



XVI Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB)
ISSN 2177-3688

GT 07 - Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação

Pôster

ANÁLISE DE CITAÇÃO: RELAÇÃO ENTRE REFERÊNCIAS E MENÇÕES EM ARTIGOS¹

CITATION ANALYSIS: RELATION BETWEEN REFERENCES AND MENTIONS IN ARTICLES

Rodrigo Aquino de Carvalho, UFRGS
racfurg@gmail.com

Sonia Elisa Caregnato, UFRGS
sonia.caregnato@ufrgs.br

Resumo: Apresenta uma perspectiva para os estudos de citação a partir da relação entre a lista de referências e as menções desses documentos no texto. Os objetivos são: i) identificar a frequência de menção no texto dos artigos aos documentos listados nas suas referências; e ii) comparar um ranking de autores mais citados com autores mais mencionados no campo da Recuperação da Informação no Brasil. Trata-se de um estudo cientométrico e exploratório, para o qual foram utilizados 66 artigos, publicados entre 2008 e 2012, indexados na base de dados BRAPCI e recuperados a partir de 12 termos relacionados à “recuperação da informação”. Os artigos apresentaram 1467 referências, sendo que 60% (n=881) são mencionadas uma única vez, 31% (n=450) são mencionadas duas ou mais vezes e 9% (n=136) não são mencionadas. O número de menções é de 2246, não considerando 136 menções não referenciadas. O *ranking* elaborado pelo estudo mostrou que dos 21 autores mais citados (oito ou mais citações), 15 estão entre os mais mencionados. Conclui-se que há um núcleo de autores que são mais mencionados e que são basicamente os mais citados e isso valida os estudos de citação a partir da lista de referências. Sugere-se a ampliação do recorte temporal desse teste, assim com a sua aplicação em outras áreas do conhecimento.

Palavras-chave: Análise de citação. Cientometria. Comunicação científica.

Abstract: It presents a perspective to citation studies through the relation between references list and

¹ O conteúdo textual deste artigo, os nomes e e-mails foram extraídos dos metadados informados e são de total responsabilidade dos autores do trabalho.

mentions to those documents in the text. The objectives are: i) identify if the frequency articles mention documents listed in their references; and ii) compare a ranking of authors most cited with authors most mentioned in the field of Information Retrieval in Brazil. It is a scientometric and exploratory study, to which were used 66 articles, published between 2008 and 2012, indexed in the BRAPCI database and retrieved using 12 terms related to "information retrieval". The articles presented 1467 references, of which 60% (n=881) are mentioned once, 31% (n=450) are mentioned two or more times and 9% (n=136) are not mentioned. The number of mentions is 2246, not considering 136 mentions not referenced. The ranking prepared by the study showed that of the 21 most cited authors (eight or more citations), 15 are also the most frequently mentioned. The conclusion is that there is a core of authors who are most frequently mentioned and cited and this evidence validates citation studies carried out through the reference list. It suggests expanding the time and its application in other areas of knowledge.

Keywords: Citation analysis. Scientometrics. Scientific communication.

1 INTRODUÇÃO

Os estudos de citação são fundamentais para entender o comportamento de produção e uso de informação na ciência. Utilizar outros trabalhos para gerar conhecimento e apresentar uma lista desse material é elemento indispensável na publicação de uma tese, artigo de pesquisa etc.

O presente trabalho tem como tema os estudos de citação e busca responder os dois objetivos que seguem: i) identificar se existe empiricamente a tendência de um artigo mencionar um documento mais de uma vez; e ii) comparar um ranking de autores mais referenciados com autores mais mencionados no campo da Recuperação da Informação no Brasil.

Vale salientar que “menção” nesse trabalho é entendida como o uso efetivo, no texto de um artigo, de um documento presente na sua lista de referências, ou seja, um documento citado (ou referenciado) deve ser mencionado no texto ao menos uma vez para caracterizar a citação. Essa forma de nomear a relação citação/referência com citação/menção é utilizada por Silveira e Caregnato (2014), já Ding *et al* (2013)² utilizam *citação* para uma menção e *recitação* para outras menções da mesma referência. Nenhum dos trabalhos pretendeu discutir questões terminológicas que envolvem a análise de citação.

2 ASPECTOS CONCEITUAIS E TRABALHOS RELACIONADOS

A citação, por Reitz (2014) e Cunha e Cavalcanti (2008), é entendida como uma menção à opinião ou argumento de uma autoridade sobre algum assunto, e nesse sentido, a

² “[...]cited/re-cited in the body of the text [...]”, p.591.

citação deve, também, indicar o “caminho” para verificação dessa menção, a referência bibliográfica.

