

GOOGLE ACADÊMICO COMO FERRAMENTA PARA OS ESTUDOS DE CITAÇÕES: Avaliação da Precisão das Buscas por Autor

RESUMO -. O Google Acadêmico surgiu na web no final de 2004 com o propósito de ser uma ferramenta de busca de informações de caráter acadêmico. No entanto, por incorporar dados extraídos das referências bibliográficas dos documentos, passou a trazer informações relativas ao impacto dos trabalhos medido pelo número de citações recebidas. Com isso, o uso do mecanismo como fonte de dados para estudos cientométricos passou a ser considerada, especialmente no contexto das ciências que não são cobertas nos índices de citação tradicionais. O objetivo deste trabalho é analisar a validade, o potencial e a cobertura do mecanismo de busca Google Scholar nos estudos de citação das ciências humanas, sociais e sociais aplicadas do Brasil. Focando especificamente o campo da Ciência da Informação no período de 2004 a 2008, buscou-se determinar se a produção dos pesquisadores PQ1 e PQSR do CNPq na área está representada na ferramenta e quais são os limites dessa representação. Para validar os resultados oriundos do Google Scholar, as referências recuperadas foram comparadas à produção do Lattes de cada pesquisador. Os 305 documentos foram agrupados segundo categorias que buscam avaliar a completude e exatidão das referências recuperadas, as fontes de dados e os índices de precisão dos resultados. A qualidade da seleção das fontes de dados e dos registros é uma área na qual o Google Scholar deveria imprimir mais esforços, se o seu objetivo for servir como instrumento de análise de citações e não somente como mecanismo de busca da produção acadêmica.

Palavras-chave: Google acadêmico, índices de citação, avaliação, precisão.

**Sonia Elisa
Caregnato**

Doutora em Ciência
da Informação
Professora
Associada,
Universidade Federal
do Rio Grande do Sul

sonia.caregnato@ufrgs.br

GOOGLE SCHOLAR AS A TOOL FOR CITATION ANALYSIS: Precision in Author Searches

ABSTRACT – Google Scholar appeared on the web in late 2004 with the purpose of being an information retrieval tool of academic nature. However, by incorporating data from the references of documents, allowed impact to be measure by the number of citations a document received. Thus, using the searching engine as a data source for scientometric studies started do be considered, especially in the context of the sciences that are not covered in traditional citation indexes. The objective of this study is to analyze the validity, the potential and the coverage of Google Scholar for citation studies in the humanities and social sciences in Brazil. Focusing specifically the field of information science in the period 2004 to 2008, we sought to determine whether the production of researchers is represented in the tool and what are the limits of its representation. To validate the results from Google Scholar, the references retrieved were compared to the production of each researcher in their Lattes. The 305 documents retrieved were grouped according to categories that seek to assess the completeness and accuracy of the retrieved references, the data sources, and precision rates. Google Scholar should put more efforts in the selection of data sources and in the quality of records, if the goal is to serve as a tool for citation analysis and not just as a search engine for academic production.

Keywords: Google scholar, citation indexes, evaluation

INTRODUÇÃO

Na última década verificou-se o surgimento de novos índices de citação, quebrando aquilo que parecia ser o monopólio absoluto do, então, Institute of Scientific Information (ISI), atualmente Thomson / ISI, desde a década de 1960. Da mesma forma, e provavelmente influenciado pelo impacto dessas novas ferramentas, observou-se o crescimento da aplicação dos métodos e técnicas bibliométricos no estudo da ciência e, especialmente, nos processos de avaliação científica realizados em âmbito nacional, regional, institucional ou individual. Produtividade e impacto são os dois elementos centrais de interesse nos estudos avaliativos que podem ser levantados a partir do uso dos índices de citação.

Tradicionalmente, a cientometria privilegiou o estudo das ciências exatas e biomédicas, pela visibilidade internacional das publicações e pela existência de ferramentas apropriadas. Isso é especialmente válido quando se busca estudar o impacto das publicações, calculado por meio das citações recebidas, o que só podia ser realizado, até recentemente, através da única base estruturada para contabilizar este fator, a base ISI, publicamente comercializado como Web of Science (WoS).

