

GRUPO DE PESQUISA MHTX: DADOS BIBLIOMÉTRICOS DA PRODUÇÃO CIENTÍFICA

Filipi Miranda Soares¹
Benildes Coura Moreira dos Santos Maculan²
Gercina Ângela Borém de Oliveira Lima³

Resumo: Apresenta os resultados quantitativos da coleta de dados que serão utilizados no estudo bibliométrico para analisar a produção científica (teses e dissertações) do Grupo de Pesquisa denominado Modelagem Conceitual para Organização Hipertextual de Documentos - MHTX (Grupo de Pesquisa MHTX). Teve como objetivo a identificação dos autores mais citados pelos integrantes do Grupo de Pesquisa MHTX. Os seguintes dados foram coletados das referências bibliográficas, coletados utilizando planilhas do *Microsoft Office Excel*: autores; título; origem do texto; data; tipologia documental; editora ou periódico. Os resultados demonstraram que a maioria dos autores foi citada na maior parte dos documentos analisados. Isso revelou que os integrantes do grupo seguem uma mesma base teórico-metodológica, desde a tese que deu origem às pesquisas subsequentes.

Palavras-chave: Bibliometria. Grupo de Pesquisa MHTX. Produção científica. Teses e dissertações.

RESEARCH GROUP MHTX: BIBLIOMETRIC DATA OF SCIENTIFIC PRODUCTION

Abstract: It presents the quantitative results of data collection to be used in bibliometric study to analyze the scientific production (theses and dissertations) of the Research Group called “Conceptual Modeling for Hipertextual Organization of Documents: MHTX” (MHTX Research Group). It aimed to identify the authors mostly cited by members of MHTX Research Group. The data collected from references and recorded in spreadsheets using Microsoft Office Excel program were: authors; title; text origin; date; document type; publisher or journal. The results showed that the majority of authors were mentioned in most of the documents analyzed. This revealed that the members of the group follow the same theoretical-methodological base, since the thesis that gave origin to the Research Group MHTX.

Keywords: Bibliometric. MHTX Research Group. Scientific production. Theses and dissertations.

1 INTRODUÇÃO

A produção científica mundial cresceu bastante nas últimas décadas e a comunicação científica se tornou mais fácil e rápida devido às tecnologias de informação e comunicação, pois pesquisadores podem se comunicar e compartilhar informações sem que haja barreiras geográficas. Um pesquisador é formado a partir de sua inserção em contextos científicos, ao ter contato com lacunas e questões não elucidadas sobre problemas em uma área do conhecimento. Em sua busca por soluções, esse pesquisador se depara com grande diversidade de teorias e metodologias que poderão respaldar o caminho que será adotado em seu estudo. Visando a facilitar as suas escolhas, a participação do pesquisador em um grupo de pesquisa é um elemento importante para a sua formação como cientista, uma vez que permitirá a ele vivenciar

¹ Graduando em Biblioteconomia pela Escola de Ciência da Informação (ECI) da Universidade Federal de Minas Gerais (UFMG); bolsista de Iniciação Científica da FAPEMIG; Av. Fleming, 394, Bairro Ouro Preto, Belo Horizonte/MG; Tel.: 88525636. filipivgp2011@hotmail.com.

² Doutora em Ciência da Informação e Professora na ECI/UFMG; Av. Presid. Antônio Carlos, 6627, Pampulha, Belo Horizonte/MG, ECI/UFMG, Sala 4030, CEP 31270-901, Tel.: 3409-6126; benildes@gmail.com.

³ Pós-Doutora em Ciência da Informação e Professora na ECI/UFMG; Av. Presid. Antônio Carlos, 6627, Pampulha, Belo Horizonte/MG, ECI/UFMG, Sala 4030, CEP 31270-901, Tel.: 3409-6126; limagercina@gmail.com.

experiências, dificuldades e ricas discussões sobre os temas que envolvem a área. Essa participação favorece ao pesquisador, pois há grande interação de saberes e habilidades, e ainda auxilia a consolidação desse grupo de pesquisa, a partir da produção científica de seus integrantes.

Os grupos de pesquisas são compostos por pesquisadores que se organizam em torno de uma linha de pesquisa, de uma área específica do conhecimento, buscando desenvolver investigações científicas para a produção e publicação de conhecimento. Vickery (1972, p. 33) define a investigação científica como toda e “qualquer investigação metódica, desenvolvida para fornecer informações e conhecimentos que foram utilizados para solucionar um problema”. Portanto, as investigações dos grupos de pesquisa são responsáveis pela validação de teorias e metodologias de pesquisa em uma área do saber.

Nesse sentido, Alves (1987, p. 149) afirma que a publicação científica é suporte para o “processo de comunicação da produção científica e cultural, transforma-se em força motriz, na medida em que é recuperada e divulgada, impulsionando o desenvolvimento intelectual e realimentando o ciclo de geração de conhecimento”. Por isso, conhecer o “estado da arte” da produção científica dos grupos de pesquisa se tornou importante e relevante para a ciência. A análise dessa produção permite determinar as principais tendências desses estudos e delinear estratégias que possam fortalecer ainda mais os grupos de pesquisas e as áreas de conhecimento às quais eles estão vinculados.

Uma técnica que pode ser utilizada para analisar a produção científica dos integrantes de um grupo de pesquisa é a Bibliometria, que é uma “técnica quantitativa e estatística de medição dos índices de produção e disseminação do conhecimento científico” (ARAÚJO, 2006, p. 12). Dentro da Bibliometria, é provável que a análise de citações seja o elemento mais importante, pois a partir dessa análise é possível descobrir, segundo Araújo (2006):

Autores mais citados, autores mais produtivos, elite de pesquisa, frente de pesquisa, fator de impacto dos autores, procedência geográfica e/ou institucional dos autores mais influentes em um determinado campo de pesquisa; tipo de documento mais utilizado, idade média da literatura utilizada, obsolescência da literatura, procedência geográfica e/ou institucional da bibliografia utilizada; periódicos mais citados, “core” de periódicos que compõem um campo (ARAÚJO, 2006, p. 18-19).

Portanto, a Bibliometria permite a extração de uma gama de dados, o que possibilita inúmeras análises e interpretações em relação a autores e textos científicos. Devido à diversidade de dados, é necessário que se reduza o universo de estudos. Seria inviável, por exemplo, fazer um estudo bibliométrico exaustivo a partir de todos os documentos científicos que tratam sobre organismos geneticamente modificados, disponíveis na web. Por conseguinte, é sugerido que, ao realizar um estudo deste caráter, seja feito um recorte para compor a amostra da pesquisa. Este recorte pode estar definido dentro de um período de tempo, uma base de dados, um periódico científico, um portal institucional, uma biblioteca digital, um grupo de pesquisa, ou outro recorte qualquer estabelecido na metodologia.

