

EXTRAÇÃO E COMPARTILHAMENTO DE DADOS DE CITAÇÕES NO GOOGLE SCHOLAR PARA A EXPLORAÇÃO COLABORATIVA

Tiago Rodrigo Marçal Murakami (USP)
trmurakami@gmail.com

Ronaldo Ferreira de Araujo (UFAL)
ronaldfa@gmail.com

Sibele Fausto (USP)
sibelefausto@gmail.com

EIXO TEMÁTICO: Métodos, técnicas e ferramentas para estudos bibliométricos e cientométricos
MODALIDADE: Pôster

1 1 INTRODUÇÃO

É notória a insuficiência de indexação de títulos de revistas científicas das áreas de Humanidades e Ciências Sociais Aplicadas nas grandes bases de dados comerciais (CAREGNATO, 2011). Essa carência é ainda mais aguda quando se trata de revistas dessas áreas publicadas em outros idiomas que não o inglês e oriundas de países em desenvolvimento (ARCHAMBAULT; LARIVIÈRE, 2010), impondo sérias barreiras à investigação sobre a visibilidade e impacto desses periódicos.

Essa situação restritiva está mudando face às novas possibilidades proporcionadas pelo surgimento do Acesso Aberto, de novas ferramentas como o motor de busca Google Acadêmico (*Google Scholar - GS*) e de aplicativos de tratamento de dados como o *Publish or Perish* (HARZING, 2007). A crescente migração das revistas das áreas de Humanidades e Ciências Sociais Aplicadas - incluindo as revistas da Ciência da Informação (CI), para o meio *online* está incrementando sobremaneira sua visibilidade na web, permitindo buscas minuciosas através do GS e ainda recuperando as citações ao artigo, sendo esta última funcionalidade considerada de grande potencial para estudos bibliométricos alternativos às bases de dados tradicionais sobre o impacto da produção científica publicada dessas áreas.

No entanto, esse contexto aparentemente favorável para a investigação bibliométrica de áreas por muito tempo desconsideradas pelas grandes bases de dados ainda enfrenta desafios relacionados à confiabilidade do GS como fonte de dados (AGUILLO, 2012; BAR-ILAN, 2010). Frente a esse contexto, este estudo propõe o compartilhamento de dados de citação das revistas de CI extraídos pelo Google Scholar com o objetivo de facilitar e

promover mais investigações pela comunidade de cientometria, no esforço de qualificação do GS nas pesquisas bibliométricas, em especial dos periódicos de CI.

2 2 METODOLOGIA DE EXTRAÇÃO E COMPARTILHAMENTO DE DADOS DE CITAÇÕES DAS REVISTAS BRASILEIRAS DE CI

Para concretizar a ideia de compartilhamento livre e aberto dos dados de citações extraídos do *Google Scholar*, facilitando e estimulando sua exploração, foram eleitas as revistas nacionais de Ciência da Informação como piloto, com os seguintes procedimentos:

- a. Levantamento dos nomes dos periódicos de CI, usando como base o WebQualis da CAPES¹, comparando com listas de periódicos de CI existentes na Internet;
- b. Consulta utilizando o *software Publish or Perish* para Windows, tendo como parâmetro de busca o título do periódico e quando presente, sua variação de nome ou abreviatura oficial, no período de 28/01/2014 a 03/02/2014;
- c. Exportação dos resultados em planilhas do *Google Drive*, uma para cada título de revista recuperado;
- d. Criação de uma planilha consolidando todas as planilhas, filtrando neste consolidado apenas os trabalhos que receberam ao menos uma citação;
- e. Elaboração de consolidados estatísticos utilizando as tabelas dinâmicas do Excel e o *Tableau Public*², que é uma ferramenta muito interessante para criação, edição e tratamento de dados.

O *Google Drive* permite o compartilhamento, e as planilhas elaboradas com os dados pré-tratados foram hospedadas de maneira dinâmica e compartilhadas publicamente através do link: https://docs.google.com/spreadsheets/d/19kcMMnf_5Ohe60_mev-myFc85FkppqRJy-HhXpfB_Q/edit.

3 3 RESULTADOS PRELIMINARES

¹ WebQualis: base de dados de revistas mantida pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), órgão do Ministério da Educação. Disponível em: <http://qualis.capes.gov.br/webqualis/principal.seam>.

² Tableau Public: ferramenta gratuita para criação, edição e visualização de dados. Disponível em: <http://www.tableausoftware.com/public>.