A crescente disponibilidade de publicações científicas em formato eletrônico disponível na web e a grande capacidade de processamento digital têm ampliado a oportunidade de se obter uma visão mais complexa e pormenorizada das redes formadas pelas comunidades científicas das diferentes áreas, por meio das citações que os pesquisadores fazem aos trabalhos uns dos outros. Aliado a essas possibilidades, observa-se um interesse crescente dos indivíduos, das instituições e mesmo da sociedade nos resultados das pesquisas realizadas com financiamento público, seja para conhecer e aperfeiçoar as formas e instrumentos de comunicação científica, seja para fins de avaliação da ciência, ou, então, como instrumento de gestão. Como resultado, tem-se o ressurgimento do interesse pela bibliometria em geral, e pela análise de citação em particular, a qual é descrita por Meho e Rogers (2008, p.1711) como um “[...] método

poderoso e popular para examinar e mapear o impacto intelectual dos cientistas, projetos, periódicos, disciplinas e nações”.

A WoS é ainda a fonte mais utilizada e conhecida para a análise de citações, mas além do prestígio, a base de dados recebe também muitas críticas, especialmente quando utilizada para a avaliação da atividade científica. Entre essas críticas destaca-se o número relativamente reduzido de títulos indexados frente ao que se estima haver no mundo, o predomínio da literatura em inglês, o número reduzido de revistas dos países em desenvolvimento, a cobertura limitada de outros tipos de documentos além do artigo de periódico, e a representação exagerada de algumas áreas das ciências, especialmente as duras, em detrimento das ciências sociais e humanidades (ETXEBARRIA; GOMEZ-URANGA, 2010; HARZING; WAL, 2008; MEHO; YANG, 2007; MUGNAINI, JANNUZZI, QUONIAN, 2004). É importante salientar, no entanto, que a base ISI não foi concebida para ser utilizada como instrumento de avaliação da ciência ou dos cientistas, mas para ser um índice da literatura científica que permitisse a recuperar documentos a partir da associação de ideias por meio das citações (GARDIELD, 1955; GARFIELD; WELLJAMS-DOROF, 1992). A existência da base permitiu que se criasse o fator de impacto dos periódicos, cujo objetivo era auxiliar no processo de seleção de periódicos (GARFIELD, 2006)

Assim, encontram-se dificuldades para a realização de estudos de citação para a avaliação de impacto das áreas das ciências humanas, sociais e sociais aplicadas, especialmente nos países periféricos (incluindo o Brasil). Isso se deve, principalmente, ao fato de que estas áreas não contam com a tradição de publicação em revistas indexadas nas principais bases de dados internacionais, pois privilegiam outras formas de publicação além do artigo (livros, capítulos, etc.), que não são normalmente indexadas na WoS nem em outras bases de dados, e por publicarem essencialmente em suas línguas maternas.

GOOGLE ACADÊMICO: ALGUMAS CONSIDERAÇÕES

Melho e Yang (2007) argumentam que o surgimento recente de novas bases de dados de citação multidisciplinares, como a Scopus e o Google Acadêmico, desafiam a

dominação da WoS e levantam dúvidas sobre sua utilização exclusiva em estudos bibliométricos, de citação e sobre comunicação científica. Adiciona-se a isso que o Google Acadêmico, diferentemente das duas bases de dados multidisciplinares (WoS e Scopus), é uma ferramenta gratuita, que permite localizar trabalhos acadêmicos de vários tipos (por exemplo, artigos de congressos, teses e dissertações, além de artigos de periódicos de acesso aberto ou pagos), em múltiplas línguas (inclusive português), disponibilizadas em repositórios na web ou sites acadêmicos, além de determinar a frequência com que foram citados em outras publicações acadêmicas.

