em questão
E-ISSN 1808-5245
Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS
v. 21, n. 1 – Jan./Abr. 2015

Análise de cocitação de autores: uma aplicação em estudos de indexação

Luciana Beatriz Piovezan
Mestranda; Universidade Estadual Paulista (UNESP);
lbpiovezan@gmail.com

Mariângela Spotti Lopes Fujita
Doutora; Universidade Estadual Paulista (UNESP);
fujita@marilia.unesp.br

Resumo: Neste artigo se apresenta a avaliação da indexação, enquanto ferramenta que permite a melhoria contínua dos sistemas de informação por meio do aperfeiçoamento da indexação de assuntos. Surge, assim, a necessidade de situar os estudos de avaliação da indexação, localizando seu núcleo de pesquisadores e correntes ideológicas. Buscou-se identificar a produção brasileira no tema "avaliação da indexação" e realizar análise de cocitação de autores de modo a compreender a interlocução entre os pesquisadores inseridos nessa temática. Para tanto, realizou-se um estudo teórico sobre indexação e avaliação do processo de indexação e pesquisa bibliométrica por meio de estudo de colaboração científica apoiado pela análise de cocitação de autores. Descobriu-se que há um interesse crescente pela temática em Ciência da Informação. A Análise de Redes Sociais permitiu visualizar a presença das três principais correntes do tratamento temático da informação influenciando a literatura brasileira da área, destacando-se os nomes Lancaster e Chaumier.

Palavras-chave*:* Indexação de assuntos. Avaliação da indexação de assuntos. Análise de cocitação. Análise de Redes Sociais.

1 Introdução

Entre os processos de tratamento temático da informação, localizamos a indexação enquanto processo de representação do conteúdo temático de documentos de modo a permitir e facilitar sua recuperação temática. No processo de indexação são identificados e selecionados os conceitos essenciais de documentos, posteriormente traduzidos em termos de indexação que servirão à sua representação e recuperação.

Percebemos a necessidade de situar os estudos de avaliação da indexação, localizando seu núcleo de pesquisadores e correntes ideológicas. Assim, apresentamos como objetivo geral deste estudo compreender a estrutura em que se organizam os pesquisadores atuantes na área de Indexação, mais especificamente os que trabalham com temáticas relativas à avaliação do processo de indexação em âmbito qualitativo e quantitativo. De modo mais específico, buscamos identificar a produção brasileira no tema "avaliação da indexação" e realizar análise de cocitação de autores de modo a compreender a interlocução entre os pesquisadores inseridos nessa temática.

A pesquisa caracterizou-se como um estudo teórico, mediante revisão de literatura sobre indexação e avaliação da indexação, e pesquisa bibliométrica, efetivada com estudo de colaboração científica apoiado pela análise de cocitação de autores.

Por estudos bibliométricos compreendemos a investigação "[...] dos aspectos quantitativos da produção, disseminação e uso da informação registrada. Desenvolve padrões e modelos matemáticos para medir esses processos, usando seus resultados para elaborar previsões e apoiar tomadas de decisão." (TAGUE-SUTCLIFFE, 1992, p. 1).

A escolha pela análise científica respaldada nos estudos bibliométricos deve-se ao fato de que os mesmos permitem "[...] uma abordagem objetiva e confiável que, associada às análises contextuais, oferece um diagnóstico real, amplo e verdadeiro da produção científica de uma área de especialidade, de um grupo, instituições ou países, produtores da ciência e tecnologia". (GRÁCIO; OLIVEIRA, 2011, p. 252).

2 Indexação e avaliação da indexação

Neste item são conceitualmente definidos por literatura fundamental os temas de indexação e a avaliação da indexação enquanto processo e método de organização do conhecimento em Ciência da Informação que influem na recuperação temática da informação.

Barité (1999) entende a indexação como o

> [...] processo de descrição e representação do conteúdo de um documento, mediante um número limitado de conceitos extraídos do texto dos documentos (palavras-chave) ou de vocabulários controlados (descritores, termos ou cabeçalhos de assunto). (BARITÉ, 1999, p. 73-74).

De acordo com Rubi (2009), a indexação é a ação de descrever documentos em seu conteúdo temático, utilizando-se da análise de assunto e da representação desse conteúdo por meio de conceitos que, em seguida, são traduzidos em termos de uma linguagem de indexação. O objetivo desse procedimento é permitir a intermediação entre usuário e documento no momento de busca e recuperação da informação em um sistema de informação. Dessa forma a indexação é uma das principais atividades do profissional de informação, e de seu bom desempenho depende também o sucesso da missão das unidades de informação.

A importância atribuída à indexação, então, é devida ao seu objetivo, aperfeiçoar a recuperação de documentos quando realizada por meio de buscas por assunto em unidades de informação (FUJITA, 2003). Sendo assim, podemos afirmar que a indexação pode garantir a recuperação temática eficaz se realizada com identificação de termos pertinentes ao conteúdo temático do documento – suas potencialidades informativas, e com a seleção de termos que considere relevante para a comunidade a que se destina.

O processo de indexação consiste de análise de assunto, onde se realiza a leitura e segmentação do texto analisado para a identificação e seleção de conceitos; síntese, construção do texto documental com os conceitos selecionados; e representação, a tradução dos conceitos selecionados em termos de uma linguagem, que pode ser construída ou natural (KOBASHI, 1996; GUIMARÃES, 2003).

