

Organização e representação do conhecimento: contribuições aos estudos métricos



Cibele Araújo Marques dos Santos
Universidade de São Paulo
cibeleac@usp.br

1 Introdução

Este estudo tem por objetivo identificar a contribuição da Organização e Representação do Conhecimento (ORC) para os estudos métricos da produção científica em Ciência da Informação no Brasil que posteriormente possibilitem desenvolver metodologias para indicadores temáticos de pesquisa.

A produção científica cresce em proporção ilimitada na sociedade atual devido aos recursos tecnológicos e a multiplicação de ferramentas para divulgação destas informações, incluindo o acesso a informação permitido pela web. Do mesmo modo, percebem-se desafios expressivos quanto ao desenvolvimento de políticas em pesquisa, principalmente nas universidades que são grandes produtoras de informação científica e tecnológica no Brasil.

Neste contexto, a análise da produção científica torna-se recurso fundamental para o planejamento de políticas científicas nacionais e institucionais e os indicadores bibliométricos podem identificar padrões e tendências das áreas. São estas métricas que permitem medidas de atividade, impacto, colaboração e dinâmica da comunicação de resultados de pesquisa que são representados por indicadores, analisando de forma quantitativa os dados de publicação e citação (SANTIN, BRAMBILLA, STUMPF, 2013).

Por outro lado, a análise da temática da produção científica permite uma visão mais qualitativa na medida que com o mapeamento dos temas podemos verificar a cobertura da pesquisa e as lacunas existentes. Porém, as abordagens para este tipo de análise geram debate em relação aos métodos mais adequados para representar um domínio disciplinar a partir de uma estruturação temática, no que se refere aos níveis de subjetividade, validade e atualização do modelo de classificação utilizado (LIBERATORE e HERRERO-SOLANA, 2013). A dispersão decorrente da variabilidade e granularidade de assuntos abordados em um conjunto de artigos de periódicos é grande quando analisamos as palavras-chaves de cada artigo, além de ocorrerem inconsistências, uso de sinônimos, variações linguísticas, bem como critérios diferenciados de nivelamento do tratamento da informação pelos autores, principalmente em áreas interdisciplinares. As ferramentas de área de Organização e Representação do Conhecimento, podem auxiliar na superação destas dificuldades para realizar mapeamentos mais consistentes.

2 Metodologia

Foi realizada pesquisa exploratória em artigos de revistas brasileiras para identificar a metodologia utilizada na definição das temáticas em Ciências da Informação nos estudos métricos.

O levantamento bibliográfico foi realizado através de pesquisa na base Brapci com estratégia de busca, realizada em maio de 2015, que utilizou o campo de palavras-chaves pesquisando os termos: estudos métricos (resultou em 9 artigos), bibliometria (120 artigos), análise bibliométrica (18 artigos), análise temática (2 artigos, apenas um em bibliometria) e mapeamento (15 artigos), sendo que estes resultados trouxeram muitos artigos, nem sempre pertinentes à bibliometria temática.

Os artigos foram selecionados a partir do título e do resumo, visando análise dos recursos da área de Organização e Representação do Conhecimento utilizados pelos autores para realizar os mapeamentos temáticos.

2.1 Estudos métricos: características e temáticas

Existem diversos artigos, trabalhos publicados em anais de eventos, dissertações e teses com abordagens bibliométricas quantitativas e qualitativas que analisam a produção científica em Ciência da Informação a partir de indicadores relacionados aos autores, como afiliação e co-autoria, referências e citações de autores, revistas e co-citações, porém não são muitos os trabalhos que incluem análise temática.

A produção científica, como apontado por Liberatore et al (2007), considerando docentes e pesquisadores em nível de pós-graduação, apresenta-se analisada por fontes, metodologias, temáticas, linhas de pesquisa e formas de difusão. Os autores em trabalho de revisão consistente informam que as teses defendidas nos programas foram estudadas também nas questões relacionadas a seu desenvolvimento, problemáticas, objetos de estudo e modos de citação.