Mesmo que tenha sido projetada para ser uma ferramenta de recuperação da informação e não uma ferramenta para a avaliação da atividade de pesquisa - assim como o foi o ISI/WoS - há sugestões na literatura de que o Google Acadêmico tem o potencial de servir como um índice de citações para estudos bibliométricos (NORUZI, 2005), que pode se tornar a fonte principal para mensuração do impacto das pesquisas (VAUGHAN; SHAW, 2008), e que é especialmente apropriado para a análise do impacto/visibilidade de periódicos de acesso aberto (MUKHERJEE, 2009).

Assim, parte-se do pressuposto de que o Google Acadêmico pode ser instrumento extremamente útil para os estudos de citação, especialmente nas ciências humanas, sociais e sociais aplicadas, mas esta promessa necessita ser testada, especialmente porque há, paralelamente aos elogios, fortes críticas a suas inconsistências, *performance* e cobertura duvidosa (JACSÓ, 2008). É fundamental reconhecer os limites na cobertura das bases de dados antes de se conduzir pesquisas sérias, pois ela representa um elemento importante nos estudos bibliométricos e de citação, podendo causar distorções e até invalidar os resultados (FRANSEN; NICOLAISEN, 2008).

Embora já existam alguns estudos na literatura internacional sobre o Google Acadêmico, especialmente os que discutem os seus problemas e potencialidades, a maioria é de natureza exploratória e se utiliza de amostras pequenas (KOUSHA; THELWALL, 2008). Adiciona-se a isso o fato que grande parte deles busca avaliar os resultados produzidos no Google Acadêmico pela comparação com as outras duas bases multidisciplinares (WoS e Scopus), o que não é viável para produção das ciências

humanas, sociais e sociais aplicadas, originadas em países como o Brasil, cujo percentual de indexação nestas bases é ínfimo. Finalmente, a maioria deles se concentra em contar as citações e não em analisá-las do ponto de vista da sua validade e exatidão (BAR-ILLAN, 2010), como é a proposta seguida neste trabalho.

As principais características do Google Acadêmico podem ser detalhadas na comparação com os demais índices de citação multidisciplinares. Diferentemente daqueles, que indexam os artigos de revistas selecionadas, o Google Acadêmico coleta dados a partir de um software que rastreia a web e reconhece automaticamente os campos que compõem os documentos científicos e suas referências. Com isso, viu-se surgir uma ferramenta sem precedentes, que oferece acesso gratuito a dados bibliográficos e também a milhões de artigos na íntegra (JACSÓ, 2010)

Para rastrear os dados bibliográficos desse número espantoso de textos acadêmicos, o Google Acadêmico recebe o consentimento dos editores de periódicos, das sociedades científicas e dos provedores de bases de dados e repositórios para utilização dos metadados de suas coleções, quando se tratam de serviços protegidos. Mas os dados podem vir também de arquivos de *preprints* ou até de páginas pessoais, se pertencentes a um domínio (ou URL) acadêmico (BAR-ILLAN, 2008).

Contudo, conforme salientam Torres-Salinas *et al.* (2009), tal qual todo método automático, esse também apresenta problemas, como reconhecer a autoria em outros elementos constitutivos do corpo do trabalho. Jacsó, na sua série de estudos sobre o Google Acadêmico, documentou extensivamente erros desse tipo (por exemplo, o emprego das palavras “Background” ou “Password” como autores fantasmas), o que ele entende ser um dos maiores problemas do mecanismo de busca (JACSÓ, 2010).

Considerando que grande parte da sua produção científica nas ciências humanas, sociais e sociais aplicadas do Brasil não está indexada nas principais bases (WoS ou Socups), é veiculada também como tipos de documentos alternativos (por exemplo, trabalhos em eventos) e é registrada em língua portuguesa, delimitou-se a seguinte questão norteadora do trabalho: o Google Acadêmico é um instrumento adequado para os estudos de citação nas ciências humanas, sociais e sociais aplicadas do Brasil?

A partir do exposto, o objetivo deste trabalho é analisar a validade, o potencial e a cobertura do mecanismo de busca Google Acadêmico nos estudos de citação das ciências humanas, sociais e sociais aplicadas do Brasil. Focando especificamente o campo da Ciência da Informação no período de 2004 a 2008, busca-se determinar se a produção dos pesquisadores da área está representada na ferramenta e quais são os limites dessa representação.

