

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

Ítalo Rodrigues Castro
Escola de Ciência da Informação - UFMG
italocastro@gmail.com

Marlene Oliveira
Escola de Ciência da Informação - UFMG
marlene@eci.ufmg.br

Recebido em: 30/07/16 Aceito em: 03/08/16
--

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 coautores), 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, IBICT, UEL, UNESP e UFBA. Foi verificado que a UFMG é a instituição vinculada mais participativa na produção científica sobre esta temática apresentando 36,11% 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. Os assuntos predominantes no período foram indexação, Organização do Conhecimento, Tesouros, Organização da Informação, Recuperação da Informação e Metadados que representam 12,65%, 12,02%, 8,22%, 7,59%, 6,96% e 6,32% respectivamente do total de 158 assuntos abordados.

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

1 INTRODUÇÃO

A Bibliometria é um subcampo 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 leis que

são: a Lei de Bradford (1934), para indicação da produtividade de periódicos; a Lei de Lotka (1926), pela qual se mede a produtividade científica de autores e, finalmente, a Lei de Zipf (1949) que trata da medição da frequência de palavras em um texto (ARAÚJO *et al.* 2010; 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 tanto, utiliza-se a indexação para a criação de índices ou resumos para organizar documentos. Nesse sentido, o resumo refere-se à representação de forma sucinta das ideias principais de um documento de forma clara e objetiva. 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).

Dessa forma, no processo de indexação cria-se uma lista de termos sintetizando o conteúdo dos documentos com o objetivo de referenciá-los para que sejam recuperados de forma mais eficiente. Na gestão dos documentos, realizada pelo indexador, analisa-se os documentos através do assunto codificando-os para uma linguagem de indexação ou linguagem documentária de acordo com o interesse do usuário. Assim, a linguagem traduz de maneira simples e objetiva o conteúdo do documento utilizando um vocabulário controlado como, por exemplo: lista de cabeçalhos de assuntos, tesouros (descritores), tabelas ou símbolos de classificação (sistemas de classificação bibliográfica). No entanto, pode ser utilizado um vocabulário simples, em que os termos são tirados do próprio

documento. Após a realização da tradução do documento, armazenam-se os índices em uma base de dados em formato impresso ou eletrônico através do indexador. Já os documentos físicos são armazenados normalmente em uma base de dados do acervo de documentos.

O critério utilizado para indexar um documento, deve-se à capacidade do documento ser acessado como fonte de informação em resposta às demandas específicas em índices e bases de dados bibliográficas e, mais recentemente, nos motores de busca na Web. A visibilidade de um documento publicado em periódicos está determinada pela inclusão de palavras-chave dos seus documentos nas linguagens específicas de uma base de dados ou bibliotecas digitais que inclui ou indexa o periódico.

Ao ser indexado, o periódico fornece aos índices, os metadados que contém a descrição referencial básica do conteúdo dos artigos e sua localização, incluindo informações como o nome do autor, título, resumo, palavras-chave que descrevem o conteúdo e, mais recentemente, o link para o texto completo. Assim, os metadados são estruturados nos índices bibliográficos no mesmo idioma e indexam periódicos de diferentes áreas geográficas ou temáticas. Um periódico é mais visível quando ele está indexado nos índices referenciais com mais visibilidade. Por exemplo, no Brasil a principal base de dados bibliográfica é a da *Scientific Electronic Library Online* (SCIELO) (POBLACION *et al.*, 2006).

Este trabalho, portanto, 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, na qual foi realizada sua XIII, XIV, XV e XVI edições nas cidades do Rio de Janeiro, Florianópolis, Belo Horizonte e João Pessoa respectivamente. O ENANCIB

existe desde 1994 e em 2015 completou 11 anos de existência. 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 REFERENCIAL TEÓRICO

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.

Segundo Guedes e Borschiver (2005), o principal marco do desenvolvimento da Bibliometria foi a Lei de Lotka, pois permite a medição da produtividade científica de pesquisadores. Considera-se, porém, a frequência do aparecimento dos diversos autores em um conjunto de documentos. Desta maneira, os pesquisadores que possuem maior prestígio em um domínio do conhecimento produzem muito e aqueles que possuem menor prestígio produzem pouco. Portanto, os cientistas mais produtivos de universidades conceituadas são mais reconhecidos do que os cientistas igualmente produtivos, de outras universidades. De acordo com Alvarado (2007), a lei de Lotka também é conhecida como modelo estatístico do quadrado inverso, em que o número de autores que fazem (n) contribuições em um determinado campo científico é aproximadamente $(1/n^2)$ daqueles que fazem apenas uma. Em outras palavras, o número de autores que escreveram três artigos corresponde a 1/9 do número de autores que escreveram apenas um.