Nesse sentido, este estudo apresenta os resultados da coleta de dados cujo levantamento bibliométrico foi realizado para analisar a produção científica dos integrantes do Grupo de Pesquisa Modelagem Conceitual para Organização Hipertextual de Documentos - MHTX (Grupo de Pesquisa MHTX), que é registrado no Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), desde 2004. A produção científica analisada é composta pelas teses e dissertações defendidas desde 2004, todas orientadas pela coordenadora do Grupo de Pesquisa MHTX. As teses e dissertações estão vinculadas ao Programa de Pós-Graduação em Ciência da Informação (PPGCI), da Escola de Ciência da Informação (ECI) da Universidade Federal de Minas Gerais (UFMG), dentro da linha de pesquisa Organização e Uso da Informação. Esses resultados fazem parte de um estudo mais amplo que tem como objetivo identificar as bases teórico-metodológicas que vêm sendo empregadas nos estudos dos pesquisadores do grupo, a partir da determinação dos autores mais citados nessa produção científica. Dessa forma, o estudo não pretende verificar o fator de impacto da publicação dos integrantes do grupo, mas averiguar quais autores esses pesquisadores vêm utilizando em suas pesquisas.

O Grupo de Pesquisa MHTX teve início com a defesa da tese de doutorado intitulada “Mapa Hipertextual (MHTX): um modelo para organização hipertextual de documentos” (LIMA, 2004), que teve como objetivo criar um modelo computacional para a organização de documentos acadêmicos (teses e dissertações), em contexto semântico. Esse protótipo digital consiste de mapa semântico e de um sumário expandido, ao qual foram inseridos *links* e pontos de acesso. Dessa forma, a criação de um sistema de hipertexto de um documento acadêmico (hiperdocumento) exige, já em sua origem (produção), a organização semântica do documento e a construção de um sistema de conceitos inter-relacionados.

O ponto de partida teórico da pesquisa dessa primeira tese se fundamentou na Teoria da Análise Facetada, desenvolvida por Ranganathan (1967), que orientou a modelagem da estrutura do documento acadêmico. A aplicação dessa teoria “satisfez a não linearidade do hipertexto, facilitando a criação de estrutura fluída e interativa como a do Mapa Hipertextual” (LIMA, 2007, p. 39), pois permitiu uma análise na área de assunto do documento, com a identificação de conceitos que puderam ser agrupados de maneira analítica, permitindo ao usuário ver o assunto sob diferentes pontos de vista.

Desde que foi implantado, em 2004, o protótipo MHTX serve como base para pesquisas de alunos do PPGCI-ECI-UFMG, sob a orientação de sua coordenadora. Os resultados dessas pesquisas têm como principal meta contribuir com a melhoria do Protótipo MHTX. Ademais, visam à simplificação dos processos de organização de informações, que são processos comumente complexos, visando a facilitar o acesso e a recuperação da informação contida nas versões impressas de teses e dissertações (LIMA, 2007).

Sob a coordenação da professora Dra. Gercina Lima, que orienta os alunos de pós-graduação, atualmente, o Grupo de Pesquisa MHTX já possui onze dissertações e cinco teses defendidas, tendo sete pesquisas em andamento (duas dissertações e cinco teses). Com a ampliação do número de mestrands e doutorandos, que tentam suprir as lacunas percebidas na implantação do Protótipo MHTX, as bases teóricas e metodológicas que fundamentaram o projeto piloto se expandiram, uma vez que faz interface com as áreas da Ciência da Computação, da Linguística e da Terminologia.

2 REFERENCIAL TEÓRICO

Os estudos bibliométricos tiveram início como uma abordagem quantitativa, mas, aos poucos, a partir dos dados estatísticos foi possível desenvolver análises mais complexas e qualitativas, ampliando o seu uso no meio acadêmico. Pritchard (1969) afirma que o termo bibliometria aparece pela primeira vez em 1922, como “bibliografia estatística”⁴ (p. 348), usado por E. Wyndham Hulme para definir processos da ciência e tecnologia que tinham por fim a contagem de documentos. O termo ficou esquecido por volta de vinte anos, até Gosnell⁵ retomá-lo no artigo “Obsolescência de livros em bibliotecas universitárias”⁶. Em 1962, Raisig aborda o termo em um ensaio crítico sobre estudo de citações. Posteriormente, Pritchard (1969) relata que o termo “bibliografia estatística”, até então adotado, não era satisfatório. Para ele, “o termo é confuso, não é muito descritivo e pode ser confundido com a própria estatística ou bibliografias sobre estatísticas”⁷. O autor passa a usar, então, o termo “bibliometria” (*bibliometrics*, em inglês) para denominar a aplicação da matemática e de métodos estatísticos para analisar a publicação de livros e outros meios de comunicação.

Alvarado (2007) afirma que na Ciência da Informação e na Biblioteconomia adotou-se o termo “Bibliometria”, criado por Pritchard em 1969, mas que os fundamentos empregados nesses estudos foram influenciados também por outras acepções dessa métrica. Para Alvarado (2007), bem antes de 1969 a bibliometria já era aplicada com outras denominações, tais como, Biometria (na Biologia), Sociometria

⁴ Tradução livre do termo *statistical bibliography*.

⁵ GOSNELL, C. F. Obsolescence of books in college libraries. *Coll. Res. Libs.*, v. 5, n. 2, Mar. 1944, p. 115-125.

⁶ Título original: *Obsolescence of books in college libraries*.

⁷ Traduzido pelos autores de *The term is clumsy, not very descriptive, and can be confused with statistics itself or bibliographies on statistics*.

(nas Ciências Sociais), Antropometria (na Antropologia), Psicometria (na Psicologia) e Econometria (na Economia). Ademais, o autor aponta que Otlet, em 1934, introduziu o termo “bibliometrie”, visando à criação de uma nova disciplina que ele denominou de Bibliologia:

Otlet estava interessado na construção de uma nova disciplina científica, a qual chamou de, e definiu como “uma ciência geral que compreende o conjunto sistemático dos dados relativos à produção, conservação, circulação e uso dos escritos e dos documentos de toda espécie”. Otlet não só estabeleceu as bases conceituais da Bibliologia, mas também o seu método científico: a Bibliometria. Considerava a Bibliometria como “a parte definida da Bibliologia que se ocupa da medida ou quantificação aplicada aos livros (Aritmética ou Matemática Bibliológica)” (ALVARADO, 2007, p. 186).

Percebe-se que o interesse de Otlet estava direcionado para a criação de medidas referentes ao livro e ao documento. De forma semelhante a Otlet, Ranganathan, pesquisador da área da Biblioteconomia, criou o termo “Librametry” ou “Librametrics”. De acordo com Alvarado (2007), Ranganathan usou os seguintes procedimentos:

coletou dados variados sobre o uso de livros, *browsing* nas estantes, etc. e, finalmente, em 1948, durante *The Annual Conference of the Association for Information Management* (ASLIB) em Leanington, Inglaterra, propôs o termo “Librametry”, que se traduziria como livrometria ou também como “bibliotecometria”; Ranganathan não oferece uma definição precisa do termo librametry. Em 1969, porém, no mesmo ano que Pritchard introduziu o termo Bibliometria, apresentou no Seminário Anual do Centro de Documentação para a Pesquisa e Treinamento, em Bangalore, Índia, a aplicação prática das técnicas bibliotecométricas, todas dirigidas à organização e criação dos processos e serviços da atividade bibliotecária (ALVARADO, 2007, p. 193-194).

