

Análise de tendências da produção científica nacional na área de Ciência da Informação: estudo exploratório de mineração de textos

Trend analysis of the Brazilian scientific production in information science area: exploratory study of text mining

Caio Cesar Trucolo¹, Luciano Antonio Digiampietri¹

¹ Universidade de São Paulo (USP), São Paulo, SP, Brasil

Autor para correspondência/Corresponding author: Caio Cesar Trucolo [trucolo@gmail.com]

Financiamento/Funding: O trabalho apresentado neste artigo foi parcialmente financiado pela CAPES (bolsa de mestrado) e pelo CNPq (Projeto Universal e bolsa de produtividade em pesquisa).

Recebido/Submitted: 17 Out. 2014

ACEITO/Approved: 15 Nov. 2014



Copyright © 2014 Trucolo & Digiampietri. Todo o conteúdo da Revista está sob uma licença Creative Commons Atribuição-NãoComercial-Compartilhável 3.0 Não Adaptada. Ao serem publicados por esta Revista, os artigos são de livre uso em ambientes educacionais, de pesquisa e não comerciais, com atribuição de autoria obrigatória. Mais informações em <http://www.atoz.ufpr.br/index.php/atoz/about/submissions#copyrightNotice>.

Resumo

Introdução: A análise de tendências pode ser utilizada como uma estratégia para identificar assuntos ou áreas de pesquisa com potencial de popularidade mas que ainda não são muito disseminados. Este trabalho consiste em identificar tendências por mineração de texto e análise histórica das produções científicas (artigos científicos) de doutores da área de Ciência da Informação.

Método: De natureza exploratória, este trabalho foi construído em três etapas. A primeira etapa foi a da obtenção dos dados dos currículos cadastrados na plataforma Lattes. A segunda etapa consistiu na extração automática dos termos mais importantes inseridos nos títulos das publicações e, na terceira etapa, foram feitas regressões lineares e não lineares dos índices de importância baseados em frequência dos termos extraídos.

Resultados: Informações gerais sobre as tendências identificadas para a área de Ciência de Informação para curto, médio e longo prazo são apresentadas.

Conclusão: Este trabalho apresenta e aplica uma metodologia de identificação de tendências que ainda pode ser considerada um primeiro passo ante ao potencial da análise de tendências para a produção científica nacional. Além disso, informações gerais sobre as tendências identificadas e os comportamentos dessas tendências ao longo do tempo foram discutidas.

Palavras-chave: Análise de tendências. Ciência da informação. Redes sociais.

Abstract

Introduction: Trend analysis can be used as a strategy to identify subjects or research areas with potential of popularity which are not very widespread. This work consists of trend identification by text mining and historic analysis of the scientific productions (scientific papers) of the Information Science area PhDs.

Method: This work, having an exploratory basis, was built in three steps. The first step was the data gathering of the curricula registered in Lattes platform. The second one consisted of automatic extraction of the most important terms inside the publications titles and, in the third step, linear and non linear regression of the frequency based importance index of the extracted terms were executed.

Results: Identified trends from the Information Science area for short, medium and long time were presented.

Conclusion: This work presents and applies a trend identification method that can be seen as a first step considering all the potential of the national scientific production trend analysis. Moreover, trend analysis general information and the trends behavior over time were discussed.

Keywords: Trend analysis. Information science. Social networks.

INTRODUÇÃO

Estratégias e políticas públicas têm sido inseridas no país para melhorar a qualidade e aumentar a produtividade da pesquisa científica. Muitas vezes essas políticas são escolhidas de acordo com áreas de pesquisa já consolidadas e populares, nas quais se sabe que haverá retorno, ou ainda, identificadas como tendências mundiais. Um país com dimensões continentais como o Brasil – tanto em extensão geográfica quanto em diversidade cultural – poderia investir em áreas e temas com potencial de crescimento, ampliando o potencial de retorno da investigação científica.

