



XVII Encontro Nacional de Pesquisa em Ciência da Informação (XVII ENANCIB)

GT 2 – Organização e Representação do Conhecimento

ANÁLISE BIBLIOMÉTRICA DA PRODUÇÃO CIENTÍFICA SOBRE AS LINGUAGENS DE INDEXAÇÃO PUBLICADA NOS ANAIS DE CONGRESSO DO ENANCIB NO PERÍODO DE 2012 A 2015

BIBLIOMETRIC ANALYSIS OF SCIENTIFIC PRODUCTION ABOUT THE INDEXING LANGUAGES PUBLISHED IN THE ANNALS OF CONGRESS OF THE ENANCIB FROM 2012 TO 2015

Italo Rodrigues Castro¹ e Marlene Oliveira²

Modalidade da apresentação: Pôster

Resumo: O objetivo da pesquisa foi analisar a produção científica veiculada nos Anais dos Encontros Nacionais de Pesquisa em Ciência da Informação e Biblioteconomia - ENANCIB sobre as linguagens de Indexação, no período de 2012 a 2015. Para análises dos artigos utilizaram-se técnicas da Bibliometria. Para análise dos dados utilizou-se as categorias: autores mais produtivos e suas respectivas instituições; instituições mais produtivas; titulação dos autores mais produtivos; anos mais produtivos; linguagens de indexação mais utilizadas, e assuntos predominantes. Foi analisada uma amostra de 73 artigos, com o total de 102 autores (foram considerados também os co-autores), os 10 autores mais produtivos foram responsáveis por 36 artigos, que corresponde a 49,31% da produção dos artigos. Estes autores estão vinculados a seis instituições de ensino: UFMG, UFF, UNESP, IBICT, UEL e UFBA. Foi verificado que a UFMG é a instituição vinculada mais participativa na produção científica sobre esta temática apresentando 34,61% do total dos artigos desenvolvidos. Os autores mais produtivos publicaram de 3 a 5 artigos no período estudado. A linguagem mais investigada nas publicações no período foi o Sistema de Classificação Bibliográfica no ano de 2014, que foi o ano em que mais se estudou sobre o tema no ENANCIB.

Palavras-chave: Bibliometria. Indexação. Linguagens de Indexação.

Abstract: *The objective of the research was to analyze the scientific production conveyed in the Annals of the National Research in Information Science and Library - ENANCIB about the Indexing languages, from 2012 to 2015. Techniques of Bibliometrics were used for analysis of the papers. For data analysis were used the categories: the most productive authors and their affiliations; the most productive institutions; the instruction degree of the most productive authors; the most productive years; the most used indexing*

¹Escola de Ciência da Informação

²Escola de Ciência da Informação

languages and predominant issues. A sample of 73 papers was analyzed with a total of 102 authors (co-authors were also considered), the 10 most productive authors were responsible for 36 papers, which corresponds to 49,31% of the production of papers. These authors are linked to six educational institutions: UFMG, UFF, UNESP, IBICT, UEL and UFBA. It was verified that the UFMG is the most participatory institution in the scientific production about this theme, presenting 34,61% of total developed papers. The most productive authors published 3 to 5 articles in the period studied. The most investigated language in publications in the period was the Bibliographic Classification System in 2014, which was the year this subject was the most studied in ENANCIB.

Keywords: *Bibliometrics. Indexing. Indexing Languages.*

1 INTRODUÇÃO

A Bibliometria é um sub-campo da Ciência da Informação que analisa a produção científica e tecnológica por meio de estudos quantitativos e estatísticos das publicações. O estudo da literatura científica de uma área ou de uma instituição permite a identificação de autores mais produtivos, títulos de periódicos científicos preferidos por pesquisadores, o desenvolvimento de indicadores, análise de citações entre outros. A Bibliometria reúne três principais leis: a Lei de Lotka (1926), pela qual se mede a produtividade científica de autores; a Lei de Bradford (1934), para indicação da produtividade de periódicos; e, finalmente, a Lei de Zipf (1949) que trata da medição da frequência de palavras em um texto (ARAÚJO 2006; GUEDES, 2012).

Com o aumento constante de informações surge a necessidade de ampliar o gerenciamento do conteúdo de documentos disponibilizados, tanto em formato impresso, quanto no formato eletrônico. Para isso, utiliza-se a indexação para a criação de índices para organizar documentos. Entretanto, o conceito de índice pode ser representado por uma lista de documentos ordenados de forma alfabética ou sistemática dos itens do conteúdo de um documento, possibilitando ao usuário identificar ou localizar o documento. Este mecanismo é usado tanto na armazenagem, como na busca e na recuperação da informação (CAMPOS, 2006).