Assim, nota-se que a bibliometria é um método métrico que tem origem em diferentes áreas do conhecimento, utilizada para a produção de indicadores bibliométricos.

Existem diversos *softwares* que auxiliam na análise bibliométrica, como o Bibexcel, Citespace, CopalRed, ReVis, entre outros. Porém, cada *software* apresenta limitações e problemas, pois alguns *softwares* bibliométricos só aceitam tipologias documentais específicas, tais como formato “txt” ou “doc”, o que impede que sejam feitos estudos com documentos em outros formatos (como o “pdf”). Por outro lado, geralmente, cada *software* é desenvolvido para ser compatível com uma determinada base de dados ou banco de dados. Eles também apresentam problemas de leitura quando os autores não seguem um padrão de normalização, ou seja, caso o autor do documento tenha redigido incorretamente algum campo de metadado na referência bibliográfica. Nesses casos, o *software* não efetuará a leitura, uma vez que é preciso que os campos estejam bem delimitados para que o programa faça uma leitura correta.

Os estudos bibliométricos sobre coautoria, que visam identificar as redes colaborativas, representam um desafio de análise, considerando-se autoria como a publicação assinada por um autor e coautoria ou autoria múltipla como um texto científico assinado por mais de um autor. Esses estudos podem ser realizados por meio de contagem absoluta de autoria, quando se atribui o peso de uma frequência para cada autor, ou pode-se utilizar a contagem relativa, ou seja, o método fracionado. Maltrás-Barba (2003) discute sobre diferentes abordagens para a análise de coautoria ou autoria múltipla: (1) completo ou absoluto: determina o peso de uma frequência para cada autor, indicando que a colaboração é maior quando o número de colaboradores também é maior; (2) fracionado ou relativo: determina diferentes pesos de frequência para cada autor, indicando que a colaboração se enfraquece à medida em que aumenta o número de colaboradores.

Na literatura sobre bibliometria encontram-se “leis” que norteiam a análise da produção científica em diferentes áreas do conhecimento que, segundo Chen, Chong e Tong (1994), as principais são: *Bradford* (produtividade de periódicos), *Lotka* (produtividades de autores) e *Zipf* (frequência de ocorrência de palavras).

2.1 Lei de Bradford

A Lei de *Bradford* surgiu a partir de observações feitas pelo bibliotecário S. C. Bradford. Durante uma análise de cerca de 1.600 referências periódicas, que eram compartilhadas pela biblioteca dos laboratórios de pesquisa de *Butterwickde* com bibliotecas externas. Os resultados da análise foram comparados com estudos sobre a dispersão de artigos em periódicos realizados na *Science Library*. O estudo foi publicado originalmente como artigo pela revista *Engineering* em 1934, sendo retomado em 1948 no livro *Documentation*, escrito pelo mesmo autor (VICKERY, 1948).

Bradford observou que se uma busca por literatura abrangente fosse feita em algum assunto cobrindo um período de tempo específico, poder-se-ia notar que a literatura estava espalhada em um padrão regular ao longo de um enorme número de fontes. Ele descobriu que quando estas fontes são organizadas em ordem decrescente de produtividade, com periódicos que produzem mais artigos no topo da lista e com os que produzem menos na parte inferior da lista, existe sempre um núcleo menor de periódicos relacionados de maneira próxima ao assunto e um núcleo ou zona maior de periódicos relacionados de maneira estreita, de tal maneira que o número de periódicos em cada zona aumenta, enquanto a produtividade diminui (ARAÚJO, 2006; CHEN; LEIMKUHNER, 1986).

2.2 A Lei de Lotka

A Lei de *Lotka* ou Lei do Quadrado Inverso foi desenvolvida por Alfred Lotka, estatístico americano, em 1926. O pesquisador estabeleceu “que o número de autores que fazem n contribuições em um determinado campo científico é aproximadamente $1/n^2$ daqueles que fazem uma só contribuição e que a proporção daqueles que fazem uma única contribuição é de mais ou menos 60%” (ALVARADO 2002, p. 14). Após observações de publicações das áreas de física e química, Lotka percebeu a existência de certa regularidade no padrão de publicação de artigos científicos. Dessa forma, Lotka mostrou uma distribuição assimétrica entre os autores, uma vez que um número pequeno de autores concentrava a maior parte das publicações, sendo o restante das publicações estava distribuído entre muitos autores.

Basicamente, a Lei de *Lotka* estipula que,

para cada 100 (cem) autores que escrevessem um artigo, 25 iriam contribuir com apenas dois ($n= 2$; $100/n^2 = 25$), cerca de onze iriam contribuir com três artigos cada ($n= 3$; $100/n^2 = 11$), cerca de seis autores iriam contribuir com quatro artigos cada ($n = 4$; $100/n^2 = 6$), de forma que o número de autores com " n " contribuições é $1/n^2$ daqueles que contribuem com apenas um artigo (OLIVEIRA, 1983, p. 126).

A partir desses princípios, Lotka identificou que a correlação estabelecida entre os autores e a sua produtividade era negativa (cerca de -2). Assim, essa Lei foi o primeiro modelo que relacionou o tamanho-frequência da distribuição de itens (publicações) por autores (TAGUE-SUTCLIFFE, 1992), que permite medir a produtividade de autores em publicações científicas.

2.3 A Lei de Zipf

A Lei de *Zipf* ou Lei do Mínimo Esforço foi criada por George Kingsley Zipf, um linguista americano. Apesar de linguista, a maior parte de seus trabalhos foi publicada em periódicos da área de psicologia, interessado que era sobre o comportamento humano no uso da linguagem, com a produção de discursos, tema que tem estreita relação com a psicologia. O autor é referência em vinte disciplinas científicas, aproximadamente, e ficou conhecido por seus métodos como fundador da linguística quantitativa (ADAMIC; HUBERMAN, 2002).

A Lei de *Zipf* ocupa-se em medir a frequência com que as palavras ocorrem em diversos textos, gerando uma lista ordenada de termos de uma determinada disciplina ou assunto (VANTI, 2002). Assim, as palavras mais citadas tendem a indicar o assunto do documento. Zipf afirmou a existência de uma “regularidade na seleção e uso de palavras e que um pequeno grupo de palavras é usado muito mais frequentemente” (ARAÚJO, 2006, p. 16-17). Seguindo a lógica da Lei,

se forem listadas as palavras que ocorrem num texto em ordem decrescente de frequência, a posição de uma palavra na lista multiplicada por sua frequência é igual a uma constante.

A equação para esse relacionamento é: $r \times f = k$, onde r é a posição da palavra, f é sua frequência e k é a constante (ARAÚJO, 2006, p. 16-17).

Assim, a Lei de Zipf permite obter uma lista ordenada de termos de um determinado campo do conhecimento ou assunto a partir da frequência da ocorrência de palavras em um conjunto de textos.

