves); dispor de um modelo computacional de análise semântica dos principais conceitos de interesse da indústria; explicar, prever e prospectar assuntos relacionados a exploração de petróleo, gás natural e biocombustíveis, através do modelo proposto e identificar quais os canais de comunicação (eg.

Periódicos científicos e Relatórios) mais adequados para a publicação de suas pesquisas.

Referências

Caldeira SMG, Petit Lobão TC, Andrade RFS, Neme A, Miranda JGV (2006) The network of concepts in written texts. EPJ B 49:523 –529

Cunha, M., Rosa, M. G., de Sousa Fadigas, I., Miranda, J. G. V., and Pereira, H. B. B. (2013). Redes de títulos de artigos científicos variáveis no tempo. In BraSNAM - II Brazilian Workshop on Social Network Analysis and Mining, pages 1744–1755.

Cunha, M.V. (2013) Redes de títulos baseadas em títulos de artigos científicos. Dissertação de mestrado. Programa de pós graduação em modelagem computacional e tecnologia industrial, Senai, Salvador–BA, Brasil.

Fadigas, I. S. e Pereira, H. B. B. (2013) “A network approach based on cliques”, Physica A, v. 392, p. 2576–2587.

Holme and J. Saramäki. Temporal Networks, edited by P. (Springer, New York, 2013).

Pereira, H. B. B., Fadigas, I., Senna, V., and Moret, M. (2011). Semantic networks based on titles of scientific papers. Physica A: Statistical Mechanics and its Applications, 390(6):1192– 1197.

Pineda, J. Otávio C. (2010) Entropia e Teoria da informação: Os fundamentos científicos da era digital. Ed. Annablume.

Shannon, C. E. (1948). A Mathematical Theory of Communication. The Bell System Technical Journal, 27:379–423, 623–656.

Teixeira GM, Aguiar MSF, Carvalho CF, Dantas DR, Cunha MV, Morais JHM, Pereira HBB, Miranda JGV (2010) Complex semantic networks. Int J Mod Phys C 21:333–347.

Wang, B., Tang, H., Guo, C. (2006). Zhilong, X. Entropy Optimization of Scale-Free Networks Robustness to Random Failures. Physica A 363, 591 – 600.

Zebende, G.F. (2011). DCCA cross -correlation coefficient: Quantifying level of crosscorrelation. Physica A 390 (4) 614 - 618.