A produção científica no Brasil vem crescendo exponencialmente nas últimas décadas (Digiampietri et al., 2012a) o que só faz crescer o interesse em entender as características da pesquisa no País. Tal análise pode beneficiar da utilização da mineração de texto em tais produções com o objetivo de tentar identificar áreas e temas de pesquisa nas quais os pesquisadores de determinada área trabalham (Miyata, Kano, & Digiampietri, 2013). Assim, analisar tendências a partir das produções científicas para áreas específicas se configura como uma estratégia para encontrar temas de pesquisa com potencial de impacto (Trucolo & Digiampietri, 2014b).

O Brasil possui uma base de dados ímpar de cadastro de currículos de pesquisadores chamada Plataforma Lattes. Nessa base existem mais de três milhões de currículos cadastrados com informações importantes para análise de pesquisadores e redes acadêmicas.

Este artigo consiste em um trabalho de metaestudo para a área de Ciência da Informação com o intuito de identificar tendências de termos de pesquisa para curto, médio e longo prazo. O objetivo é desenvolver e aplicar uma metodologia para identificar tendências de assuntos e ramos de pesquisa a partir das informações dos currículos cadastrados na Plataforma Lattes dos doutores que atuam em Ciência da Informação. É de interesse deste trabalho explorar a metodologia de forma a identificar e prever tendências a partir de análises históricas.

O restante deste artigo está organizado da seguinte forma: a seção 2 sumariza os trabalhos correlatos, a seção 3 apresenta a plataforma Lattes e sua importância informativa sobre a produção científica nacional, a metodologia é descrita na seção 4, na seção 5 os resultados são apresentados e, a seção 6, que contém as considerações finais.

Trabalhos correlatos

Análises de tendência vêm sendo estudadas ao longo dos últimos anos e diversas são as áreas de aplicação. Os trabalhos que mais se aproximam desta proposta são aqueles que assumem como método a mineração de texto e, como aplicação, documentos textuais históricos.

No trabalho de Boletti, Ertekin, Zhou e Giles (2009), o *Llatent Dirichlet Allocation*, que é um modelo generativo probabilístico e a *Gibbs Sampling*, algoritmo gerador de amostras que se aproximam de distribuições de probabilidade específicas, foram utilizados conjuntamente com a ordem temporal dos documentos para a criação de um modelo generativo que aprende as distribuições de autor, tópico e palavra. Em uma aplicação sintética, houve uma taxa de acerto de aproximadamente 72%.

O trabalho de Kawamae e Higashinaka (2010) consistiu em tentar predizer as distribuições dos tópicos em artigos científicos levando o tempo em consideração. Com a mesma ideia, e em um novo trabalho, Kawamae (2012) estabeleceu uma diferença entre tópicos estáveis (que não possuem variação significativa ao longo do tempo) e tópicos dinâmicos, tentando rebater outros modelos que apenas levam em consideração as explosões de tópicos (aumentos súbitos de aparecimento de tópicos em determinados períodos). Para a avaliação, o autor compara o modelo proposto com outros dois modelos utilizando medida de perplexidade e medida de erro L1, que são meios de medir e comparar a acurácia de predição de modelos em relação as amostras utilizadas. Por fim, o modelo apresenta a medida de perplexidade e a taxa de erro L1 menores que a dos outros dois modelos, sendo 2,44 a média de L1.

Jayashri e Chitra (2012) propuseram um modelo com rede *Adaptive Resonance Theory* (ART), que é um tipo de rede neural que propõem resolver o problema de “esquecimento” de aprendizado conforme a apresentação de novas informações, para identificar tópicos em documentos científicos de diferentes bases e detectar tendências considerando os picos de frequência desses tópicos extraídos. A abordagem utilizada foi capaz de detectar os tópicos em alta para diferentes bases de dados.