3 METODOLOGIA

O objetivo inicial deste estudo foi realizar uma coleta de dados para apoiar uma análise qualitativa da produção científica dos integrantes do Grupo de Pesquisa MHTX. Sendo assim, este estudo é parte de uma pesquisa mais ampla, composta por duas fases: a primeira fase, que ora apresentamos neste artigo, trata-se da coleta de dados, com a aplicação de uma abordagem quantitativa, e, a segunda fase, será a análise qualitativa dos dados levantados. A segunda fase, da análise qualitativa dos dados, está ainda em andamento e visa a identificar as bases teórico-metodológicas que vêm sendo adotadas nas pesquisas que foram desmembradas a partir dos estudos apresentados na tese que deu origem ao grupo, conforme já indicado na seção 1.1 deste artigo. O Grupo de Pesquisa MHTX continua tendo como objetivo solucionar as lacunas observadas na investigação inicial, e, posteriormente, as lacunas verificadas nas pesquisas subsequentes. A análise da produção científica do grupo fará um mapeamento conceitual e de autoria que retratará a evolução das pesquisas já realizadas e permitirá apontamentos que podem agregar novas ideias e orientar o desenvolvimento de pesquisas em andamento ou de futuras pesquisas.

Esta primeira fase da pesquisa é caracterizada como documental, descritiva e de abordagem quantitativa, na qual foram coletados os dados das produções científicas dos integrantes do Grupo de Pesquisa MHTX, referentes às teses e dissertações defendidas no período entre junho de 2004 (origem do Grupo) a junho de 2015.

Para o desenvolvimento deste estudo, foram analisados dezesseis documentos, sendo cinco teses e onze dissertações. Para a coleta de dados foram compiladas as referências bibliográficas de todas essas publicações, totalizando-se 1551 referências. Desse total, foram descartadas 99 referências, após aplicar os seguintes critérios de exclusão:

- a) descartar as referências dos dicionários “não-especializados”, como Aurélio, Houaiss e outros dicionários de língua e gramática;
- b) descartar referências de tesouros, glossários e enciclopédias não-especializados;
- c) descartar referências de portais, bibliotecas digitais e outros sites sem cunho científicos;
- d) descartar referências de normas técnicas e verbetes, referentes a publicações tais como dicionários de línguas (ex.: português-inglês, de sinônimos), tesouros, portais online, e de bibliotecas digitais.

Os dados das referências bibliográficas foram analisados e coletados de forma manual (intelectual), utilizando-se o programa editor de planilhas Microsoft Office Excel, de acordo com os seguintes campos: nome de autores; título da publicação; indicação se o texto é original, traduzido, ou se é uma adaptação; ano de publicação; tipologia documental; editora ou periódico que publicou o texto.

Além, desses campos mencionados anteriormente, foram adicionadas cinco colunas para possibilitar a inserção da atribuição de diferentes pesos para a citação recebida por um autor. Conforme Maltrás-Barba (2003), a análise de coautoria, para estabelecer redes de colaboração, pode ser realizada pelos métodos completo ou fracionado. Contudo, a análise da produção científica do Grupo de Pesquisa MHTX não inclui identificar as ligações relacionais entre os diferentes autores, mas conhecer aqueles autores que são mais citados pelos integrantes do grupo. Sendo assim, decidiu-se por aplicar uma metodologia híbrida, com a qual valorizam-se os trabalhos em colaboração, ao dar o peso de uma frequência para o primeiro autor, sem tirar deste a importância da responsabilidade pela publicação, fracionando pela metade o peso da frequência para os demais coautores. Dessa forma, cada autor recebeu um peso de citação de acordo com a posição ocupada, considerando-se as variáveis:

- (a) um único autor: foi atribuído o peso de frequência 1 para cada vez que uma obra é citada;

(b) coautoria: o primeiro autor: foi atribuído o peso de frequência 1 para o primeiro autor;
(c) coautoria: demais coautores: foi atribuído o peso de frequência 0,5 a cada coautor;
(d) para capítulos de livros: foi atribuído o peso de frequência 1 ao primeiro autor e o peso de frequência 0,5 a cada coautor do capítulo referenciado; não foram computadas frequências para o(s) autor(es) do livro(s);

(e) em casos nos quais apareceu a expressão “et al.”, buscou-se a autoria das obras nas bases de dados dos periódicos, editoras e na Internet;

(f) foram descartadas da contagem as referências de obras tais como dicionários de sinônimos (Aurélio, Houaiss) e outros dicionários de línguas (traduções) e gramática;

(g) foram descartadas da contagem as referências de obras tais como tesouros, assim como de glossários e enciclopédias gerais, não-especializados;

(h) foram descartadas da contagem as referências de portais, bibliotecas digitais e outros sites utilizados somente para consulta geral;

(i) foram descartadas da contagem as referências de normas técnicas e verbetes;

Portanto, para todo autor e primeiro autor citado foi atribuído um peso de frequência 1. Para os demais coautores, ou seja, ao segundo e aos subsequentes coautores, foi atribuído um peso de frequência 0,5 por cada coautoria. Ao final, foram somados os pontos de citações obtidos por cada autor listado nas referências.

Na compilação dos dados, a análise das referências evidenciou alguns problemas:

(1) a inconsistência na grafia dos nomes dos autores;

(2) a identificação dos autores ora pelo nome completo, ora pelas iniciais de todos os nomes e ora somente com o primeiro nome por extenso e o restante pelas iniciais.

Por causa da diversidade e variação dos nomes dos autores, fez-se necessário criar um arquivo para fazer uma padronização e o controle de autoridade das autorias, uniformizando, desta forma, as entradas dos nomes dos autores:

(a) padrão de entrada: o último sobrenome de cada autor, acompanhado pelas iniciais de todos os prenomes;

(b) compilação, em tabela própria, todas as variações dos nomes dos autores encontradas na coleta de dados;

(c) nomes de autores informados pela expressão “et al.”: foram registrados o nome de todos os autores na planilha;

O problema de inconsistência na grafia também foi identificado nos nomes de periódicos científicos. Considerou-se que esse problema poderia atrapalhar análises futuras, uma vez que essas incongruências poderiam impedir que essas referências fossem agrupadas, sendo tratadas como itens diferentes. Para estabelecer um padrão que permitisse a análise posterior de dados, estabeleceu-se também um controle de autoridade para os nomes de periódicos:

(a) padrão de entrada para periódico nacional: adotou-se a forma abreviada autorizada pelos próprios periódicos. Essa informação foi obtida a partir do Catálogo Coletivo Nacional (CCN), disponibilizado gratuitamente pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), no qual constam informações de diversos periódicos nacionais e internacionais;

(b) padrão de entrada para periódico internacional: o mesmo da letra (a).

O mesmo procedimento foi também adotado na identificação, nas referências, das instituições responsáveis pela publicação dos periódicos. Da mesma maneira, para padronizar essas entradas, criou-se um arquivo de controle de autoridade com as formas preferidas dos nomes dos responsáveis institucionais:

(a) padrão de entrada para instituições nacionais: a sigla do responsável, geralmente composta pelo nome da instituição. Por exemplo: a Editora da Universidade Federal da Paraíba teve como forma autorizada a sigla EDUFBA;

(b) padrão de entrada para instituições internacionais: decidiu-se por não usar abreviaturas, mas utilizar o nome por extenso. Essa decisão se justifica porque os responsáveis internacionais muitas vezes não adotam formas abreviadas dos nomes ou porque não são conhecidos por suas abreviaturas.