Nesse sentido, Macedo (1987) publicou artigo sobre pesquisa em Ciência da Informação e Biblioteconomia que abordava entre outras questões a temática de teses, sendo que os dados foram coletados nos Catálogos de Dissertação e Teses em Ciência da Informação e Biblioteconomia publicados entre 1982 a 1985 pelo IBICT. Os assuntos identificados para as teses e dissertações foram retirados da indexação realizada pela fonte bibliográfica destes dados, os catálogos citados, e utilizavam tesauro construído a partir do documento “Percepção estrutural da Ciência da Informação e áreas correlatas” com as seguintes divisões: Generalidades; Ensino, pesquisa e atividade profissional; Organização das atividades de informação e de bibliotecas; Estudos da literatura e do documento; Entrada, tratamento, armazenamento da informação; Recuperação e disseminação da informação, Transferência e uso da informação e áreas correlatas. No artigo citado, as teses foram distribuídas por grandes assuntos com titulação e dados geográficos e também analisadas pelos assuntos mais específicos do tesauro.

Esta divisão de assuntos também era utilizada segundo o Instituto Brasileiro de Informação em Ciência e Tecnologia (1984) pela Bibliografia Brasileira de Ciência da Informação que referencia artigos de periódicos na década de 80 do século 20.

Poblacion e Noronha realizaram diversos estudos sobre a produção da pós-graduação no Brasil a partir do Núcleo de Produção Científica da ECA-USP. Em artigo publicado em 2002 analisam a produção caracterizando-a tematicamente pelas linhas de pesquisas institucionalizadas de seus autores nos programas de pós-graduação de

cinco universidades brasileiras e incluíam 22 linhas de pesquisas diferentes que continham temáticas comuns, porém dispersavam os assuntos dos documentos.

Em outro artigo Queiroz e Noronha (2004), mapearam tematicamente as dissertações e teses do programa de pós-graduação da ECA-USP utilizando para classificação temática o *Broad Subject Headings*, a lista de cabeçalho de assunto do *Library and Information Science Abstracts* (LISA), que em sua versão impressa apresenta 19 categorias temáticas subdivididas em assuntos específicos. Informa que cada um dos 114 itens foi classificado em apenas umas das 19 categorias da lista, mais uma subcategoria dentro deste assunto principal, além de uma das linhas de pesquisas do programa. As categorias utilizadas foram: Biblioteconomia e Ciência da Informação; Profissões; Bibliotecas e centros de informação; Uso de bibliotecas e usuários; Materiais; Organização; Prédios de bibliotecas; Tecnologia em bibliotecas; Serviços técnicos; Difusão da informação; Controle bibliográfico; Registro bibliográfico; Armazenamento e recuperação da informação; Tecnologia da informação e comunicação; Leitura; Mídia; Conhecimento e aprendizado; Gerenciamento de registros; Outros assuntos correlatos.

Mueller e Pecegueiro (2001), analisaram artigos da revista *Ciência da Informação* de 1990-1999, que consideraram como elementos significativos da produção científica da área, indicando os temas predominantes e comparando a frequência destes e os descritores registrados na base LISA, que apresentaram diferenças expressivas. Foi utilizada uma tabela com 9 macrotemas: Ciência da Informação, Biblioteconomia e Documentação; Ensino, atividade profissional e pesquisa; Organização e gerência de atividades de informação, de bibliotecas e centros de pesquisa; e Estudo de usuários, transferência e uso da informação e da biblioteca; Estudos da literatura e do documento; Prédios de bibliotecas; Serviços técnicos; Entrada, tratamento, armazenamento, recuperação e disseminação da informação, Outros assuntos correlatos ou adicionais.

Percegueiro (2002) publicou outro trabalho sobre a temática dos artigos de seis revistas brasileiras em *Ciência da Informação* que foram classificados pela tabela anterior agora com o acréscimo do tema “Política de informação e política científica e tecnológica”. A classificação foi comparada novamente com a ocorrência de temas na base LISA e foram também encontradas diferenças. Neste trabalho, a autora chama a atenção para a oscilação de determinados temas ao longo dos anos como por

exemplo o pico de publicação sobre “Ensino, atividade profissional e pesquisa em 1990 caindo para 5 artigos em 1991 e 2 em 1992. O tema Prédios não teve publicações no período e Serviços Técnicos teve poucos artigos publicados.