Romancini (2010) diz que a dimensão básica e fundamental da citação é relacionar dois participantes de um circuito científico, “o par ‘autor (texto) citado/autor (texto) citante” (ROMANCINI, 2010, p. 20). Essa relação se mostra muitas vezes política (social) e assim “forçada” e conveniente, mas também se apresenta com caráter cognitivo (epistemológico).

O certo é que, somente estudando a lista de referências é mais difícil perceber algumas facetas que caracterizam as citações efetivamente, como mostram Brambilla, Vanz e Stumpf (2006), quando mapeiam as razões de citação de um determinado artigo, Silveira e Caregnato (2014), quando propõem uma classificação para entender as razões que levam a citação, Ding *et al* (2013), quando apresentam a distribuição das menções das referências na estrutura textual dos artigos de um periódico e Jeonga, Songa e Ding (2014), que apresentam uma análise de co-citação de autores baseado em conteúdo, mostrando que os pares formados, a partir das menções, são mais específicos e favorecem uma análise de domínio mais precisa.

Evidente que os estudos de citação a partir das referências são necessários e devem continuar a ser realizados, pois, como indicam Vanz e Caregnato (2003), possibilitam identificar características e mapear a comunicação científica, além de servirem para atividades de gestão para políticas científicas. Mas parece necessário avançar em técnicas que possam transpor essa tendência, desde que isso seja necessário para mensurar a produção científica, identificando com mais especificidade e clareza as características que moldam a produção e uso da informação.

Assim, esse trabalho quer mostrar se há diferenças quantitativas significativas entre referências e menções e ampliar a visão para rankings de citações feitos a partir das referências.

3 PROCEDIMENTOS METODOLÓGICOS

O trabalho se caracteriza por ser empírico, cientométrico e exploratório. O *corpus* do estudo foi constituído de 66 artigos, recuperados na base BRACPI³, nos meses de fevereiro e março de 2015, publicados entre 2008 e 2012 (cinco anos). Documentos publicados entre 2013 e 2014 não foram incluídos, pois apresentaram poucos resultados, indicando que números atrasados de alguns periódicos ainda não estavam na base.

³ Base de dados referencial de artigos de periódicos em Ciência da Informação. Disponível em: <http://www.brapci.ufpr.br/>

As expressões que seguem foram utilizadas, como frase (entre aspas), para realizar as buscas, com ocorrência dos termos no “título” e nas “palavras-chave”: a) RECUPERAÇÃO DA INFORMAÇÃO; b) RECUPERAÇÃO DE INFORMAÇÃO; c) RECUPERAÇÃO INTELIGENTE DA INFORMAÇÃO; d) RECUPERAÇÃO INTELIGENTE DE INFORMAÇÃO; e) RECUPERAÇÃO DE DADOS; f) ARMAZANAMENTO DE DADOS; g) ARMAZENAGEM DE DADOS; h) COLETA DE DADOS; i) BUSCA DE INFORMAÇÃO; j) BUSCA DA INFORMAÇÃO; k) *INFORMATION RETRIEVAL*; e, l) *RETRIEVAL OF INFORMATION*.

Após a definição do *corpus*, as referências de cada artigo foram colocadas em um arquivo de texto criado para cada ano, onde foram destacados os autores e as datas de publicação dos documentos referenciados (AUTOR A 1999), além dos seguintes dados: número de referências; número de menções às referências no corpo do texto, separando as que apresentavam uma menção das que apresentavam duas ou mais menções; número de referências não mencionadas; e número de menções não referenciadas. A Figura 1 ilustra como os dados foram organizados.

Figura 1 - Exemplo de organização dos dados coletados.

ARTIGO 01 de 2008			
Referências (lista no final do artigo)		menção do texto	
1. AUTOR A 1900		01	
2. AUTOR B 1999		02	
	AUTOR A 1999		
Nº de referências:	08	Nº de menções no corpo do texto:	11
Nº de referências com 01 menção:	06	Nº de ref. com 02 ou + menções:	02
Nº de menções não referenciadas:	00	Nº de referências não mencionadas:	00

Fonte: dados da pesquisa, 2015.

O trabalho foi realizado manualmente nos 66 artigos e o *software Excel* foi utilizado para agrupar e tratar os dados coletados. A Tabela 1 apresenta a distribuição dos artigos, que foram publicados em 19 diferentes periódicos.

Tabela 1 - Corpus da pesquisa, por ano de publicação e estrato dos periódicos na base *Qualis*.