A linguagem natural é a linguagem de uso corrente, presente no discurso do autor (manifestada no documento), do indexador e do usuário. As linguagens construídas, linguagens documentais, têm por princípio o controle terminológico e o propósito de aperfeiçoar o acesso à informação. São formas artificiais de linguagem, construídas tendo por alicerce um conjunto de regras pré-estabelecidas. Utilizam-se as linguagens documentais para expressar o conteúdo potencialmente informativo de documentos de forma a representá-los em catálogos e sistemas de recuperação da informação (CINTRA et al., 2002; LARA, 2004).

Buscando assegurar o alcance do objetivo da indexação, a recuperação da informação por meio de busca por assunto, realizam-se ensaios de avaliação do processo e do produto da indexação em unidades de informação.

A avaliação da indexação, segundo Gil Leiva, pode ser realizada pelos métodos de avaliação intrínseca e de avaliação extrínseca. A avaliação intrínseca da indexação "[...] é o conjunto de tarefas centradas no resultado da indexação (descritores, cabeçalhos, sub-cabeçalhos ou identificadores) com a finalidade de conhecer sua qualidade". (GIL LEIVA, 2008, p. 385, tradução nossa). Essa avaliação pode ser realizada qualitativa ou quantitativamente.

A avaliação extrínseca da indexação considera o resultado da indexação, os termos atribuídos a cada documento, podendo ser realizada também de duas formas: a avaliação extrínseca mediante a interconsistência; e a avaliação extrínseca mediante a recuperação (GIL LEIVA, 2008, p. 391).

De acordo com Gil Leiva, a avaliação extrínseca mediante a recuperação "[...] consiste em interrogar duas bases de dados que contém conteúdos idênticos, salvo os campos que abrigam a indexação. Com os resultados obtidos se fazem índices de exaustividade e precisão na recuperação". (GIL LEIVA, 2008, p. 391, tradução nossa).

A exaustividade refere-se à quantidade de termos atribuídos a um documento, uma indexação exaustiva atribui tantos termos quanto podem ser identificados no documento. Já uma indexação menos exaustiva, ou mais seletiva, atribuirá menor número de termos de assunto ao documento. Martínez Tamayo e Valdez (2008, p. 42, tradução nossa) lembram que "[...] quanto mais termos se assinalem a um registro bibliográfico, mais possibilidades terá o usuário de recuperá-lo", porém, em contrapartida este usuário terá que lidar com uma busca pouco precisa.

O índice de precisão é determinado pelo resultado da divisão do número de itens relevantes recuperados em uma questão posta ao sistema pelo número total de itens recuperados nessa mesma questão (LANCASTER, 2004). A precisão da representação relaciona-se com sua especificidade, "[...] a extensão em que o sistema nos permite ser precisos ao especificarmos um documento que estejamos processando [...]" (FOSKETT, 1973, p. 12). Atuando no lado contrário ao da

exaustividade, esse parâmetro permite que o usuário tenha uma recuperação mais precisa em sua busca, porém com uma resposta quantitativa menor.

A definição desses parâmetros tem por função equilibrar as respostas do sistema de recuperação da informação, reduzindo os níveis de ruído, a recuperação de registros que não são compatíveis com a questão proposta pelo usuário, e silêncio, a não recuperação de registros pertinentes à questão de busca do usuário ainda que a coleção os contemple. Buscando, dessa forma, alcançar resultados adequados aos pedidos de informação dos usuários.

Segundo Gil Leiva (2008), a avaliação extrínseca mediante a interconsistência observa a indexação realizada por dois ou mais indexadores (ou sistemas de indexação ou ainda indexadores humanos e automáticos) ao indexarem um mesmo documento. Tem por objetivo realizar uma comparação dos termos atribuídos para verificar se há consistência nos resultados.

De acordo com Moreiro González (2004, p. 51, tradução nossa), a consistência "[...] busca que um conceito ou tema apareça expresso sempre da mesma forma". Tal afirmação suscita uma problemática na análise da consistência da indexação, que seria a possibilidade de que se encontre "[...] uma indexação consistentemente incorreta [...]" como coloca Gil Leiva (1999, p. 31).

Em termos sintéticos, a avaliação da indexação se coloca como uma ferramenta que tem por objetivo a melhoria contínua da recuperação da informação dos sistemas de informação. Tal procedimento procura o aperfeiçoamento da indexação de assuntos ao apontar seus fatores de sucesso, assim como suas falhas.

3 Estudos bibliométricos em Ciência da Informação

Os estudos métricos possuem três ênfases: a ênfase aplicada, utilizada para o estudo de outras áreas da Ciência, visando sua avaliação e o planejamento de ações; a ênfase teórica, que tem por objetivo o avanço dos estudos métricos enquanto campo científico; e a terceira ênfase, que abarca o estudo dos métodos que são utilizados para realização destes estudos. Estes estudos apresentam seus resultados na forma de indicadores, que podem, igualmente, ser visualizados em três tipos: indicadores de produção, a contagem de publicações por tipologia; indicadores de ligação, como a

coautoria, a cocitação e a coocorrência de palavras, além da análise de redes de colaboração científica; e indicadores de citação, tais como o Fator de Impacto, vida média da publicação, média de citações recebidas e índice de Hirsch, derivados da contagem do número de citações (OLIVEIRA; GRÁCIO, 2011).

A produção científica consiste no conjunto de resultados originados do trabalho do cientista ou grupo de cientistas durante e após o processo de investigação. A publicação destes resultados em diferentes tipologias documentais e registrada em diferentes suportes garante a validação dos esforços do pesquisador – por meio da avaliação realizada por seus pares –, assim como a comunicação dos resultados obtidos para uso comum.