Já a Lei de Bradford (1934), conhecida como Lei da Dispersão, mede a produtividade científica de periódicos,

permitindo estimar o grau de relevância de periódicos em uma determinada área do conhecimento. Assim, os periódicos que produzem o maior número de artigos estabelecem o núcleo e as áreas de dispersão de maior qualidade e relevância sobre um determinado assunto em um conjunto de periódicos. Caso os periódicos científicos sejam ordenados em ordem decrescente de produtividade sobre um determinado assunto, estes podem ser agrupados em um núcleo de periódicos específicos da temática em questão. Portanto, os periódicos devem ser listados com o número de artigos de cada um, em ordem decrescente. O total de artigos deve ser somado e dividido por dois, o grupo que tiver o maior número de artigos, até o total de metade dos artigos, é considerado o periódico com maior visibilidade sobre aquele assunto (ARAÚJO, 2006).

Por último, considera-se a Lei do Mínimo Esforço, a Lei de Zipf (1949), que consiste em medir a distribuição e frequência do aparecimento de palavras em um texto científico. A concentração de termos ou palavras-chave podem poucas ou várias vezes em um determinado texto. Dessa forma, a palavra de maior frequência tem ordem 1 e a de menor frequência tem ordem 2 e assim sucessivamente, pois gera-se uma lista ordenada de palavras em ordem decrescente de frequência de um determinado assunto. Assim, a posição de uma palavra na lista multiplicada pela sua frequência é igual a uma constante, em que a equação é: $r \times f = k$, em que r é a posição da palavra, f é a sua frequência e k é a constante. A partir daí formulou-se a lei do menor esforço, ou seja, se existe uma economia do uso de palavras, e se usar o mínimo significa que elas não vão se dispersar, pelo contrário, uma mesma palavra será usada muitas vezes. As palavras mais usadas irão indicar o assunto do documento, por outro lado, se os autores dos documentos usarem palavras

diferentes, a lei não irá servir (ARAÚJO *et al.*, 2010).

Dessa maneira, a Bibliometria exerce o papel de informar sobre o número de autores, trabalhos, periódicos, países que existem em cada categoria de produtividade, entre outros indicadores. Inicialmente, a bibliometria era medida sobre a quantidade de livros, edições, exemplares, palavras contidas nos livros, espaço ocupado pelos livros nas bibliotecas, estatísticas relativas às publicações do livro.

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, estudo de citações, a análise da produção científica e a busca de benefícios práticos imediatos para bibliotecas tais como desenvolvimento de coleções e gestão dos serviços prestados para a biblioteca, aumento do controle bibliográfico, possibilitando conhecer o tamanho e as características dos acervos, elaborar previsões de crescimento, entre outros.

2.2 Linguagens de indexação

Conhecida como 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 (análise conceitual ou de assuntos) 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).

Já para Fujita *et al.* (2009), as linguagens de indexação representam o conteúdo do documento de forma consistente, permitindo a tradução entre o vocabulário do indexador e a do usuário. Além disso, proporcionam meios através

dos quais o usuário modela sua busca e assim, obtém um maior número de documentos existente no sistema.

Nesse sentido, para que o usuário possa obter uma maior precisão em sua busca a maior parte dos sistemas adotam linguagens controladas com normas previamente estabelecidas. Na etapa inicial da indexação encontra-se a análise de assunto, em que o indexador procura compreender de maneira geral o documento para identificar e selecionar 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