Park et al. (2011) utilizaram uma abordagem de detecção de tendências usando seleção de características baseada em IG-I (*Improved Gini – Index*). Para cada tópico dado como entrada, subtópicos foram extraídos para então se analisar o comportamento temporal de cada subtópico e identificá-los como em alta ou em baixa. Para os quatro tópicos dados como entrada para o modelo, foram realizados testes de medida F1 para avaliação dos subtópicos com SVM (*Support Vector Machine*), técnica supervisionada de inteligência artificial, e kNN (*k – nearest neighbours*), técnica não supervisionada de inteligência artificial. O resultado de F1 para SVM foi de 0,982 enquanto para kNN foi de 0,916. Uma limitação do método utilizado é que o processo não é totalmente automatizado.

Abe e Tsumoto (2009) selecionaram termos de importância por TF-IDF (*Term Frequency – Inverse Document Frequency*), que é um índice que mede a importância de determinada palavra para um documento dentro de

uma coleção de documentos, e coeficiente de Jaccard, coeficiente que mede a similaridade entre conjuntos de amostras, utilizando regressão linear *a posteriori* para detectar tendências emergentes. Todas as tendências detectadas foram confirmadas como tendências reais por especialistas dos domínios.

Trucolo e Digiampietri (2014a) analisaram tendências de termos extraídos automaticamente a partir de métodos de regressões lineares e não lineares para a rede de doutores inseridos em programas de pós graduação strictu sensu avaliados pela CAPES em Ciências da Computação. Além disso, foi realizada uma análise da rede social baseada em coautoria entre os professores permanentes de cada programa para verificar se existia correlação entre as principais tendências identificadas e as coautorias entre os programas. Nesse trabalho, os autores não conseguiram identificar correlações significativas.

Além dos documentos textuais utilizados como base para a análise de tendências, nos últimos anos as redes sociais começaram a ser também utilizadas para auxiliar nesta análise. Cimenler, Reeves e Skvoretz (2014) analisam algumas métricas de redes sociais, em sua maioria métricas de centralidade, para entender quão significativas são essas métricas para prever o desempenho de pesquisadores baseando-se em índices derivados de citação como o índice h. O método de regressão de Poisson é utilizado para analisar a importância das métricas para alguns tipos de redes formadas por pesquisadores de uma faculdade de engenharia.

Uma revisão sistemática sobre técnicas de identificação e análise de tendências para outras aplicações além de documentos textuais históricos pode ser encontrada em Trucolo e Digiampietri (2014b).

Uma das principais contribuições do presente trabalho é o aprofundamento relevante para com a aplicação porque os dados extraídos da plataforma Lattes proporcionam uma análise bastante rica sobre a condição científica nacional. Pode-se dizer, portanto, que esse trabalho se diferencia dos demais citados pela característica de aproximação da análise à realidade científica do país. Outra característica importante é a capacidade de a identificação e predição de tendências utilizada aqui não necessitar de esforço humano, ou seja, ela é realizada a partir de termos e expressões extraídos automaticamente da base de dados sem a necessidade de manualmente se estabelecer a importância dos termos ou valores limítrofes para as tendências.

MÉTODOS

Todo o processo de análise de tendências foi realizado em três fases: obtenção dos dados; extração automática dos termos; e análise de tendência dos termos extraídos.

Obtenção dos dados

Para a obtenção dos dados, inicialmente foram identificados todos os doutores que atuam na área de Ciência da Informação e que possuem currículos cadastrados na plataforma Lattes. As informações dos currículos destes pesquisadores foram tabuladas e armazenadas em um banco de dados seguindo a metodologia apresentada em Digiampietri et al. (2012a), para a realização dos testes, foram extraídos os termos de 34.289 títulos de artigos publicados entre os anos de 1991 e 2012.