4 RESULTADOS

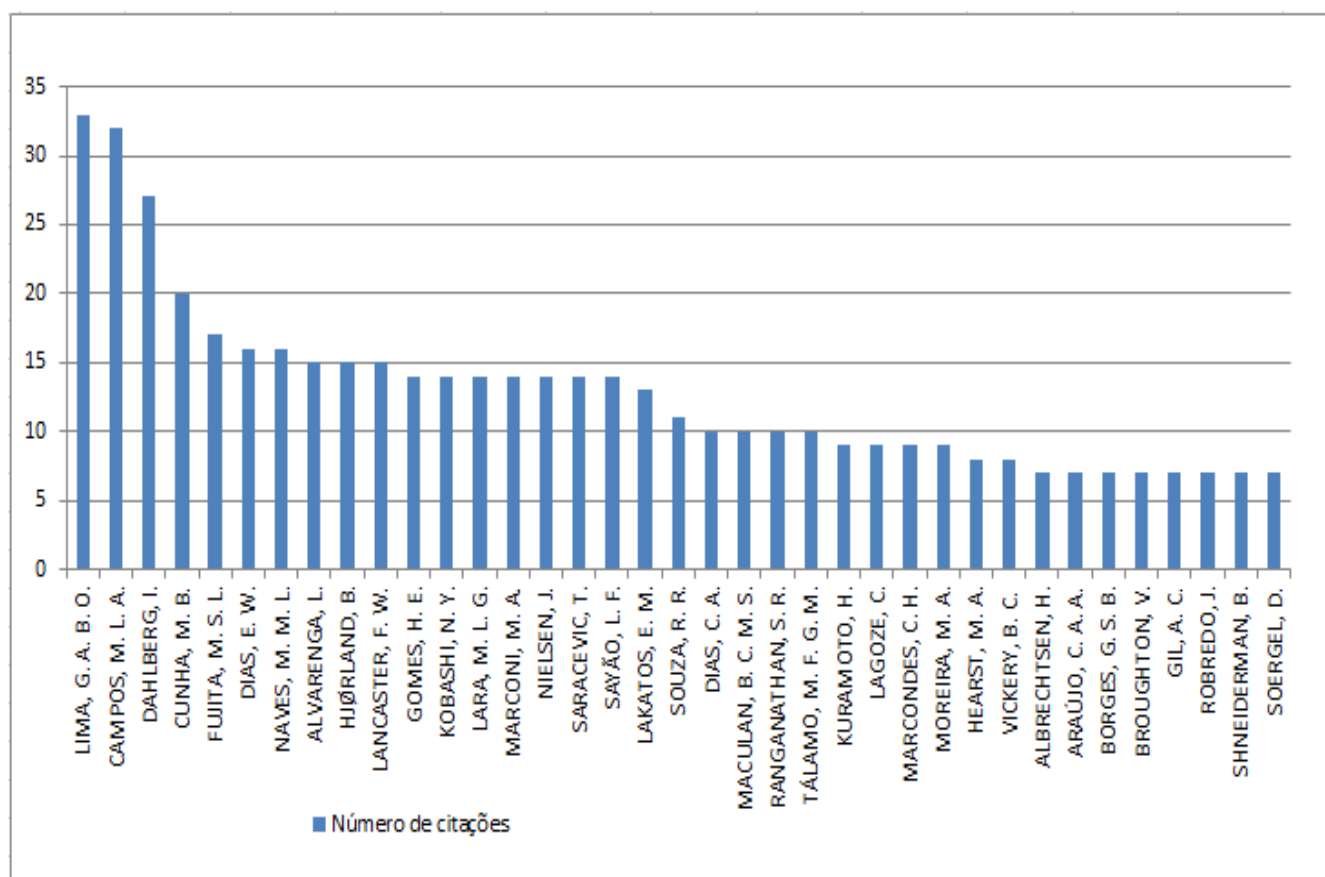
Nesta seção apresentam-se os resultados da coleta de dados, mostrando os dados quantitativos obtidos nesse levantamento: (1) autores mais citados; (2) documentos mais citados; e (3) datas das publicações dos documentos citados no universo pesquisado.

4.1 Autores mais citados

Para verificar quais os autores foram mais citados foram utilizadas duas formas de abordagem:

(1) pela contagem absoluta de valores: cada vez que um autor foi citado nas referências dos documentos analisados recebeu um peso de citação 1, conforme apresentado no Gráfico 1.

Gráfico 1 – Número de citações por autor, em valores absolutos



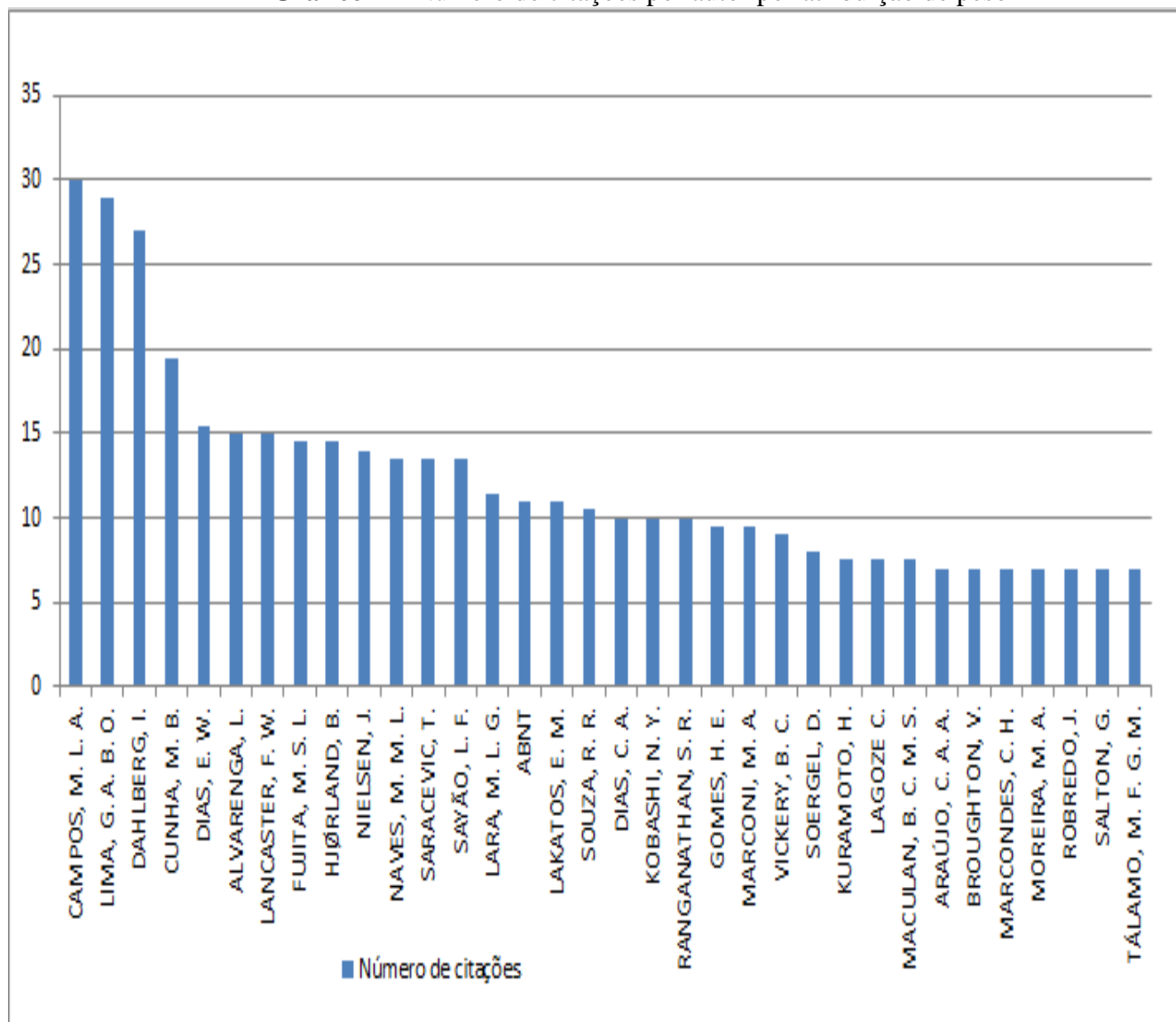
Fonte: resultado da pesquisa, 2015.