- As listas de cabeçalhos de assuntos (*Library of Congress Subject Headings*) são usadas principalmente pelas bibliotecas públicas e de universidades. Exemplo: “Gestão de pessoas” e “Administração de pessoal”. Um exemplo seria ao definir o termo autorizado “Gestão do conhecimento”, o sistema não permite que o indexador utilize outro termo como, por exemplo, “Gerenciamento do conhecimento”.
- Os tesouros são voltados para documentos de uma área do conhecimento. Trata-se de uma lista de termos que representam de maneira normalizada o conteúdo dos documentos e das buscas em um determinado sistema para apoiar o usuário na indexação dos documentos. Os termos utilizados nos tesouros possuem um significado semântico que estabelece relações semânticas, tais como: relações de equivalência, como por exemplo, sinônimos; de associação como, por exemplo: Catalogação → Classificação, e de hierarquia como, por exemplo: (Animal → Cachorro).
- As linguagens simbólicas são voltadas apenas para os sistemas de classificação bibliográfica (também chamado de tabelas ou esquemas de

usuário vai procurar sobre o assunto, se a informação retornada é satisfatória para o usuário. Para isso, os documentos devem ser representados de forma que os termos da indexação sejam o mais semelhante possível aos termos utilizados pelos usuários para que a informação recuperada seja satisfatória.

Entre as principais linguagens de indexação estão as alfabéticas e simbólicas (que possuem códigos ou símbolos). Os primeiros tipos de linguagens alfabéticas são as listas de cabeçalhos de assuntos e os tesouros. No entanto, segundo Vanti *et al.* (2011), as linguagens simbólicas são representadas pelos Sistemas de Classificação Bibliográfica. Cada uma dessas linguagens de indexação é discutida abaixo:

classificação), que utilizam uma simbologia composta por letras e números para representar os assuntos dos documentos. Pode ser visto como um mapa completo de qualquer área do conhecimento, com seus conceitos e suas relações, 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 linguagens de indexação são classificadas como pré-coordenadas ou pós-coordenadas. Os índices pré-coordenados referem-se aqueles que são relacionados entre si no momento da indexação ou catalogação, como as listas de cabeçalhos de assunto e os sistemas de classificação bibliográfica (FUJITA, 2013).

Já os índices pós-coordenados buscam combinações de termos na saída, ou seja, durante a recuperação da informação. Para estes índices, temos a lista alfabética de termos autorizados, na qual constitui-se por palavras que indexará o conteúdo dos documentos e das buscas em um determinado sistema. São exemplos de linguagens pós-coordenadas: Tesouros, o Sistema Unitermo (indexação por termos

simples extraídos do conteúdo do documento).

a) Exemplos do sistema Unitermo:

- 1) Praça aparece nos documentos de número: 3 – 10 – 45 – **50** – 90– 100
- 2) Tiradentes aparece nos documentos de número: 1 – 7 – 11 – 40 – **50** – 200

O documento 50 satisfaz o assunto Praça Tiradentes.

b) Exemplos de termos de um tesouro:

- 1) UP (Usado Para) representa o termo equivalente que não é considerado como descrito;
- 2) TG representa o Termo Genérico;
- 3) TE representa o Termo Específico;
- 4) TR representa o Termo Relacionado.
- 5) NE (Nota de escopo): nota explicativa para a compreensão do termo. É utilizada somente para termos difíceis de serem identificados conceitualmente.

A estrutura semântica deste tipo de linguagem de indexação destina-se a controlar as palavras que são sinônimas para serem padronizadas. Devem-se diferenciar de homógrafos, como por exemplo, manga (camisa) é um termo bastante diferente de manga (fruta).

a) Exemplos de um Tesouro:

1) Frutas

UP Sucos
TG Produtos agrícolas
TE Frutas Cítricas

Frutas Cítricas

TG Frutas
TE Abacaxi
TE Laranja

Abacaxi

TG Frutas Cítricas

Laranja

TG Frutas Cítricas

Produtos Agrícolas

TE Frutas

2) Biblioteconomia

UP Documentação
TG Ciência da Informação
TE Bibliotecário
TR Arquivologia

Todavia, uma linguagem de indexação é formada pelo vocabulário e a sintaxe. O vocabulário se refere à relação dos descritores usados para a identificação do conteúdo de um documento e a sintaxe se refere às regras empregadas para a combinação dos descritores usados para a identificação do conteúdo de um documento.

Segundo Rubi e Fujita (2010), a lista de cabeçalhos de assuntos, em formato de um tesouro é o exemplo mais característico na catalogação de assuntos com o rigor científico e metodológico da indexação para obter a especificidade e precisão na análise e representação de assuntos quanto na recuperação da informação. Os cabeçalhos de assunto, anteriormente pré-coordenados por catalogadores não são necessários após a evolução dos catálogos manuais em catálogos on-line que permitem buscar quaisquer palavras relacionadas com outras.