Extração automática de termos

A técnica de extração automática de termos utilizada consiste em determinar os termos mais importantes de um conjunto de documentos pela frequência adjacente das palavras que compõem esses termos. A fórmula (1) utilizada para o cálculo dos pesos de cada termo candidato é a seguinte:

$$FED(TC) = f(TC) \times \left(\prod_{i=1}^T (FE(N_i) + 1) (FD(N_i) + 1) \right)^{1/T} > 1.0 \quad (1)$$

Em que $f(TC)$ é a frequência do termo candidato TC, e $FE(N_i)$ e $FD(N_i)$ indicam a frequência dos candidatos da esquerda e da direita, respectivamente. Esta fórmula é detalhadamente descrita por Nakagawa e Mori (2002).

Observou-se que os termos compostos tinham mais significado do que os termos simples em relação aos assuntos abordados pelas publicações. Desta forma, os termos utilizados na fase de análise de tendências foram os termos compostos e de maiores pesos.

Análise de tendências

Com base nos termos extraídos, foram calculados os índices de importância dos termos para cada ano. O índice de importância utilizado nesse trabalho é o TF-IDF (*Term Frequency divided by Inverse Document Frequency*), que é um dos índices mais utilizados para aferir a importância de termos. Com esses índices calculados, foram utilizadas análises de regressão do tipo linear e não linear. Análises de regressão seguem o formato (2)

$$Y \approx f(X, \beta) \quad (2)$$

em que a variável dependente Y pode ser aproximada pelas variáveis independentes X e seus respectivos parâmetros β para determinada função $f()$. Nas análises de regressão desse trabalho a variável dependente é o índice TF-IDF do termo e variável independente é o tempo (os anos de publicações dos artigos).

O método de mínimos quadrados foi utilizado para determinar as curvas de tendência que mais se adequavam as séries temporais de cada termo. Os tipos de regressão utilizados foram linear, exponencial, logarítmica, *power law* e polinomial de grau 2 a 5. Posteriormente, foi calculado o erro quadrático para cada curva de tendência gerada para se determinar a curva mais adequada à série temporal de cada termo.

A classificação dos termos como tendências foi baseada na previsão, a partir da curva de tendência mais adequada (isto é, com menor erro quadrático) para alguns anos específicos que indicassem curto, médio e longo prazo. A análise histórica dos termos foi realizada entre os anos de 1991 e 2012, com isso, a análise de curto prazo foi realizada para o ano de 2013, a de médio prazo para o ano de 2015 e a de longo prazo para o ano de 2020. Optou-se por utilizar dados das publicações apenas até o final de 2012 pois a partir de 2013 muitos currículos não estão atualizados (Digiampietri et al., 2014).

RESULTADOS

A partir da extração automática de termos, conforme explicado na seção anterior, foram selecionados três termos entre os mais importantes para exemplificar o comportamento temporal de cada um. O gráfico 1 ilustra esses comportamentos e é possível visualizar comportamentos bem diferentes entre eles. Os termos **ciência da informação** e **gestão do conhecimento** vinham em uma crescente parecida até o ano de 2006, quando se separaram. **Ciência da informação** continuou crescendo até 2009 quando se estabilizou e teve uma queda significativa em 2012. Já **gestão do conhecimento** começou a decrescer em 2006 e se estabilizou a partir de 2008. O termo **meio ambiente**, diferentemente dos outros dois termos, tem um comportamento estável com altos e baixos não muito significativos durante o período.