A opção por analisar a ferramenta a partir da produção recuperada por autor é explicada pela importância que as medidas de reconhecimento científico de indivíduos têm alcançado na avaliação da pesquisa. Exemplo disso é prestígio que o índice H alcançou como um indicador simples e robusto, que pode ser estendido para a avaliação de periódicos (BRAUN; GLANZEL; SCHUBERT, 2006), e cuja agregação permite medir também instituições e países (GARCIA-PEREZ, 2010).

O recorte temporal desta pesquisa, 2004 a 2008, justifica-se, pois neste período observou-se o surgimento do mecanismo Google Acadêmico (em 2004) e o aumento da publicação científica na web. O estabelecimento do ano de 2008 como data limite, por outro lado, permitiu que se abrisse uma janela de dois anos desde a publicação do trabalho, tempo padrão para as avaliações de impacto.

PROCEDIMENTOS METODOLÓGICOS

O primeiro passo consistiu em definir a unidade de análise, no caso pesquisador, e localizar um instrumento para validação dos resultados encontrados. Assim, o grupo dos pesquisadores brasileiros da área da Ciência da Informação foi delimitado a partir do reconhecimento institucional do CNPq, na forma de concessão de bolsa de produtividade em pesquisa (PQ). Para fins de operacionalização, foram selecionados apenas aqueles com bolsas classificadas nos níveis mais elevados (PQ1 ou PQSR) no ano de 2009, num total de 13 pesquisadores.

Nos demais estudos realizados, o instrumento de validação tem sido usualmente uma ou as duas bases de citação multidisciplinares: WoS ou Scopus. No caso das ciências

humanas, sociais e sociais aplicadas do Brasil, esses instrumentos não são os mais apropriados, assim como também não é o único sistema nacional que mede citações – Scielo – pela baixa representatividade da disciplina selecionada. Dessa forma, foi necessário construir um conjunto de dados de referência. Esse foi composto pelas publicações dos pesquisadores acima referidos, no período de 2004 a 2008, recuperada de seus respectivos Currículo Lattes no mês de junho de 2010.

A coleta de dados no mecanismo Google Acadêmico foi realizada em maio de 2010. A partir da pesquisa avançada, buscou-se o nome dos pesquisadores, no campo autor, a partir de várias formas de expressão do nome, entre aspas, unidas pelo operador booleano OR, a fim de ampliar ao máximo o número de resultados. Por exemplo, a estratégia de busca para obtenção da produção do professor Jaime Robredo ficou assim definida:

autor:"j robredo" OR autor:"robredo j" OR autor:"jaime robredo"

Além da busca no campo autor e da utilização do operador, a restrição de ano foi utilizada para delimitar o período de 2004 a 2008.

Para validar os resultados recuperados pelo Google Acadêmico, as referências recuperadas foram comparadas à lista da produção obtida através dos currículos Lattes de cada pesquisador.

RESULTADOS E DISCUSSÃO

A produção total dos 13 pesquisadores representada no Google Acadêmico durante o período foi de 305 documentos, os quais receberam 580 citações até a data da coleta de dados, ou seja, até maio de 2010.

Após a checagem com os dados do Lattes, os 305 registros foram organizados em categorias que levam em conta a sua completude e exatidão.

As categorias foram assim delimitadas:

- a) “correta”: a referência recuperada está de acordo com os dados do Lattes;

- b) “incompleta”: na referência recuperada faltam alguns dados, por exemplo, título da publicação, data, páginas, entre outros;
- c) “imprecisa”: a referência recuperada apresenta que algumas informações (por exemplo, datas e/ou número de páginas) não condizem com os dados do Lattes;
- d) “incompleta/imprecisa”: a referência recuperada omite alguns dados, ao mesmo tempo que apresenta dados que não condizem com aqueles do Lattes;
- e) “duplicata”: referência recuperada está em duplicata, sem considerar a existência ou não de caracteres incorretos, entre outros problemas;
- f) “equivocada”: a referência recuperada não é do autor que se buscava;
- g) “errada”: a referência recuperada não fornece nenhum tipo de indicação de autor, (por exemplo, um endereço no campo de autor), o que Jacsó (2010) chama de fantasmas,.

