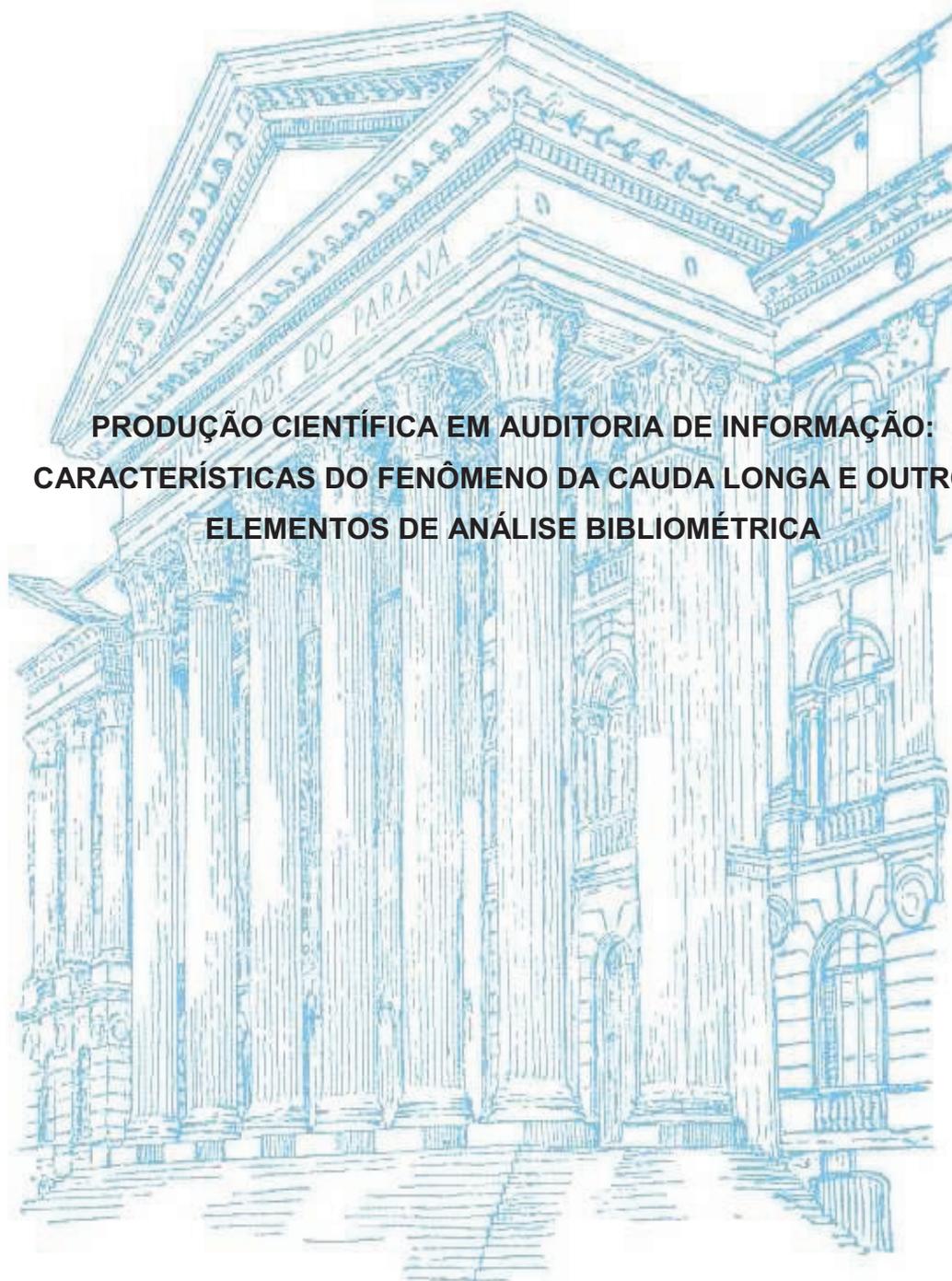


UNIVERSIDADE FEDERAL DO PARANÁ

MARCELO MAIA



**PRODUÇÃO CIENTÍFICA EM AUDITORIA DE INFORMAÇÃO:
CARACTERÍSTICAS DO FENÔMENO DA CAUDA LONGA E OUTROS
ELEMENTOS DE ANÁLISE BIBLIOMÉTRICA**

CURITIBA

2016

MARCELO MAIA

**PRODUÇÃO CIENTÍFICA EM AUDITORIA DE INFORMAÇÃO:
CARACTERÍSTICAS DO FENÔMENO DA CAUDA LONGA E OUTROS
ELEMENTOS DE ANÁLISE BIBLIOMÉTRICA**

Dissertação apresentada ao Curso de Pós Graduação em Ciência, Gestão e Tecnologia da Informação, Linha Gestão da Informação e do Conhecimento, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como parte das exigências para obtenção do título de Mestre em Ciência, Gestão e Tecnologia da Informação.

Orientadora: Prof^a. Dr^a. Patrícia Zeni Marchiori

CURITIBA

2016

UNIVERSIDADE FEDERAL DO PARANÁ. SISTEMA DE BIBLIOTECAS.
CATALOGAÇÃO NA FONTE

Maia, Marcelo

Produção científica em auditoria de informação: características do fenômeno de cauda longa e outros elementos de análise bibliométrica / Marcelo Maia. - 2016.

105 f.

Orientadora: Patrícia Zeni Marchiori.

Dissertação (Mestrado) – Universidade Federal do Paraná.

Programa de Pós- Graduação em Ciência, Gestão e Tecnologia da Informação, do Setor de Ciências Sociais Aplicadas.

Defesa: Curitiba, 2016

1. Bibliometria. 2. Auditoria de informação – Produção científica. 3. Fenômeno da cauda longa. I. Marchiori, Patricia Zeni, 1964- II. Universidade Federal do Paraná. Setor de Ciências Sociais Aplicadas. Programa de Pós- Graduação em Ciência. Gestão e Tecnologia da Informação. III. Título.

CDD 001.422



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
Setor CIÊNCIAS SOCIAIS APLICADAS
Programa de Pós Graduação em CIÊNCIA, GESTÃO E TECNOLOGIA DA INFORMAÇÃO
Código CAPES: 40001016058P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIA, GESTÃO E TECNOLOGIA DA INFORMAÇÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **MARCELO MAIA**, intitulada: "**Produção científica em auditoria de informação: características do fenômeno da cauda longa e outros elementos de análise bibliométrica.**", após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO.

CURITIBA, 16 de Dezembro de 2016.

PATRICIA ZENI MARCHIORI
Presidente da Banca Examinadora (UFPR)

CICERO APARECIDO BEZERRA
Avaliador Interno (UFPR)

LEILAH SANTIAGO BUFREM
Avaliador Interno (UFPR)

AGRADECIMENTOS

À minha família pelo apoio no desenvolvimento desta pesquisa, em especial, à minha namorada Fernanda, minha mãe Dalva, meu irmão Marcos e ao meu pai Teodoro.

Aos meus amigos do mestrado em nossas inúmeras conversas e discussões sobre temas correlatos. Aos amigos da Companhia Paranaense de Gás, em especial ao gerente contábil (Marcelo), gerente da assessoria de governança corporativa e gestão de risco (Paulo), pelo apoio incondicional desde o ingresso no programa.

Ao Programa de Pós Graduação em Ciência, Gestão e Tecnologia da Informação da Universidade Federal do Paraná, em especial à minha orientadora, Professora Doutora Patrícia Zeni Marchiori, por ter contribuído imensamente com seu conhecimento para a pesquisa.

Aos membros da banca que tanto colaboraram com questionamentos e críticas construtivas para a qualidade deste trabalho.

RESUMO

Considerando-se a emergente área da auditoria da informação, este trabalho explora as características das manifestações de “cauda longa”, derivadas de distribuições bibliométricas no domínio de auditoria de informação. Coletou-se material disponível na EBSCOhost, Google Scholar, ISI/Web of Science^(C), SciELO, e SCOPUS, compondo-se um *corpus* inicial de estudo de 346 itens, publicados entre 1970 e 1º de setembro de 2016. O *corpus* foi submetido à fórmula estatística de amostragem de população finita, tendo sido recuperados, em texto completo (incluindo as referências), 173 itens, os quais foram inseridos e processados no MS-Excel^(T) e na ferramenta Stikis para a geração de relatórios de evolução quantitativa desta produção, tais como: local de publicação dos itens; veículos de publicação; e, o idioma de maior predomínio. Adicionalmente, os itens foram submetidos às fórmulas e pressupostos teóricos da bibliometria, identificando-se e explorando-se elementos de destaque em algumas das dimensões reconhecidas como sendo “de cauda longa”, nas distribuições obtidas. Os resultados indicaram, para o universo de publicações, que há concentração em poucos autores, evidenciando-se a lei de Lotka. Com a aplicação da Lei de Zipf e, mais especificamente, o Ponto de Transição de Goffman, a palavra-chave que poderia carregar um componente semântico significativo se encontra na cauda longa (zona de ruído). O veículo de publicação com maior frequência - *International Journal of Information Management* - não é o que apresenta maior nível de impacto, fato que corrobora um dos pressupostos da investigação. Aplicando-se os cálculos relativos à correlação de Pearson, Cosseno de Salton, e índice de Jaccard para 117 artigos e 39 comunicações em eventos, a análise de cocitação de autores associados a tais índices (absolutos e relativos) demonstra um domínio científico concentrado em poucos pesquisadores. Futuras pesquisas podem replicar o estudo em outras bases de dados e possibilitar a comparação de resultados.

Palavras-chave: Auditoria de informação. Análise de domínio. Fenômeno da cauda longa. Bibliometria.

ABSTRACT

Considering the emerging area of auditing information, the objective is to explore the characteristics of 'long tail' manifestations derived from distributions bibliometric in the field of audit information. Material was collected from EBSCOhost, Google Scholar, ISI / Web of Science (C), SciELO, and SCOPUS, the initial corpus of study of 346 items was published between 1970 and September 1, 2016. The corpus was submitted to the statistical formula of finite population sampling, and 173 items were retrieved, (including references), which were inserted and processed in MS-Excel (T) and in the Stikis tool for the generation of the quantitative evolution of production such as: where the items were published; which vehicles to publish; and the most predominant language. In addition, the items are submit to the theoretical formulas and assumptions being identified and explored elements of prominence in some of the dimensions identified as "long tail" in the obtained distributions. The results indicates, for the universe of publications, that there is concentration in few authors, evidencing the law of Lotka. With the applying Zipf's Law and, more specifically, Goffman's Transition Point, the keyword that could carry a significant semantic component lies in the long tail (noise zone). The most frequent vehicle – International Journal of Information Management – is not the one that presents greater level of impact, fact that corroborates one of the presuppositions of the investigation. Applying the calculations related to Pearson correlation, Salton's cosine, and Jaccard index for 117 articles and 39 communications the analysis of cocitation of authors associated to such (absolute and relative) indexes demonstrates of a scientific domain concentrated in few researchers. Future research can replicate the study in other databases and enable comparison of results.

Keywords: Information Audit. Domain analysis. Long Tail phenomenon.
Bibliometrics.

LISTA DE FIGURAS

FIGURA 1 – DIAGRAMA REPRESENTANDO GRAFICAMENTE A INTER- RELAÇÃO DE MÉTRICAS DE COMUNICAÇÃO	30
FIGURA 2 – COCORRÊNCIA DE CITAÇÃO DE DOIS AUTORES	38
FIGURA 3 – CAUDA LONGA UM NOVO MERCADO (NICHOS DE MERCADO)	44
FIGURA 4 – TRÊS FORÇAS PODEROSAS DA CAUDA LONGA.....	46
FIGURA 5 – FLUXO DE PROCESSAMENTO DE DADOS NO PROGRAMA STIKIS	59

LISTA DE GRÁFICOS

GRÁFICO 1 – DISTRIBUIÇÃO DE FREQUÊNCIA E LEI DE ZIPF	34
GRÁFICO 2 – AS TRÊS ZONAS DE DISTRIBUIÇÃO NA CURVA DE ZIPF	35
GRÁFICO 3 – EVOLUÇÃO DA PRODUÇÃO CIENTÍFICA NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO POR TIPO DE PUBLICAÇÃO – 1970 - 2016/SETEMBRO	62
GRÁFICO 4 – CAUDA LONGA DE AUTORES <i>VERSUS</i> ARTIGOS PRODUZIDOS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	67
GRÁFICO 5 – AS TRÊS ZONAS DE DISTRIBUIÇÃO NA CURVA DE ZIPF NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	73
GRÁFICO 6 – APLICAÇÃO DA PRIMEIRA LEI DE ZIPF NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	74
GRÁFICO 7 – EVOLUÇÃO DA PRODUÇÃO CIENTÍFICA NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	76
GRÁFICO 8 – GRÁFICO DE ANÁLISE DE REDES SOCIAIS: AUTORES E COAUTORES, NA MATRIZ DO COSSENO DE SALTON NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	82