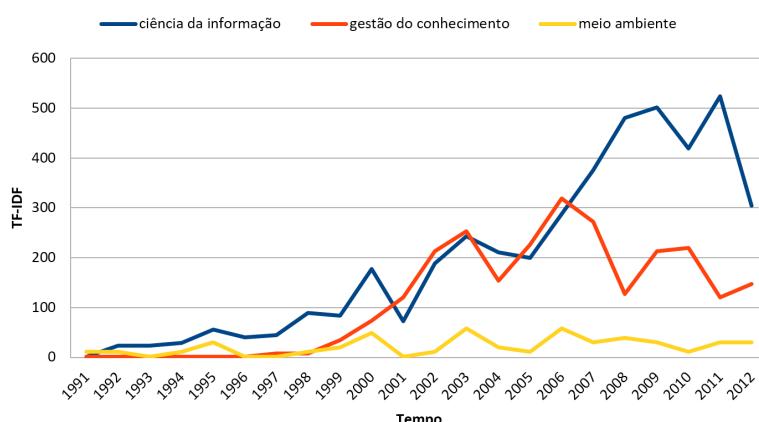
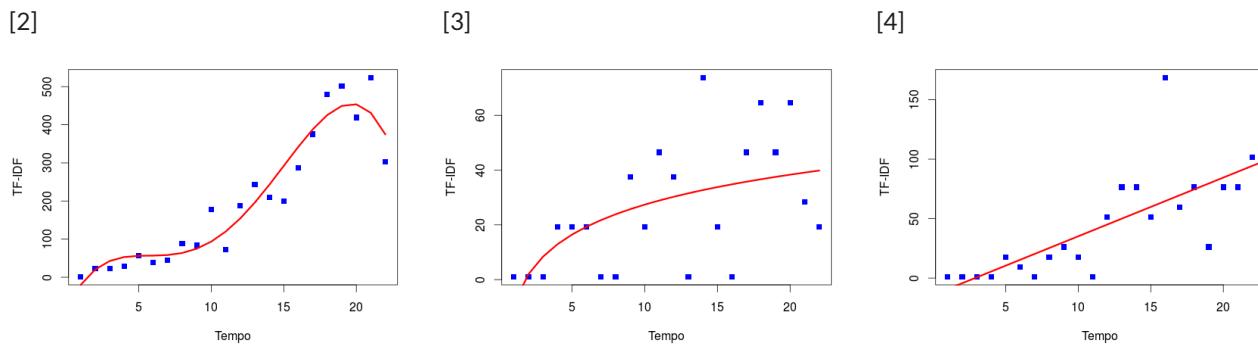


Gráfico 1. Comportamento temporal de três termos
Fonte: os autores.

Como explicado anteriormente, foram calculadas as curvas de tendência das séries temporais de cada termo extraído. As figuras 2, 3 e 4 mostram as curvas de tendência baseadas nas regressões polinomial de grau 4, logarítmica e linear, respectivamente, para os termos **ciência da informação**, **latin america** e **comunicação científica**. A forma algébrica das curvas de tendência são, respectivamente (3)

$$\begin{aligned} Y &= -84,76 - 0,3 \times X^4 + 1,28 \times X^3 - 15,89 \times X^2 + 79,48 \times X \\ Y &= -8,99 + 15,80 \times \log(X) \\ Y &= -14,24 + 4,94 \times X \end{aligned} \quad (3)$$

Pelos gráficos 2, 3 e 4 é possível notar uma tendência de aumento do índice TF-IDF para os termos **latin america** e **comunicação científica** enquanto que **ciência da informação**, que teve um aumento substancial no meio da série temporal, está em uma fase de queda. O comportamento temporal do termo **latin america** é interessante pelo motivo de que mesmo tendo grandes variações entre os anos, a curva de tendência aponta um aumento baixo do índice para os anos seguintes.



Gráficos 2, 3 e 4. [2] Curva de tendência gerada pela regressão não linear polinomial de grau 4 para os termo ciência da informação. [3] latin america. [4] comunicação científica.

Fonte: os autores.

A tabela 1 apresenta as vinte principais tendências a curto, médio e longo prazo em ordem decrescente. Nesta tabela já é possível notar comportamentos de alguns dos termos. **Information retrieval**, por exemplo, não aparece entre as principais tendências a curto prazo, mas em 2015 já aparece entre as duas principais. Para uma visão melhor desses comportamentos, a tabela 2 mostra o deslocamento de posições para médio e longo prazo entre as vinte principais tendências de 2013. A tabela 2 igualmente confirma a queda do índice TF-IDF para médio e longo prazo do termo **ciência da informação** (verificado na curva de tendência da figura 2). O termo **comunicação científica** apesar de ter uma tendência positiva, como visto na figura 4, teve um deslocamento negativo das posições para médio e longo prazo.