Este trabalho objetivou-se realizar uma análise bibliométrica da produção científica e tecnológica sobre as linguagens de indexação no conteúdo dos anais do Encontro da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB), o ENANCIB no período de 2012 a 2015. A comunidade científica da área de Organização e Representação do Conhecimento (ORC) está representada pelo Grupo de Trabalho Organização e Representação do Conhecimento - GT2, que anualmente são publicados artigos dos anais de congressos apresentados por pesquisadores dos Programas de Pós-Graduação em Ciência da Informação (PPGCI), motivo pelo qual foi escolhido fazer a extração dos dados desta pesquisa.

Diante do exposto, a análise bibliométrica possibilitou a identificação da produção bibliográfica dos autores, os títulos destes artigos, bem como trouxe um comparativo da produtividade dos autores e das tendências de pesquisa sobre as linguagens de indexação.

2 FUNDAMENTOS TEÓRICOS

2.1 BIBLIOMETRIA

O desenvolvimento da Bibliometria iniciou-se por E. Wyndham Hulme em 1922, a partir da elaboração de leis empíricas sobre o comportamento da produção científica e tecnológica em diversas áreas do conhecimento. A Bibliometria consiste na aplicação de técnicas estatísticas para medir a informação sobre a produção científica de um pesquisador, de uma área científica, de uma instituição ou de um país (VANTI, 2002; GUEDES; BORSCHIVER, 2005). Alvarado (2007), ao abordar a história, a estrutura e a legitimação da Bibliometria apresenta uma síntese, destacando alguns autores e suas conceituações:

Potter (1981) define a Bibliometria como o estudo e a forma de medir os padrões de publicação da comunicação escrita e de seus autores. Já para Ikpaahindi (1985), a Bibliometria é um termo genérico que descreve uma série de técnicas que buscam quantificar o processo de comunicação escrita. Essas técnicas têm sido usadas na identificação dos autores mais produtivos, na identificação de paradigmas na ciência, na fusão e fissão de disciplinas científicas e na identificação dos periódicos mais produtivos em diferentes campos, etc. (ALVARADO, 2007, p. 28).

Para Vanti (2002), a bibliometria refere-se a um conjunto de métodos de pesquisa que utiliza análises quantitativas e estatísticas para o mapeamento do conhecimento de um campo científico, e para análises do comportamento dos pesquisadores e da literatura produzida. Tague-Sutcliffe (1992) *apud* Macias-Chapula (1998) define a bibliometria como o estudo dos aspectos quantitativos da produção, disseminação e uso da informação registrada.

Com o passar do tempo, a análise bibliométrica passou a ter o formato de produção bibliográfica, como artigos de periódicos, no qual é medida a produtividade de autores, o estudo de citações, a análise da produção científica e a busca de benefícios práticos imediatos para bibliotecas tais como o desenvolvimento de coleções, a gestão dos serviços prestados para a biblioteca e o aumento do controle bibliográfico, o que possibilita conhecer o tamanho e as características dos acervos, elaborar previsões de crescimento, entre outros.

2.2 LINGUAGENS DE INDEXAÇÃO

Conhecida como o produto da criação de índices, o termo indexação (natural do latim, *index*) refere-se ao processo que possibilita o usuário ter acesso à informação através da análise documental ou análise de assunto, e pela tradução da linguagem natural (linguagem do usuário) para a linguagem documentária (linguagem de indexação), que permite traduzir de maneira objetiva o conteúdo do documento de acordo com a sintaxe de indexação (DA SILVA; FUJITA, 2012).

Na opinião de Fujita (2009), as linguagens de indexação representam o conteúdo do documento de forma consistente, permitindo a tradução entre o vocabulário do indexador e a linguagem do usuário. Além disso, proporcionam meios através dos quais o usuário modela sua busca e assim, obtém um maior número de documentos existentes no sistema. Na etapa inicial da indexação encontra-se a análise de assunto, em que o indexador procura compreender de maneira geral o documento para identificar os termos representativos na recuperação da informação.

De acordo com Lancaster (2004), o processo de indexação deve contemplar o assunto do documento, o motivo da sua inclusão no acervo e quais os aspectos são de interesse para os usuários, como o usuário vai procurar sobre o assunto e se a informação retornada é satisfatória. Entre as principais linguagens de indexação estão as alfabéticas e as simbólicas (códigos ou símbolos).

Os primeiros tipos de linguagens alfabéticas são as listas de cabeçalhos de assuntos e os tesouros. As linguagens simbólicas são representadas pelos Sistemas de Classificação Bibliográfica, como por exemplo, a Classificação Decimal Universal (CDU), a Classificação Decimal de Dewey (CDD) ou as classificações facetadas, como a *Colon Classification* de Ranganathan.