	2008	2009	2010	2011	2012	Total por extrato <i>Qualis</i> (%)	
Revistas A1	02	02	06	05	02	17	25,75
Revistas B1	06	06	07	11	09	39	59,09
Revistas B2, B3 e B5	02	03	02	02	01	10	15,16
Total por ano	10	11	15	18	12	66	100,00
%	15,15	16,67	22,73	27,27	18,18		

Fonte: dados da pesquisa, 2015.

Para colaborar na análise de comparação entre o número de referências com o número de menções, foram utilizadas outras duas categorias: “número de artigo citantes”; e “pico de menção de um único documento referenciado”. Foi medido o coeficiente de correlação de

*Spearman*⁴ para as quatro categorias no *software* SPSS⁵, considerando nível de significância igual a 0,05.

4 PRINCIPAIS RESULTADOS E DISCUSSÃO

Os resultados iniciais mostram que existe o comportamento de mencionar mais de uma vez um documento referenciado (Tabela 2). As 1467 referências geraram 2246 menções. Além disso, existe uma proporção significativa maior de documentos mencionados apenas uma vez em relação aos que são mencionados duas ou mais vezes. Os dados discrepantes são as “menções não referenciadas” (n=66) e as “referências não mencionadas” (n=136), mas esses resultados não parecem ser uma tendência, uma vez que a mediana e a moda das categorias foi igual a zero.

Tabela 2 - Distribuição do número de referências e menções no corpus da pesquisa.

Categorias	Total	Média	Mediana	Desvio padrão	Moda
Número de referências	1467	22,23	19	11,86	17
Número de menções no texto completo	2246	34,03	30	21,30	18
Número de referências com uma menção	881	13,35	12	8,17	06
Número de referências com duas ou mais menções	450	6,82	06	5,57	03
Número de menções não referenciadas	66	1,00	00	1,90	00
Número de referências não mencionadas	136	2,06	00	4,07	00

Fonte: dados da pesquisa, 2015.

A Figura 2 mostra que 31% das referências foram mencionadas mais de uma vez. Parece necessário estudar outros contextos (áreas de conhecimento e tipos de publicação) para verificar se isso é um comportamento. Voos e Dagaev (1976⁶ *apud* DING *et al*, 2013) sugerem que o número de menções de uma referência indica a relevância para o assunto do documento.

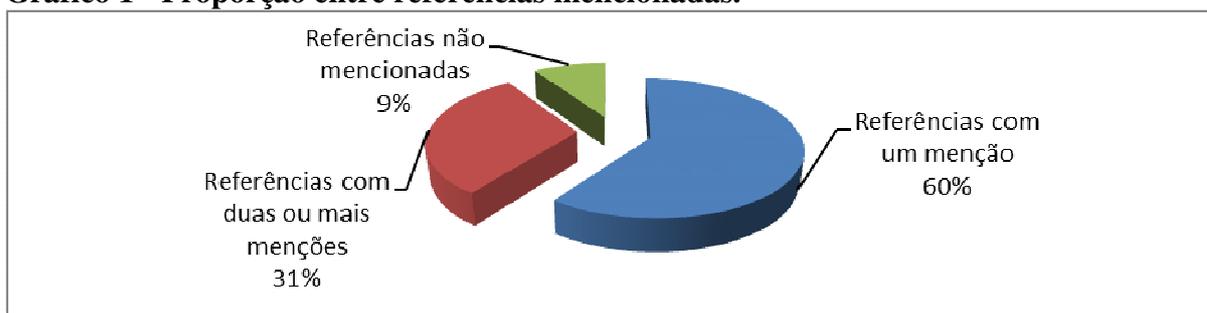
Os 9% de referências não mencionadas se apresentam como uma prova empírica que Romancini (2010) reivindica quando comenta a afirmação de Wouters sobre muitas referências não serem citadas. Essa taxa próxima de 10% pode ser um caso isolado, mas indica que os editores precisam estar atentos para diminuir essa prática.

⁴ “[...] medida padronizada da força do relacionamento entre duas variáveis que não depende das hipóteses de um teste paramétrico” (FIELD, 2009, p.643).

⁵ *Statistical Package for the Social Sciences for Windows*.

⁶ Voos, H., & Dagaev, K. S. (1976). Are all citations equal? Or Did we op. cit. your idem? *J. of Acad. Librarianship*, 1, 20–21.

Gráfico 1 - Proporção entre referências mencionadas.



Fonte: dados da pesquisa, 2015.

Para atender o segundo objetivo do trabalho foi gerado um ranking com os pesquisadores mais citados e mais mencionados. A essas duas categorias juntaram-se “número de artigos citantes” e “pico de menção de um único documento referenciado”.