LISTA DE QUADROS

QUADRO 1 – CRONOLOGIA DOS PRINCIPAIS PESQUISADORES DE AUDITORIA DE INFORMAÇÃO	20
QUADRO 2 – PRINCIPAIS PESQUISADORES EM AUDITORIA DA INFORMAÇÃO POR PERÍODO E OBJETIVO ABORDADO NA INVESTIGAÇÃO DE CADA METODOLOGIA	21
QUADRO 3 – TIPOLOGIA DE ESTUDOS MÉTRICOS DA INFORMAÇÃO, CONFORME MACIAS-CHAPULA 1998	29
QUADRO 4 – TIPOS DE INDICADORES BIBLIOMÉTRICOS.....	30
QUADRO 5 – LEIS E PRINCÍPIOS BIBLIOMÉTRICOS	31
QUADRO 6 – CARACTERÍSTICAS NOS DADOS QUE COMPÕEM O GRÁFICO SEGMENTADO NA CABEÇA E NA CAUDA	50
QUADRO 7 – PECULIARIDADES NA DEFINIÇÃO DE BUSCA NAS BASES DE DADOS SELECIONADAS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 – 2016/SETEMBRO	55
QUADRO 8 – PRODUTIVIDADE DE AUTORES PRINCIPAIS E COAUTORES (CONTAGEM AJUSTADA/FRACIONADA) NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	70
QUADRO 9 – FREQUÊNCIA DAS PALAVRAS-CHAVE NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	72
QUADRO 10 – FREQUÊNCIA DAS PALAVRAS-CHAVE PARA APLICAÇÃO DA PRIMEIRA LEI DE ZIPF NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO.....	73
QUADRO 11 – CLASSIFICAÇÃO DOS VEÍCULOS DE PUBLICAÇÃO PELA QUANTIDADE DE PUBLICAÇÕES, FATOR DE IMPACTO E ÍNDICE H - ITENS PRESENTES NA CABEÇA NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	77
QUADRO 12 – MATRIZ DE VALORES ABSOLUTOS (10 X 10) NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	80
QUADRO 13 – INDICADOR NORMATIVO - COSSENO DE SALTON - NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	81

QUADRO 14 – INDICADOR NORMATIVO – ÍNDICE DE JACCARD - NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	82
QUADRO 15 – INDICADOR NORMATIVO – CORRELAÇÃO DE PEARSON - NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	83
QUADRO 16– COEFICIENTE DE DETERMINAÇÃO R ² - NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	83

LISTA DE TABELAS

TABELA 1 – QUANTIDADE DE ITENS QUE COMPÕEM O UNIVERSO DA PESQUISA APÓS A ELIMINAÇÃO DOS DUPLICADOS NA PRÓPRIA BASE NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 – 2016/SETEMBRO	56
TABELA 2 – QUANTIDADE DE ITENS QUE COMPÕEM O UNIVERSO DA PESQUISA, POPULAÇÃO DOS ITENS COMPLETOS RECUPERADOS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 – 2016/SETEMBRO	57
TABELA 3 – PUBLICAÇÃO POR ANO E CONTINENTE EM COMUNICAÇÃO E EVENTO NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 A 2016/SETEMBRO.....	63
TABELA 4 – LOCAL DE PUBLICAÇÃO DA PRODUÇÃO CIENTÍFICA NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	64
TABELA 5 – IDIOMA DA PRODUÇÃO CIENTÍFICA NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	65
TABELA 6 – NÚMERO DE ITENS PUBLICADOS SEGUNDO O NÚMERO DE AUTORES PRINCIPAIS E COAUTORES NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	65
TABELA 7 – FREQUENCIA DE PUBLICAÇÃO DOS AUTORES PRINCIPAIS E COAUTORES, POR ANO, NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO.....	66
TABELA 8 – PRODUTIVIDADE DE AUTORES PRINCIPAIS (CONTAGEM DIRETA) NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	67
TABELA 9 – AUTORES PRINCIPAIS MAIS PRODUTIVOS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	68
TABELA 10 – PRODUTIVIDADE DE AUTORES PRINCIPAIS E COAUTORES (CONTAGEM COMPLETA) NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO.....	68

TABELA 11 – PRODUTIVIDADE DE AUTORES PRINCIPAIS E COAUTORES (CONTAGEM AJUSTADA/FRACIONADA) NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	69
TABELA 12 – FREQUÊNCIA DAS PALAVRAS-CHAVES POR ÁREA DE CONHECIMENTO ESTABELECIDA PELA CAPES NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	74
TABELA 13 – PALAVRAS-CHAVE CLASSIFICADAS POR ÁREA DE CONHECIMENTO ESTABELECIDA PELA CAPES NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	75
TABELA 14 – NÚMERO DE VEÍCULOS DE PUBLICAÇÕES, SEGUNDO A QUANTIDADE DE ITENS PUBLICADOS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	76
TABELA 15 – FREQUÊNCIA DAS PALAVRAS-CHAVE NA CAUDA LONGA DOS VEÍCULOS DE PUBLICAÇÕES NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO.....	78
TABELA 16 – PESQUISADORES DE MAIOR IMPACTO E VISIBILIDADE IGUAL E SUPERIOR A VINTE E CINCO PUBLICAÇÕES NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	79
TABELA 17 – AUTORES E VALORES EXTREMADOS DOS INDICADORES ABSOLUTOS E NORMALIZADOS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO.....	84
TABELA 18 – AUTORES MAIS PRODUTIVOS, AUTORES E COAUTORES MAIS CITADOS E AUTORES MAIS COCITADOS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO	85

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
AD	Análise de domínio
AI	Auditoria de Informação
APA	<i>American Psychological Association</i>
ARS	Análise de rede
Capes	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CDs	<i>Compact Discs</i>
CEO	<i>Chief Executive Officer</i>
C&T	Ciência e Tecnologia
CS	Cosseno de Salton
ISI	<i>Institute of Scientific Information</i>
r	Correlação de Pearson
SCI	<i>Science Citation Index</i>
SNA	<i>Social Network Anallsys</i>
UFPR	Universidade Federal do Paraná
WoS	ISI/Web of Science [©]

SUMÁRIO

1	INTRODUÇÃO	15
2	REFERENCIAL TEÓRICO.....	19
2.1	AUDITORIA DE INFORMAÇÃO: CONCEITOS, CARACTERÍSTICAS E MODELOS.....	19
2.2	ANÁLISE DE DOMÍNIO: ALGUNS ELEMENTOS E ABORDAGENS.....	23
2.3	OS ESTUDOS BIBLIOMÉTRICOS: RECORTE PARA A ANÁLISE DE DOMÍNIO	27
2.4	INDICADORES BIBLIOMÉTRICOS DE INTERESSE PARA A PESQUISA.	32
2.4.1	Lei de Lotka	32
2.4.2	Lei de Zipf.....	33
2.4.3	Ponto de Transição de Goffman	35
2.4.4	Análise de Citação.....	36
2.4.5	Análise de Cocitação	38
2.4.5.1	Correlação de Pearson.....	39
2.4.5.2	Cosseno de Salton	40
2.4.5.3	Índice de Jaccard.....	41
2.5	FENÔMENO DE CAUDA LONGA: UMA MANIFESTAÇÃO DAS LEIS DE POTÊNCIA	43
3	PROCEDIMENTOS METODOLÓGICOS.....	54
3.1	COLETA DE PUBLICAÇÕES NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO	54
3.2	TÉCNICAS BIBLIOMÉTRICAS SELECIONADAS PARA O ESTUDO	58
4	ANÁLISE DOS DADOS E RESULTADOS	62
4.1	CITAÇÃO E COCITAÇÃO	78
5	CONSIDERAÇÕES FINAIS	86
	REFERÊNCIAS.....	89
	APÊNDICE 1 – PROGRAMAS DE ANÁLISE BIBLIOMÉTRICA DE COLABORAÇÃO CIENTÍFICA E DE VISUALIZAÇÃO DE DADOS	100
	APÊNDICE 2 – ÁREAS DO CONHECIMENTO DA CAPES E PALAVRAS-CHAVE PRESENTES NA CAUDA LONGA DO TEMA AUDITORIA DA INFORMAÇÃO	102

1 INTRODUÇÃO

A relação do crescimento da ciência e tecnologia (C&T), resultante das atividades científicas, técnicas correlatas e das atividades de pesquisa e desenvolvimento, financiadas por investimentos públicos e de empresas privadas (estimulados pela renúncia fiscal do Governo Federal) são ações voltadas à geração de novas tecnologias e aumento do conhecimento científico (MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO, 2015).

O conhecimento científico incorpora atividades associadas à produção, disseminação e uso da informação, visando agregar credibilidade na troca de informações com os pares. Essa troca emerge como uma peça fundamental para a equidade no acesso à informação científica.

Esta “informação científica representa insumo básico para o desenvolvimento científico e tecnológico de um país” (KURAMOTO, 2006, p. 91), o que implica em esforços para entender os processos de produção desta informação e compreender a metodologia científica aplicada.

A metodologia científica é utilizada por pesquisadores em diversos campos científicos e em áreas distintas. Cada área do conhecimento apresenta suas especificidades, seja para avaliar e analisar a produção científica, como também, no dimensionamento de um determinado domínio de estudo. Para Hjørland (2004, p. 4, tradução nossa), o domínio “[...] pode ser uma disciplina, um campo escolar. Pode ser uma comunidade discursiva conectada a um partido político, à religião, ao comércio, ou a um lazer”. Como uma das diversas abordagens indicadas por Hjørland para a análise de domínio, destacam-se os estudos bibliométricos.

Os estudos bibliométricos surgiram no século XX, derivados do termo bibliometria que, para Spinack (1996), refere-se à análise do curso da comunicação escrita e o curso de uma disciplina, aplicando-se métodos quantitativos (matemático e estatístico).

Os estudos bibliométricos apresentam variadas formas de analisar e avaliar a produção científica, como por exemplo, os indicadores de produção, de citação e de ligação (ALVES, 2013); o que permite identificar as tendências da comunidade estudada, bem como as correntes da área, temática abordada, principais autores, tipo de autoria, produtividade dos autores, instituições, periódicos, países, tipo de

publicação predominante, hegemonia de algum idioma, coautoria, citação, cocitação, entre outros (GREGOLIN, 2005).

As publicações (artigos de periódicos, comunicações em eventos, livros, dissertações, entre outros) são canais de comunicação científica, apresentados como parte integrante na construção do conhecimento, no qual a utilização de citações e referências evidencia a sustentação do estudo em propostas teóricas e metodológicas anteriores. Assim, a lista de referências reflete as relações entre os pesquisadores e seus pares no processo da construção do conhecimento (SILVA; MENEZES, 2005).

Estudos bibliométricos são realizados em distintas áreas do conhecimento, com o objetivo de quantificar, descrever e prognosticar o processo da comunicação escrita e a auditoria de informação tem como objetivo analisar a situação atual e ajudar no processo de melhoria do fluxo e necessidade da informação.

A área de auditoria de informação é entendida como uma técnica de gestão e uma boa prática de gestão da informação em um contexto voltado à melhoria dos serviços baseados em informação, e da informação como um recurso (SOY, 2013). Tem sido estudada por Guitián (2009) e Guitián, Pérez e Porra (2015) como um domínio sendo que Guitián (2009), caracteriza o domínio analisando sua produção científica na base de dados Lisa¹, no período de 1979 e 2008. Guitián, Pérez e Porra (2015) aproximam os domínios de auditoria de informação e do conhecimento a partir da análise da produção científica das bases de dados Lisa, *Scopus* e *Web of Knowledge*, entre os anos 1994 a 2013. Ambos os estudos exploram o quantitativo da produção por ano, por revista; países que obtiveram maior frequência de produtividade, análise da produção de autor e coautor e as linhas de investigação cujos resultados, em geral, mostram que, em um domínio determinado, existe a reunião de um pequeno número de artigos que são altamente citados, enquanto que outros artigos são citados poucas vezes, demonstrando que a distribuição é comum nas chamadas leis de potência.

A primeira utilização das leis de potência foi na Ciência Social, por meio do economista e sociólogo Vilfredo Pareto, considerado como o estudo da distribuição individual de riqueza e rendimento, uma caracterização da vida econômica de diferentes países. Atualmente, são descritas e analisadas em diferentes áreas, tais

¹ URL: < <http://search.proquest.com/lisa>>.

como: economia e finanças, educação, demografia, geologia, história, climatologia, bibliometria e infometria, terrorismo e guerra, corrupção, turismo, esportes, artes, agronomia, ecologia, biologia, linguística, ciência da computação, ciências cognitivas, ciências sociais, astronomia, mecânica dos sólidos, física e química (BORTOLOSSI; QUEIROZ; SILVA; [s.n]).²

A Lei de Pareto, também conhecida como a Regra 80-20, apresenta-se, contemporaneamente, aliada ao chamado “fenômeno de cauda longa” (*long tail phenomenon*) – entendido como o efeito verificado em mercados *online*, nos quais a soma de vendas de produtos de “baixa popularidade” rivaliza – quando não ultrapassa – as vendas de itens “campeões” (ANDERSON, 2006)

Para Heidorn (2008), o fenômeno é adequado às discussões relativas aos dados científicos, e “um paralelo pode ser traçado com a bibliometria, na qual artigos de alto impacto não são encontrados, necessariamente, nos periódicos de alto impacto” (SEGLEN, 1997; SUN & GILES, 2007 *apud* HEIDORN, 2008).