O resultado da categorização dos 305 registros recuperados no Google Acadêmico é mostrado na Figura 1.

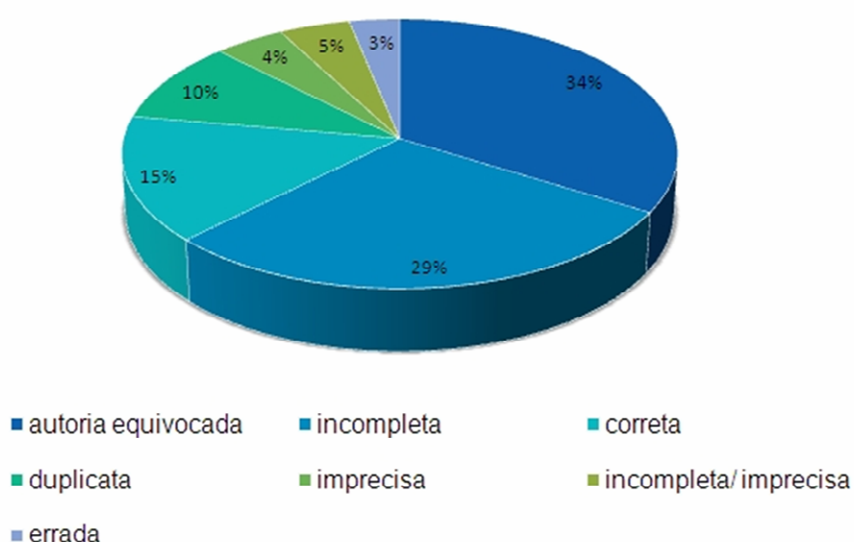


Figura 1: Distribuição das referências bibliográficas dos pesquisadores brasileiros da CI, recuperadas a partir do Google Acadêmico, 2004-2008. (n=305)

Observa-se a grande incidência de referências com problemas, já que somente 15% delas estão totalmente corretas. Isso decorre, em parte, da utilização de técnica de processamento automático das referências (HARZING, 2008). Como a maior incidência de registros que apresentaram problemas é na categoria “autoria equivocada” (34%), ou seja, são de autores homônimos aos buscados, entende-se que o problema maior, no caso estudado, deve-se à falta de controle de autoridade – situação também presente nos demais índices de citação – e à limitada interface de busca do Google Acadêmico, que não permite que se construam estratégias com vários delimitadores de campo ou de operação nas buscas avançadas (JACSÓ, 2008; GARCIA-PEREZ, 2010; TORRES-SALINAS *et al.*, 2009) de maneira a reduzir a recuperação de itens falso positivos.

Outro aspecto relevante é a existência de referências duplicadas, o que representou 15% dos documentos recuperados neste estudo. Tal proporção não é similar às encontradas em outros estudos, como por exemplo, os conduzidos por Bar-Illan (2010) e Chen (2010) que encontraram poucas referências deste tipo.

Chen (2010) avalia que a política de remoção de referências duplicadas do Google Acadêmico permite a diminuição do problema, mas como ela não é clara, pode ocasionar a manutenção de registros “ruins” em detrimento dos “bons”. O autor explica que “bons”, neste caso, são os registros originais dos editores (por exemplo, Emerald, ScienceDirect e SpringerLink) ou de bases de dados confiáveis (por exemplo, CAT.INIST, JSTOR, e PubMed). Registros “ruins” variam muito, mas geralmente são provenientes de páginas pessoais na web que pertençam a um domínio acadêmico, de base de dados não estável ou são registros de citações breves sem *link* para o texto completo. No caso dos pesquisadores brasileiros da Ciência da Informação, essa política não parece estar sendo completamente implementada.

No que tange aos tipos de documentos referenciados, os 163 registros que correspondem às categorias corretas, imprecisas, incompletas ou incompleto-imprecisas são 76 artigos, 6 orientações, 23 participações em bancas, 7 apresentações de trabalhos, 6 livros, 22 capítulos de livros, 19 anais e 3 produções diversas, conforme ilustra a Figura 2.

