

Análise sobre Citações de Pesquisadores e Autores de Eventos Científicos

Daniel M. Cunha¹, Leandro O. da Silva¹, Raimundo S. Moura¹

¹Departamento de Computação – Universidade Federal do Piauí(UFPI)
64.049-550 – Teresina – PI – Brasil

daniel.mesquita97@outlook.com, oliveira.leandro826@gmail.com,
rsm@ufpi.edu.br

Abstract. *The purpose of this paper is to make an analysis about Researches' quotes in scientific works obtained through Google Scholar. The paper consists in a script developed in Python that returns the index of quotes. From those indexes was done an analysis of scientific events and graduate programs, obtaining the indexes of events and programs. Additionally, we made an analysis to verify which of the events and programs have the most experienced researchers. Lastly, the results allowing the new students and researchers to do an analysis of the events to see which of them are more appropriated to publish their works.*

Resumo. *O objetivo deste trabalho é fazer uma análise de citações de pesquisadores em trabalhos científicos obtidos através do Google Acadêmico. O trabalho consiste em um script desenvolvido em linguagem Python que retorna os índices de publicação dos pesquisadores. A partir desses índices foi realizada uma análise de eventos científicos e programas de pós-graduação (PPG), obtendo-se os índices gerais dos eventos e dos PPGs. Adicionalmente, realizou-se uma análise para verificar quais eventos e programas possuem pesquisadores mais experientes. Por fim, os resultados possibilitam a alunos e pesquisadores novatos fazerem uma análise dos eventos para se ter uma noção de qual o melhor evento para publicar o seu trabalho.*

1. Introdução

Pesquisadores de diversas áreas publicam artigos em eventos de seus respectivos campos de atuação para compartilhar conhecimentos com o maior número possível de pessoas interessadas em determinado tema. Porém, a escolha do evento deve levar em consideração a relevância do trabalho, pois determinados eventos são mais apropriados para trabalhos iniciais, enquanto outros para trabalhos mais avançados. Este presente trabalho visa fazer uma análise de eventos científicos para se ter a possibilidade de identificar a grandiosidade do evento a partir da experiência dos pesquisadores. Adicionalmente, iremos analisar programas de pós-graduação (PPG) da área de computação e informática. De maneira geral, pretende-se analisar a reputação de um evento ou programa de pós-graduação.

Para Jones, Hesterly e Borgatti (1997), a reputação envolve uma estimativa do caráter, habilidades e confiabilidade de um indivíduo. Segundo os autores, a reputação reduz o comportamento de incerteza por prover informações a respeito da confiabilidade e boa vontade dos outros. Na área de governança corporativa, uma boa reputação traz consequências econômicas para as empresas.

Deephouse (2000) define reputação como a avaliação de uma empresa por seus interessados em termos de seus efeitos, estima e conhecimento. Hall (1993) descreve reputação como ativo intangível de uma empresa. Já para Caudron (1997), a reputação é construída não somente por palavras, mas principalmente por meio de atos. Segundo Heinonen (2011), uma medida bem definida do grau de reputação não existe. Neste trabalho iremos usar o número de citações dos autores de um evento científico ou pesquisadores de um programa de pós-graduação para definir a reputação do mesmo.

Portanto, este artigo apresenta uma análise sobre citações de pesquisadores e autores de eventos científicos como uma alternativa para a análise da grandiosidade de eventos ou de programas de pós-graduação baseada na quantidade de citações que os autores possuem. A partir das citações dos pesquisadores em trabalhos anteriores pode-se ter noção da grandiosidade do evento pois se o autor possuir muitas citações é possível deduzir que há uma experiência maior na área e que o trabalho desenvolvido vem sendo aprimorado cada vez mais.

O restante do artigo está organizado da seguinte forma: a Seção 2 destaca alguns conceitos e trabalhos sobre reputação e citações. A seção 3 apresenta a abordagem proposta para classificação de eventos e/ou programas de pós-graduação. A seção 4 descreve os experimentos realizados para análise de PPGs nordestinos e eventos científicos da área de computação. Finalmente, a seção 5 conclui o artigo e apresenta sugestões de trabalhos futuros.

2. Sobre Reputação e Citações

Atualmente, não existe uma definição formal para reputação de autor na Web aceita pela comunidade. Porém, algumas medidas têm sido propostas em sistemas de avaliação. Segundo Li et al. (2014), a reputação de um usuário em um sistema de avaliação pode ser medida de acordo com as informações postadas por tal usuário, desta forma, quanto mais avaliações justas e confiáveis forem escritas, melhor a reputação do autor.

Carlos SÁ (2017) examinou um conjunto de medidas definidas em redes sociais online (RSO), ambientes colaborativos *Wiki* e fóruns *online* para inferir a reputação do autor de comentários Web, utilizando Redes Neurais Artificiais (RNA). Ele observou que a quantidade de votos positivos que um autor recebe tem um peso significativo em sua reputação.