O ranking ficou com 299 autores (com duas ou mais citações). A Tabela 3 apresenta o coeficiente de correlação de *Spearman*. Ding e Cronin (2011) utilizaram o mesmo coeficiente para mostrar a relação de três rankings em um estudo de citação. A análise apresentou valor de significância igual à zero para todos os pares de correlação, o que, segundo Field (2009), é aceitável visto que precisa ser $< 0,05$.

Em relação às categorias, “número de citação” apresentou correlação alta com “número de artigos citantes” e “número de menções” (as três mais importantes). A categoria “pico de menções” apresentou correlação baixa com “número de citação” e “número de artigos citantes” e, portanto, foi a mais insignificante, mostrando apenas a distribuição das menções nos artigos citantes.

Tabela 3 - Distribuição das taxas de “coeficiente de correlação de *Spearman*” das categorias

CATEGORIAS	Nº Citação	Nº Artigos citantes	Nº Menções	Pico de menções
Nº Citação	1,000			
Nº Artigos citantes	0,742	1,000		
Nº Menções	0,735	0,577	1,000	
Pico de menções	0,412	0,341	0,860	1,000

* O valor de significância de todos os pares da correlação foi igual a zero (0,000), ou seja, menor que 0,05.

Fonte: dados da pesquisa, 2015.

Na Tabela 4 são apresentados os autores mais citados do ranking elaborado (primeira coluna), com oito ou mais citações. As demais colunas apresentam os resultados das outras categorias. Os números apresentados após o nome dos autores indicam a posição do mesmo em cada categoria. Assim, o autor “Saracevic [3,1,9,42]” foi o terceiro mais citado, o primeiro

em número de artigos citantes, o nono em número de menções e 42º na categoria pico de menção.

Tabela 4 - Distribuição dos autores mais citados no corpus da pesquisa.

	Número de citações (referências)		Número de artigos citantes		Número de menções		Pico de menção de um único artigo referenciado	
01	Hjorland [1,7,1,1]	22	Saracevic [3,1,9,42]	13	Hjorland [1,7,1,1]	66	Hjorland [1,7,1,1]	19
02	Kuhlthau [2,3,2,8]	20	Lancaster [4,2,5,19]	12	Kuhlthau [2,3,2,8]	59	Moura [134,117,63,2]	10
03	Saracevic [3,1,9,42]	16	Kuhlthau [2,3,2,8]	10	Ellis [9,13,3,3]	36	Ellis [9,13,3,3]	09
04	Lancaster [4,2,5,19]	14	Kobashi [6,4,16,103]	10	Wilson [5,12,4,9]	31	Cox [81,69,22,4]	09
05	Wilson [5,12,4,9]	14	Barreto [7,5,17,43]	10	Lancaster [4,2,5,19]	27	Hall [82,70,23,5]	09
06	Kobashi [6,4,16,103]	12	Souza [11,6,14,20]	09	Capurro [14,15,6,7]	27	Taylor [135,118,41,6]	09
07	Barreto [7,5,17,43]	11	Hjorland [1,7,1,1]	08	Ingwersen [16,40,7,21]	24	Capurro [14,15,6,7]	08
08	Spink [8,8,13,44]	11	Spink [8,8,13,44]	07	Fujita [10,25,8,11]	23	Kuhlthau [2,3,2,8]	07
09	Ellis [9,13,3,3]	11	Belkin [13,9,10,45]	07	Saracevic [3,1,9,42]	21	Wilson [5,12,4,9]	07
10	Fujita [10,25,8,11]	11	Alvarenga [18,10,24,12]	07	Belkin [13,9,10,45]	20	Almeida Júnior [270,269,68,10]	07
11	Souza [11,6,14,20]	10	Cunha [22,11,96,182]	07	Ferreira [20,27,11,13]	19	Fujita [10,25,8,11]	06
12	Lima [12,14,32,104]	10	Wilson [5,12,4,9]	06	Järvelin [27,56,12,23]	18	Alvarenga [18,10,24,12]	06
13	Belkin [13,9,10,45]	09	Ellis [9,13,3,3]	06	Spink [8,8,13,44]	16	Ferreira [20,27,11,13]	06
14	Capurro [14,15,6,7]	09	Lima [12,14,32,104]	06	Souza [11,6,14,20]	16	Heidegger [73,96,19,14]	06
15	Crespo [15,26,26,60]	09	Capurro [14,15,6,7]	06	Choo [25,28,15,22]	16	Rubi [74,97,35,15]	06
16	Ingwersen [16,40,7,21]	09	Jansen [19,16,27,46]	06	Kobashi [6,4,16,103]	15	Gadamer [136,119,42,16]	06
17	Campos [17,41,37,105]	09	Castells [23,17,49,107]	06	Barreto [7,5,17,43]	15	Stambaugh [137,120,48,17]	06
18	Alvarenga [18,10,24,12]	08	Marcondes [24,18,59,183]	06	Guarino [43,43,18,25]	15	Tibbo [139,122,67,18]	06
19	Jansen [19,16,27,46]	08	Baeza-yates [31,19,20,48]	06	Heidegger [73,96,19,14]	15	Lancaster [4,2,5,19]	05
20	Ferreira [20,27,11,13]	08	Ribeiro-neto [32,20,21,49]	06	Baeza-yates [31,19,20,48]	14	Souza [11,6,14,20]	05
21	Jardim [21,42,69,106]	08	Ferneda [33,21,33,61]	06	Ribeiro-neto [32,20,21,49]	14	Ingwersen [16,40,7,21]	05