Considerando-se, portanto, que a área de auditoria de informação tenha sido já investigada com o auxílio de algumas técnicas bibliométricas (Gutián (2009); Gutián, Pérez e Porra (2015)) estas foram limitadas a uma base (Lisa) no caso de Gutián (2009) e, três bases (Lisa, *Scopus* e *Web of Knowledge*) no estudo de Gutián, Pérez e Porra (2015). Percebeu-se, ainda, que nestes estudos, os resultados obtidos apresentaram o “efeito da cauda longa”, o qual não foi analisado. Portanto a presente pesquisa apresenta a seguinte questão: “como se configura o domínio de auditoria de informação ao se ampliar não apenas o universo de bases de dados, mas igualmente as técnicas bibliométricas utilizadas acompanhadas da análise do “efeito de cauda longa?”

Nesse sentido, o presente estudo tem como objetivo explorar as características das manifestações de “cauda longa”, derivadas de distribuições bibliométricas, no domínio da auditoria da informação. Este objetivo geral desdobra-se em três outros específicos, a saber:

- a) apresentar as publicações relativas ao domínio “auditoria de informação”;
- b) definir técnicas bibliométricas a serem aplicadas do material coletado sobre “auditoria de Informação”;

² Uma variação de lei de potência é a Lei de Zipf, que analisa a frequência com que certa palavra aparece num resumo, em um texto, nas palavras-chave. (SILVA; 2008)

- c) examinar os “alongamentos de cauda longa”, derivados das técnicas bibliométricas aplicadas, explorando potenciais relações.

Justifica-se o estudo pelo fato deste auxiliar na ampliação do conjunto de técnicas bibliométricas utilizadas para análise de domínio, avançando na exploração das características dos elementos que compõem a cauda longa, resultante dos cálculos realizados, mais especificamente, nas relações de citação e cocitação, como suporte para estudos de impacto da comunicação científica. Para o domínio da temática de auditoria de informação, contribui na identificação de quais os autores e as temáticas que apresentam maior relevância. Para o Programa em Ciência, Gestão e Tecnologia da Informação, justifica-se pelo fato de estimular o estudo de método e técnicas bibliométricas em um contexto interdisciplinar.

Para atingir os objetivos propostos, esta dissertação está organizada em cinco seções (ou capítulos): introdução, referencial teórico, procedimentos metodológicos, análise dos dados e resultados e considerações finais.

Na primeira parte, descreve as considerações sobre o financiamento dos investimentos públicos quanto ao crescimento da ciência e tecnologia, conhecimento científico como informação e a utilização da lei de potência.

O capítulo dois apresenta subsídio teórico metodológico, expondo os principais fundamentos que orientam a investigação da pesquisa. Sua organização está distribuída em cinco partes, sendo o primeiro bloco dedicado a conceitos, abordagens da auditoria da informação; o segundo relata a contribuição para a análise do domínio do conhecimento; o terceiro e o quarto apresentam os estudos bibliométricos utilizados na investigação. Na quinta e última seção deste capítulo, apresenta-se a definição do princípio/efeito da cauda longa.

No capítulo três, é tratada a caracterização da pesquisa e descrição dos procedimentos e recursos utilizados para o desenvolvimento.

O capítulo quatro demonstra os resultados da pesquisa, por meio da exposição de figuras, quadros e tabelas, criadas com base na coleta de dados.

As considerações finais trazem o destaque dos principais autores, e palavras chaves, o local de maior destaque de publicação, o idioma, a construção de rede dos autores mais cocitados, as limitações da pesquisa e a contribuição para pesquisas futuras.

2 REFERENCIAL TEÓRICO

O presente capítulo aborda a temática de auditoria da informação, quanto aos conceitos, tipos, métodos, abrangência e contribuições, caracterizando-a como “domínio”. Na análise de domínio, destacam-se os estudos bibliométricos como um campo da ciência da informação que trabalha com o estudo da produção e comunicação científica, em especial, o comportamento da literatura. Apresentam-se, o desenvolvimento no recorte das leis – Lotka e Zipf, entre outras. Elas caracterizam-se como leis de potência e o “fenômeno da cauda longa” é uma das suas manifestações.

2.1 AUDITORIA DE INFORMAÇÃO: CONCEITOS, CARACTERÍSTICAS E MODELOS

A auditoria surgiu no final do século XIII, na Inglaterra, com Eduardo I, em decorrência do desenvolvimento dos negócios e hoje é separada “em vários segmentos, proporciona aos seus usuários uma maior segurança sobre as atividades que estão sendo realizadas em uma organização” (MARCELO; CORDEIRO, 2013, p. 3). De acordo com Carneiro (2009), a auditoria poderá ter sua aplicabilidade em diversos domínios: gestão geral (as políticas, os objetivos, as estratégias e a coordenação global); econômico-financeiro (análise dos custos, rentabilidade, indicadores de gestão); técnicos (equipamentos, *software*, produção e linhas de inovação); valor humano (funções, hierarquias, motivação e formação profissional); informação (organização dos sistemas de informação, segurança e fluxos de informação) e funcional (resultados, adequação dos sistemas de informação e na capacidade de resposta da organização).

Estas auditorias auxiliam na tomada de decisão na gestão da organização e, acompanhando o segmento funcional, o presente trabalho explora o tema da Auditoria da Informação (AI). A Auditoria da Informação “disponibiliza um modelo que permite examinar a forma como as organizações gerem o seu ativo mais importante, a informação, através de processos de criação, captura, distribuição, partilha e armazenamento da informação” (GUIMARÃES; QUARESMA, 2015, p. 1).

Para Buchanan e Gibb (1998, p. 35, tradução nossa), o termo auditoria da informação se refere ao “processo de descobrir, monitorar e avaliar os recursos de

informação de uma organização para programar, manter e melhorar a gestão de informação dela”. Desta forma, a

Auditora de Informação depende de repositórios de informação descritiva sobre a própria informação a auditar e métodos de auditoria, para garantir o objetivo de identificar anomalias de integridade, consistência e segurança, além de visão mais estratégica da informação necessária para suporte às necessidades das organizações. [...] é fundamental conhecer a localização e estrutura da informação, e a semântica da informação através de regras de negócio modeladas com base na informação. (GUIMARÃES; QUARESMA, 2015, p. 2)

Devido à quantidade e acúmulo de informações, Von-Tran (2011) considera que as organizações têm a tendência para perder de vista o “como gerir a informação” e que informações existem (ou são necessárias) ao ambiente organizacional. Antes de uma organização, efetivamente, gerir a informação que cria, possui e distribui, deve, primeiro, efetuar uma auditoria sobre as práticas correntes de gestão de informação. Assim, para o autor, uma ferramenta para auxiliar o processo é a auditoria de informação (VON-TRAN, 2011).

A literatura sobre a auditoria de informação, de acordo com Pestana (2014, p. 52), “remonta a meados da década de 70” (quadro 1).

QUADRO 1 – CRONOLOGIA DOS PRINCIPAIS PESQUISADORES DE AUDITORIA DE INFORMAÇÃO

Século XX	Século XXI
Anos 1970	Anos 2000
Riley (1976)	Henczel (2000, 2001)
Quinn (1979)	Botha and Boon (2003)
Anos 1980	Orna (2004)
Henderson (1980)	Buchanan e Gibb (2007, 2008)
Reynolds (1980)	Anos 2010
Gruber (1983)	González (2011)
Gillman (1985)	
Burk e Horton (1988)	
Anos 1990	
Barker (1990)	
Orna (1990,1999)	
Ellis et al. (1993)	
Robertson (1994)	
Haynes (1995)	
Swash (1997)	
Buchanan e Gibb (1998)	

FONTE: PESTANA (2014, p. 70).

Vários autores tem trabalhado o conceito de AI, definindo diversos modelos e com enfoques diferentes. Um modelo, para se efetivar a AI, deve permitir

[...] examinar a forma como as organizações geram, [...] a informação, através de processos de criação, captura, distribuição, partilha e armazenamento da informação. Estes processos dependem de estruturas de informação para entender o que existe, como está estruturado, quem utiliza e quando necessita da informação, em particular no processo de captura da informação (GUIMARÃES; QUARESMA, 2015, p. 2).

Assim, a AI “é uma avaliação sistêmica do uso, recurso e fluxo da informação, como uma verificação referente tanto às pessoas como aos documentos existentes, para estabelecer a extensão no que contribuem nos objetivos da organização” (GUITIÁN, 2009, p. 3-4, tradução nossa).

Conforme descreve Saint Clair (1995, p. 4, tradução nossa), “o objetivo da auditoria de informação é a identificação das necessidades de informação dos usuários, assim como, a prevenção do grau em que o serviço de informação atende a essas necessidades”. Desta forma, a auditoria tem objetivos mais amplos, tais como, “a obtenção de informação, sua validade, sua organização, sistematização e análise” (DANTE, 2008, p. 6-7).

De 1970 a 2011, observa-se uma gama de estudos, entre proposta de vários enfoques, modelos e metodologias, apresenta por grupo de pesquisador, quanto ao objetivo, a saber (quadro 2).

QUADRO 2 – PRINCIPAIS PESQUISADORES EM AUDITORIA DA INFORMAÇÃO POR PERÍODO E OBJETIVO ABORDADO NA INVESTIGAÇÃO DE CADA METODOLOGIA

PERÍODO	AUTOR	OBJETIVO
1975	Riley e Anderson	ênfase na medição quantitativa de custos;
1980	Reynolds	análise nas fraquezas dos relatórios dos sistemas;
1983	Gruber	considera a eficiência e eficácia como, organizações usam, manejam e protegem a informação;
1985	Gilman	ênfata mais a análise do sistema de informação que o custo benefício dos recursos de informação;
1987	Worlock	discute a estrutura do processo de auditoria antes de testes em distintos lugares;
1988	Burk e Horton	identifica, mapeia e avalia os recursos de informação;
1990	Barker	ênfata a confiabilidade do sistema de informação;
1992	Stanat	avalia a efetividade das redes de sistemas de informação da organização;
1993	Booth e Haines	desenvolve uma nova política de informação para autoridade de saúde no Reino Unido;
1993	Ellis, Barker, Potter and Pridegeon	reconhece as diferentes abordagens de auditoria de informação;
1994	Webb	usado para auditar o sistema eficiente quanto aos recursos que estão sendo usados;
1997	St Clair	agrupado em cinco principais áreas;
1998	Buchanan & Gibb	avalia as fontes e correspondência das tarefas que apoiam os fatores chaves para o êxito, os objetivos;

QUADRO 2 – PRINCIPAIS PESQUISADORES POR PERÍODO NA AUDITORIA DA INFORMAÇÃO E O OBJETIVO ABORDADO

(conclusão)

PERÍODO	AUTOR	OBJETIVO
1989	Orna	ênfatiza a importância do estudo dos fluxos dentro da organização;
2001	Susan Henczel	analisa a organização estratégia de cima para baixo;
2003	Cristina Soy i Aumatell	propõem método específico no software para o processamento e análise dos dados;
2003	Botha & Boon	classifica em custo benefício, enfoque geográfico, informação gerencial, operações assessorias, auditoria híbrida;
2005	Villaderfrancos, Álvarez	analisa os processos de informação na organização;
2011	González, Guitián	avalia, quantitativamente e qualitativamente, a gestão de informação nas organizações.

FONTE: O autor (2016)

Em 2007, com a publicação da obra de Buchanan e Gibb, percebe-se a ampliação da abrangência da AI, a qual

não deve apenas incidir sobre dados e informação, mas deve cobrir a análise do conhecimento e investigar os especialistas, redes e comunidades que suportam ou podem fornecer conhecimentos para a empresa, a experiência documentada, os resultados que podem ser compartilhados e a propriedade intelectual (BUCHANAN, GIBB, 2007, p. 168, tradução nossa).

Como ferramenta para a gestão da informação nas organizações a auditoria da informação ocorre mediante: a identificação e verificação da existência de serviços, racionalização de recursos, o controle dos custos de recursos, a visibilidade dos serviços com o aperfeiçoamento do mercado e o cumprimento das recomendações de melhoria oriunda do mercado, conscientiza a organização do que constituem esses recursos e serviços para o desenvolvimento do trabalho (DANTE, 2008).

Além destes pontos, Buchanan e Gibb (2007, p. 35, tradução nossa) descrevem a identificação dos recursos da organização e de suas necessidades de informação, tais como:

- a) identificação de custos e benefícios dos recursos de informação;
- b) identificação de oportunidades para o emprego dos recursos de informação para vantagens estratégicas competitivas;
- c) integração dos investimentos em tecnologia de informação com as iniciativas estratégias de negócio;
- d) identificação de fluxos e processos de informação;
- e) desenvolvimento de uma estratégia integrada e/ou política de informação;
- f) criação de uma ciência da importância da gestão de informação e a definição de seu papel na gestão;

- g) monitoramento e avaliação da correspondência com normas, leis e políticas.