De acordo com Lancaster (2004), para que se possa fazer a organização das informações de forma eficaz foram desenvolvidos programas de computador que geram automaticamente os índices a partir de uma sequência de termos. Alguns exemplos de automação de índices já realizados são os modelos de índices KWIC (*Key Word in the Context*), que utiliza títulos dos documentos, no qual se

faz a indexação de palavras, outro exemplo é o modelo KWOC (*Key Word Out of Context*), que utiliza palavras-chave entre outros. Assim, nos dias atuais, torna-se imprescindível a automação da indexação de documentos para aumentar a velocidade e precisão na recuperação de informações que sejam relevantes para o usuário.

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. Enancib são os 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 utilizam-se palavras-chave relacionadas ao tema, tais como: “Vocabulário Controlado”, “Linguagem Documentária”, “Linguagem de Indexação”, “Metadados”, “Tesauro”, “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.

Dos artigos selecionados foram levantados os dados:

- Autores e
- Coautores,
- Titulação,

- Afiliação,
- Título do artigo,
- Assunto e
- Ano do artigo.

Os dados analisados foram reunidos pelas seguintes categorias:

- Autores mais produtivos;
- Instituições;
- Instituições mais produtivas;
- Titulação dos autores mais produtivos;
- Anos mais produtivos;
- Linguagens de indexação mais utilizadas, dentre elas: Listas de cabeçalhos de assuntos, Tesouros (descritores) e Sistemas de Classificação Bibliográfica;
- Assuntos predominantes.

4 ANÁLISE DE DADOS E RESULTADOS

Os artigos dos anais de congresso do ENANCIB no período de 2012 a 2015 forma 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.

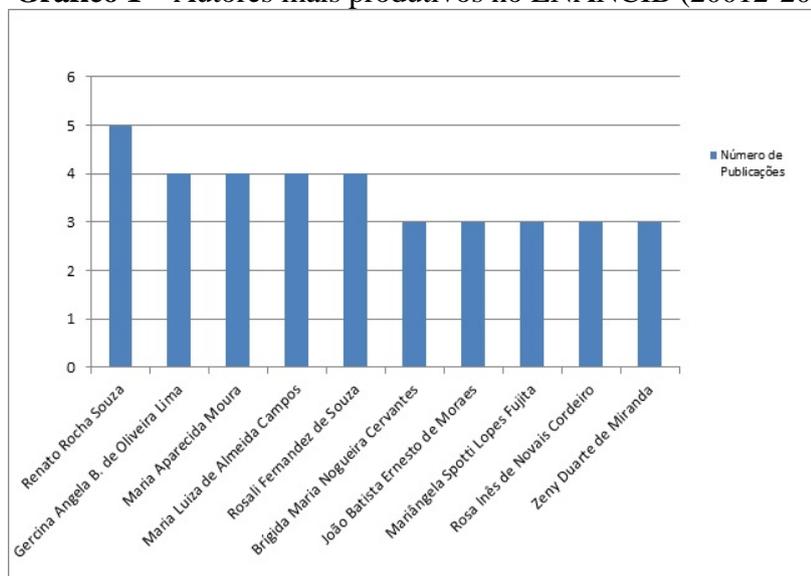
4.1 Produtividade dos autores

Ao realizar a contagem dos autores, consideraram-se também os coautores e foi atribuída 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 estudado. 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% do total de autores neste período. O gráfico 1, apresenta os autores mais produtivos que tiveram o total de 3, 4 e 5 artigos

publicados, o que corresponde à 4,9%, 3,92% e 0,98% respectivamente do total de autores. O total de publicações dos autores

mais produtivos representam 36 artigos, que correspondem a 49,31% do total das publicações.

Gráfico 1 – Autores mais produtivos no ENANCIB (20012-2015).



Fonte: Dados da pesquisa

4.2 Instituições mais produtivas

As instituições que publicaram nos anais de congresso do ENANCIB no período de 2012 a 2015 estão apresentadas na tabela

1. 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.

Tabela 1 – Instituições que publicaram no ENANCIB (20012-2015).