A fim de se avaliar a acurácia do modelo proposto, foram comparados os resultados previstos para os 20 termos com maiores tendências de popularidade e os resultados reais TF-IDF dos mesmos para o ano de 2011. Observou-se um erro padrão médio de aproximadamente 38,6% e um grau de correlação positivo entre os valores previstos e os valores reais em torno de 0,55.

As oscilações de tendências ao longo do tempo se devem aos diferentes modelos de regressão determinados automaticamente, ou seja, o modelo mais adequado baseado no método dos mínimos quadrados. O gráfico 5 exemplifica o comportamento das curvas de regressão para os termos **information retrieval** e **comunicação científica**. Analisando-se o gráfico 5 juntamente com a tabela 2, nota-se que comunicação científica está tendo um crescimento, porém, esse crescimento é baixo em relação a termos como **information retrieval**, que é a principal tendência para 2020.

Tendências em 2013	Tendências em 2015	Tendências em 2020
ciência da informação	sustentabilidade ambiental	information retrieval
santa catarina	information retrieval	sustentabilidade ambiental
comunicação científica	santa catarina	marc21 bibliográfico
minas gerais	marc21 bibliográfico	multiwords expressions
sustentabilidade ambiental	comunicação científica	bioactive glass
comportamento informacional	multiwords expressions	doenças negligenciadas
américa latina	doenças negligenciadas	médio porte
information retrieval	médio porte	direitos autorais
biblioteca universitária	betim paes	buenos aires
marc21 bibliográfico	betim paes leme	diretrizes curriculares
matto grosso	diretrizes curriculares	betim paes
região metropolitana	minas gerais	betim paes leme
doenças negligenciadas	direitos autorais	febre amarela
inteligência competitiva	classificação facetada	classificação facetada
são paulo	febre amarela	suggestion box
multiwords expressions	suggestion box	ivan illich
neural network	ivan illich	claims derived
médio porte	claims derived	crenicichla menezesi
língua portuguesa	crenicichla menezesi	rheumatoid arthritis study
betim paes	rheumatoid arthritis study	rheumatoid arthritis

Tabela 1. Vinte principais tendências para curto, médio e longo prazo

Fonte: os autores.

Posição	2013	2015		2020	
1°	ciência da informação	sustentabilidade ambiental	4	information retrieval	1
2°	santa catarina	information retrieval	6	sustentabilidade ambiental	-1
3°	comunicação científica	santa catarina	-1	marc21 bibliográfico	1
4°	minas gerais	marc21 bibliográfico	6	multiwords expressions	2
5°	sustentabilidade ambiental	comunicação científica	-2	doenças negligenciadas	2
6°	comportamento informacional	multiwords expressions	10	médio porte	2
7°	américa latina	doenças negligenciadas	6	betim paes	2
8°	information retrieval	médio porte	10	inteligência competitiva	8
9°	biblioteca universitária	betim paes	11	santa catarina	-6
10°	marc21 bibliográfico	minas gerais	-6	comunicação científica	-5
11°	matto grosso	américa latina	-4	língua portuguesa	2
12°	região metropolitana	matto grosso	-1	matto grosso	0
13°	doenças negligenciadas	língua portuguesa	6	américa latina	-2
14°	inteligência competitiva	comportamento informacional	-8	região metropolitana	1
15°	são paulo	região metropolitana	-3	minas gerais	-5
16°	multiwords expressions	inteligência competitiva	-2	biblioteca universitária	1
17°	neural network	biblioteca universitária	-8	neural network	1
18°	médio porte	neural network	-1	comportamento informacional	-4
19°	língua portuguesa	ciência da informação	-18	são paulo	1
20°	betim paes	são paulo	-5	ciência da informação	-1

Tabela 2. Alteração das posições a médio e longo prazo das vinte principais tendências de curto prazo

Fonte: os autores.