3 Organização e Representação do Conhecimento e métricas

A contribuição da área de ORC aos estudos métricos, nas pesquisas destacadas acima, mostra procedimentos de classificação por macrotemas ou por assuntos específicos, o uso do tesouro do IBICT, da lista de termos da base LISA e de vocabulário controlado adaptado para estudos métricos temáticos. A utilização destas linguagens documentárias ocorreu através da indexação e às vezes reindexação dos dados coletados, porém não foram discutidas, na maioria dos estudos, as questões linguísticas e terminológicas envolvidas na escolha dos vocabulários e as metodologias dos procedimentos de análise documentária.

Em estudo sobre a revista *Ciência da Informação*, Silva (2002), discute rapidamente aspectos relacionados ao vocabulário controlado e questões de linguagem nos índices para análise temática de produção científica. Vanz e Stumpf (2010), discutem os processos de avaliação da produção incluindo fontes de coleta de dados, bases de dados nacionais e internacionais, Google Acadêmico e outros, softwares utilizados, procedimentos para limpeza e padronização de dados, considerados imprescindíveis para eliminar as inconsistências, bem como a organização da produção em áreas do conhecimento. As autoras ponderam que a inserção da produção em grandes áreas torna-se uma forma de evitar a sobreposição de assuntos, como o uso de esquemas de classificação como as 15 divisões do conhecimento do ISI ou as tabelas de áreas do CNPq-CAPES.

Na pesquisa realizada para a aplicação bibliométrica na análise de dissertações e teses, Kobashi e Santos (2010), propõem abordagem interdisciplinar dos Estudos Sociais da Ciência, da Organização e Representação do Conhecimento e dos métodos bibliométricos avançados. Os autores lembram que os estudos bibliométricos fundamentam-se tradicionalmente na descrição e quantificação e propõem combinar abordagem de estudos quantitativos e qualitativos. Nestes estudos, os indicadores de relação são construídos pela co-ocorrência, seja autoria, citações ou palavras, e podem ser aplicados na elaboração de mapas estruturados de

conhecimento e redes de relacionamento, considerando que estes mapas são representações da produção científica da área analisada. Existe também a possibilidade de utilizar conceitos produzidos pela própria área, que traz “garantia literária na geração das representações cartográficas da institucionalização cognitiva de um dado campo do conhecimento”.

A utilização das palavras-chave dos autores nas análises temáticas foi recorrente, sendo que Araujo e Melo (2011) na análise dos artigos publicados na revista *Perspectivas em Ciência da Informação*, optou por utilizá-las e observou que a falta de controle terminológico das palavras-chave, gera uma “variação imensa na maneira de expressar determinados conceitos da área”. Comentam que, quando foi possível, realizaram a padronização terminológica e que a dispersão dos termos e as diferentes categorias usadas como palavras-chave, impediram uma análise mais aprofundada possibilitando apenas a visualização de um quadro sugestivo de temas abordados.

Por sua vez, Liberatore e Herrero-Solana (2013) analisaram os temas de investigação científica em *Ciência da Informação* entre 2000 e 2009 estudando artigos de 4 revistas da área, e em sua abordagem metodológica optaram por não adotar um esquema específico, utilizando as palavras-chaves dos autores por entenderem ser uma representação ajustada a percepção dos próprios autores ao seu campo de atuação semântico. Obtiveram uma lista de 2498 palavras pertencentes a 965 artigos, analisaram este universo pelo cálculo da co-ocorrência, fixando um mínimo de 8 ocorrências como uma representação significativa.

Realizaram também nesta pesquisa uma análise dos artigos pelos macrotemas dos grupos de trabalho da ANCIB. A matriz de dados resultante dessas análises em conjunto, foi apresentada por meio de uma rede social cujos nós indicados pelo número de relações entre as ocorrências permitiu identificar o grau e a centralidade destas relações. Nos resultados, os autores chamam a atenção para os componentes da subjetividade dos autores nas temáticas, a questão do nível de consenso na terminologia empregada pela comunidade científica e a dispersão das expressões utilizadas.