Nesse sentido, Meadows (1999[1] apud VANZ; CAREGNATO, 2003, p. 247) defende que a comunicação é vital para a ciência, "[...] pois a ela não cabe reivindicar com legitimidade esse nome enquanto não houver sido analisada e aceita pelos pares". Em outras palavras, a comunicação é um dos aspectos mais visíveis e relevantes da inovação em ciência, é por meio da comunicação que a ciência se legitima e se reconhece (SPINAK, 1998).

A fim de obter melhores níveis de investigação, pesquisadores têm sido levados a colaborar cada vez mais, Katz e Martin (1997) apontam, inclusive, ao interesse crescente em colaboração motivada por fatores políticos, além dos interesses científicos. As aproximações entre pesquisadores podem acontecer em congressos, conferencias, reuniões de associações de pesquisa, entre outros. Balancieri et al. (2005, p. 66) afirmam que essa auto-organização "[...] transcende os limites do departamento, da instituição de um país e abrange cientistas de todos os lugares do mundo onde houver atividade científica relevante na especialidade em questão", formando o que se conhece por colégios invisíveis.

Colégios invisíveis são um tipo de organização informal formada por indivíduos que compartilham dos mesmos interesses de investigação e que mantêm entre si um fluxo constante de comunicação também informal, como troca de correspondência - cartas, telefonemas, e-mails. Segundo as palavras de Vanz (2009, p. 42), colégios invisíveis são "[...] as comunidades informais de pesquisadores que se comunicam, trocam informações e experiências e também publicam formalmente seus resultados de pesquisa". Para isso, os membros do grupo "[...] afirmam estar

razoavelmente em contato com todo mundo que contribui significativamente para pesquisa neste assunto não apenas em escala nacional, mas em geral, incluindo todos os outros países em que essa especialidade é forte". (PRICE; BEAVER, 1966, p. 1.011, tradução nossa).

A visualização dos colégios invisíveis pode acontecer com a utilização da Análise de Redes Sociais. Este tipo de análise tem sido realizada desde a década de 1960 e, de acordo com Watts (1999, p. 494), tem oferecido

> [...] medidas tanto de significância individual quanto de centralidade e de eficiência de redes que podem elucidar fenômenos não-óbvios como os '*key-players*' em uma organização ou sua estrutura favorável para difusão de informação.

A Análise de Redes Sociais

> [...] conceitualiza a estrutura social como uma rede, com nós conectando membros e canalizando recursos, foca nas características dos nós, em detrimento das características individuais dos membros e visualiza comunidades como 'comunidades pessoais', ou seja, como redes de relações individuais que pessoas criam, mantêm e usam no curso de suas vidas diárias (OTTE; ROUSSEAU, 2002, p. 442).

Em Ciência da Informação, as publicações, citações, cocitações e as colaborações são insumos para a elaboração de estudos com Análise de Redes Sociais (OTTE; ROUSSEAU, 2002). O diferencial destas análises encontra-se em sua interlocução com as pesquisas de Análise de Domínio. Hjørland (2002) ao elencar as onze abordagens da Análise de Domínio situa os estudos métricos como uma das principais metodologias para a compreensão de um domínio. Por domínio podemos compreender as "[...] comunidades de pensamento ou discurso, que são partes de uma divisão da sociedade de trabalho.". (HJØRLAND; ALBRECHTSEN, 1995, p. 400).

Entende-se que a frequência de citações de autores ou documentos e a coocorrência desses itens permite que o domínio seja observado. Dessa forma, a análise de citações pretende visualizar um campo científico enquanto um domínio, por meio da forma principal com a qual a comunidade desse domínio se legitima – as citações.

A frequência com que dois autores são citados conjuntamente em produções

científicas posteriores é captada pela análise de cocitação de autores. A citação nesse contexto é percebida como uma medida objetiva da efetivação da comunicação científica – posto que evidência que um relato de pesquisa foi publicado, divulgado, acessado e utilizado para fins de nova investigação. A partir da citação é possível perceber a inserção dos pesquisadores na comunidade observada, demonstrando "[...] os pesquisadores de maior impacto na área, apontando seus paradigmas, procedimentos metodológicos pertinentes, bem como os pesquisadores de 'vanguarda' que constroem o novo conhecimento na área". (GRÁCIO; OLIVEIRA, 2013, p. 199).

A Análise de Redes Sociais enfatiza os atores da rede e seus relacionamentos ao mapear as conexões estabelecidas em uma dada comunidade. Nesse contexto, a análise de cocitações verifica pares de documentos citados conjuntamente, demonstrando a maneira como uma área de conhecimento se estrutura para seus pesquisadores. Como já afirmava Carvalho (1975), a cocitação pode ser usada para identificar o desenvolvimento de áreas científicas, pois que o fato de dois autores serem citados conjuntamente em trabalhos posteriores leva a crer que há relacionamento entre o trabalho de pesquisa realizado por tais pesquisadores.

Os parâmetros para Análise de Redes Sociais são os seguintes:

a) *"densidade*: é um indicador para o nível geral de conectividade do grafo e define-se como o número de ligações dividido pelo número de vértices em um grafo completo com o mesmo número de nós". (OTTE; ROUSSEAU, 2002, p. 442);

b) *"centralidade de grau*: em um grafo de coautoria, o grau de centralidade de um ator é o número de autores no grafo com quem ele colaborou em pelo menos um artigo". (OTTE; ROUSSEAU, 2002, p. 442);

c) *"proximidade*: também é um indicador de centralidade, porém com característica mais geral que o anterior, a proximidade de um nó é igual à distância total deste nó aos outros nós no grafo". (OTTE; ROUSSEAU, 2002);

d) *"centralidade de intermediação*: o número de caminhos mais curtos entre dois atores que passam por um determinado nó". (GRÁCIO; OLIVEIRA, 2011, p. 260); e

e) *"cliques*: um clique é um subgrafo no qual todos os nós estão diretamente conectados". (OTTE; ROUSSEAU, 2002, p. 443).