Para classificar a produção científica de pesquisadores e de programas de pós-graduação, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES¹), fundação vinculada ao Ministério da Educação do Brasil, tem utilizado o sistema Qualis Web. No entanto, esse sistema tem recebido críticas e outras medidas devem ser consideradas nos próximos anos.

Uma alternativa a ser considerada é utilizar as citações do Google Escolar e de outras plataformas, que proveem um jeito simples dos autores manterem um caminho de citações para seus artigos. Você pode checar quem está citando suas publicações (<https://scholar.google.com/intl/en/scholar/citations.html>). De acordo com o próprio site Google Escolar, esse serviço é um banco de dados de publicações de fundamentação acadêmica que tem propriedades como citações, que é o dado utilizado na pesquisa.

¹ <http://www.capes.gov.br/>

No nosso trabalho, a ferramenta utilizada para obter as citações foi a biblioteca *Scholarly*² que está sustentada na linguagem Python 3.0³. Destaca-se que as citações retornadas foram apenas de usuários que estão cadastrados na plataforma, pois, há pesquisadores que possuem citações, mas não foram retornados por falta de cadastro no sistema do Google Acadêmico.

Destacamos que boa parte dos estudos relacionados a esta área envolvem coleta e análise de dados de programas de pós-graduação para realizar comparações a lista de classificação. Digiampietri et al. (2016) faz uma comparação entre os programas de pós-graduação brasileiros de níveis 6 e 7 e programas melhores colocados em Rankings internacionais e obtém como resultado que os programas brasileiros não possuem as mesmas características dos melhores programas internacionais.

O trabalho de Laender et al. (2008) fez uso de dados do DBLP para comparar os indicadores de programas de pós-graduação brasileiros que possuem níveis 6 ou 7 (considerados de padrão internacional) com programas da Europa e América do Norte chegando a conclusão que as pós-graduações brasileiras estão crescendo e ganhando maturidade.

O sistema CSIndexbr Brasil⁴ fornece dados significativos e transparentes sobre a produção científica brasileira na área de Ciência da Computação. Os artigos são coletados da biblioteca digital DBLP e são coletados apenas artigos completos, aceitos na trilha principal de pesquisa das conferências consideradas. Esse serviço é mantido pelo *Applied Software Engineering Research Group* da Universidade Federal de Minas Gerais (UFMG).

Nossa intenção é medir, de acordo com as citações do Google Escolar programas de pós-graduação de instituições nordestinas, bem como de alguns eventos científicos da área de processamento de linguagem natural e mídias sociais.

3. Método Proposto

A Figura 1 apresenta uma visão geral da abordagem proposta. Inicialmente, coleta-se os nomes dos autores do PPG ou evento científico a ser avaliado. Em seguida, com o *Script* feito em linguagem Python, usa-se a biblioteca *Scholarly* para obter as citações de cada autor do artigo ou professor do programa. Desta forma, obtêm-se o número de citações dos pesquisadores cadastrados no Google Acadêmico. A etapa final consiste em totalizar os resultados e apresentar de forma sumarizada.

² <https://pypi.org/project/scholarly/0.2.1/>

³ <https://docs.python.org/3/>

⁴ <http://csindexbr.org/>

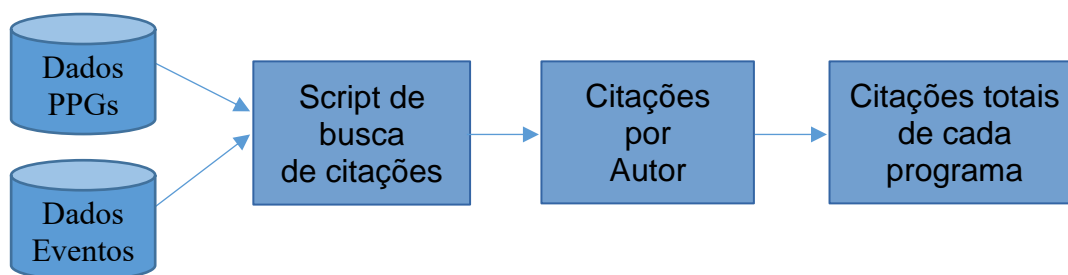


Figura 1. Visão geral do método proposto

4. Experimentos

A pesquisa foi executada seguindo em duas direções: i) análise dos programas de pós-graduação dos cursos de computação de Instituições de Ensino Superior (IES) da região nordeste; e ii) análise de eventos científicos, através dos autores dos artigos publicados.

4.1. Análise de PPGs

Os dados foram coletados dos principais programas de pós-graduação em ciência da computação / ciência da informação, de instituições de ensino superior (IES) nordestinas. A Tabela 1 mostra os programas com os pesquisadores permanentes e o número de citações. A última coluna representa o total de citações do programa.

