

ANÁLISE DE COCITAÇÃO DE AUTORES: uma aplicação em estudos de indexação

Luciana Beatriz Piovezan (UNESP)

lbpiovezan@gmail.com

Mariângela Spotti Lopes Fujita (UNESP)

fujita@marilia.unesp.br

EIXO TEMÁTICO: Análise de citação

MODALIDADE: Apresentação oral

1 INTRODUÇÃO

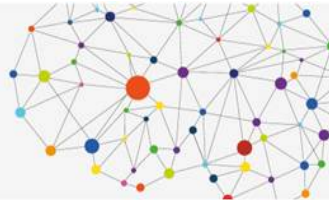
Entre os processos de tratamento temático da informação, localizamos a indexação enquanto processo de representar o conteúdo temático de documentos de modo a permitir e facilitar sua recuperação temática. No processo de indexação são identificados e selecionados os conceitos essenciais de documentos,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

Identifica-se como problema desta pesquisa a necessidade de situar os estudos de avaliação da indexação, localizando seu núcleo de pesquisadores, correntes ideológicas e tendências de pesquisa. Apresentamos como objetivo geral a busca por compreender a estrutura em que se organizam os pesquisadores atuantes na área de Indexação, mais especificamente 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.

2 INDEXAÇÃO E AVALIAÇÃO DA INDEXAÇÃO

Barité 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 a 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.



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 (FUJITA, 2003).

3 ESTUDOS BIBLIOMÉTRICOS EM CIÊNCIA DA INFORMAÇÃO

A análise de redes sociais enfatiza características da estrutura social enquanto rede, onde os nós integram indivíduos sociais, ressaltando que as comunidades analisadas são “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) situa os estudos métricos como uma das principais metodologias para a compreensão de um domínio.

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 (GRÁCIO; OLIVEIRA, 2013).

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. O princípio orientador dessa tipologia de análise é que “[...] quando dois documentos ou autores são citados juntos em um trabalho posterior, existe, na perspectiva do autor citante, uma proximidade de assunto entre os citados” (GRÁCIO; OLIVEIRA, 2013, p. 197).

4 METODOLOGIA

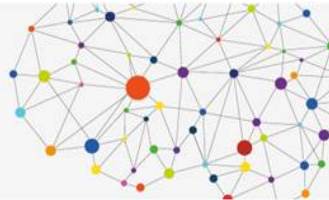
A princípio realizou-se levantamento dos periódicos nacionais em Ciência da Informação, para tal teve como apoio a relação *qualis* publicada pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) em 2012, chegando a um total de 30 periódicos. Em seguida realizou-se coleta bibliográfica nas páginas *online* de cada periódico, utilizando-se do recurso de pesquisa oferecido pelo próprio periódico. Foi feita consulta pelos termos “avaliação indexação”, retornando um total de 48 artigos.

Em seguida foi realizada uma consulta interna aos artigos recuperados, com uso de ferramenta de localização de palavras do *software* visualizador de documentos empregado, buscando confirmar o assunto tratado pelo artigo. Foram utilizados os termos “indexação”, “consistência”, “exaustividade”; “precisão”, definindo-se que o artigo deveria conter ao menos dois dos termos consultados. Este filtro resultou na seleção de 26 artigos.

Passou-se então ao levantamento dos dados, iniciando-se por uma súmula dos artigos selecionados, contendo o título e seu conjunto de referências. Levantou-se o conjunto total das referências, foram desdobradas as referências contendo autoria múltipla de modo a contemplar todos os autores do artigo. A seguir fez-se a contagem de citações para verificação dos autores mais citados. Identificou-se um total de 433 referências, tendo sido mencionados um total de 288 autores. Para realização da análise de cocitação foram considerados os autores que foram citados em ao menos três trabalhos, correspondendo a um total de 31 autores.

Construiu-se, então, 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, 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. 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.



5 RESULTADOS E DISCUSSÕES

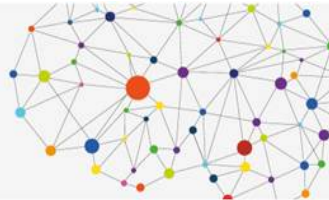
A partir da contagem de citações recebidas foi possível obter os autores mais citados nos artigos analisados. Foram selecionados os autores que receberam menção em ao menos 4 trabalhos (aproximadamente 15% dos artigos), o que somou um total de 17 autores que encontram-se distribuídos na Tabela 1. Vale ressaltar que dos 26 artigos analisados apenas 3 (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 o grupo de autores selecionados constitui-se representativo para a análise.

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
PIEIDADE, Maria Antonietta R. (Brasil)	4
ROBREDO, Jaime (Brasil)	4
SILVA, Maria dos Remédios (Brasil)	4
VAN SLYPE, Georges (França)	4

Fonte: Elaboração própria

Podemos observar, entre os pesquisadores mais citados, que a literatura brasileira sobressai-se entre a internacional e, ao observarmos que dos 12 pesquisadores mais citados, 6 são de nacionalidade diversa da brasileira, sendo que destes 6, apenas 2 ainda se mantêm em atividade, entendemos que o tempo de absorção da literatura estrangeira pela área parece ser relativamente grande. 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 observada.