Tendo sido explicitado o entendimento prévio das temáticas abordadas para realização desta pesquisa, passaremos à enumeração dos procedimentos metodológicos adotados para a efetivação desta análise.

4 Metodologia

Os procedimentos metodológicos realizados constaram de duas etapas, a primeira constou de levantamento da produção brasileira relativa à avaliação da indexação e a segunda constou de procedimentos bibliométricos de análise de cocitação.

Durante a primeira etapa realizou–se levantamento dos periódicos nacionais em Ciência da Informação, para tal teve como apoio a relação Qualis publicada pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) em 2012, chegando a um total de 30 periódicos. Em seguida realizou-se coleta bibliográfica nas páginas online de cada periódico, utilizando-se do recurso de pesquisa oferecido pelo próprio periódico. Foi feita consulta pelo termo "avaliação indexação", buscando em "palavras no título"; "resumo", "palavras-chave" e "texto completo". Esta consulta retornou um total de 48 artigos provenientes de 14 periódicos.

Em seguida foi realizada uma consulta, interna aos artigos recuperados, buscando confirmar se o assunto abordado pelo artigo tratava-se efetivamente da temática de interesse desta pesquisa. Pra tal utilizou-se a ferramenta de localização de palavras (Ctrl+F ou Ctrl+L) do software visualizador de documentos utilizado para acessar o texto do artigo. Foram utilizados os termos "indexação", "consistência", "exaustividade", "precisão" e "seletividade". Definindo-se que o artigo deveria conter ao menos dois dos termos consultados. Este filtro resultou na seleção de 26 artigos provenientes de 11 periódicos.

A segunda fase metodológica consistiu na aplicação dos procedimentos bibliométricos. Foi feito o levantamento dos dados iniciando-se por uma súmula dos artigos selecionados, contendo o título do artigo e seu conjunto de referências. Em seguida foi levantado o conjunto total das referências; foram desdobradas as referências contendo autoria múltipla de modo a contemplar todos os autores do

artigo. Em seguida foi feita a contagem de citações para verificação dos autores citados em maior número de artigos. Identificou-se um total de 433 referências, tendo sido mencionados um total de 288 autores. Para realizar a análise de cocitação, foram considerados os autores que foram citados em ao menos três trabalhos, correspondendo a um total de 31 autores. Ressalte-se que um total de 221 autores foram citados por apenas um trabalho.

Tendo sido selecionados os 31 autores, construiu-se uma matriz assimétrica 26x31 de ocorrência de citação entre autores citantes (por linha: 26 linhas) e autores citados (por coluna: 31 colunas). Para cada autor citado, em sua respectiva coluna, foi registrado o valor um na linha correspondente ao artigo em que ele foi citado e zero quando o autor não foi citado pelo trabalho.

A partir dessa matriz de ocorrência (26x31), construiu-se, também no software Excel com apoio da função matemática "SOMARPRODUTO", a matriz quadrada e simétrica de tamanho 31x31, cujo resultado constituiu a frequência de cocitação entre os autores mais citados nos artigos analisados. A Figura 1 apresenta um recorte da matriz 31x31 de cocitação entre esses autores. Com o uso do software Ucinet, traçou-se a rede de cocitação de autores e calcularam-se os indicadores de densidade e de centralidade – sendo eles centralidade de grau, de intermediação e de proximidade – da rede gerada.

Figura 1 – Recorte da matriz de frequência de cocitação

AH33

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
1		LANCASTER	FUJITA	FOSKETT	BOCCATO	KOBASHI	CHAUMIER	NAVES	ROBREDO	RUBI	CARNEIRO	GUIMARÃES	MOREIRO GON	MOYA ANEG	NARDI	PIEDADE	SILVA, M.R.	VAN SLYPE	AFFONSO	ALMEIDA, M	CABRÉ	CAMPELLO	CESARINO	CUNHA	CUTTER	DIAS	FERNÁNDEZ M	GIL LEIVA	HJØRLAND	PINHEIRO	TARGINO	VAN DER LA
2	LANCASTER	0																														
3	FUJITA	10	0																													
4	FOSKETT	7	6	0																												
5	BOCCATO	6	4	5	0																											
6	KOBASHI	3	2	1	0	0																										
7	CHAUMIER	4	5	2	3	1	0																									
8	NAVES	3	3	1	2	1	3	0																								
9	ROBREDO	4	3	1	1	0	1	1	0																							
10	RUBI	5	5	4	4	1	3	2	1	0																						
11	CARNEIRO	4	2	2	1	1	2	1	1	2	0																					
12	GUIMARÃES	4	4	2	2	1	2	1	1	3	1	0																				
13	MOREIRO GONZÁLEZ	3	2	0	1	0	1	1	2	1	0	2	0																			
14	MOYA ANEGÓN	4	4	4	4	0	1	1	1	2	1	0	0	0																		
15	NARDI	4	4	3	3	1	1	1	0	2	1	2	1	2	0																	
16	PIEDADE	4	2	2	1	0	0	0	1	0	0	0	0	1	1	0																
17	SILVA	4	4	2	3	0	3	2	2	3	1	3	2	1	1	0	0															

Freq. cit. | matriz 1 | Matriz cocit | Plan3

Pronto

Fonte: elaboração própria.