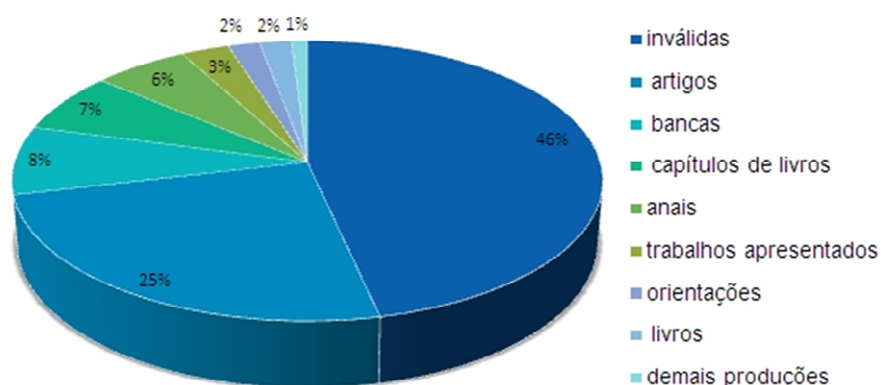


Figura 2: Distribuição das referências bibliográficas dos pesquisadores brasileiros da CI, recuperadas a partir do Google Acadêmico, segundo a tipologia, 2004-2008. (n=163)

Há aqui a evidência de que itens relativos às atividades acadêmicas de um pesquisador, mas não “literatura acadêmica” (GOOGLE ACADÊMICO, [2011]), também são indexados pela ferramenta, como os itens relativos à participação em bancas e orientações. Esse tipo de resultado é, muito provavelmente, proveniente dos repositórios institucionais ou bibliotecas digitais de teses e dissertações, já comuns nas universidades brasileiras. Esse é certamente um ponto que merece atenção dos desenvolvedores do Google Acadêmico e que demonstra, conforme pontuou Jacsó (2010), a falta de testes de controle da qualidade dos dados.

Para fins de verificação do desempenho do Google Acadêmico na recuperação da informação, calculou-se o índice de precisão para cada autor. O índice de precisão é uma medida da relevância da recuperação da informação, que é representado pela porção dos documentos recuperados considerados relevantes na busca. Para fins de cálculo, foram excluídos dados relativos à participação em bancas e orientações. O resultado é mostrado na Tabela 1.

Pesquisador	Nº Doc. Recuperados	Nº Doc. Relevantes	Índice de Precisão¹
1	20	4	20%
2	9	7	78%
3	21	16	77%
4	24	15	62%
5	64	1	1%
6	7	2	30%
7	16	8	50%
8	34	27	80%
9	23	14	60%
10	8	6	75%
11	17	11	60%
12	20	9	45%
13	42	14	33%
Total	305	134	52%

Tabela 1: Precisão da recuperação das referências bibliográficas de cada pesquisador brasileiro da CI, obtidas no Google Acadêmico, 2004-2008

Para um autor, o índice de precisão chegou a ser tão baixo quanto 1%. Não está clara a razão para isso, exceto o fato de que se trata de pessoa que utiliza duas iniciais como prenomes e que tem um sobrenome bastante comum. A recuperação para os demais autores, em geral, foi mais precisa, chegando a 80% para um deles.

Walter (2010) também utilizou o índice de precisão para avaliar os resultados das buscas de um conjunto de 155 artigos previamente selecionados no Google Acadêmico e em outras 11 bases de dados. Para ele, o Google Acadêmico obteve um desempenho melhor do que a maioria das outras bases de dados. Na verdade, ele chega a esta conclusão porque considera o número de resultados de busca a serem examinados. Ou seja, se forem considerados somente os primeiros 30 resultados, então a precisão do Google Acadêmico é de 53%, ao passo que se forem considerados todos os 300 resultados, então o índice cai para 21% e a ferramenta ficaria na oitava posição em precisão, entre as 12 bases de dados estudadas.