Instituições	Total de Produção
UFMG	18
UFF	12
UNESP	11
UNB	6
UFBA	5
UEL- Londrina	4
IBICT	4
UFSCAR	4
UFPB	4
UNIRIO	3
UFPE	3
USP	2
UFRJ	1
UIUC	1
Faculdade Estácio de Sá	1
Faculdade Pitágoras	1

FIOCRUZ	1
IBGE	1
UFSC	1
Total	89

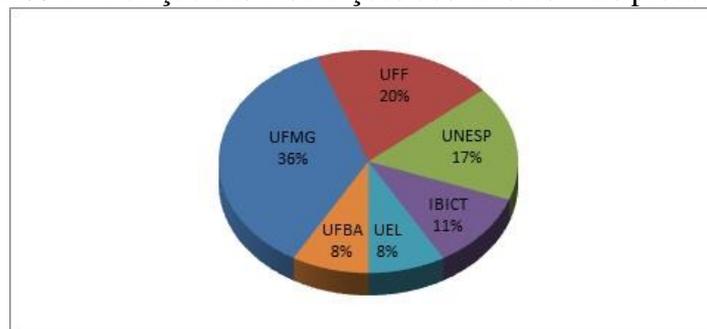
Fonte: Dados da pesquisa

4.3 Instituições dos autores mais produtivos

No gráfico 2 apresentado abaixo, foram apresentadas as Instituições vinculadas aos autores mais produtivos. Analisou-se um total de 36 publicações em que os autores que mais publicaram, tiveram a quantidade de 3, 4 ou 5 artigos publicados dentro do período e todos possuem a titulação de Doutor (D), apresentado na tabela 2. Do

total dos artigos publicados pelos autores mais produtivos, 13 artigos estão relacionados com os autores vinculados à UFMG, 7 artigos com autores vinculados a UFF, 6 artigos com autores vinculados à UNESP, 4 artigos com autores vinculados ao IBICT e 3 artigos com autores vinculados à UEL e UFBA. Outro achado importante refere-se às instituições. Entre as cinco mais produtivas, duas estão com nota seis na CAPES, a UFMG e a UNESP.

Gráfico 2 - Relação das Instituições dos autores mais produtivos.



Fonte: Dados da pesquisa

Tabela 2 - Publicações, instituição e titulação dos autores mais produtivos.

Autor	Nº publicações	Instituição	Titulação
Renato Rocha Souza	5	UFMG	D
Gercina Angela 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

A tabela 3 mostra o total de publicações por ano no período de 2012 a 2015. 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.

Tabela 3 – Total de produções no ENANCIB (2012-2015).

Ano	Total de Produção
2015	20
2014	27
2013	19
2012	7
Total	73

Fonte: Dados da pesquisa

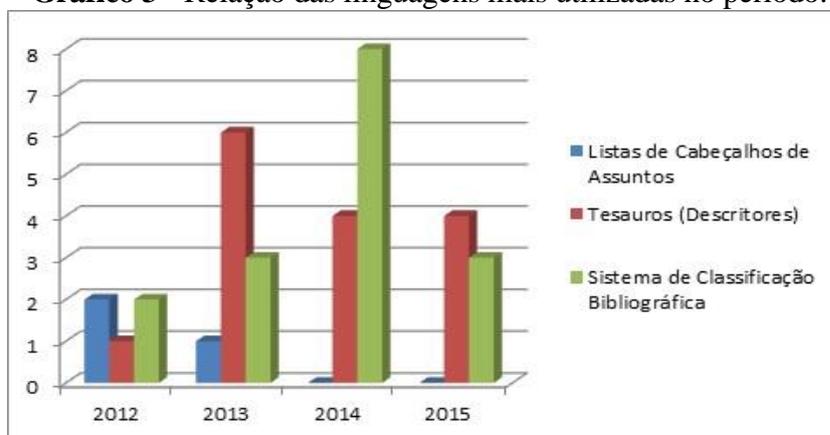
4.5 Linguagens de indexação mais utilizadas

As principais linguagens de indexação pesquisadas nos artigos foram às listas de cabeçalhos de assuntos, os tesauros (descritores) e os sistemas de classificação bibliográfica. O gráfico 3, apresenta a relação das linguagens mais utilizadas no período.

As 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 foi o ano que mais se estudou sobre esta linguagem que apareceram em 12 artigos. Em segundo lugar os Tesauros (descritores), que foi identificado 5 artigos no ano de 2013 e 4 artigos nos anos de 2014 e 2015 respectivamente. A linguagem menos estudada foi as Listas de Cabeçalhos de Assuntos, e o maior número de artigos que foi identificado foram nos anos de 2012 e 2013 que apareceram em 2 e 1 artigos por ano respectivamente.