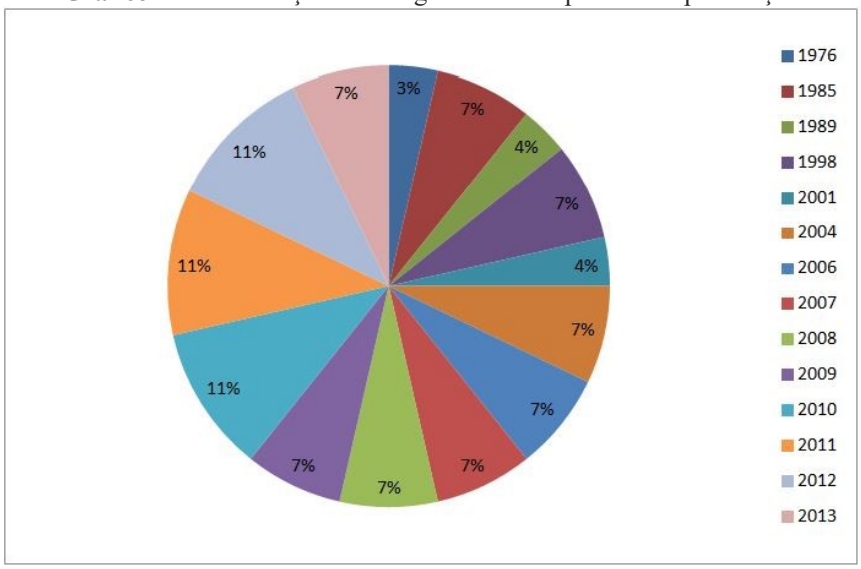
Tendo sido apresentados os procedimentos metodológicos realizados, em seguida iremos expor os resultados alcançados e hipóteses que podem ser colocadas para a interpretação desses resultados tendo por embasamento a revisão de literatura exposta nas seções iniciais.

5 Resultados e discussões

Para seleção do corpus a ser analisado, esta pesquisa não realizou um corte temporal, optando por limitar a busca apenas para a temática que se procura observar. Ainda assim, houve uma recuperação bastante limitada de artigos. Esses dados demonstram que a temática "Avaliação da indexação", embora importante para os desdobramentos práticos da área, está sendo pouco discutida no âmbito teórico da Ciência da Informação no Brasil.

A distribuição de publicações por ano, dentre os artigos coletados, encontra-se no Gráfico 1. Neste gráfico podemos perceber que o maior percentual de trabalhos publicados encontra-se no intervalo dos anos de 2010 a 2012, 11% em cada ano. Nos últimos cinco anos foram publicados 50% dos trabalhos analisados, demonstrando também que há uma tendência ao aumento do interesse da área pela temática.

Gráfico 1 – Distribuição dos artigos coletados por ano de publicação



Fonte: elaboração própria.

A partir da contagem de citações recebidas foi possível obter os autores mais citados nos artigos analisados. Foram selecionados especificamente os autores que receberam menção em ao menos quatro trabalhos (aproximadamente 15% dos artigos verificados), o que somou um total de 17 autores. Os mesmos encontram-se distribuídos na Tabela 1.

Tabela 1 – Autores mais citados

AUTOR	N. de trabalhos em que foi citado
LANCASTER, F. W. (Inglaterra)	16
FUJITA, Mariângela Spotti L. (Brasil)	11
FOSKETT, Antony Charles (Inglaterra)	7
BOCCATO, Vera Regina C. (Brasil)	6
KOBASHI, Nair Yumiko (Brasil)	6
CHAUMIER, Jacques (França)	5
RUBI, Milena Polsinelli (Brasil)	5
CARNEIRO, Marília Vidigal (Brasil)	4
GUIMARÃES, José Augusto C. (Brasil)	4
MOREIRO GONZÁLEZ, José A. (Espanha)	4
MOYA ANEGÓN, F. (Espanha)	4
NARDI, Maria Isabel A. (Brasil)	4
NAVES, Madelena Martins L. (Brasil)	4
PIEDADE, Maria Antonietta R. (Brasil)	4
ROBREDO, Jaime (Brasil)	4
SILVA, Maria dos Remédios (Brasil)	4
VAN SLYPE, Georges (França)	4

Fonte: elaboração própria.