Já para Pires (2012, p. 6), “a auditoria da informação tem a ver com o exame sistemático dos recursos de informação, o uso de informação, os fluxos de informação em uma organização”. Implica na identificação das necessidades de informação dos usuários e quanto, eficazmente ou não, são atendidos. Já para Guitián (2009, p. 4, tradução nossa), a auditoria da informação busca:

- ✓ assegurar que a informação que circula através do sistema seja ideal para alcançar os objetivos da empresa;
- ✓ determinar os pontos fortes, pontos fracos e lacunas em recursos de informação da organização;
- ✓ especificar o valor, desempenho e utilidade dos sistemas de informação;
- ✓ avaliar a informação da organização de serviço completo;
- ✓ determinar a relação entre o esforço realizado para a obtenção da informação e utilização.

Há diferentes tipos de auditoria de informação, tais como, as auditorias de comunicação, mapeamento, sistema de informação, conhecimento e inteligência (BOTHA; BOON, 2003).

A Auditoria da Informação, portanto, representa uma contribuição para a identificação, aquisição, geração, tratamento e uso dos recursos de informação interna, correspondente a missão e objetivo da organização, possibilitando determinar o comportamento dos fluxos e canais que passam as informações, com a possibilidade de determinar certas irregularidades na relação com a utilização (GUITIÁN, 2009).

Como o tema está sendo discutido, é possível identificar que a auditoria da informação tem se configurado como um domínio (GUITIÁN, 2009); (GUITIÁN, PÉREZ, PORRA, 2015).

2.2 ANÁLISE DE DOMÍNIO: ALGUNS ELEMENTOS E ABORDAGENS

Os pesquisadores da comunidade científica, seus discursos e a cultura da pesquisa, em diferentes áreas, podem contribuir para delimitar o domínio do conhecimento. Um domínio pode ser entendido como “[...] uma área de especialização, um corpo de literatura, ou um grupo de pessoas trabalhando juntas em uma organização” (MAI, 2005, p. 605, tradução nossa).

O domínio “pode ser uma disciplina científica, uma área do conhecimento científico ou uma comunidade discursiva, [...], interesse, em que se demarca determinado conhecimento com limites definidos, com profissionais ou grupos” (OLIVEIRA, 2013, p. 26).

A Análise de Domínio (AD) teve seu início, em meados de 1980, sendo utilizada por Neighbors, numa “tentativa de identificar os objetos, operações e relações entre os peritos em um determinado domínio” (NEIGHBORS, 1981 *apud* KERR, 2003, p. 11), na área de Engenharia da Computação.

Em 1995, com a publicação do artigo: *Toward a new horizon in information science: domain analysis*, no periódico *Journal of the American Society for Information Science*, de autoria de Birger Hjørland e Hanne Albrechtsen, a análise de domínio teve sua contextualização na Ciência de Informação.

Para Hjørland (2004, p. 4, tradução nossa), “domínio pode ser uma disciplina, um campo escolar. Pode ser uma comunidade discursiva conectada a um partido político, à religião, ao comércio, ou a um lazer”. Desta forma,

a análise de domínio é um processo pelo qual se pode verificar o que é importante ou significativo sobre algum campo do esforço humano, identificando elementos que permitem analisar um contexto científico, tais como tendências, padrões, objetos (pessoas), processos e relacionamentos existentes (OLIVEIRA e GRACIO, 2009, p. 50)

Com o objetivo de conhecer, estudar, examinar um domínio, Hjørland (2002) destaca onze abordagens, ou orientações, para a análise de domínio, a saber:

1. Produção de guias de literatura: literatura de referência (bibliografias, dicionários, enciclopédia, etc.) são “publicações que listam e descrevem os recursos da informação em uma ou mais áreas, [...] é uma espécie de bibliografia de documentos num domínio”, guia que concentra explicação a fim de fornecer referências da literatura, orienta na seleção e auxíla o usuário na consulta à literatura, informações e banco de dados, “podem ser combinados com outras abordagens do tipo: produção de classificações especializadas; estudos de documentos e gêneros; estudos epistemológicos e crítico e estudos de estruturas e instituições de comunicação científica” (HJØRLAND, 2002, p. 426, tradução nossa).

2. Elaboração de classificações especiais e tesouros: voltados à classificação dos termos específicos, os tesouros consistem em conceitos centrais, organizado de acordo com relações semânticas, tais como as genéricas e de sinonímia. Esta

abordagem pode ser combinada com outras do tipo: pesquisa sobre indexação e recuperação nas especialidades; estudos bibliométricos, históricos, epistemológicos e críticos, e terminológicos” (HJØRLAND, 2002, p. 429, tradução nossa).

3. Indexação e recuperação da informação: esta abordagem pode aumentar a visibilidade da capacidade de recuperação, avaliando as práticas pouco desenvolvidas e auxiliar na recuperação de informações. “Esses estudos para análise de domínio são favorecidos, em conjunto com: elaboração de classificações especializadas e tesouros; estudos bibliométricos, epistemológicos e críticos, terminológicos e de discurso” (HJØRLAND, 2002, p. 430, tradução nossa).

4. Estudos empíricos dos usuários: medem as tendências conforme as necessidades do usuário, “organizando a preferência ou comportamento ou modelos mentais de seus usuários” (OLIVEIRA, 2013, p. 28). Essa abordagem pode “ser compartilhada com: estudos bibliométricos, epistemológicos e críticos; e estudos de estruturas e instituições de comunicação científica” (HJØRLAND, 2002, p. 431, tradução nossa).

5. Estudos bibliométricos: ferramentas de análise de domínio que representam a ligação entre pesquisadores, instituições ou países, demonstrando correntes dominantes, por meio de análise de citações ou de cocitação. A “interpretação dessas análises deve contar com o auxílio de estudos históricos, epistemológicos e críticos” (HJØRLAND, 2002, p. 436, tradução nossa).

6. Estudos históricos: mostram tradições organizadas, nas quais o conhecimento de fatos anteriores auxilia na amplitude e aprofundamento de questões. Trata-se de “compreender documentos, organizações, sistemas, conhecimento e informação, em uma perspectiva histórica e métodos históricos, capazes de fornecer uma perspectiva mais profunda e mais coerente, comparado com os tipos de pesquisas não históricas de natureza mecanicista” (HJØRLAND, 2002, p. 436, tradução nossa).

7. Estudos de documentos e estilos: diferentes estudos que abrangem aspectos qualitativos e/ou quantitativos, apresentando diferentes tipos, em diferentes comunidades. Essa “abordagem deve-se apoiar em pesquisas sobre indexação e recuperação em áreas especializadas, estudos históricos, estudos epistemológicos e críticos” (HJØRLAND, 2002, p. 438, tradução nossa).

8. Estudos epistemológicos e críticos: juntamente com a abordagem de estudos históricos, “as teorias da epistemologia [...] fornecem orientação para seleção,

organização e recuperação da informação”, além do “nível mais alto de generalidade sobre as necessidades de informação” (HJØRLAND, 2002, p. 440, tradução nossa).

9. Estudos terminológicos, linguagens para propósitos específicos (LSP):

“linguagem e terminologia são objetos muito importantes para a ciência da informação porque afetam o pensamento e, portanto, as questões que se colocam nos bancos de dados, bem como os textos de busca” (HJØRLAND, 2002, p. 440).

Para análise de domínio, “devem ser combinados com estudos bibliométricos, históricos, epistemológicos e críticos” (HJØRLAND, 2002, p. 436, tradução nossa).

10. Estruturas e instituições da comunicação científica:

“as fontes primárias, secundárias e terciárias são as principais categorias de literatura, que devem ser complementadas com literatura da fonte e, finalmente, popularizações e livros-textos, que têm a função de informar as pessoas de fora do domínio. [...] Muitos dados quantitativos e qualitativos são necessários para mapear essa estrutura de maneira precisa. [...] O estudo das estruturas da divisão interna do trabalho dentro dos domínios e a troca de informação entre os domínios, fornecem informação útil para a compreensão da função de tipos específicos de serviços de documentos” (HJØRLAND, 2002, p. 446-448, tradução nossa).

11. Cognição científica, conhecimento especializado e inteligência artificial:

tem como foco a ciência cognitiva na ciência da informação voltada a “fornecer aos usuários, informações que possam ajudá-los a avaliar a validade das diferentes alegações de conhecimento e a estabelecer sua própria visão em alguma questão, baseada nos estudos de todos os argumentos disponíveis” (HJØRLAND, 2002, p. 450, tradução nossa).

Hjørland “destaca que a combinação de mais de uma abordagem, das 11 abordagens apresentadas, fortalece os argumentos, dando um caráter de maior consistência à análise de um domínio” (ALVES, 2013, p. 23).

Para Valério *et al* (2013, p. 6), “a análise de domínio não é uma metodologia, é, pois, um olhar, uma perspectiva e uma verificação do contexto sobre um determinado comportamento das comunidades discursivas”. Porém, qualquer estudo neste segmento deve esclarecer as definições, o escopo e alcance, destacando-se a extensão e a intensão do domínio sob análise, no propósito de deixar claro a que se destina (TENNIS, 2012).

Assim, deve-se nomear o domínio, descrevendo as especificidades, detalhando as extensões e as exclusões e apresentar uma análise de domínio com

claro propósito. Tennis explora a análise de domínio descritiva, que visa a uma pesquisa básica; e a análise de domínio instrumental, que visa ao planejamento e à construção de um sistema de informação (2012, p. 9).

A análise de domínio descritiva apresenta três imperativos em sua concepção. Segundo Tennis (2012, p. 10), é necessário “(1) seguir os métodos das Ciências Sociais; (2) observar os atores sociais; (3) apoiar-se em um coletivo que seja externo às posturas subjetivas”. Já a análise de domínio instrumental “surge a partir da necessidade de compreender um domínio para se construir ou revisar um sistema de informação” (TENNIS, 2012, p. 10).

Para o mesmo autor, “devemos levar em conta a função do tempo quando analisarmos as evidências dos conceitos em um domínio, assim como as diferentes posturas epistêmicas potenciais” (TENNIS, 2012, p. 12). Identificar um domínio de interesse está no escopo dos estudos infométricos.

Assim, entre os métodos de análise de domínio, destacam-se os estudos bibliométricos e infométricos. Os estudos bibliométricos são entendidos como “um campo da ciência da informação que trabalha com o estudo da produtividade dos autores através da análise e medição da produção bibliográfica” (DIAS, 2015, p. 12). Alguns elementos da bibliometria contribuem para o desenvolvimento desta pesquisa.

2.3 OS ESTUDOS BIBLIOMÉTRICOS: RECORTE PARA A ANÁLISE DE DOMÍNIO

A avaliação dos pares e a visibilidade do conhecimento produzido pelos pesquisadores assume papel de importância na comunidade científica. Para a mensuração desta produção, são utilizados métodos quantitativos aplicados à bibliografia, “especialmente nos últimos 40 anos, quando a explosão documentária impôs de certa forma, a necessidade da criação de instrumentos para a avaliação da ciência” (OLIVEIRA; GRACIO, 2009, p. 2040). Estas publicações podem ser avaliadas por algumas variáveis, tais como: tipo de documento, idioma, temporalidade, tipo de autoria, autores citados, títulos de periódicos, palavras-chave, entre outras.

Este método quantitativo (matemático e estatístico), voltado à análise do comportamento da informação, “foi utilizado, pela primeira vez, por W. Hulme, em

duas conferências na Universidade de Cambridge” (SPINACK, 1996, p. 33, tradução nossa), com a denominação: bibliografia estatística. No decorrer dos anos, este termo passou por diversas denominações, tais como: ‘livrometria’ e ‘bibliotecometria’, mas foi por volta de 1969, que o termo ‘bibliografia estatística’, “considerado insatisfatório devido ao que poderia ser interpretado, erroneamente, como bibliografia sobre estatística” (SPINACK, 1996, p. 33, tradução nossa) deu lugar à utilização do termo “bibliometria”. Spinack apresenta as seguintes definições da bibliometria:

- ✓ aplicação da análise estatística para estudo das características do uso e criação de documentos;
- ✓ estudo quantitativo da produção de documentos, refletido nas bibliografias;
- ✓ aplicação de métodos matemáticos e estatísticos para o estudo da utilização dos livros e outros meios de comunicação dentre e entre sistemas de bibliotecas;
- ✓ estudos quantitativos de unidades físicas publicadas, bibliográficas ou sua substituta unidade; e
- ✓ estuda a organização dos setores científicos e tecnológicos a partir das fontes bibliográficas para identificar os autores, suas relações e tendências. (1996, p. 34-35, tradução nossa)

Assim, a avaliação por meio de indicadores bibliométricos, voltados para a mensuração do desempenho da comunidade de uma determinada área, pode ser considerada “parte integrante do processo da construção de novos conhecimentos, alcançados por parte dos pesquisadores, com a finalidade de atribuir credibilidade e qualidade do conhecimento produzido” (CUSTÓDIO, 2012, p. 51). Neste contexto, a bibliometria

[...] tem como objetos de estudo os livros ou as revistas científicas, cujas análises se vinculam à gestão de bibliotecas e bases de dados. A cienciometria preocupa-se com a dinâmica da ciência, como atividade social, tendo como objetos de análise a produção, a circulação e o consumo da produção científica. A infometria, por sua vez, abarca as duas primeiras, tendo desenvolvido métodos e ferramentas para mensurar e analisar os aspectos cognitivos da ciência. (SANTOS, KOBASHI, 2009, p. 159)

Com o desenvolvimento e a utilização do método (bibliométrico), surgem novas técnicas de mensuração, tais como a cienciometria, que

[...] vai além das técnicas bibliométricas, pois também examina o desenvolvimento e as políticas científicas. As análises quantitativas da cienciometria consideram a ciência como uma disciplina da atividade econômica. Por esta razão, a cienciometria pode estabelecer comparações

entre as políticas de investigação entre os países, analisando seus aspectos econômicos e sociais. (SPINACK, 1996, p. 49, tradução nossa)

A infometria é considerada a base das investigações da bibliometria e cienciometria que, para Spinack (1996, p. 131-132, tradução nossa), “compreende assuntos como o desenvolvimento de modelos teóricos e as medidas de informações, para regularidade dos atos associados à produção e ao uso da informação registrada”.