Tabela 1. Programas de pós-graduação de IES nordestinas e número de citações

Tipo	Instituição	Curso	Programa	Citações	Total Citações
Permanente	UFPI	Ciência da Computação	PPGCC	1730	1730
Permanente	UFMA	Ciência da Computação	PPGCC	4568	4568
Permanente	UFC	Ciência da Computação	PPGCC	6366	6366
Permanente	UFPE	Ciência da Informação	PPGCC	3965	3965

Obs: Apesar de os programas terem os professores colaboradores, os resultados de suas citações foram 0, portanto, foi preferível não incluir essa linha na tabela.

4.2. Análise de Eventos Científicos

Com relação aos eventos, foram coletadas as citações de todos os autores de artigos de cinco eventos científicos da área de computação, realizados no ano de 2017, a saber: BRACIS, BRASNAM, ENIAC, STIL e WEBMEDIA. É importante destacar que os

autores são pesquisadores de instituições do Brasil e de outros países, incluindo alunos de iniciação científica, mestrandos, doutorandos e professores. Primeiramente, para cada evento os artigos foram organizados em um arquivo texto com a relação de todos os autores. Através de um *script* desenvolvido em Python, realizou-se a análise das citações de cada autor individualmente, exibindo as citações que cada um possui no Google Acadêmico ou informando que o respectivo autor não possui citação. Em seguida, realizou-se uma comparação geral de todo o evento, mostrando a soma das citações de todos os pesquisadores, a média dos índices dos artigos e o percentual de autores que não possuem citação.

A Tabela 2 apresenta os dados comparativos dos cinco eventos analisados. Com os resultados obtidos, é possível se ter uma ideia da grandiosidade do evento baseado na experiência que os pesquisadores possuem na área científica.

Tabela 2. Eventos científicos realizados em 2017 e número de citações

Evento	Artigos	Autores	Citações	MédiaCitações	AutoresSemCitação
BRACIS	74	230	57852	251,53 %	36,7 %
BRASNAM	21	69	50312	739,8 %	50,0 %
ENIAC	63	187	37491	200,48 %	42,7 %
STIL	28	77	39654	514,99 %	38,0 %
WEBMEDIA	90	217	43567	200,77 %	46,1 %

5. Conclusões e Trabalhos Futuros

Este artigo aponta para a possibilidade de analisar a experiência de pesquisadores em geral, através de suas citações em trabalhos científicos e conclui-se que o trabalho pode continuar a ser desenvolvido em busca de melhoras.

Não pretendemos que o sistema de classificação de programas de pós-graduação e eventos e periódicos científicos usado pela CAPES seja substituído pela nossa proposta. No entanto, alertamos aos jovens pesquisadores sobre eventos, que ainda não foram classificados pela CAPES, mas que têm a participação de grandes pesquisadores da área.

Como trabalhos futuros, destaca-se a possibilidade é desenvolver um *Script* que faça a análise automática dos dados buscando os nomes dos autores e pesquisadores diretamente dos respectivos websites através das urls dos programas a serem analisados, evitando assim o trabalho manual. Além da realização de novos experimentos.

Referências

- CAUDRON, S. (1997) *The Search for Meaning at Work*. [S.l.]: Training and Development.
- DEEPHOUSE, D. L. (2000) Media reputation as a strategic resource: An integration of mass communication and resource-based theories. *Journal of management*, Sage Publications, v. 26, n. 6, p. 1091–1112.

- DIGIAMPIETRI, L., LINDEN, R., BARBOSA, L. (2016) Caracterizando Departamentos e Programas de Computação Utilizando Análise de Redes Sociais e Bibliometria. Proc. of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM).
- HALL, R. (1993) A Framework Linking Intangible Resources and Capabilities to Sustainable Competitive Advantage. *Strategic Management Journal*, v. 14, p. 607–618.
- HEINONEN, K. (2011) Consumer activity in social media: Managerial approaches to consumers' social media behavior. *Journal of Consumer Behaviour*, Wiley Online Library, v. 10, n. 6, p. 356–364.
- JONES, C.; HESTERLY, W. S.; BORGATTI, S. P. (1997) A general theory of network governance: Exchange conditions and social mechanisms. *The Academy of Management Review*, Academy of Management, v. 22, n. 4, p. 911–945.
- LAENDER, A., LUCENA, C., MALDONADO, J., DE SOUZA E SILVA, E., ZIVIANI, N. (2008) Assessing the Research and Education Quality of the Top Brazilian Computer Science Graduate Programs. *SIGCSE Bulletin*, 40 (2): 135-145.
- LI, Baichuan, LI, Rong-Hua, KING, Irwin, LYU, Michael, YU, Jeffrey Xu. (2015) A topic-biased user reputation model in rating systems. *Knowledge and Information Systems*, v. 44 (3), p. 581-607.
- SÁ, Carlos A. (2017) Abordagem para Definir a Reputação do Autor em Comentários de Produtos na Web Utilizando Redes Neurais Artificiais. Dissertação (Mestrado) – Universidade Federal do Piauí, Teresina-PI.