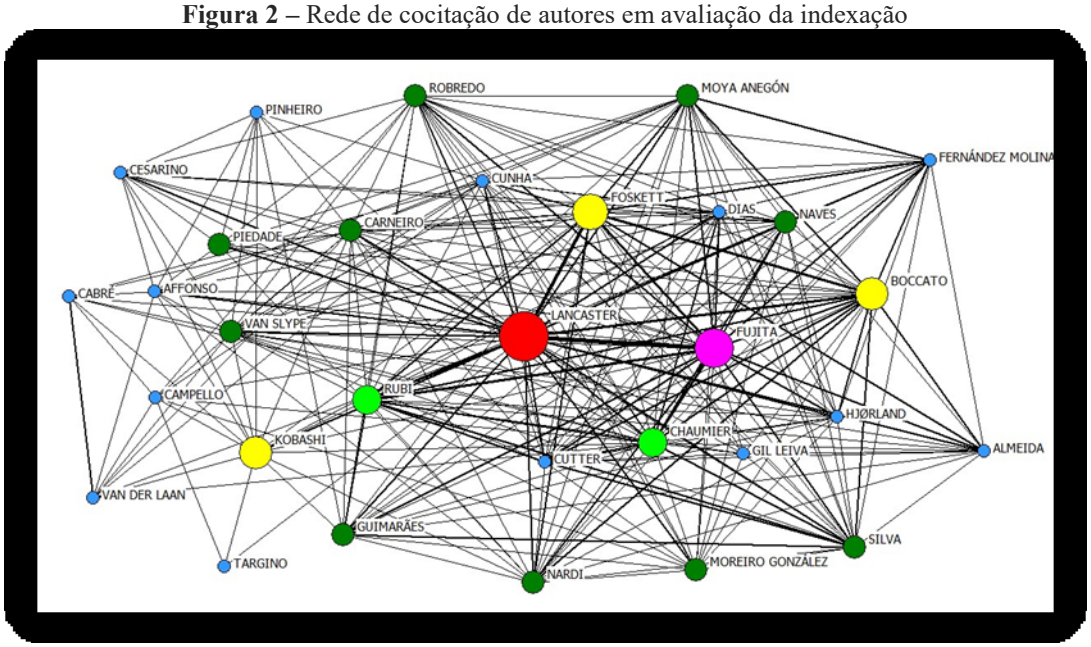
Vale ressaltar que, dos 26 artigos analisados, apenas três (11,5%) não mencionaram em seu corpo teórico nenhum dos 31 autores selecionados para a análise de cocitação, e, em média, cada artigo analisado citou 5,3 autores dentre este grupo. Ou seja, 88,5% dos artigos analisados tem seu embasamento teórico no grupo de pesquisadores selecionados, valores que demonstram que, mesmo sendo uma temática pouco consolidada, o grupo de autores selecionados constitui-se representativo para sua análise.

Podemos observar, entre os pesquisadores mais citados, que a literatura brasileira se sobressai ante a internacional, e, ao observarmos que dos 12 pesquisadores mais citados, seis são de nacionalidade diversa da brasileira, sendo que destes seis, apenas dois ainda se mantêm em atividade, entendemos que o tempo de absorção da literatura estrangeira pela área parece ser alto. Observa-se também que dentre os 31 pesquisadores do grupo de análise, aproximadamente 90% das citações recebidas são de obras originalmente escritas em português ou espanhol ou que foram traduzidas para esses idiomas. Relacionando estes dois dados, podemos inferir a existência de uma barreira idiomática entre os pesquisadores brasileiros na temática obervada.

Entre as temáticas associadas às citações, podem ser destacadas: estudos históricos da indexação, análise documental, análise de assunto, catalogação de assuntos, terminologia, leitura documental, linguagens documentais, política de indexação e indexação automática.

Apresentamos, na Figura 2, uma visualização da organização da temática analisada explorando a cocitação de autores. Observe-se que a área dos círculos é proporcional à frequência de citação dos autores, e a espessura dos segmentos de reta é proporcional à frequência de cocitação entre os autores que liga.

Figura 2 – Rede de cocitação de autores em avaliação da indexação



Fonte: elaboração própria.

A análise da rede exposta na Figura 2 destaca a atuação do pesquisador Lancaster, que, além de figurar como o mais citado nos artigos analisados, aparece também como o único pesquisador que possui cocitação com todos os demais 30 pesquisadores presentes na seleção, demonstrando a inserção de seus escritos no campo analisado. Destaca-se que, assim como Lancaster, são mencionados também os pesquisadores Chaumier, Foskett e Cutter, pioneiros nos estudos de tratamento temático da informação; Chaumier, representante da corrente teórica francesa denominada *Analyse documentaire*; Foskett e Lancaster, como representantes da corrente teórica inglesa, *indexing*; e Cutter, da denominada corrente teórica *subject cataloguing*, de origem norte-americana (GUIMARÃES, 2008, 2009; GUIMARÃES; SALES, 2010).

Pode-se ressaltar também a proximidade das pesquisas brasileiras com a literatura espanhola, representada pelos pesquisadores Moreiro González, Moya Anegón, Fernandez Molina e Cabré, esta última, ressalta-se, possui suas pesquisas na temática terminologia, com a qual a Ciência da Informação possui interface.

A média de cocitação entre os 31 pesquisadores selecionados é de 1,06 entre pares de autores. Podemos destacar, entre os valores absolutos de frequência de cocitação, os valores encontrados nos pares Lancaster e Fujita, 10 cocitações;

Lancaster e Foskett, sete cocitações; Lancaster e Boccato, seis cocitações; Fujita e Foskett, seis cocitações; Foskett e Boccato, cinco cocitações; Lancaster e Rubi, cinco cocitações; Fujita e Chaumier, cinco cocitações – sendo que, entre estas frequências mais altas, nenhuma se trata de coautoria –; e no par Fujita e Rubi, cinco cocitações, em que se observa a existência de efetiva colaboração entre as autoras por meio de coautoria.

Entre as colaborações científicas explicitadas pela coautoria, destacamos, como mencionado, as autoras Fujita e Rubi, com dois trabalhos em coautoria, a coautoria existente entre as autoras Fujita, Rubi e Boccato, com também dois trabalhos citados no escopo desta pesquisa, e a coautoria entre Fujita e Silva, em que todas as citações recebidas pela segunda revertem para esta coautoria. Os três casos apresentados tratam-se, também, da relação de orientador-orientado existente entre as pesquisadoras.

Para o cálculo dos indicadores de rede, a princípio determinou-se a densidade da rede, que apresentou valor 58,9% de coesão entre os pesquisadores desta temática.