A webmetria é o termo designado para estudos métricos, a partir de dados da *web*, e se configura como o estudo dos aspectos quantitativos da construção e uso dos recursos de informação (GOUVEIA; LANG; 2013, p. 173). O quadro 3 demonstra as dinâmicas do estudo métrico da informação.

QUADRO 3 – TIPOLOGIA DE ESTUDOS MÉTRICOS DA INFORMAÇÃO, CONFORME MACIAS-CHAPULA 1998

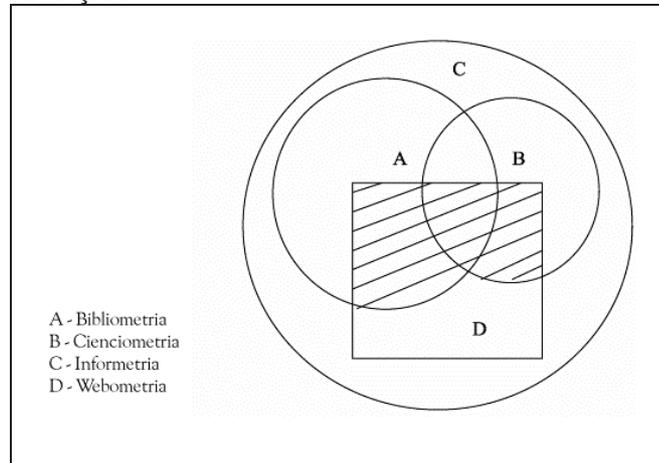
Tipologia	Bibliometria	Cienciometria	Infometria
Objetos de estudo	Livros, documentos, revistas, artigos, autores, usuários	Disciplinas, assunto, áreas, campos	Palavras, documentos, bases de dados
Variáveis	Número de empréstimos (circulação) e de citações, frequência de extensão de frases etc.	Fatores que diferenciam as subdisciplinas. Revistas, autores, documentos. Como os cientistas se comunicam	Difere da cienciometria no propósito das variáveis: por exemplo, medir a recuperação, a relevância, a revocação etc
Métodos	<i>Ranking</i> , frequência, distribuição	Análise de conjunto e de correspondência	Modelo vetor-espaço, modelos booleanos de recuperação, modelos probabilísticos; linguagem de processamento, abordagens baseadas nos conhecimentos tesouros
Objetivos	Alocar recursos: tempo, dinheiro, etc.	Identificar domínios de interesse. Onde assuntos estão concentrados. Compreender como e quanto os cientistas se comunicam	Melhorar a eficiência da recuperação

FONTE: MACIAS-CHAPULA (1998, p. 135).

Na figura 1, Vanti representa, em forma de diagrama, as relações entre os diferentes tipos de estudos métricos “que considera que a infometria é mais ampla e compreende a bibliometria e a cienciometria. [...] A webmetria [...] representada como um subcampo dentro da infometria, já que nasceu e é parte integrante desta” (VANTI, 2002, p. 161). As relações significam “a mensuração dos fluxos de

informação, da comunicação científica e da difusão do conhecimento” (GUEDES, 2012, p. 80).

FIGURA 1 – DIAGRAMA REPRESENTANDO GRAFICAMENTE A INTER-RELAÇÃO DE MÉTRICAS DE COMUNICAÇÃO



Fonte: VANTI (2002, p. 161).

O conjunto de indicadores bibliométricos, voltado para a análise da produção científica, pode ser dividido em indicadores de produção, de citação e de ligação (GREGOLIN, 2005). Os indicadores são medidas de instrumentos para avaliação, com o objetivo de mensurar o resultado e comparação de desempenho (quadro 4).

QUADRO 4 – TIPOS DE INDICADORES BIBLIOMÉTRICOS

INDICADORES DE PRODUÇÃO	
Definição	Construídos pela contagem do número de publicações, por tipo de documentos (livros, artigos, publicações científicas, relatórios, etc.); por instituição; por área de conhecimento; por país etc. O indicador básico é o número de publicações.
Resultados	- indicadores de participações percentuais; - taxas de crescimento ou rateios; - produtividade de autores (lei de <i>Lotka</i>); - frequência de palavras (distribuição de <i>Zipf</i>), - produtividade dos periódicos (distribuição de <i>Bradford</i>).
INDICADORES DE CITAÇÃO	
Definição	Baseiam-se na medida do número de citações recebidas por uma determinada publicação.
Resultados	- fator de impacto; - número de citações; - índice de imediatez; - índice de impacto; - índice de atividade; - índice de afinidade; - índice de atração; - índice h.
INDICADORES DE LIGAÇÃO	
Definição	Baseados em coocorrência de autoria, de citações e de palavras, e aplicados para o mapeamento de conhecimento e de redes de relacionamento entre pesquisadores, instituições e países, empregando-se, inclusive, técnicas de análise de estatística e de agrupamento.

QUADRO 4 – TIPOS DE INDICADORES BIBLIOMÉTRICOS

(conclusão)

INDICADORES DE LIGAÇÃO	
Resultados	<ul style="list-style-type: none"> - “análise de coautoria” é medida pelo número de publicações de coautores, utilizada para mensuração dos esforços de colaboração científica, sendo empregada para identificar e mapear a cooperação nacional, internacional ou regional em diferentes áreas do conhecimento; - “análise de cocitações” tem como medida, o número de cocitações que autores e publicações frequentemente cocitados estão mais próximos em especialização científica do que outros; - “análise de coocorrência de palavras” é realizada a partir de palavras-chave, utilizada para descrever artigos, palavras do título, do resumo, do texto integral, palavras presentes na classificação dos artigos ou do próprio código de classificação do artigo.

FONTE: Adaptado de GREGOLIN (2005).

Os indicadores bibliométricos, no desenvolvimento de uma ciência, podem identificar:

- a) o crescimento cronológico da produção científica;
- b) a produtividade de autores e instituições;
- c) a colaboração entre pesquisadores e instituições;
- d) o impacto das publicações;
- e) a análise e avaliação de fontes difusoras de trabalhos; e
- f) a dispersão da produção científica entre diversas fontes. (PEREIRA, 2015, p. 143)

Para Danuello (2007, p. 56), “as técnicas bibliométricas podem oferecer valiosas informações sobre uma disciplina e sobre as relações entre disciplinas diversas, podendo revelar alguns aspectos da realidade, tais como os padrões sociais na comunidade científica”, e são aplicadas em diferentes situações (quadro 5).

QUADRO 5 – LEIS E PRINCÍPIOS BIBLIOMÉTRICOS

Leis/ Princípios/ Indicadores	Foco de Estudo	Principais Aplicações
Lei dos 80/20	Demanda de informação	Composição, ampliação e redução de acervos
Lei de Lotka	Autores	Estimar o grau de relevância de autores em dada área do conhecimento
Leis de Zipf	Palavras	Indexação automática de artigos científicos e tecnológicos
Ponto de Transição de Goffman	Palavras	Indexação automática de artigos científicos e tecnológicos
Cocitação	Citações	Estimar o grau de ligação de dois ou mais artigos
Fator de impacto	Citações	Estimar o grau de relevância de artigos, cientistas e periódicos científicos, em determinada área do conhecimento
Correlação de Pearson (r)	Rede de autores	Identificar a evidência, a similaridade do perfil de cocitação dos dois autores cocitados em relação aos outros autores cocitados
Cosseno de Salton (CS)	Rede de autores	Medir a proximidade de dois autores cocitados, independente do volume da literatura referenciada

QUADRO 5 – LEIS E PRINCÍPIOS BIBLIOMÉTRICOS

(conclusão)

Leis/ Princípios/ Indicadores	Foco de Estudo	Principais Aplicações
Índice de Jaccard	Rede de autores	Medir a proximidade de dois autores, altamente cocitado ou menos intensa a associação
Frequência Esperada	Rede de autores	Medir a colaboração esperada relacionada à frequência observada

FONTE: Adaptado de GUEDES E BORSCHIVER (2005).

Parte destas leis e princípios será abordada neste estudo, para atingir o objetivo proposto.

2.4 INDICADORES BIBLIOMÉTRICOS DE INTERESSE PARA A PESQUISA

Os indicadores bibliométricos sinalizam a importância ou o que é significativo dentro do campo ou contexto científico, “constituindo um dos instrumentos metodológicos que contribuem para a visualização do comportamento da ciência em uma determinada área do conhecimento” (ALVES, 2013, p. 27). Os seguintes princípios: Lei de Lotka, Leis de Zipf, ponto de transição de Goffman, cocitação, correlação de Pearson, Cosseno de Salton e Índice de Jaccard serão detalhados a seguir.

2.4.1 Lei de Lotka

Também conhecida como a lei do quadrado inverso, formulada por Lotka, em 1926, visa a apoiar a análise da produção científica dos autores, tendo como premissa que o número de autores que tenham publicado, exatamente, (n) trabalho é inversamente proporcional a (n²). Desta maneira, analisando-se o número de cientistas que escrevem dois artigos, a proporção seria igual a ¼ do número de cientistas que escrevem um; e o número de cientistas que escrevem três artigos, seria igual a 1/9 do número de cientistas que escreveram um; e assim, sucessivamente.

Na aplicação da Lei, de acordo com Beuren e Silva (2014, p. 40), “merecem destaque três formas de contagem de autoria: a) direta, quando se atribui crédito apenas ao autor nomeado em primeiro lugar; b) completa, em que se atribui crédito a todos os autores; c) contagem ajustada/ fracionada, onde o crédito é fracionado

entre os autores”. Contudo, considera-se que métodos de contagem direta e ajustada não produzem diferenças essenciais. (URBIZAGASTEGUI, 2008).

A aplicabilidade da Lei de Lotka “se verifica na avaliação da produtividade de pesquisadores, na identificação dos centros de pesquisa mais desenvolvidos e no reconhecimento da “solidez” de uma área científica” (GUEDES, 2012, p. 84) e o resultado é obtido pela aplicação da equação 1.

$$a_n = \frac{1}{n^2} \quad (1)$$

Onde, “n é igual à quantidade de artigos” (MACHADO, J. R., 2016, p. 114), voltada para a representação que expressa o número de autores que tenham publicado exatamente (n).

2.4.2 Lei de Zipf

De acordo Rajneesh e Rana (2015, p. 223, tradução nossa), “a lei de Zipf foi usada para analisar a relação de palavras-chave e sua frequência; as palavras-chave de mapeamento da literatura da ciência e as palavras-chave de alta frequência”. Ao analisar a obra de Ulisses de James Joyce, Zipf encontrou uma correlação entre o número de palavras diferentes e a frequência de seu uso, e que existe uma regularidade fundamental na seleção e uso das palavras, na qual, um pequeno número de palavras é usado muito mais frequentemente. (ARAÚJO, 2006, p. 16). Esta correlação é obtida, seguindo a frequência de dada palavra ocorrida e a posição constante, na lista de palavras ordenadas. Para Guedes e Borschiver (2005, p. 6),

[...] essa lista era confeccionada, levando-se em conta a frequência de ocorrência. À posição nesta lista dá-se o nome de ordem de série (rank). Assim, a palavra de maior frequência de ocorrência tem ordem de série 1, a segunda maior frequência de ocorrência, ordem de série 2 e, assim, sucessivamente. Na análise, constatou que a ordem de uma série de palavras (r) multiplicados pela frequência (f), produz uma constante denominada (k), conhecida como a primeira lei de Zipf, [conforme a equação 2].