4 Considerações finais

Foi realizado um levantamento bibliográfico não exaustivo selecionando artigos que tenham feito mapeamento temático em estudos bibliométricos, analisando na metodologia destes trabalhos os procedimentos da Organização e Representação e Organização do Conhecimento empregados. Observou-se que os autores, normalmente, não se referem ao processo de indexação de forma detalhada com a indicação da metodologia e critérios, por exemplo, mostrando apenas a lista de termos utilizada, se a mesma pertence a um tesouro ou vocabulário institucional, não discutindo também as categorias presentes.

Não houve uma lista de assuntos predominante, mas foram localizados artigos que mostraram questões quanto a linguagem e uso de grandes áreas temáticas nos procedimentos, a importância de uma abordagem interdisciplinar e o uso dos recursos da ORC.

No entanto, os artigos não apresentaram discussões sobre a influência das terminologias utilizadas nos resultados da pesquisa, embora em um trabalho tenha ocorrido uma padronização terminológica e em outro um artigo que trabalhou com as palavras-chaves dos autores, tenha sido ressaltado o aspecto da subjetividade que a escolha envolveu.

Podemos observar que contribuição da ORC para os estudos métricos não se esgota nos pontos levantados e nem nos trabalhos identificados e que existe ainda um campo de pesquisa conjunta que pode ser mais explorado permitindo dessa forma o aprimoramento das metodologias e o aprofundamento dos mapeamentos temáticos em Ciência da Informação.

Referências

ARAÚJO, C. A. A.; MELO, M. O. T. Análise dos quinze anos do periódico *Perspectivas em Ciência da Informação*. **Perspectivas em Ciência da Informação**, v.16, n.4, p.243-256, 2011.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Bibliografia Brasileira de Ciência da Informação** v.6 (1980;83). Brasília, IBICT, 1984.

KOBASHI, N. Y.; SANTOS, R. N. M. dos. Arqueologia do trabalho imaterial: uma aplicação bibliométrica a análise de dissertações e teses. **Encontros Bibli**, n. esp. p. 106-115, 2008.

LIBERATORE, G.; HERRERO-SOLANA, V. Caracterización temática de la investigación en Ciencia de la Información en Brasil en el período 2000-2009. **Transinformação**, v.25, n.3, p.225-235, 2013.

LIBERATORE, G.; HERRERO-SOLANA, V.; GUIMARÃES, J. A. C. Análise bibliométrica do periódico brasileiro Ciência da Informação durante o período 2000-2004. **Brazilian Journal of Information Science**, v.1, n.2, p. 3-21, 2007.

MACEDO, N. D. de. Pesquisa em Ciência da Informação e Biblioteconomia: implicações na pós-graduação; análise temática. **Ciência da Informação**, v.16, n.2, p.129-144, 1987.

MUELLER, S. P. M.; PECEGUEIRO, C. M. P. de. O periódico Ciência da Informação na década de 90: um retrato da área refletido em seus artigos. **Ciência da Informação**, v. 30, n. 2, p.47-63, 2001.

PECEGUEIRO, C. M. P. de. Temática dos artigos de periódicos brasileiros na área da ciência da informação na década de 90. **Transinformação**, v.14, n.2, p.117-131, 2002.

POBLACION, D. A.; NORONHA, D. P. Produção das literaturas “brancas” e “cinzenta” pelos docentes-doutores dos programas de pós-graduação em ciência da informação no Brasil. **Ciência da Informação**, v.31, n.2, p.98-106, 2002.

QUEIROZ, F. M.; NORONHA, D. P. Temática das dissertações e teses em Ciência da Informação no Programa de pós-graduação em Ciências da Comunicação da USP. **Ciência da Informação**, v.33, n.2, p. 132-142, 2004.

SILVA, F. M. e. Análise da revista Ciência da Informação disponibilizada na SciELO a partir do seu vocabulário controlado. **Transinformação**, v.14, n.2, p.133-138, 2002.

VANZ, S. A. de S.; STUMPF, I. R. C. Procedimentos e ferramentas aplicados aos estudos bibliométricos. **Informação & Sociedade**, v.20, n.2, p.67-75.

Apoio: PROJETO FAPESP 2013-04109-7