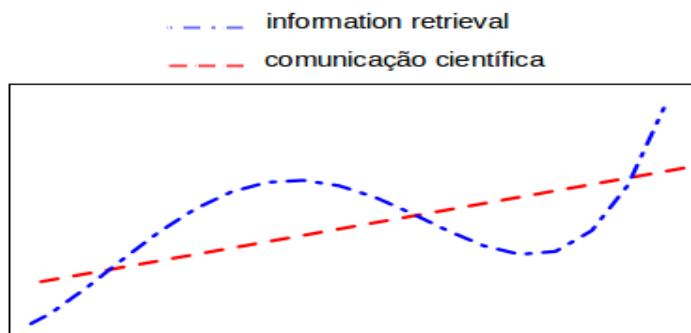


Gráfico 5. Comportamento das curvas de regressão dos termos **information retrieval** e **comunicação científica**
Fonte: os autores.

CONSIDERAÇÕES FINAIS

A aplicação de estratégias e políticas públicas mais acuradas para o aumento da qualidade e produtividade da ciência no país dependem de análises mais profundas em relação a realidade acadêmica brasileira. O Brasil, por ser um país com dimensões continentais tanto em extensão geográfica quanto em diversidade cultural, necessita de estudos específicos para identificar áreas e assuntos com grande potencial de impacto científico e social.

Com caráter informacional, este trabalho apresentou informações gerais sobre as tendências de assuntos e termos para a área de Ciência da Informação. Foram mostradas as principais tendências de curto, médio e longo prazo e o comportamento de alguns desses termos ao longo desse intervalo. Dessa forma, foi possível vislumbrar quais assuntos estarão em alta para curto, médio e longo prazo.

Este trabalho não contempla o agrupamento dos termos extraídos em tópicos para se, então, analisar a tendência desses tópicos mais gerais que podem ser relevantes para a área de Ciência da Informação. Neste trabalho também não foi realizada uma análise de correlação entre as principais tendências entre programas de pós-graduação e as principais coautoriais inter programas como feito por Trucolo e Digiampietri (2014a) pelo fato de nem todos os doutores da análise estarem inseridos em programas de pós-graduação.

Os resultados deste trabalho ainda podem ser considerados iniciais, considerando-se todo o potencial da análise de tendências da produção científica nacional. Em trabalhos futuros, a estrutura das fontes de informação, ou seja, as características das redes sociais, serão agregadas. Métricas das redes serão inseridas de forma que auxiliem na explicação do comportamento das séries temporais. Com isso, objetiva-se aumentar o poder de acurácia do modelo de predição de tendências.