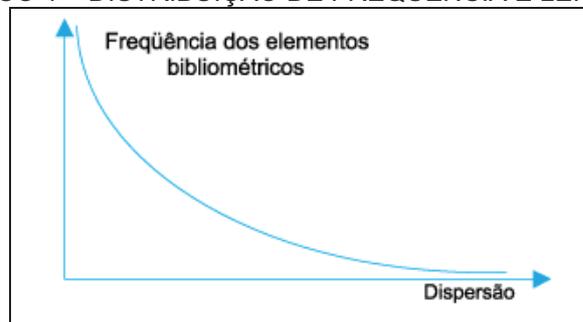
$$k = r \cdot f \quad (2)$$

Esta primeira lei é aplicável a palavras com elevado índice de ocorrência no texto. Para as palavras de menor ocorrência (um pequeno índice), Zipf apresentou a segunda lei, por Booth, no ano de 1967, expressada pela equação 3.

$$\frac{l_1}{l_n} = \frac{n(n+1)}{2} \quad (3)$$

Onde, "l₁ é o número de palavras que têm frequência 1, l_n é o número de palavras que têm frequência n, sendo a constante válida para a língua inglesa". (GUEDES, BORSCHIVER, 2005, p. 6). O comportamento das palavras de alta frequência e de baixa frequência resulta em duas extremidades (gráfico 1), da distribuição de palavras de um texto, tendo a representatividade, uma região crítica à transição destes comportamentos, ou seja, o meio da frequência (informação interessante) sendo determinado como hipótese que, nessa região, encontram-se as palavras de maior conteúdo semântico de um texto (GUEDES, BORSCHIVER, 2005, p. 7).

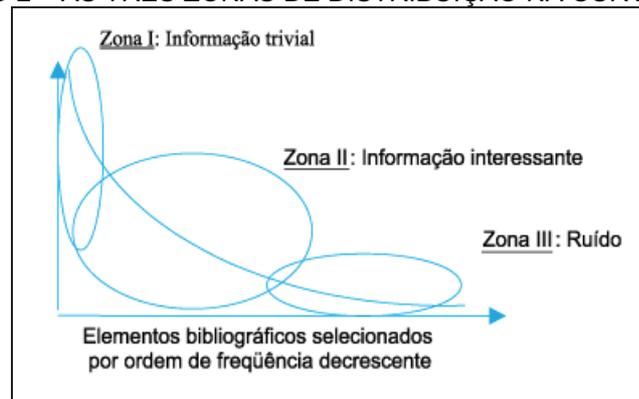
GRÁFICO 1 – DISTRIBUIÇÃO DE FREQUÊNCIA E LEI DE ZIPF



FONTE: QUONIAM et al (2001, p. 24).

No gráfico 2 – Zonas de distribuição observam-se os extremos (zona I e zona III), e a intermediária (zona II), para o desenvolvimento em saber quais são as palavras de maior conteúdo, relevância, para o estudo.

GRÁFICO 2 – AS TRÊS ZONAS DE DISTRIBUIÇÃO NA CURVA DE ZIPF



FONTE: QUONIAM et al (2001, p. 24).

Cada zona (I, II e III) apresenta informação determinante para a área da pesquisa, a saber:

Zona I – Informação trivial ou básica: define os temas centrais da análise bibliométrica;

Zona II – Informação interessante: localiza-se entre as zonas I e III e mostra ora os temas periféricos, ora a informação potencialmente inovadora. É aí que as transferências de tecnologia relacionadas aos novos temas devem ser consideradas;

Zona III – Ruído: tem como característica possuir conceitos ainda não emergentes, onde é impossível afirmar se eles serão emergentes ou se são apenas ruídos estatísticos. (QUONIAN, et al, 2001, p. 23-24)

A partir da segunda lei de Zipf, Goffman admite que, entre as palavras de alta e baixa frequência, há uma região denominada de “ponto de transição”.

2.4.3 Ponto de Transição de Goffman

Para atingir a região que compreende a transição, palavras de alto conteúdo semântico, a segunda lei de Zipf foi utilizada, fornecendo o número de palavras que apresentam frequência n que tenderia a um (unidade). Assim,

a expressão da 2ª Lei de Zipf modificada por Booth, teria que fornecer o comportamento típico das palavras de alta frequência [...]. Substituindo-se na expressão da 2ª Lei de Zipf-Booth, l_n por 1 obtém-se:

$$\frac{l_1}{1} = \frac{n(n+1)}{2} \text{ ou ainda, rearranjando: } n^2 + n - 2l_1 = 0, \text{ cujas raízes são } n = \frac{-1 \pm \sqrt{1+8l_1}}{2} \text{ (GUEDES, BORSCHIVER, 2005, p. 8)}$$

O resultado expresso (n), denominado de Ponto de Transição (T) de Goffman, demonstra a posição onde, ao redor do ponto, existe a probabilidade da

localização das palavras com alto conteúdo semântico. Nesta região, “ao redor do ponto, existe a probabilidade de concentrar as palavras de alto conteúdo semântico, que seriam utilizadas na indexação de um texto em análise” (GUEDES, 2012, p. 88). Pao (1978), no teste de hipótese do referido ponto de transição de Goffman, delimitou a região, projetando para baixo o mesmo número de palavras que ocorre acima do ponto T.

Assim, o ponto T de Goffman relaciona-se com a representação da informação, representado pelas palavras de maior conteúdo semântico.

2.4.4 Análise de Citação

Com a função de estabelecer a relação de dois documentos, a análise de citação teve seu contexto histórico no início do século XVII, com a contagem de referências utilizado por P. Gross e E. Gross, no ano de 1927. Dois anos após, esse procedimento foi também realizado por Allan e, posteriormente, em 1931, por Gross e Woodford. Mas, foi apenas na década de 1960, com a utilização do computador, o marco do surgimento do índice de citação, denominado de *Science Citation Index – SCI*, com Eugene Garfield, o fundador do *Institute of Scientific Information – ISI* (ARAÚJO, 2006).

Por citação, entende-se, “quando um documento menciona ou refere-se a um documento anterior, diz que o documento está citado. A referência da citação encontra-se no texto principal, em nota de rodapé, no final do documento, ou como bibliografia” (SPINACK 1996, p. 54, tradução nossa). Esta bibliografia, organizada nas referências, adotam normas e padrões estabelecidos por organizações como a *American Psychological Association – APA*, e a Associação Brasileira de Normas Técnicas – ABNT, órgãos responsáveis pela normatização de técnicas para as normas bibliográficas.

As citações são utilizadas por razões históricas, tendo como fatores motivadores:

- a) passar informação histórica;
- b) descrever outros trabalhos relevantes;
- c) fornecer dados ou informações, por outras razões que não sejam comparações;
- d) fornecer dados ou informações para realizar comparações;
- e) usar equações teóricas;

- f) uso de metodologia; e
- g) mencionar teorias e métodos nas aplicações, as melhores. (SPINACK, 1996, p. 53, tradução nossa)

As citações são informações retiradas de publicações consultadas na realização do trabalho, introduzidas no texto para esclarecer e/ ou complementar a ideia do autor, a partir de textos transcritos ou parafraseados.

A análise da citação é utilizada como uma ferramenta para avaliar e comparar as fontes utilizadas pelos pesquisadores (GINN, 2003). Conforme afirma Job (2006, p. 31), “têm sido utilizado como forma de dar crédito e reconhecimento aos autores [...], uma vez que o caminho realizado pelos pesquisadores através das citações poderá ser refeito por outra pessoa, averiguando a veracidade das informações”.

Destaca-se a análise de citação pela contagem de referências ao final de um trabalho, com o objetivo de identificar as características e mapear a comunidade científica igualmente, com base em métodos estatísticos e métricas da análise de produção.

Para Garfield (1979), a análise de citação apresenta problemas quanto à contagem na quantidade de periódicos utilizados, à distinção entre os pesquisadores de publicação única, as pessoas com o mesmo nome, ainda que afirme que este estudo contribui para medir o nível de colaboração do pesquisador ou da instituição a ciência.

A análise de citação, em conjunto com outros métodos, possibilita a indicação da tendência de pesquisa da área de conhecimento, sendo “utilizada como ferramenta de avaliação para subsidiar investimentos, principalmente, no setor público, uma vez que se propõe a fazer diagnóstico e prognóstico de fatos que permeiam a comunicação científica em uma determinada área ou instituição” (KROEFF, 2015, p. 45).

O fator de impacto (FI) é verificado com a análise de citação, caracterizando as publicações pela sua idade em áreas ativas, ou não, e o desenvolvimento científico, comparando o desempenho de um em relação a outros da mesma área (GARFIELD, 1979). Conforme Kroeff (2015, p. 45), o fator de impacto é também conhecido como “fator de imediatismo, que estima a concentração de citações de um artigo ou documentos publicados nos últimos anos. Mostra o quanto cada autor

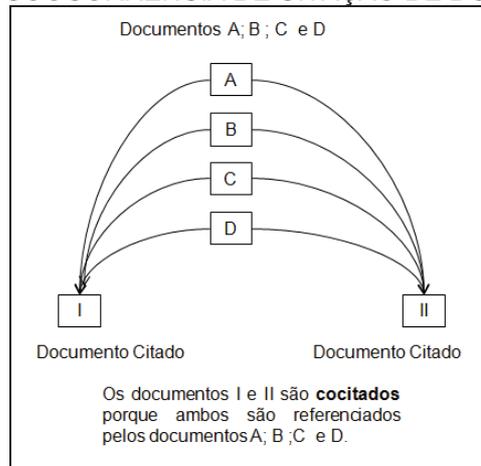
contribuiu na produção dos demais autores, ou quanto cada artigo publicado individualmente foi citado nos demais” (equação 4).

$$\text{Fator de impacto}(x) = \frac{\text{citações em } x \text{ para artigos publicados em } x - 1 \text{ e } x - 2}{\text{número de artigos publicados em } x - 1 \text{ e } x - 2} \quad (4)$$

2.4.5 Análise de Cocitação

Consiste na frequência em que dois documentos, periódicos ou autores são citados, juntos, na literatura, e se constitui em um indicador de aproximação de conteúdos dos documentos, permitindo visualizar uma área de conhecimento comum aos pesquisadores. Este indicador remete na identificação das redes de pesquisa que são referenciadas em conjunto, demonstrando a estrutura de um campo do conhecimento (figura 2).

FIGURA 2 – COCORRÊNCIA DE CITAÇÃO DE DOIS AUTORES



FONTE: Adaptada de SPINACK (2006, p. 70).

A cocitação, para Small (2004, p. 72), tem como objetivo verificar se “os cientistas concordam quanto ao que constitui a literatura relevante a priori, incluindo o que é significante naquela literatura, estão de fato definindo as estruturas de suas comunidades”. Para o autor, “a cocitação pode ser usada para estabelecer o núcleo da literatura dentro de determinado tema ou área em particular” (SMALL, 2004, p. 76).

É por meio da visualização dinâmica de uma determinada comunidade que a análise de cocitação apoiada pelo mapeamento da estrutura intelectual verifica a composição de pesquisadores que estudam um determinado assunto.

No mapeamento da literatura científica que inclui, para a análise, os documentos e os pesquisados, a intensidade de aproximação e a análise de rede – ARS, ou *Social Network Anallsys* – SNA, tem sido um instrumento de apoio e visualização gráfica.

A sociometria (berço da SNA), busca identificar as estruturas sociais das redes, através das ligações das organizações e o fluxo da informação e do conhecimento, gerados pelos autores. Dessa maneira, “o uso das cocitações como instrumento para a visualização das proximidades temáticas dos pesquisadores em determinado tema ou área, torna possível a identificação da frente de pesquisa determinada por meio da construção de redes sociais de comunicação científica” (LOPES, et al, 2013).

Assim, a cocitação deriva de citações, sendo tratada como a frequência com que dois autores, ou documentos, são citados de forma conjunta, podendo ser documentos, periódicos, autores, entre outros.

As medidas absolutas apresentam os atores centrais na rede e as medidas normalizadas demonstra a intensidade das relações. Dentre os índices relativos, citam-se o Coeficiente de Correlação de Pearson (r), Cosseno de Salton (CS) e Índice de Jaccard, em que consistem na visualização das similaridades, proximidades entre os autores.

Desta maneira, a “análise de cocitação de autores baseia-se em matrizes de valores absolutos ou relativos que mensuram a cocorrência entre os autores citados” (GRACIO; OLIVEIRA, 2013, p. 200), que indica quanto um autor está próximo ou distante dos pesquisadores citantes.

2.4.5.1 Correlação de Pearson

Índice relativo, de medida global, que “identifica e evidência a similaridade do perfil de cocitação dos dois autores cocitados em relação aos outros autores cocitados” (GRACIO, OLIVEIRA, 2015, p. 107).

Para Leydesdorff e Vaughan (2006), este indicador representa mais similaridade no estudo de cocitação de autores, “mede a tendência de associação linear de duas variáveis quantitativas, [...] a associação de pares de autores cocitados, dois a dois, considerando o conjunto de valores de cocitações de cada

um deles com os demais em estudo” (GRACIO, OLIVEIRA, 2015, p. 107). O resultado da equação matemática 5, em que:

$$r = \frac{\sum x_i \cdot y_i - \frac{\sum x_i \cdot \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right) \cdot \left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}} \quad (5)$$

x_i = a frequência de cocitação de um autor X com os demais autores;

y_i = a frequência de cocitação de um autor Y com os demais autores;

i = variando de 1 a n ; e

n = igual quantidade de autores no estudo.