Pela frequência absoluta, nota-se que o autor mais citado é LIMA. Isso se justifica porque o trabalho desta autora foi o que deu origem ao grupo e, assim, às pesquisas subsequentes. Observam-se, também, as ocorrências de autores tais como CAMPOS e DAHLBERG, pesquisadores cujos trabalhos servem como base teórico-metodológica para o desenvolvimento das investigações do Grupo de Pesquisa MHTX. Destaca-se que esses dados se referem à primeira fase do estudo, e que, numa segunda fase, esses

dados serão ainda analisados qualitativamente, quando pretendemos revelar as bases teórico-metodológicas do grupo.

(2) pela contagem relativa de valores: foram consideradas diferentes atribuições de peso para cada autor: foi atribuído o peso 1 para o autor citado em primeiro lugar na referência, e o peso 0,5 para os demais coautores.

Gráfico 2 – Número de citações por autor por atribuição de peso



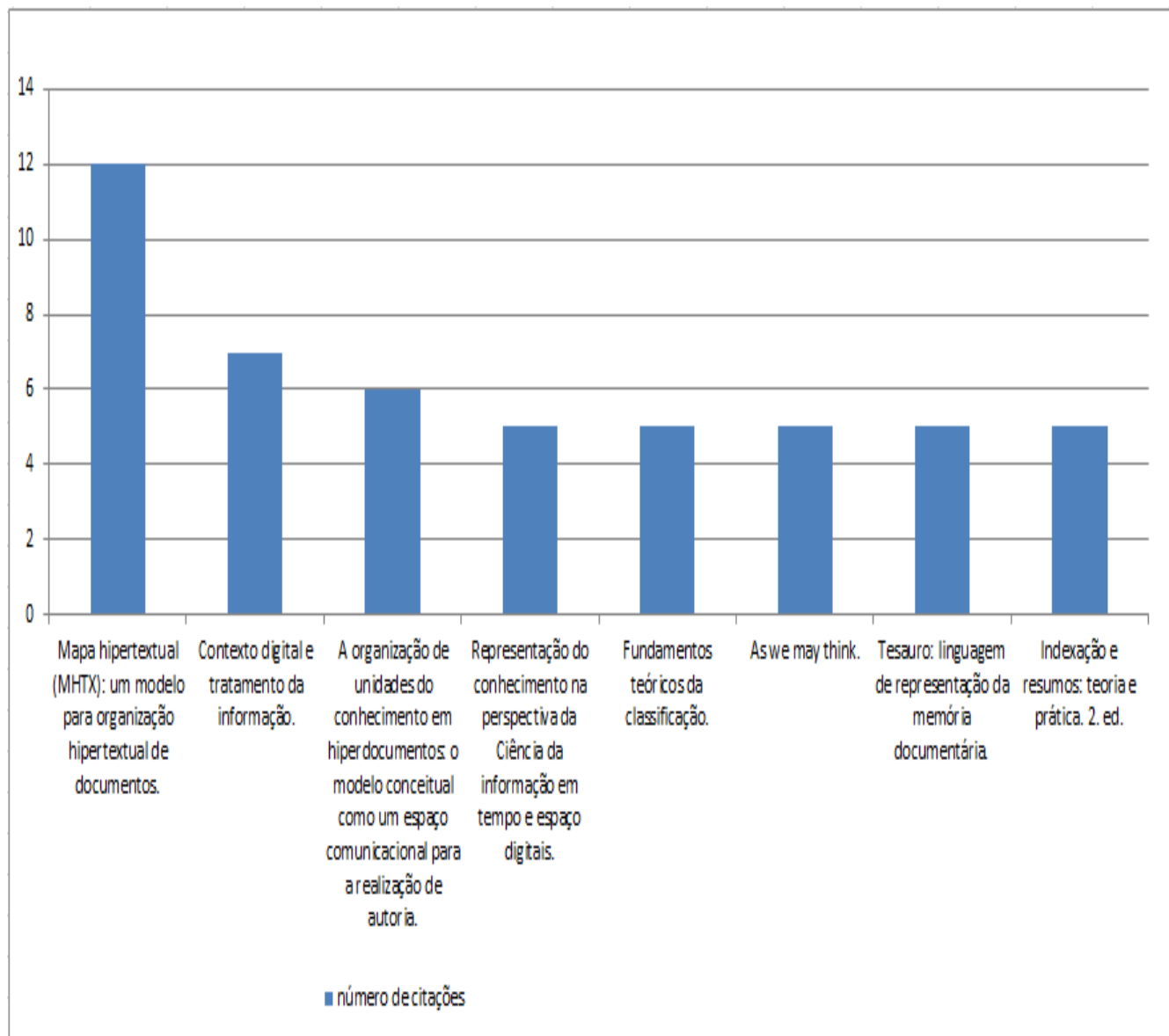
Fonte: resultado da pesquisa, 2015.

Observando e comparando os dois gráficos (1 e 2), percebemos algumas mudanças nos resultados, quando identificamos os autores que mais contribuíram nos documentos utilizados. Nota-se que LIMA agora ocupa o segundo lugar, e, CAMPOS, ocupa o primeiro lugar, confirmando que LIMA citou o trabalho de CAMPOS em sua tese. O confronto com os dois gráficos permitirá análises qualitativas que poderão indicar quais autores que mais aparecem em primeiro lugar nas referências estão realmente ligados às bases teórico-metodológicas empregadas nas pesquisas.

4.2 Documentos mais citados

Para verificar os documentos mais citados nas teses e dissertações defendidas pelos membros do Grupo de Pesquisa MHTX foi feita uma contagem absoluta das referências bibliográficas mais citadas.