As listas de cabeçalhos de assuntos (*Library of Congress Subject Headings*) são usadas principalmente pelas bibliotecas públicas e de universidades. Já os tesouros são voltados para documentos de uma determinada área do conhecimento. Trata-se de uma lista de termos que representam de maneira normalizada o conteúdo dos documentos analisados em um sistema para apoiar o usuário na indexação e recuperação de documentos. Finalmente, os Sistemas de Classificação Bibliográfica podem ser vistos como um mapa completo de qualquer área do conhecimento, com seus conceitos e suas relações.

3 PROCEDIMENTOS METODOLÓGICOS

Este trabalho trata-se de uma pesquisa descritiva. A coleta de dados realizou-se nos Anais do Enancib no período de 2012 a 2015. Os ENANCIBs são encontros anuais de Ciência da Informação e Biblioteconomia no âmbito da Associação Nacional de Pesquisa e Pós-graduação em Ciência da Informação e Biblioteconomia – ANCIB.

Na localização do tema, utilizaram-se as palavras-chave relacionadas ao tema tais como: “Vocabulário Controlado”, “Linguagem Documentária”, “Linguagem de Indexação”, “Metadados”, “Tesouro”, “Recuperação da Informação”, “Ontologia”, “Indexação Automática”, “Organização da Informação”, “Recuperação da Informação”, “Tratamento Temático da Informação”, “Representação do Conhecimento”, “Classificação Facetada”, “Análise

Documentária”, “Análise de Assunto”, “Folksonomia”, “Taxonomia” entre outras.

A amostra engloba apenas os artigos de comunicação oral do grupo de trabalho Organização e Representação do Conhecimento - GT2, sendo que o critério de exclusão proposto inclui os artigos que não tratam sobre linguagens de indexação, pôsteres ou que não apresentam nenhum dos descritores que não pertençam ao tema escolhido. Assim, dos artigos selecionados foram levantados os dados: os autores e co-autores; titulação; afiliação; título do artigo; assunto e ano do artigo. Os dados analisados foram reunidos pelas seguintes categorias: autores mais produtivos; instituições mais produtivas; instituições dos autores mais produtivos; titulação dos autores mais produtivos; anos mais produtivos; linguagens de indexação mais utilizadas, dentre elas: Listas de cabeçalhos de assuntos, Tesouros e Sistemas de Classificação Bibliográfica.

4 ANÁLISE DOS DADOS E RESULTADOS

Os artigos dos anais de congresso do ENANCIB no período de 2012 a 2015 constituem o corpus da pesquisa. A quantidade de artigos encontrados após a coleta totalizaram 73 (setenta e três artigos), os quais foram analisados e apresentados a seguir.

4.1 PRODUTIVIDADE DOS AUTORES

Ao realizar a contagem dos autores, consideraram-se também os co-autores, atribuindo uma unidade para cada autor do artigo referenciado. Observou-se um total de 102 autores, e tais autores participaram em 73 artigos publicados no ENANCIB no período. Considerando o total de autores, verificou-se que 76 publicaram apenas um artigo, o que corresponde a 74,5% do total de autores no período. O número de autores que publicaram dois artigos foram 16, representando 15,68%. A Tabela 1 apresenta os autores mais produtivos que tiveram o total de 3, 4 e 5 artigos publicados, o que corresponde a 4,9%, 3,92% e 0,98% respectivamente do total de autores. O total de publicações dos autores mais produtivos foram 36 artigos, representando 49,31% do total das publicações.

4.2 INSTITUIÇÕES MAIS PRODUTIVAS

As instituições mais produtivas que publicaram nos anais de congresso do ENANCIB no período de 2012 a 2015 foram a UFMG, a UFF, a UNESP, a UNB e a UFBA com a publicação de 18, 12, 11, 6 e 5 artigos, representando um total de 52 publicações. A UFMG destaca-se como a principal instituição vinculada à produção científica, o que representa 20,22% do total de participações em artigos relacionada com a temática proposta.

4.3 INSTITUIÇÕES DOS AUTORES MAIS PRODUTIVOS

Analisou-se um total de 36 artigos em que os autores que mais publicaram, tiveram a quantidade de 3, 4 ou 5 artigos publicados dentro do período estudado e todos possuem a titulação de Doutor (D), conforme apresentado na Tabela 1. A partir desta amostra, verificou-se que os autores mais produtivos publicaram 13, 7, 6, 4 e 3 artigos e estão vinculados à UFMG, UFF, IBICT, UEL e UFBA respectivamente. Outro achado importante refere-se às instituições, das cinco mais produtivas, duas estão com a nota seis na CAPES, a UFMG e a UNESP.