Apresentamos, na Figura 1, 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.

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, o que demonstra a inserção de seus escritos no campo analisado. Destaque-se também os pesquisadores Chaumier, Fosskett e Cutter, pioneiros nos estudos de tratamento temático da informação, Chaumier representante da corrente teórica francesa denominada *Analyse documentaire*; Fosskett e Lancaster, como representantes da corrente teórica inglesa, *indexing*; e Cutter da denominada corrente teórica *subject cataloguing*, de origem norte-americana (GUIMARÃES, 2008). Pode-se ressaltar, também, a proximidade das pesquisas brasileiras com a literatura espanhola, representada pelos pesquisadores Moreiro González, Moya Anegón, Fernandez Molina e Cabré.

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 Fosskett, 7 cocitações; Lancaster e Boccato, 6 cocitações; Fujita e Fosskett, 6 cocitações; Fosskett e Boccato, 5 cocitações; Lancaster e Rubi, 5 cocitações; Fujita e Chaumier, 5 cocitações; sendo que, entre estas frequências mais altas, nenhuma se trata de coautoria.

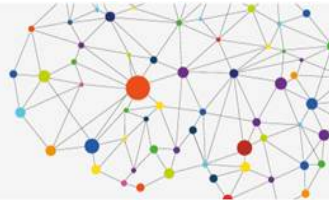
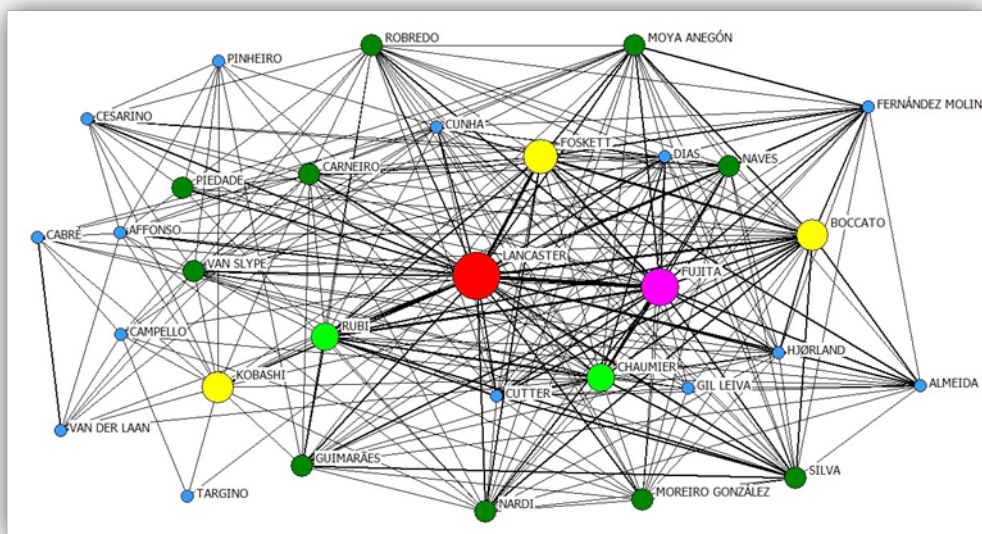


Figura 1: Rede de cocitação de autores em avaliação da indexação



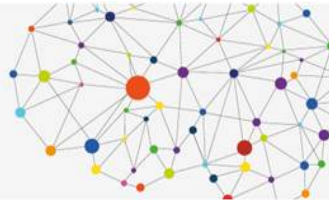
Fonte: Elaboração própria

Para o cálculo dos indicadores de rede a princípio determinou-se a densidade da rede, que apresentou valor 58,9%, valor relativamente alto, demonstrando que há coesão entre os pesquisadores desta temática. Quanto ao indicador centralidade de grau, destaca-se o autor Lancaster que foi cocitado com 100% dos atores da rede. Em seguida destaca-se o autor Chaumier que foi cocitado com 86,7% dos demais, seguido por Fujita cocitada com 83% dos atores. Percebe-se a prevalência da corrente inglesa entre os autores mais citados, também com a menção à corrente francesa.

No índice de centralidade de intermediação destacou-se 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 à zero. Em razão desta rede não ser completamente conectada, não foi possível calcular a centralidade de proximidade, dado que esta medida aplica-se apenas a redes fechadas.

6 CONSIDERAÇÕES FINAIS

Por meio do estudo realizado foi possível perceber 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 se trata de uma rede densa e coesa, tendo como



atores mais centrais Lancaster, que além de haver sido o autor mais citado possui também os maiores índices de centralidade de grau e de intermediação, e Chaumier, que embora tenha sido apenas o 5º 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, onde foi possível observar a organização dos pesquisadores da temática. Consideramos necessária a ampliação de estudos dessa natureza dada a centralidade das questões de Organização do conhecimento para a Ciência da Informação.

REFERÊNCIAS

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

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.

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**, v. 1, p. 77-99, 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.

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

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

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... São Paulo: Cultura Acadêmica. 2009. p. 81-92.