A extração de dados do GS com o *Publish or Perish* resultou no total de 24 revistas nacionais da área de Ciência da Informação, sendo todas em acesso aberto. A busca por título recuperou dados imprecisos - neste levantamento tivemos dois casos, o primeiro da revista “Em questão”, uma vez que existem diversos periódicos que adotam a expressão “em questão” no título; e o segundo da revista “Biblios”, que possui um homônimo publicado em Portugal. Nestes casos, analisamos artigo por artigo e retiramos aqueles fora do escopo da busca.

Os dados obtidos foram dispostos em planilhas dinâmicas do Tableau Public através de vários exercícios preliminares, permitindo sua exploração neste link compartilhado: http://public.tableausoftware.com/views/EstudodascitaesrecebidasporperidicosdaCI/Citaesrec_ebidaspoperidicos?:embed=y&:display_count=no. Como exemplo de um exercício preliminar, a Figura 1 mostra a distribuição das citações obtidas por título de revista.

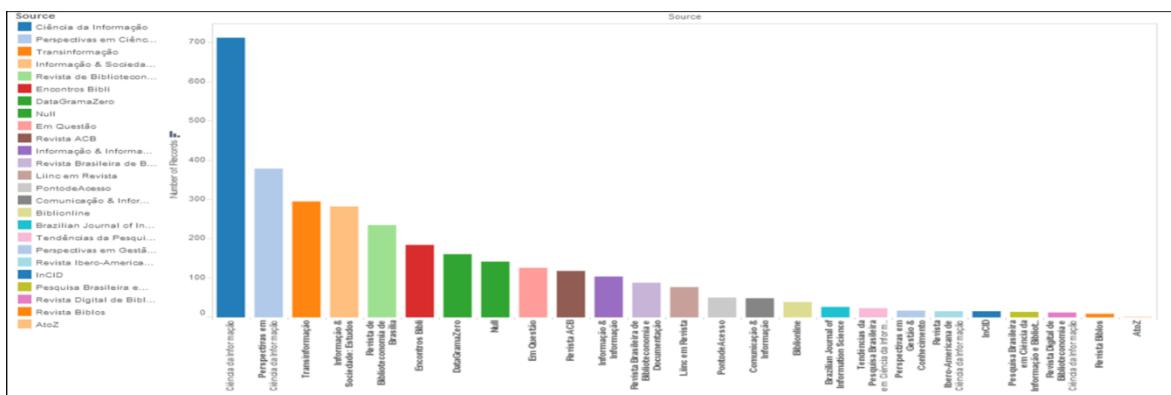


Figura 1: Distribuição das citações recebidas por título de revista (Fonte: Google Scholar)

4 4 CONSIDERAÇÕES FINAIS

O surgimento de aplicativos que facilitam a extração e o tratamento dos dados do GS, como o *Publish or Perish* e ferramentas como o *Google Drive* e *Refine* e o *Tableau Public* descomplica e favorece a tarefa de validação dos dados extraídos do GS, e acreditamos que o compartilhamento público e aberto de dados pré-tratados de citações estimule mais investigações colaborativas pela comunidade de cientometristas brasileiros, visando uma desejada qualificação do *Google Scholar* como fonte alternativa nos estudos métricos de revistas nacionais, minimizando a dependência de fontes comerciais restritas.

5 REFERÊNCIAS

AGUILLO, I. F. Is Google Scholar useful for bibliometrics? A webometric analysis. **Scientometrics**, v. 91, p. 343-351, 2012.

ARCHAMBAULT, E.; LARIVIÈRE, V. The limits of bibliometrics for the analysis of the social sciences and humanities literature. In: UNESCO (Ed.). **2010 World Social Science Report: Knowledge Divides**. Paris: UNESCO; International Social Science Council, 2010. p. 251-254. Disponível em: <http://unesdoc.unesco.org/images/0018/001883/188333e.pdf>. Acesso em 20 fev. 2014.

BAR-ILAN, J. Citations to the “Introduction to informetrics” indexed by WOS, Scopus and Google Scholar. **Scientometrics**, v.82, n. 3, p. 495-506, 2010.

CAREGNATO, S. E. Google Acadêmico como ferramenta para os estudos de citações: avaliação da precisão das buscas por autor. **Ponto de Acesso**, Salvador, v.5, n.3, p. 72-86, dez. 2011.

HARZING, A-W. **Publish or Perish**. 2007. Disponível em <http://www.harzing.com/pop.htm>. Acesso em 20 jan. 2014.