Quanto ao indicador centralidade de grau, destaca-se o autor Lancaster, que foi, como dissemos, cocitado com 100% dos atores da rede. Em seguida destaca-se o autor Chaumier, que foi cocitado com 86,7% dos demais autores, seguido pela autora Fujita, cocitada com 83% dos autores, e pelos autores Carneiro e Foskett, que foram cocitados, cada um, com 80% dos autores presentes na rede. Percebe-se a prevalência da corrente inglesa entre os autores mais citados, também com a menção à corrente francesa.

Já entre os autores que apresentaram menor índice de centralidade de grau, mencionemos Cabré, Pinheiro e van der Laan, cocitados com 30% dos autores selecionados para esta análise, e Targino, que foi cocitada com apenas 10% dos demais autores analisados.

No índice de centralidade de intermediação destacou-se novamente o pesquisador Lancaster, com índice 10,2%, indicando que intermedeia 44 “caminhos mais curtos” entre pares de autores; seguido de Chaumier, que apresentou índice 3,9%, indicando que está presente em 16 caminhos mais curtos, e Kobashi, que com índice 3,1% intermedeia 13 caminhos mais curtos. Apenas a autora Targino

apresentou índice de intermediação igual a zero.

Em razão de esta rede não ser completamente conectada – há atores que não estão conectados por cocitação –, não foi possível calcular a centralidade de proximidade, dado que esta medida aplica-se apenas a redes fechadas.

6 Considerações finais

Embora pouco consolidada na área de Ciência da Informação, nota-se o recente crescimento do interesse de pesquisa na temática avaliação da indexação, representado pelo aumento das publicações referentes a esse tema. Destacamos, também, a prevalência da literatura nacional em relação à literatura estrangeira no corpus analisado.

Quanto à rede de cocitação em avaliação da indexação, podemos afirmar que tem como atores mais centrais os pesquisadores Lancaster, que, além de haver sido o autor mais citado, possui também os maiores índices de centralidade de grau e de intermediação, e o pesquisador Chaumier, que, embora tenha sido apenas o quinto mais citado, recebeu os segundos maiores índices nas duas medidas de centralidade calculadas. Esse dado revela que um autor pode ser muito citado, porém não ser um ator central; hipótese para esses casos pode ser que o autor exerce influência para apenas um grupo restrito de pesquisadores, ou que trabalhe com aspectos específicos da temática.

Avaliamos como positivo o estudo da temática avaliação da indexação por meio dos estudos métricos, abordagem pela qual foi possível observar a organização dos pesquisadores da temática. Consideramos necessária a ampliação de estudos dessa natureza, tratando, por exemplo, da obsolescência dessa literatura, dada a centralidade das questões de organização do conhecimento para a Ciência da Informação.

Referências

BALANCIERI, R. et al. A análise de redes de colaboração científica sob as novas

tecnologias de informação e comunicação: um estudo na Plataforma Lattes. **Ciência da Informação**, Brasília, v. 34, n. 1, p. 64-77, jan./abr. 2005. Disponível em: <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/619>. Acesso em: 10 abr 2014.

BARITÉ, M. **Glosario sobre organización y representación del conocimiento**: clasificación, indización, terminologia. Montevidéu: Universidad de la República, 1999.

CARVALHO, M. Análises bibliométricas da literatura de química no Brasil. **Ciência da Informação**, Brasília, v. 4, n. 2, dez. 1975. Disponível em: <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1618/1229>. Acesso em: 30 maio 2014.

CINTRA, A. M. M. et al. **Para entender as linguagens documentárias**. 2. ed. rev. e atual. São Paulo: Polis, 2002.

FOSKETT, A. C. **A abordagem temática da informação**. São Paulo: Polígono, 1973.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003. Disponível em: <http://www.sbu.unicamp.br/seer/ojs/index.php/rbci/article/view/287>. Acesso em: 30 jun 2014.

GIL LEIVA, I. **La automatización de la indización de documentos**. Gijón: Trea, 1999.

GIL LEIVA, I. **Manual de indización**. Gijón: Trea, 2008.

GRÁCIO, M. C. C.; OLIVEIRA, E. F. T. de. Análise de cocitação de autores: um estudo teórico-metodológico dos indicadores de proximidade, aplicados ao GT7 da ANCIB. **Liinc em Revista**, Rio de Janeiro, v. 9, n. 1, p. 196-213, maio 2013. Disponível em: <http://www.ibict.br/liinc>. Acesso em: 21 jun 2013.

GRÁCIO, M. C. C.; OLIVEIRA, E. F. T. Produção e comunicação da informação em CT&I – GT7 da ANCIB: análise bibliométrica no período 2003/2009. **Liinc em Revista**, Rio de Janeiro, v.7, n.1, p. 248–263, mar. 2011. Disponível em: <http://revista.ibict.br/liinc/index.php/liinc/article/viewFile/412/289>. Acesso em: 15 maio 2012.

GUIMARÃES, J. A. C. A análise documentária no âmbito do tratamento temático da informação: elementos históricos e conceituais. In: RODRIGUES, G. M.; LOPES, I. L. (Org.). **Organização e representação do conhecimento na perspectiva da Ciência da Informação.** Brasília: Thesaurus, 2003. p. 100-117.

GUIMARAES, J. A. C. A dimensão teórica do tratamento temático da informação e

suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). **Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 1, p. 77-99, 2008.