TABELA 1 - Publicações, instituição e titulação dos autores mais produtivos

Autor	Nº Publicações	Instituição	Titulação
Renato Rocha Souza	5	UFMG	D
Gercina Ângela B. de Oliveira Lima	4	UFMG	D
Maria Aparecida Moura	4	UFMG	D
Maria Luiza de Almeida Campos	4	UFF	D
Rosali Fernandez de Souza	4	IBICT	D
Brígida Maria Nogueira Cervantes	3	UEL	D
João Batista Ernesto de Moraes	3	UNESP	D
Mariângela Spotti Lopes Fujita	3	UNESP	D
Rosa Inês de Novais Cordeiro	3	UFF	D
Zeny Duarte de Miranda	3	UFBA	D

Fonte: Dados da pesquisa

4.4 ANOS COM MAIOR NÚMERO DE PUBLICAÇÕES

O total de publicações por ano no período de 2012 a 2015 foram 73 artigos. Por ano obteve-se um total de 7, 19, 27 e 20 artigos publicados respectivamente no período. Foi constatado um aumento de 16,44% do ano de 2012 para 2013, um aumento de 10,96% do ano de 2013 para 2014 e uma redução de 9,59% do ano de 2014 para 2015 do total da produção científica no período.

4.5 LINGUAGENS DE INDEXAÇÃO MAIS UTILIZADAS

As principais linguagens de indexação pesquisadas nos artigos foram às Listas de Cabeçalhos de Assuntos, os Tesouros e os Sistemas de Classificação Bibliográfica. Estas linguagens foram identificadas por meio de palavras-chave e dos assuntos retirados dos resumos dos artigos. Verificou-se que a linguagem mais abordada nas publicações no período é o Sistema de Classificação Bibliográfica, no ano de 2014, aparecendo em 12 artigos. Em segundo lugar, foram os Tesouros, em 5 artigos no ano de 2013, em 4 artigos no ano de 2014 e 2015. A linguagem de indexação menos estudada foram as Listas de Cabeçalhos de Assuntos, que apareceram em 2 artigos em 2012 e em apenas 1 artigo em 2013.

5 CONSIDERAÇÕES FINAIS

Esta pesquisa permitiu conhecer o desenvolvimento dos estudos sobre as linguagens de indexação por meio da análise Bibliométrica, o que confirma a evolução do estudo de linguagens de indexação em cada ano. O resultado sobre os autores mais produtivos mostra um elenco de pesquisadores reconhecidos e, por isso, são demandados por agências de financiamento como CAPES e CNPq como pareceristas em projetos de pesquisas, representantes de área e julgamento de novos cursos na CAPES. Dentre as linguagens mais pesquisadas nota-se pouco interesse pelos cabeçalhos de assunto, atualmente pouco utilizados em bibliotecas convencionais e nenhum uso em base de dados e bibliotecas digitais. A diminuição de comunicação oral em 2015 em relação a 2014 precisa ser melhor pesquisado uma vez que pode trazer prejuízos à temática.

Verificou-se também a contribuição dos ENANCIBs na disseminação de resultados de pesquisa da área da Organização e Representação do Conhecimento dentro da Ciência da Informação.

6 REFERÊNCIAS

- ALVARADO, Ruben Urbizagastegui. **A Lei de Lotka e a produtividade dos autores**. 2007. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.
- ARAÚJO, Carlos Alberto. Bibliometria: evolução história e questões atuais. **Em Questão**, v. 12, n.1, p. 11-32, jan/jun. 2006.
- CAMPOS, M. L. A. Indexação e descrição em arquivos: a questão da representação e recuperação de informações. **Arquivo e Administração**, v. 5, p. 17-31, 2006.
- DA SILVA, Maria dos Remédios; FUJITA, Mariângela Spotti Lopes. A prática de indexação: análise da evolução de tendências teóricas e metodológicas. **Transinformação**, v.16, n.2, p.133-161, mai/ago, 2012.
- FUJITA, M. S. L. **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.
- GUEDES, V. L. S.; BORSCHIVER, S. Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica. In: Encontro Nacional de Ciência da Informação, 6, 2005, Bahia, **Anais...** Salvador: CIFORM, 2005.
- GUEDES, V. L. S. A bibliometria e a gestão da informação e do Conhecimento científico e tecnológico: uma revisão da literatura. **Ponto de Acesso**, v.6, n.2, p.74-109, ago, 2012.
- LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2.ed. Brasília: Briquet de Lemos,

2004.

MACIAS-CHAPULA, Cesar A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, v.27, n.2, p.134-140, maio/ago, 1998.

VANTI, N. A. P. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da Informação**, v.31, n. 2, p. 152-162, maio/ago, 2002.