¹ Precisão = (número de documentos relevantes recuperados) / (número total de documentos recuperados)
* 100

O mesmo autor acredita que utilizar todos os 300 resultados para calcular a precisão não é justo, porque o Google Acadêmico recuperou muitos mais registros do que as outras bases de dados. Ora, o que não parece imparcial é excluir parte desses resultados, já que a precisão busca exatamente calcular a capacidade da ferramenta de excluir itens não-relevantes. Se o Google Acadêmico ainda não tem essa capacidade, por incluir referências duplicadas ou por incluir itens relativos às atividades acadêmicas dos autores (bancas e orientações, por exemplo) ou por outras razões, ele ainda não consegue atingir índices de precisão aceitáveis.

Tomando-se o conjunto da produção dos 13 pesquisadores brasileiros da Ciência da Informação, tem-se um índice de precisão de 52%, ou seja, quase metade dos itens recuperados não é adequada à busca pelo nome do autor. Esse número mostra que ainda nos encontramos distantes da possibilidade de utilização da ferramenta para avaliação séria da produtividade e impacto de pesquisadores e, conseqüentemente, de agrupamentos originados deles, como grupos de pesquisa, instituições, regiões ou países.

Aparentemente, se o objetivo desse trabalho fosse utilizar o Google Acadêmico para calcular o índice H dos pesquisadores, teríamos presenciado sérias distorções. A boa notícia é que estudos mostram que após cinco anos de seu nascimento o Google Acadêmico mostra uma dramática melhora na sua cobertura (CHEN, 2010). Resta esperar que esta melhora continue indefinidamente.

CONCLUSÕES

O Google Acadêmico mostrou ser capaz de recuperar um número grande de documentos não presentes nos índices de citação tradicionais, tais como livros, capítulos de livros e trabalhos acadêmicos escritos em português. Os estudos de citação na área das ciências sociais, cujas características de comunicação científica estão centradas neste tipo de publicação, poderiam ser conduzidos com a utilização desta ferramenta.

Esta característica de sensibilidade na recuperação de documentos, também foi detectada na localização de trabalhos que não estavam presentes nos currículos Lattes dos pesquisadores analisados.

Entre os problemas encontrados, pode-se destacar a ocorrência de muitos documentos duplicados e escritos de forma incorreta, o que exige do pesquisador um tempo redobrado na limpeza de dados. A qualidade dos dados oriundos do próprio material bibliográfico distribuído na rede é, muitas vezes, responsável por esses problemas. Por outro lado, o método automático de reconhecimento de metadados empregado pelo Google Acadêmico também apresenta dificuldades, por exemplo, reconhecendo autoria em elementos constitutivos do corpo do trabalho.

Portanto, a qualidade da seleção das fontes de dados e dos registros é uma área na qual o Google Acadêmico deveria imprimir mais esforços, se o seu objetivo for servir como instrumento de análise de citações e não somente como mecanismo de busca da produção acadêmica.

É preciso ressaltar que mesmo as duas base de dados de citação mais consolidadas (WoS e Scopus) apresentam inconsistências nos dados, as quais são frequentemente relatadas na literatura. A pouca maturidade do Google Acadêmico e o fato de ser um mecanismo gratuito e aberto talvez autorizem a existência de inconsistências por mais algum tempo

O Google Acadêmico como ferramenta para os estudos de citação foi, até o momento, analisado quanto às características de consistência e cobertura na recuperação de dados. É necessário dar continuidade as análises para investigar minuciosamente todo o seu potencial e adequação aos estudos de citações.

AGRADECIMENTOS

A autora agradece o apoio financeiro recebido do CNPq para a realização desta pesquisa, assim como o auxílio da bolsista de iniciação científica Natasha Morello Peixoto.

Referência

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.

BAR-ILAN, J. The h-index of h-index and of other informetric topics. **Scientometrics**, v.75, n. 3, p. 591–605, 2008.

BRAUN, T.; GLANZEL, W.; SCHUBERT, A. A Hirsch-type index for journals. **Scientometrics**, v. 69, n. 1, p. 169-73, 2006.