GUIMARAES, J. A. C. Abordagens teóricas de tratamento temático da informação (TTI): catalogação de assunto, indexação e análise documental. In: GARCÍA MARCO, F. J. (Ed.). **Avances y perspectivas en sistemas de información y documentación**. Ibersid, Zaragoza, 2009, p. 105-117.

GUIMARAES, J. A. C; SALES, R. de. Análise documental: concepções do universo acadêmico brasileiro em Ciência da Informação. **DataGramaZero**: Revista de Ciência da Informação, v.11, n.1, 2010. Disponível em: <http://www.dgz.org.br/fev10/Art_02.htm>. Acesso em: 19 jun 2013.

HJØRLAND, B. Domain analysis in information science: eleven approaches-traditional as well as innovative. **Journal of Documentation**, London, v. 58, n. 4, p. 422-462, 2002.

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in information science: domain-analysis. **Journal of the American Society for Information Science**, Washington, v. 46, n. 6, p. 400-425, 1995.

KATZ, J. S.; MARTIN, B. R. What is research collaboration?. **Research Policy**, Amsterdam, n. 26, p. 1-18, 1997. Disponível em: <http://www.sussex.ac.uk/Users/sylvank/pubs/Res_col9.pdf>. Acesso em: 15 maio 2012.

KOBASHI, N. Y. Análise documentária e representação da informação. **Informare**, São Paulo, v. 2, n. 2, p. 5-27, 1996.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. Brasília: Briquet de Lemos, 2004.

LARA, M. L. G. de Linguagen documentária e terminologia. **Transinformação**, Campinas, v. 16, n. 3, p. 231-240, set./dez., 2004. Disponivel em: <http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/710>. Acesso em: 30 maio 2014.

MARTÍNEZ TAMAYO, A. M.; VALDEZ, J. C. **Indización y classificación en bibliotecas.** Buenos Aires: Alfagrama, 2008.

MEADOWS, A. J. **A Comunicação Científica**. Brasília: Briquet de Lemos, 1999.

MOREIRO GONZÁLEZ, J. A. **El contenido de los documentos textuales**: su análysis y representación mediante el lenguage natural. Gijón: Trea, 2004.

OLIVEIRA, E. F. T. de; GRÁCIO, M. C. C. Indicadores bibliométricos em ciência da informação: análise dos pesquisadores mais produtivos no tem “estudos

métricos" na base Scopus. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 16, n. 4, p. 16-28, 2011. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1299>. Acesso em: 9 maio 2012.

OTTE, E.; ROUSSEAU, R. Social network analysis: a powerful strategy, also for the information sciences. **Journal of Information Science**, Cambridge, v. 28, n. 6, p. 441-453, 2002. Disponível em: <jis.sagepub.com>. Acesso em: 11 maio 2012.

PRICE, D. J. de S.; BEAVER, D. B. Collaboration in an invisible college. **American Psychologist**, Washington, v. 21, p. 1011-1018, 1966.

RUBI, M. P. Os princípios da política de indexação na análise de assunto para catalogação: especificidade, exaustividade, revocação e precisão na perspectiva dos catalogadores e usuários. In: FUJITA, M. S. L. (Org.). **A indexação de livros**: a percepção de catalogadores e usuários de bibliotecas universitárias: um estudo de observação do contexto sociocognitivo com protocolos verbais. São Paulo: Cultura Acadêmica. 2009. p. 81–92.

SPINAK, E. Indicadores cienciométricos. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 141-148, maio/ago. 1998. Disponível em: <http://revista.ibict.br/ciinf/index.php/ciinf/article/viewArticle/349>. Acesso em: 26 abr 2012.

TAGUE-SUTCLIFFE, J. An introduction to infometrics. **Information Processing & Management**, Oxford, v. 28, n. 1, p. 1-3, 1992.

VANZ, S. A. de S. **As redes de colaboração científica no Brasil (2004-2006)**. 2009. 204 f. Tese (Doutorado em Comunicação e Informação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009. Disponível em: <http://hdl.handle.net/10183/17169>. Acesso em: 15 maio 2014.

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://seer.ufrgs.br/EmQuestao/article/view/75>. Acesso em: 26 abr 2012.

WATTS, D. J. Networks, dynamics and the small world phenomenon. **American Journal of Sociology,** Chicago, n. 2. set. 1999. Disponível em: <http://www.twine.com/item/11jskfgbb-2397/networks-dynamics-and-the-small-world-phenomenon-1999>. Acesso em: 16 mar 2010.

em questão
E-ISSN 1808-5245
Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS
v. 21, n. 1 – Jan./Abr. 2015

Authors' co-citation analysis: an application on subject indexing studies

Abstract: This study presents the subject indexing evaluation as a tool that enables the continuous improvement of information systems through improvement of subject indexing. This raises the need to place the studies on subject indexing evaluation, locating its core researchers and ideological schools. The research aimed at identifying the Brazilian production on the theme "subject indexing evaluation" and performing authors' co-citation analysis in order to understand the dialogue between the researchers that work with this theme. For this purpose, we conducted a theoretical study on subject indexing and subject indexing evaluation, as well as a bibliometric study of scientific collaboration supported by authors' co-citation analysis. We found that there is a growing interest in this topic in Information Science. The social network analysis allowed the visualization of the presence of three mainstreams of thematic information processing influencing Brazilian literature in the field, with emphasis on two authors: Lancaster and Chaumier.

Keywords: Subject indexing. Subject indexing evaluation. Co-citation analysis. Social network analysis.

Recebido: 20/06/2014
Aceito: 04/02/2015

CC BY

[1] MEADOWS, Arthur J. **A comunicação científica.** Brasília: Briquet de Lemos, 1999.