O valor presente na diagonal pode ser:

- ✓ a maior frequência de cocitação do autor com os demais;
- ✓ adotar a média de cocitação de cada autor como valor da diagonal;
- ✓ usar o número de vezes que o autor foi cocitado consigo mesmo (número de pesquisas na literatura analisada que cita pelo menos dois trabalhos diferentes do autor); utilizar zeros na diagonal; ou
- ✓ considerar a diagonal como um conjunto de valores perdidos (*missing*). (GRACIO, OLIVEIRA, 2015, p. 108)

O resultado de r varia entre -1 (menos um) e 1 (um positivo). Quando positivo, indica a mesma tendência de crescimento e decrescimento de valores de cocitação dos dois autores com os demais analisados e, quando o valor r negativo, significa tendências opostas no crescimento e no decrescimento. Ou seja, para Gracio, Oliveira (2015), a medida que um dos dois autores analisados tende a ser cocitado com os demais, o outro tende a não ser cocitado com aqueles. O valor pode indicar ausência (igual a zero), correlação muito forte (menos um) ou até correlação perfeita (igual a um).

2.4.5.2 Cosseno de Salton

Índice relativo, de medida local, que “mede a proximidade de dois autores cocitados, independente do volume da literatura referenciada” (GRACIO, OLIVEIRA, 2015, p. 107), expresso pela equação matemática 6:

$$CS = \frac{\text{cocit}(x, y)}{\sqrt{\text{cit}(x) \cdot \text{cit}(y)}} \quad (6)$$

resultante da razão de frequência de cocitação de dois autores (x, y) e a raiz quadrada da multiplicação do número de pesquisas em que os autores x e y, recebidos individualmente, sendo:

cocit (x, y) = frequência de cocitação de dois autores (x, y)

cit (x) = número de pesquisas em que o autor x foi citado;

cit (y) = número de pesquisas em que o autor Y foi citado.

O presente índice tem como resultado, números variando entre 0 (zero) e 1 (um), que expressa:

[...] quanto mais próximo de um, mais intensa a similaridade temática pesquisada pelos dois autores, quanto mais próximo de zero, menos intensa é a associação entre eles, na perspectiva de autores citantes. Valor zero para CS significa ausência de cocitação entre os dois autores; valor igual a um indica que eles foram citados exatamente nas mesmas pesquisas, o que pode ser decorrente de os autores serem coautores ou por terem proximidade teórica e/ou metodológica. (GRACIO, OLIVEIRA, 2015, p. 107)

A fórmula do cosseno de Salton, frequentemente, de acordo com Hamers (1989), produz uma medida de similaridade, que representa duas vezes o número obtido pelo índice de Jaccard.

2.4.5.3 Índice de Jaccard

Define como a frequência de cocitação de dois atores, dividida pelo número total de citações recebidas por eles, (GRACIO, OLIVEIRA, 2013), tendo como representatividade matemática 7:

$$IJ = \frac{\text{cocit}(X, Y)}{\text{Cit}(X) + \text{Cit}(Y) - \text{Cit}(X \cap Y)} \quad (7)$$

cocit (X, Y) = números de trabalhos em que os autores X e Y foram citados conjuntamente (cocitados);

cit (x) = número de trabalhos em que o autor X foi citado;

cit (y) = número de trabalhos em que o autor Y foi citado;

cit ($X \cap Y$) = intersecção entre o conjunto de trabalhos em que X foi citado e o conjunto de trabalhos em que Y foi citado, o que é equivalente a cocit (X, Y).

O resultado varia de zero a um, “quanto mais próximo de um, mais similares (altamente cocitados) são os dois autores; quanto mais próximo de zero, mais distante (menos intensa) é a associação entre eles” (GRACIO, OLIVEIRA, 2013, p. 201).

Para possibilitar outra forma de interpretação, não apenas a obtida pela medida absoluta (relações de colaboração revelam um centro) e medidas relativas (que oferecem resposta a questões de intensidade das relações), Luukkonen, Persson e Silvertsen (1992) propuseram a fórmula que calcula a frequência esperada. É o resultado da aplicação da fórmula resultante da razão do total de coautorias entre o país X e Y multiplicado pelo total de coautorias da matriz, dividido pelo total de coautorias do país X possui na matriz multiplicado pelo total de coautorias do país Y possui na matriz (equação 8).

$$\frac{C_{x,y} \cdot T}{C_x \cdot C_y} \quad (8)$$

$C_{x,y}$ = total de coautorias entre o país X e Y;

T = total de coautorias da matriz;

C_x = total de coautorias o país X possui na matriz; e

C_y = total de coautorias o país Y possui na matriz.

Esta fórmula tem como objetivo possibilitar diferentes interpretações, em que “a frequência esperada é relacionada à frequência observada, [...] o índice deve ser calculado com base em uma matriz completa, ou seja, com a diagonal diferente de zero” (VANZ, STUMPF, 2010, p. 71). O resultado menor que 1 (um), indica uma colaboração menor do que a esperada. Quando maior que 1 (um), indica uma relação de colaboração mais forte do que o esperado. E, quando igual a 1 (um), indica uma colaboração observada de acordo com a esperada.

A análise de cocitação apresenta, não raro, o “fenômeno da cauda longa”, quando os resultados são apresentados em forma gráfica.

2.5 FENÔMENO DE CAUDA LONGA: UMA MANIFESTAÇÃO DAS LEIS DE POTÊNCIA

O princípio da “cauda longa” está diretamente associado ao “fato de que há quase infinitas possibilidades de consumo na internet e praticamente tudo atinge algum público, por menor que seja” (FRANCO, 2012, p. 17).

O termo “cauda longa” foi popularizado por Chris Anderson, “no slide 20 de sua palestra “Novas Regras”, apresentado para o *Chief Executive Officer (CEO) - Heed Hastings* - da Netflix” (ANDERSON, 2006, p. 10). No ano de 2004, Chris Anderson estava terminando um artigo com o mesmo título (“*The Long Tail*”), publicado na revista *Wired*, “tornando-se a matéria mais citada na revista, em todos os tempos” (AGUNE; CARLOS, 2008, não p.). Nesse artigo, há três principais observações descritas por Anderson (2006, p. 10):

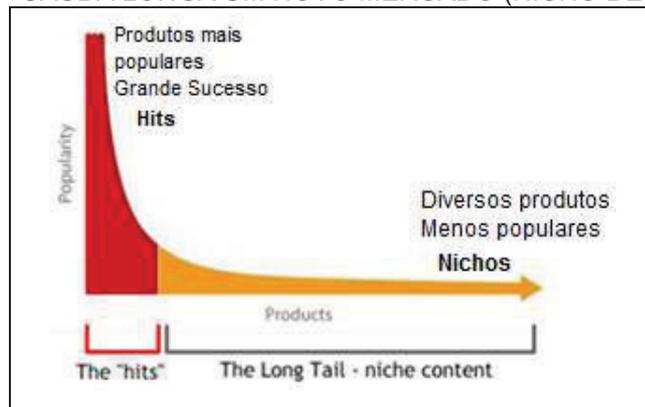
- (1) a cauda das variedades disponíveis é muito mais longa do que supomos;
- (2) ela agora é economicamente viável;
- (3) todos esses nichos, quando agregados, podem formar um mercado significativo - pareceram inquestionáveis, sobretudo com o respaldo de dados até então desconhecidos.

As observações serviram de base para diversas empresas (Amazon.com, Apple, eBay, Netflix, iTunes Store, Rhapsody, Wal-Wart) e, também, em áreas como a da comunicação de massa, das redes sociais, de marketing, entre outras; e, ainda, para outras relações que foram encontradas em partituras de músicas, jogos universitários, brinquedos, propaganda e utensílios de cozinha (ANDERSON, 2006). Assim, as empresas afastam-se do foco em um número relativamente pequeno de *hits* (produtos mais populares, de grande sucesso) para fazê-lo no grande número de nichos (diversos produtos, os menos populares), aumentando a fatia no mercado (ANDERSON, 2006, p. 53).

Os produtos “mais apreciados pela então massa de consumidores tinham prioridade com relação aos que não atingiam determinado grau de popularidade” (BELOCHIO, 2009, p. 36), sendo comercializados aqueles mais conhecidos e, os de menor reconhecimento, eram esquecidos. Anderson (2006, p. 15) acreditava ser este mecanismo, “a tirania do mínimo divisor comum, sujeitos à estupidez dos sucessos de verão e dos produtos populares”, fruto da economia, onde existe a

incompatibilidade entre oferta e demanda. Assim, o pressuposto da oferta e demanda é, em conjunto com a internet, aquele que “deu origem a um novo universo, em que a receita total de uma multidão de produtos de nicho, com baixos volumes de vendas, é igual à receita total dos poucos grandes sucessos” (ANDERSON, 2006, capa). “Nicho” é o termo que indica o que pode estar localizado na parte inferior da curva de demanda (figura 3).

FIGURA 3 – CAUDA LONGA UM NOVO MERCADO (NICHOS DE MERCADO)



FONTE: adaptada de ANDERSON (2006).

O fenômeno também é explicado devido à limitação do espaço físico em empresas – como os comércios – que concentram diversos produtos em estoque, como o caso das lojas que vendem *compact discs* (CDs), ou aquelas de alugueis de *digital versatile discs* (DVDs). Servem de exemplos, também, as livrarias e as bancas de jornal, nas quais os produtos são colocados em prateleiras que concentram diversos títulos, desde aqueles considerados *hits* e *best-sellers* (que vendem muito), quanto os que vendem pouco, ocupando o mesmo local. Ou seja, produtos de menor saída ocupando um espaço que poderia ser destinado àqueles que apresentam maior rotatividade (VALENTE, 2009). Para Valente (2009, p. 65), “o ponto chave para compreender essa discussão é exatamente a restrição apresentada pelo mundo físico”, que não há espaço suficiente para oferecer tudo a todos.

A tendência é dar destaque para os mais vendidos (representados em vermelho na figura 3), onde a popularidade é determinante, enquanto os produtos menos conhecidos, tem seu destaque diminuído gradativamente. Assim, “as pessoas se concentram obsessivamente no lado esquerdo da curva, na tentativa de adivinhar o que se aglomerará nesse pequeno espaço” (ANDERSON, 2006, p. 19).

Porém, os ditos *hits*, para Anderson (2006, capa), “já não são mais a força econômica de outrora”, sendo necessário um olhar diferenciado, o que também pode significar um olhar para a direita, em que os itens representados em laranja estão distribuídos em uma curva longa (que não chega a zero), mas que apresenta volumes significativos; ou seja, são tantos que, no conjunto, compõem um mercado significativo. Para Pontes (2014, p. 12), “em estatística, as curvas deste tipo têm o nome de “distribuições de cauda longa”, porque a “cauda” da curva é bastante mais longa em comparação com o topo da distribuição, também designada por “cabeça””.

Anderson considera que esse fenômeno é observável em variados outros produtos, e que “os consumidores em vez de avançarem como manadas numa única direção, eles agora se dispersam ao sabor dos ventos, à medida que o mercado se fragmenta em inúmeros nichos” (ANDERSON, 2006, capa). Os nichos apresentam-se como uma heterogeneidade do mercado, no intuito de oferecer mercadorias diferenciadas, para diferentes consumidores. Para Mattar e Auad (1997, p. 2), nicho é caracterizado como:

- ✓ [...] suficientemente grande e com poder de compra para gerar lucros;
- ✓ [...] tem sido desprezado ou ignorado pelo líder de mercado;
- ✓ [...] poderá ser atendido eficientemente e efetivamente pela empresa; e
- ✓ [...] poderá ser defendido contra o líder de mercado e outros competidores utilizando-se dos diferenciadores competitivos que a empresa desenvolveu junto aos seus consumidores.

A presença de nichos caracteriza-se como uma difusão de consumidores para experimentar diversos produtos de marcas não muito conhecidas, mas que, devido à grande diversidade destas, vão apresentando um valor comercial equivalente aos produtos considerados mais populares (*hits*) (ANDERSON, 2006).

Esse processo da oferta e da demanda está, para Côrtes *et al* (2008, p. 9), “[...] intimamente ligado às ferramentas da colaboratividade da internet”, e que na teoria da cauda longa está

[...] cada vez mais se afastando do foco em alguns *hits* relativamente pouco numerosos (produtos e mercados da tendência dominante), no topo da curva da demanda, e avançando em direção a uma grande quantidade de nichos na parte inferior ou na cauda da curva de demanda. (ANDERSON, 2006, p. 50)

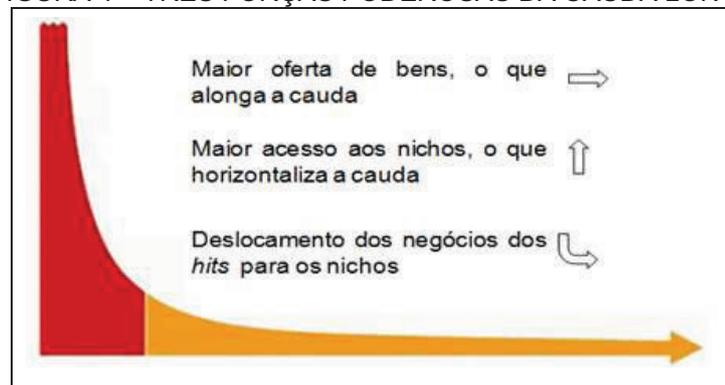
Sendo assim, a demanda definida antes como a quantidade de serviço ou produto a ser consumido (e a oferta como a quantidade de unidades que pretendem

oferecer ao mercado) é vista, por Anderson (2006, p. 50), pela ótica da “nova oferta” que, esta, se relaciona com seis condições da cauda longa:

1. [...] há muito mais nichos do que hits;
2. [...] os custos de atingir esses nichos estão caindo drasticamente [...] agora é possível oferecer muito maior variedade de produtos;
3. [...] tais “filtros” são capazes de impulsionar a demanda ao longo da cauda;
4. [...] a curva da demanda se torna mais horizontal e mais longa;
5. [...] são tantos os produtos de nicho que, como um todo, pode compor um mercado capaz de rivalizar com o dos *hits*; e
6. [...] **forma natural da curva de demanda se revela.** (ANDERSON, 2006, p. 51)

O mercado dos nichos - considerada a redução dos custos – é, segundo Anderson (2006), resultante de três “forças”, a saber (figura 4): a) a democratização da produção (maior diversidade de produtos); b) a democratização da distribuição (a internet torna mais barato o acesso a mais pessoas); e c) ligação da oferta e demanda (novos produtos disponíveis).