Gráfico 3 – Trabalhos mais citados



Fonte: resultado da pesquisa, 2015.

Após a coleta de dados, foram obtidos oito documentos mais citados, tendo em primeiro lugar a tese da coordenadora do Grupo de Pesquisa MHTX, professora Dra. Gercina Lima. O documento “Contexto digital e tratamento da informação”⁸, cujo autor é Eduardo Wense Dias, pesquisador que trabalha com o tema de indexação, foi o segundo mais citado. Esse artigo introduz uma primeira reflexão sobre as semelhanças entre a indexação de documentos impressos e digitais, que orientou as pesquisas do grupo. Em seguida, o documento intitulado “A organização de unidades do conhecimento em

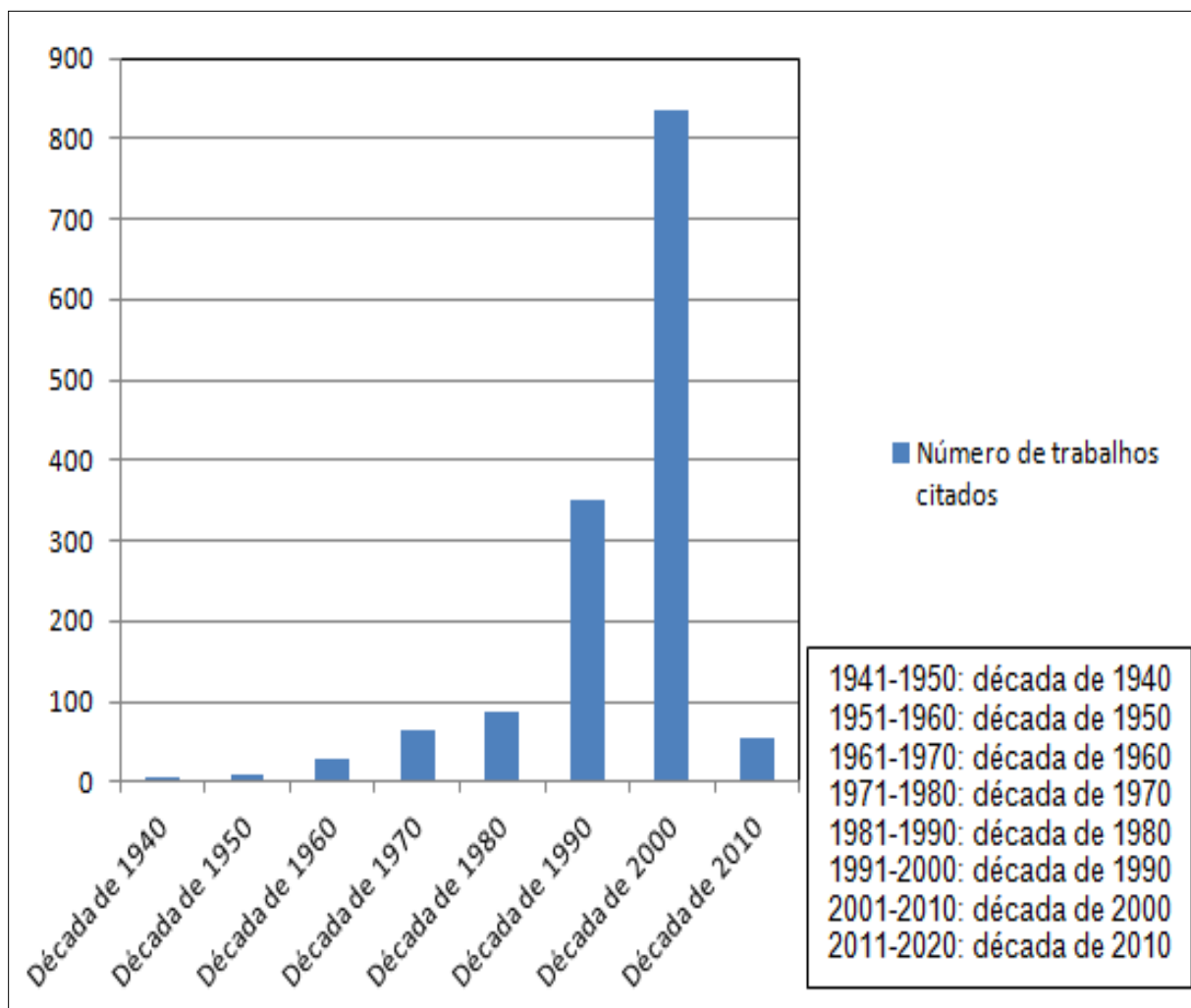
⁸ Dias, E. W. Contexto Digital e Tratamento da Informação. **DataGramaZero**, v. 2, n. 5, out. 2001. Disponível em: <http://www.dgz.org.br/out01/Art_01.htm>. Acesso em: 27 ago. 2015.

hiperdocumentos: o modelo conceitual como um espaço comunicacional para a realização de autoria”⁹, publicado por Maria Luiza de Almeida Campos, Rosali Fernandez de Souza e Maria Luiza Machado Campos. Esse artigo é resultado da tese de doutorado da primeira autora, Maria Luiza de Almeida Campos (2001), denominada “A Organização de Unidades do Conhecimento em Hiperdocumentos: o modelo conceitual como um espaço comunicacional para realização da autoria”. Essa foi uma importante referência metodológica para o desenvolvimento da tese que deu origem ao Grupo de Pesquisa MHTX, defendida por Lima (2004), sendo também ponto de apoio para as pesquisas subsequentes.

4.3 Temporalidade

Para coletar a data de publicação dos documentos referenciados nas teses e dissertações defendidas pelos integrantes do Grupo de Pesquisa MHTX, foi utilizado o campo denominado “data de publicação” da referência bibliográfica.

Gráfico 4 – Intervalos de tempo



Fonte: resultado da pesquisa, 2015.

⁹ CAMPOSI, M. L. A.; SOUZAI, R. F.; CAMPOS, M. L. M. A organização de unidades do conhecimento em hiperdocumentos: o modelo conceitual como um espaço comunicacional para a realização de autoria. **Ciência da Informação**, v. 32, n. 2, p. 7-16, maio/ago. 2003. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/111/92>>. Acesso em: 27 ago. 2015.

Observa-se que, no Gráfico 4, os períodos de tempo, divididos em décadas, representam as datas de publicação dos documentos referenciados nas bibliografias das teses e dissertações analisadas. Percebe-se que a maior parte das referências utilizadas nos documentos se insere, majoritariamente, na década de 2000, totalizando mais de 800 trabalhos citados. Percentualmente, isso significa que 58,20% do total dos documentos analisados, sendo justificado pela data da defesa da tese de LIMA e de referências utilizadas, dessa década, que tratavam do tema da pesquisa. Na década de 1990 há, também, um número significativo de publicações, correspondendo a 24,5% dos documentos citados nas referências bibliográficas. Pensamos que isso se justifica porque foi nessa época que houve a popularização do computador pessoal e da Internet, o que deu origem aos estudos sobre bibliotecas digitais e a representação da informação em ambiente *web*. Dessa forma, era esperado que os estudos dessa época fossem bastante citados, uma vez que o tema mais pesquisa pelo grupo, no início de sua existência, tinha como foco as bibliotecas digitais. Ressaltamos, porém, que esses dados quantitativos representam a primeira fase de nossos estudos e que irão respaldar as análises qualitativas, da segunda fase, que almejam revelar as características das pesquisas do Grupo de Pesquisa MHTX.