CHEN, X. Google Scholar’s dramatic coverage improvement five years after debut. **Serials Review**, v. 36, n.4, p. 221-226, 2010.

ETXEBARRIA, G.; GOMEZ-URANGA, M. Use of Scopus and Google Scholar to measure social sciences production in four major Spanish universities. **Scientometrics**, v. 82, n. 2, p. 333-349, 2010.

FRANSEN, T. F.; NICOLAISEN, J. Intradisciplinary differences in database coverage and the consequences for bibliometric research. **Journal of the American Society for Information Science and Technology** , v.59, n.10 p. 1570-1581, 2008.

GARCIA-PEREZ, M. A. Accuracy and Completeness of Publication and Citation Records in the Web of Science, PsycINFO, and Google Scholar: A Case Study for the Computation of h Indices in Psychology. **Journal of the American Society for Information Science and Technology**, v.61, n.10, p.2070-2085, 2010.

GARFIELD, E. Citation indexes to science: a new dimension in documentation through association of ideas. **Science**. v.122, p. 108-111, 1955.

GARFIELD, E. The History and Meaning of the Journal Impact Factor. **JAMA: The Journal of the American Medical Association**, v. 295, n. 1, p. 90 -93, 2006.

GARFIELD, E. WELLJAMS-DOROF, A. Citation data: their use as quantitative indicators for science and technology evaluation and policy-making. **Science and Public Policy**, v. 19, n. 5, p. 321-327, 1992.

GOOGLE ACADÊMICO. **Sobre o Google Acadêmico**. [2011] Disponível em: <http://scholar.google.com.br/intl/pt-BR/scholar/about.html> Acesso em: 10 out. 2011.

HARZING, A. W.; WAL, R. Google Scholar—a new data source for citation analysis. **Ethics in Science and environmental Politics**, v.8, p. 61-73, 2008.

JACSÓ, P. Google Scholar revisited. **Online Information Review**, v. 32, n.1, p.102-114, 2008.

JACSÓ, P. Metadata mega mess in Google Scholar. **Online Information Review**, v. 34, n. 1, p.175-191, 2010.

KOUSHA, K.; THELWALL, M. Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. **Scientometrics**, v. 74, n. 2, p273–294, 2008.

MEHO, L. I.; ROGERS, Y. Citation counting, citation ranking, and *h*-index of human-computer interaction researchers: A comparison between Scopus and Web of Science. **Journal of the American Society for Information Science and Technology**, v. 59, n.11, p. 1711–1726, 2008.

MEHO, L. I.; YANG, K. Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. **Journal of the American Society for Information Science and Technology**, v. 58, n.13, p. 2105–2125, 2007.

MIKKI, S. Comparing Google Scholar and ISI Web of Science for Earth Sciences. **Scientometrics**, v. 82, p. 321-331, 2010.

MUGNAINI, R.; ANNUZZI, P.; QUONIAM, L. Indicadores bibliométricos da produção científica brasileira: uma análise a partir da base Pascal. **Ciência da Informação**, Brasília, v.33, n.2, p.123-131, maio/ago.2004.

MUKHERJEE, Bhaskar. Do open-access journals in library and information science have any scholarly impact? A bibliometric study of selected open-access journals using Google Scholar. **Journal of the American Society for Information Science and Technology**, v. 60, n.3, p. 581-594, 2009.

NORUZI, A. Google Scholar: The new generation of citation indexes. **LIBRI**, v.55, n.4, p. 170–180, 2005.

TORRES-SALINAS, D.; RUIZ-PÉREZ, R.; DELGADO-LÓPEZ-CÓZAR, E. Google Scholar como herramienta para la evaluación científica. **El Profesional de la Informacion**, v. 18, n. 5, p. 501-510, 2009.

VAUGHAN, L.; SHAW, D. A new look at evidence of scholarly citation in citation indexes and from web sources. **Scientometrics**, v. 74, n.2, p. 317–330, 2008.

WALTERS, W. H. Google Scholar Search Performance: Comparative Recall and Precision. **portal: Libraries and the Academy**, v. 9, n. 1, p. 5-24, 2009.