REFERÊNCIAS

- Abe, H., Tsumoto, S. (2009). Evaluating a method to detect temporal trends of phrases in research documents. *8th IEEE International Conference on Cognitive Informatics*, 378-383. doi:[10.1109/COGINF.2009.5250711](https://doi.org/10.1109/COGINF.2009.5250711)
- Bolelli, L., Ertekin, S., Zhou, D., & Giles, C. L. (2009). Finding topic trends in digital libraries. *9th ACM/IEEE-CS Joint Conference on Digital Libraries*, 69-72. doi:[10.1145/1555400.1555411](https://doi.org/10.1145/1555400.1555411)
- Cimenler, O., Reeves, K. A., & Skvorets, J. (2014). A regression analysis of researchers' social network metrics on their citation performance in a college of engineering. *Journal of Informetrics*, 8(3), 667-682. doi:[10.1016/j.joi.2014.06.004](https://doi.org/10.1016/j.joi.2014.06.004)
- Digiampietri, L. A., Mena-Chalco, J. P., Pérez-Alcázar, J. J., Tuesta, E. F., Delgado, K. V., Mugnaini, R., & Silva, G. S. (2012a). Dinâmica das relações de coautoria nos programas de pós-graduação em computação no Brasil. *2012 Brazilian Workshop on Social Network Analysis and Mining*. Retirado de http://www.imago.ufpr.br/csbc2012/anais_csbc/eventos/brasnam/artigos/BRASNAM%20-%20Minerando%20e%20Caracterizando%20Dados%20de%20Curriculos%20Lattes.pdf
- Digiampietri, L., Mugnaini, R., Mena-Chalco, J., Delgado, K., & Pérez-Alcázar, J. (2014). Análise da atualização dos currículos Lattes. *IV Encontro Brasileiro de Bibliometria e Cientometria*. Retirado de http://www.uspleste.usp.br/digiampietri/bibtex/DigiampietriEtAl_EBBC2014.pdf
- Kawamae, N. (2012). Theme chronicle model: chronicle consists of timestamp and topical words over each theme. *21st ACM International Conference on Information and Knowledge Management*, 2065-2069. doi:[10.1145/2396761.2398573](https://doi.org/10.1145/2396761.2398573)
- Kawamae, N., & Higashinaka, R. (2010). Trend detection model. *19th International Conference on World Wide Web*, 1129-1130. doi:[10.1145/1772690.1772838](https://doi.org/10.1145/1772690.1772838)
- Jayashri, M., & Chitra, P. (2012). Topic clustering and topic evolution based on temporal parameters. *2012 International Conference on Recent Trends in Information Technology*, 559-564. doi:[10.1109/ICRTIT.2012.6206816](https://doi.org/10.1109/ICRTIT.2012.6206816)
- Miyata, B. K. O., Kano, V. Y., & Digiampietri, L. A. (2013). Combinando mineração de textos e análise de redes sociais para a identificação das áreas de atuação de pesquisadores. *Second Brazilian Workshop on Social Network Analysis and Mining*. Retirado de <https://drive.google.com/viewerng/viewer?a=...>
- Nakagawa, H., & Mori, T. (2002). A simple but powerful automatic term extraction method. *Second International Workshop on Computational Terminology*. doi:[10.3115/1118771.1118778](https://doi.org/10.3115/1118771.1118778)
- Park, H., Kim, E., Bae, K., Hahn, H., Sung, T., & Kwon, H. (2011). Detection and analysis of trend topics for global scientific literature using feature selection based on Gini-Index. *23rd IEEE International Conference on Tools with Artificial Intelligence*, 965-969. doi:[10.1109/ICTAI.2011.166](https://doi.org/10.1109/ICTAI.2011.166)
- Trucolo, C. C., & Digiampietri, L. A. (2014a). Análise de tendências da produção científica nacional da área de Ciência da Computação. *Revista de Sistemas de Informação da FSMA*, 14, 2-9. Retirado de http://www.fsma.edu.br/si/edicao14/FSMA_SI_2014_2_Estudantil_1.pdf
- Trucolo, C. C., & Digiampietri, L. A. (2014b). Uma revisão sistemática acerca das técnicas de identificação de análise de tendências. *X Simpósio Brasileiro de Sistemas de Informação*, 639-650. Retirado de <http://www.uspleste.usp.br/digiampietri/bibtex/TrucoloEDigiampietri2014a.pdf>

Como citar este artigo (ABNT):

TRUCOLO, C. C.; DIGIAMPIETRI, L. A. Análise de tendências da produção científica nacional na área de Ciência da Informação: estudo exploratório de mineração de textos. *AtoZ: novas práticas em informação e conhecimento*, Curitiba, v. 3, n. 2, p. 87-94, jul./dez. 2014. Disponível em: <<http://www.atoz.ufpr.br>>. Acesso em:

<http://www.atoz.ufpr.br/index.php/atoz/article/view/79>

How to cite this article (APA):

Trucolo, C. C., & Digiampietri, L. A. (2014). Análise de tendências da produção científica nacional na área de Ciência da Informação: estudo exploratório de mineração de textos. *AtoZ: novas práticas em informação e conhecimento*, 3(2), 87-94. Retrieved from <http://www.atoz.ufpr.br>