5 CONSIDERAÇÕES FINAIS

Os métodos estatísticos da bibliometria, cujas principais Leis são as Lei de *Lotka*, em 1926, Lei de *Bradford*, em 1934, e Lei de *Zipf*, em 1936, são importantes para a coleta de dados quantitativos para análises qualitativas da produção científica de um grupo de pesquisadores. Nesse contexto, as análises de citação podem dar subsídios para investigar a realidade dos estudos de um grupo de pesquisadores, dando subsídios para verificar possíveis lacunas para novos estudos. Nos estudos bibliométricos de citação, a atribuição de valores para a determinação da frequência em publicações com coautoria pode ser realizada por valores absolutos (valor de uma frequência para cada autor) ou por valores relativos, pelo método do fracionamento. Neste trabalho, adotamos uma abordagem híbrida, uma vez que não havia a intenção de verificar o fator de impacto de um autor e nem mesmo o grau de colaboração de coautorias, e, sim, os autores mais citados.

Neste artigo, apresentamos apenas a primeira fase, com os dados quantitativos que irão respaldar as análises qualitativas já em andamento, que visa a identificar as bases teórico-metodológicas que estão sendo empregadas pelos integrantes do Grupo de Pesquisa MHTX. Para tanto, a coleta de dados foi realizada a partir da lista de referências das teses e dissertações defendidas pelo grupo desde a sua origem, em 2004. Apesar de seus limites, pois mostrou apenas a fase quantitativa do estudo, este artigo já aponta indícios de que, prioritariamente, o grupo utiliza os aportes de Ranganathan (Teoria da Análise Facetada) e de Dahlberg (Teoria do Conceito e a área da Organização do Conhecimento). Verificou-se, assim, que a base teórica inicial do grupo, que foi com a Teoria da Análise Facetada, continua como ponto norteador da maior parte das pesquisas do grupo.

Agradecimentos

Agradecimento pelo auxílio financeiro da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERÊNCIAS

- ADAMIC, L. A.; HUBERMAN, B. A. Zipf's law and the Internet. **Glottometrics**, v. 3, n. 1, p. 143-150, 2002. Disponível em: <<http://www.ram-verlag.eu/journals-e-journals/glottometrics/>>. Acesso em: 28 ago. 2015.
- ALVARADO, R. U. A bibliometria: história, legitimação e estrutura. In: TOUTAIN, L. M. B. B. (Org.). **Para entender a ciência da informação**. Salvador: UFBA, 2007. p. 185-217. (Saladeaula; 5) .
- ALVARADO, R. U. A Lei de Lotka na bibliometria brasileira. **Ciência da Informação**, Brasília, DF, v. 31, n. 2, p. 14-20, maio/ago. 2002. Disponível em: <<http://www.scielo.br/pdf/%0D/ci/v31n2/12904.pdf>>. Acesso em: 28 ago. 2015.
- ALVES, M. A. M. A Biblioteca Nacional, banco de dados da produção científica e cultural brasileira. SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 5., Porto Alegre, 1987. **Anais...** Porto Alegre: UFRGS, 1987 v. 1 p. 149-166.
- ARAÚJO, C. A. A. Bibliometria: evolução histórica e questões atuais. **Em Questão**, Porto Alegre, v. 12, n. 1, p. 11-32, jan./jun. 2006. Disponível em: <<http://revistas.univerciencia.org/index.php/revistaemquestao/article/view/3707/3495>>. Acesso em: 27 ago. 2015.
- CHEN, Y.; CHONG, P. P.; TONG, M. Y. The Simon-Yule Approach to Bibliometric Modeling. **Information Processing & Management**, v. 30, n. 4, p. 535-56, 1994. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0306457394900388>>. Acesso em: 17 ago. 2015.
- CHEN, Y.-S.; LELMKUHLER, F. F. A Relationship Between Lotka's Law, Bradford's Law, and Zipf's Law. **Journal of the American Society for Information Science**, v. 37, n. 5, p. 307-314, 1986. Disponível em: <[onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-4571\(198609\)37:5%3C307::AID-ASI5%3E3.0.CO;2-8/pdf](http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(198609)37:5%3C307::AID-ASI5%3E3.0.CO;2-8/pdf)>. Acesso em: 28 ago. 2015.
- FERREIRA, A. G. C. Bibliometria na avaliação de periódicos científicos. **DataGramaZero**, v. 11, n. 3, jun. 2010. Disponível em: <http://www.dgz.org.br/jun10/Art_05.htm>. Acesso em: 27 ago. 2015.
- LIMA, G. A. B. O. **Mapa hipertextual (MHTX): um modelo para a organização hipertextual de documentos**. 2004. 199 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004. Disponível em: <<http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/LHLS-6BUPG9>>. Acesso em: 22 ago. 2015.
- MALTRÁS-BARBA, B. **Los indicadores bibliométricos: fundamentos y aplicación al análisis de la ciencia**. Gijon: Trea, 2003.
- OLIVEIRA, S. M. Aplicação da lei de produtividade de autores de Lotka à literatura de Jaca. **Revista de Biblioteconomia**, Brasília, v. 11, n. 1, p. 125-130, jan./jun. 1983. Disponível em: <http://www.brapci.inf.br/repositorio/2011/06/pdf_71c1d2ca3b_0017066.pdf>. Acesso em: 18 ago. 2015.
- PINHEIRO, L. V. R. Lei de Bradford: uma reformulação conceitual. **Ciência da Informação**, Brasília, DF, v. 12, n. 2, p. 59-80, jul./dez. 1983.

PRITCHARD, O. V. Groos Alan. Documentation notes. **Journal of Documentation**, v. 25, n. 4, p. 344-349, 1969. Disponível em: <<http://www.emeraldinsight.com/doi/pdfplus/10.1108/eb026482>>. Acesso em: 27 ago 2015.

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**, Brasília, DF, v. 31, n. 2, p. 152-162, maio/ago. 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n2/12918>>. Acesso em: 28 ago. 2015.

VICKERY, B. C. Bradford's Law of Scattering. **Journal of Documentation**, v. 4, n. 3, p. 198-203, 1948. Disponível em: <<http://www.emeraldinsight.com/doi/abs/10.1108/eb026133>>. Acesso em: 28 ago. 2015.

VICKERY, B. C. The administration of research in institution. In: THE ADVISORY BOARD ON RESEARCH OF THE LIBRARY ASSOCIATION, London. Objectives and Administration of Library research. **Anais...** London: The Library Association, 1972, p. 33-38.

WITTER, G. P. Pós-graduação e produção científica: a questão de autoria. **TransInformação**, v. 1, n. 1, p. 29-37, 1989. Disponível em: <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/1698>>. Acesso em: 20 ago. 2015.