Gráfico 3 - Relação das linguagens mais utilizadas no período.



Fonte: Dados da pesquisa

4.6 Assuntos predominantes

Os assuntos predominantes foram retirados dos títulos, palavras-chave ou resumo dos artigos. Foi realizado um agrupamento de termos do mesmo assunto e esta classificação foi feita de acordo com os assuntos descritos na tabela 4, na qual são apresentados os termos em ordem

decrecente de frequência nos artigos analisados no período. Os assuntos mais abordados foram: indexação, Organização do Conhecimento, Tesouros, Organização da Informação, Recuperação da Informação e Metadados que representam 12,57%, 11,94%, 8,17%, 7,54%, 7,38% e 6,28% respectivamente do total de 158 assuntos abordados no período.

Tabela 4 – Assuntos predominantes nos artigos publicados no ENANCIB (2012-2015).

	2012	2013	2014	2015	Total por período
Indexação	2	7	8	3	20
Organização do Conhecimento	1	6	4	8	19
Tesouros	1	5	4	3	13
Organização da Informação	3	3	3	3	12
Recuperação da Informação	2	2	6	1	11
Metadados	3	2	2	3	10
Linguagem Documentária	2	6	1	0	9
Sistemas de Classificação	0	1	7	1	9
Representação do Conhecimento	1	1	2	5	9
Tratamento temático da Informação	2	2	2	3	9
Ontologia	0	3	2	3	8
Sistemas de Recuperação da Informação	0	2	1	2	5
Vocabulário Controlado	1	0	3	1	5
Análise de Assunto	1	1	3	0	5
Classificação Facetada	0	0	4	0	4
Análise Documentária	0	0	3	0	3
Folksonomia	0	1	1	0	2
Indexação Automática	0	1	1	0	2
Taxonomias	1	1	0	0	2
Linguagem de Indexação	1	0	0	0	1
Total por ano	21	44	57	36	158

Fonte: Dados da pesquisa

5 CONSIDERAÇÕES FINAIS

Esta pesquisa permitiu conhecer o desenvolvimento dos estudos sobre 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.

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

Abstract

The objective of the research was to analyze the scientific production conveyed in the Annals of the National Research in Information Science and Librarianship - 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 languages and predominant issues. A sample of 73 papers was analyzed with a total of 102 authors (coauthors 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, IBICT, UEL, UNESP and UFBA. It was verified that the UFMG is the most participatory institution in the scientific production about this theme, presenting 36,11% 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. The predominant issues in the period were Indexing, Knowledge Organization, Thesaurus, Information Organization, Information Recovery and Metadata representing 12,65%, 12,02%, 8,22%, 7,59%, 6,96% e 6,32% respectively of the total of 158 issues addressed.

Keywords: *Bibliometrics. Indexing. Indexing Languages.*

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.

ARAÚJO, Carlos Alberto Ávila *et al.* Um retrato da Revista Escola de Biblioteconomia da UFMG. **Perspectivas em Ciência da Informação**, v.15, n.Especial, p. 134-153, nov, 2010.

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.

FUJITA, M. S. L. A representação documentária no processo de indexação com o modelo de leitura documentária para textos científicos e livros: uma abordagem cognitiva com protocolo verbal. **Ponto de Acesso**, v. 7, n.1, p.42-66, abr, 2013.

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 cienciométrica e sua perspectiva nacional e internacional. **Ciência da Informação**, v.27, n.2, p.134-140, maio/ago, 1998.

POBLACION, Dinah Aguiar; *et al.* **Comunicação & Produção Científica**. São Paulo: Editora São Paulo, 2006.

RUBI, M. P.; FUJITA, M. S. L. Política de indexação na catalogação de assunto em bibliotecas universitárias: a visão sociocognitiva da atuação profissional com protocolo verbal. **Revista Digital de Biblioteconomia e Ciência da Informação**. v.7, n.2, p.118-150, jan/jun, 2010.

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.

VANTI, Nadia *et al.* Linguagens de indexação: uso das linguagens presentes na prática da indexação. In: ENCONTRO REGIONAL DE ESTUDANTES DE BIBLIOTECONOMIA, DOCUMENTAÇÃO, CIÊNCIA DA INFORMAÇÃO E GESTÃO DA INFORMAÇÃO, 14, 2011, Maranhão. **Anais...** Maranhão: EREBD, 2011.