Fonte: dados da pesquisa, 2015.

Os resultados mostram que dos 21 autores mais citados, 15 estão entre os mais mencionados (destaques em negrito nas colunas um e três). Esse resultado mostra uma evidência da importância e confiabilidade dos estudos de citação realizados a partir das referências.

Contudo, as menções podem oferecer perspectivas mais amplas, principalmente para entender a relação desses autores mais citados com os textos citantes. Afinal, o que leva um autor a mencionar uma referência 19, 10 ou nove vezes, como mostram as três primeiras posições da quarta coluna da Tabela 4? Esse é um questionamento para trabalhos futuros.

5 CONSIDERAÇÕES FINAIS

Na produção brasileira sobre Recuperação da Informação há um núcleo de autores que são mais mencionados e que são basicamente os mais citados. Sugere-se a ampliação do

recorte temporal desse teste, assim com a sua aplicação em outras áreas do conhecimento e em comunicações orais.

As limitações tecnológicas parecem ser um impedimento para estudos com menções em larga escala, pois automatizar a extração em textos completos é mais complexo do que lidar com a lista de referências.

REFERÊNCIAS

BRAMBILLA, S. D. S.; VANZ, S. A. de S.; STUMPF, I. R. C. Mapeamento de um artigo produzido na UFRGS: razões das citações recebidas. **Encontros Bibli**, Florianópolis, n. esp., 1º sem. 2006. Disponível em: <<https://goo.gl/frk4Zp>>. Acesso em: 10 jul. 2015.

CUNHA, M. B.; CAVALCANTI, C. R. de O. **Dicionário de Biblioteconomia e Arquivologia**. Briquet de Lemos: Brasília, DF, 2008. 451p.

DING, Y.; CRONIN, B. Popular and/or prestigious? Measures of scholarly esteem. **Information Processing and Management**, n.47, p.80-86, 2011. Disponível em: <<http://goo.gl/BiShPP>>. Acesso em: 10 jul. 2015.

DING, Y. et al. The distribution of references across texts: some implications for citation analysis. **Journal of Informetrics**, n. 7, p. 583-592, 2013. Disponível em: <<http://goo.gl/qtPc70>>. Acesso em: 10 jul. 2015.

FIELD, A. **Descobrimo a estatística usando o SPSS**. 2ed. Porto Alegre: Artmed, 2009.

JEONGA, Y.; SONGA, M.; DING, Y. Content-based author co-citation analysis. **Journal of Informetrics**, n.8, p.197-211, 2014. Disponível em: <<http://goo.gl/2GHtBJ>>. Acesso em: 10 jul. 2015.

REITZ, J. M. **ODLIS - Online Dictionary for Library and Information Science**. c2014. Disponível em: <http://www.abc-clio.com/ODLIS/odlis_c.aspx>. Acesso em: 13 mar. 2015.

ROMANCINI, R. O que uma citação? Análise de citações na ciência. **Intexto**, Porto Alegre, v. 2, n. 23, p. 20-35, jul./dez. 2010. Disponível em: <<http://goo.gl/Jgeidh>>. Acesso em: 13 mar. 2015.

SILVEIRA, M.; CAREGNATO, S. E. Estudo das razões das citações na Ciência da Informação: proposta de classificação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15., 2014, Belo Horizonte, MG). **Anais...** Belo Horizonte: ECI, UFMG, 2014. Disponível em: <<http://goo.gl/Q5MfKA>>. Acesso em: 23 mar. 2015.

VANZ, S. A. de S.; CAREGNATO, S. E. Estudos de Citação: uma ferramenta para entender a comunicação científica. **Em Questão**, Porto Alegre, v. 9, n. 2, p. 295-307, jul. /dez. 2003. Disponível em: <<http://goo.gl/HdRvbB>>. Acesso em: 10 jul. 2015.