FIGURA 4 – TRÊS FORÇAS PODEROSAS DA CAUDA LONGA



FONTE: adaptada de ANDERSON (2006).

Com a utilização da maior oferta de bens, um maior acesso aos nichos e o deslocamento dos negócios dos *hits* para esses, apresenta-se um conjunto com novo potencial emergente de mercado da cauda longa, que “[...] é constatação, através de uma curva gráfica (denominada de curva de Pareto), que um conjunto pequeno de produtos tem uma procura bastante elevada (*hits*), enquanto um conjunto elevado de produtos tem uma procura mais reduzida (nichos)” (VALENTE, 2009, p. 66).

Há uma relação clara entre o “fenômeno da cauda longa” e a Lei de Pareto, ou “princípio do mínimo esforço”, “princípio do desequilíbrio”, que é um estudo do

economista Vilfredo Pareto. O estudo de Pareto refere-se à sociedade de sua época, onde “observou que uma grande parte da riqueza, cerca de 80%, se encontrava nas mãos de um número bastante reduzido de pessoas, 20%” (AGUNE; CARLOS, 2008, não p.). Ampliado para outros fenômenos econômicos, ficou conhecido como curva de Pareto ou regra 80/20, “podendo ser utilizado quando se deseja priorizar problemas ou causas relativas a um determinado assunto” (MARSHALL JUNIOR *et al*, 2008, p. 109).

Com a “acentuada queda nos preços do processamento, armazenamento e distribuição de dados [...] e às tecnologias de informação e comunicação” (AGUNE; CARLOS, 2008, não p.), há uma alteração no formato da curva esboçada por Pareto que, para Anderson (2006, p. 129-130), “é apenas o reconhecimento da existência, em determinado caso, da distribuição de Pareto, [...] a novidade da Cauda Longa é oferecer estímulo para não deixar dominar pela regra”.

A curva, originalmente esboçada por Pareto, torna-se mais alongada, no que Costa (2015, p. 1) afirma que, na “teoria da Cauda Longa considera que a venda de poucas quantidades, mas de muitos artigos, mais que compensa a venda dos artigos mais populares”. Desta forma,

os comerciantes que têm apenas os produtos que se concentram na Cabeça da curva logo descobrem que os clientes querem mais do que lhes está sendo oferecido. Já os comerciantes que dispõem os produtos que se espalham pela Cauda da curva também constataam que os clientes não têm ideia de onde começar suas escolhas. Eles não conseguem encontrar pontos de apoio no mercado, pois não estão familiarizados com nada daquilo. A importância de oferecer itens tanto da Cabeça quanto da Cauda é a possibilidade de começar no mundo já conhecido pelos clientes: produtos familiares que exploram um espaço definido. (ANDERSON, 2006, p. 147)

Observa-se, conforme afirmado por Pontes (2014, p. 38), “que a Cauda Longa pode ser composta por várias “mini-caudas”, pelas características que continuam a ter o mesmo formato”, havendo a possibilidade de “que os *hits* não mais dominem a sociedade e o comércio tanto quanto no século passado, mas ainda exercerão impacto inigualável” (ANDERSON, 2006, p. 147).

Por mais que a cabeça da cauda seja representada pelos produtos mais populares, “a maioria dos clientes está na cauda, [...], o ponto fraco do varejo tradicional. As lojas perdem negócios simplesmente porque seu modelo econômico não lhes dá condições para explorá-los” (ANDERSON, 2006, p. 162).

Um paralelo com as vendas pode ser traçado com “a necessidade de prestar mais atenção à cauda longa da ciência, em que há um crescente aumento na quantidade de cientistas e nos dados gerados por cada cientista” (HEIDORN, 2008, p. 281, tradução nossa). A cauda longa apresenta estruturas sociais que, com o auxílio da rápida expansão das ferramentas de informação, será possível gerenciá-las (HEIDORN, 2008, p. 281, tradução nossa).

Para Mueller (1995, p. 66), “a ciência moderna vem crescendo de maneira acelerada e constante. O crescimento da ciência interessa-nos porque exerce enorme influência nas atividades de comunicação entre cientistas: o volume de pesquisas e o de literatura científica crescem juntos”. Assim, observa-se que os números de publicações produzidas por ano, em uma determinada área, e os autores que publicam pela primeira vez na mesma área, mostravam que áreas específicas cumprem o estágio de crescimento: lento no início, depois exponencial, depois linear e, finalmente, em declínio gradual (CRANE, 1971).

Por mais que ocorra a presença do estágio de crescimento, “a comunicação científica representa as atividades associadas à produção, disseminação e uso da informação, desde o momento em que o cientista concebe uma ideia para pesquisar até a aceitação dos resultados” (ALELUIA, 2009, p. 38). A aceitação “de comunicação entre os cientistas não exige a revisão prévia dos documentos que por ele circulam, permitindo o fluxo de estudos, protótipos ainda em desenvolvimento” (VARELA, *et al.*, 2011).

Essa estrutura “e a evolução técnica poderia aumentar significativamente a disponibilidade e valor de dados dos cientistas de forma individual” (HEIDORN, 2008, p. 280, tradução nossa). No entanto, para Heidorn (2008), a maior parte dos projetos que os cientistas vêm desenvolvendo é em conjunto com dois ou três alunos de pós-graduação. Os dados científicos são resultados alcançados com a coleta, preservação e reutilização de dados gerados, que constitui a teoria científica, podendo os projetos ser organizados ao longo de um eixo de grande projeção (direita) à pequena (esquerda) (HEIDORN, 2008). No lado esquerdo do eixo, estão os projetos maiores, com dezenas de cientistas e grande quantidade de dados, e no lado direito, os projetos menores. Para Heidorn (2008, p. 280, tradução nossa), “a grande área do lado direito da curva é a cauda longa dos dados científicos. Estes dados não são facilmente encontrados e apresentam pouca utilização ou não são preservados”.

Os dados que não são facilmente encontrados por potenciais usuários,

[...] são publicações produzidas no âmbito das instituições governamentais, organizações privadas, instituições culturais, acadêmicas e ONGs, tais como relatórios de atividades e científicos, bem como os documentos de reuniões, congressos e outros eventos, muitos de natureza científica, mas também econômicos e de negócios, que não são publicados, ou o são em pequenas tiragens, bem como, não fazem parte do sistema convencional de controle e registro, tais como periódicos-índices e bibliografias (VARELA, *et al.*, 2011)

Essas publicações apresentam como denominação dados escuros³ ou literatura cinzenta⁴. Tais termos foram cunhados em 1978, na Inglaterra, no encontro com bibliotecários britânicos, para discutir sobre a aquisição e controle bibliográfico. De acordo com Heidorn (2008), os dados “tornam-se escuros, porque ninguém está prestando atenção”. Assim, cabe “o desafio para a política científica em desenvolver instituições e práticas, tais como repositórios institucionais que tornem esses dados úteis à sociedade” (HEIDORN, 2008, p. 281, tradução nossa).

O repositório institucional (RI), “representa proposta que amplia a visibilidade e acesso à produção científica, atuando como instrumento de ação política e assumindo relevante papel na discussão sobre direitos autorais e impactos na ciência e tecnologia” (TARGINO, *et al.*, 2014, p. 118). Para Leite e Costa (2006, p. 212), “o desenvolvimento de repositórios institucionais tem se dado, amplamente, no contexto de universidades”.

Assim, o acesso ao conjunto de dados “pode ter um impacto muito significativo na ciência. Parece provável que a ciência transformadora é mais provável de acontecer a partir da cauda do que da cabeça” (HEIDORN, 2008, p. 282, tradução nossa).

Como na cauda há uma explosão de dados, cabe medir esse conhecimento com a utilização de ferramentas que visam como resultado recolher informações a partir dos dados (McAfee; Brynjolfsson, 2012). No entanto, a cauda pode apresentar uma rica complexidade da ciência, havendo a necessidade de estudar os cientistas, quanto a uma melhor gestão dos dados, ferramentas e infraestrutura (BORGMAN; *et. al.*, 2016).

³ Dados inéditos, mas de experimentos fracassados (HEIDORN, 2008, p. 285).

⁴ Caracteriza-se segundo sua apresentação não convencional, portanto, não sendo publicada e distribuída através dos canais normais do parque editorial, que o seu difícil acesso acarreta sérios problemas para a coleta, armazenagem e recuperação [...] recebe várias denominações: literatura cinzenta, literatura fugitiva, literatura invisível, informal e mesmo efêmera (POBLACIÓN, 1992, p. 243-244).

A cauda longa da ciência “é um terreno fértil para novas ideias” (PECK, 2008), em “que os cientistas têm de lidar aparentemente com a variabilidade infinita no formato de dados diferentes, sendo desafio para novas práticas e novas tecnologias para o tratamento dos dados” (HEIDORN, 2008, p. 282, tradução nossa). A cauda corresponde a diferentes dados reunidos com menor nível de planejamento e com baixa visibilidade de outros cientistas. Esses mesmos compõem uma proporção de dados escuros, que tende a ser heterogêneo, podendo fazer parte de diversos projetos, ou seja, utilização dos mesmos dados, mas realizados em projetos diferentes. Os dados dispersos, mas semelhantes na cauda, podem contribuir em grandes realizações, e no conhecimento acumulado, quando tratados adequadamente (HEIDORN, 2008, p. 285, tradução nossa).

Dessa forma, os dados pertencentes à cauda longa são recursos importantes para a ciência no desenvolvimento da teoria e avaliação (HEIDORN, 2008, p. 285, tradução nossa). Já aqueles que pertencem à cabeça, são dados mais acessíveis, homogêneos e volumosos, com fácil armazenamento quando a base é estruturada.

Para Heidorn (2008, p. 289, tradução nossa), “a diferença entre os dados na cabeça e os dados na cauda é o impacto que o acesso a eles tem sobre a carreira das pessoas”, quadro 6.

QUADRO 6 – CARACTERÍSTICAS NOS DADOS QUE COMPÕEM O GRÁFICO SEGMENTADO NA CABEÇA E NA CAUDA

CABEÇA	CAUDA
Homogêneo e volumoso	Heterogêneo
Coleta de dados (mecanizada)	Coleta de dados (necessita que a coleta seja realizada individualizada - item por item - e manual)
Procedimentos uniformes (facilidade no armazenamento dos dados nas bases estruturadas)	Procedimento excepcional (cada base de dados tem uma forma para cadastrar os dados para serem armazenados)
Livre acesso	Protegido (necessita permissão – senha – para acesso à base de dados)
Dados mais acessíveis	Cada projeto pode ter seu próprio formato de dados

FONTE: adaptado de HEIDORN (2008).

Sendo assim, como existe uma maior frequência na concentração dos dados na cabeça do gráfico, e algum movimento para a cauda. Cabe então:

[...] abordar algumas das perguntas dos dados escuros na cauda longa:
a) Quão extensa é a cauda?

- b) Qual é a área sob a cauda longa?
- c) Os dados da cauda e da cabeça são escuros?
- d) Como podemos determinar o valor dos dados escuros?
- e) O que é diferente entre a cauda da ciência e a cabeça da ciência?
- f) Quais os dados mais propensos a contribuir para a ciência transformadora? (HEIDORN, 2008, p. 296-297, tradução nossa).

Como na cauda o tratamento e “quantidade de informação disponível cresce vertiginosamente e surgem novos comportamentos e sentimentos decorrentes deste crescimento” (RIBEIRO, 2014, p. 97). A coleta de “grande volume de dados com formato variado [...] que precisam ser geridos, possibilitará a resolução de problemas que nem sabíamos que existia” (RIBEIRO, 2014, p. 98). Assim, “a necessidade de vencer o desafio, reunindo e analisando fontes de diversas naturezas, deu origem às pesquisas que nos levaram ao tema “*Big Data*”” (RIBEIRO, 2014, p. 99).

Com a quantidade expressiva de informação e a necessidade de vencer desafio, cabe verificar o histórico do tema “Big Data”, que está “enraizada, principalmente, no início de 1970. Uma das iniciativas foi em 1974, com o livro de Peter Naur, “*Concise Survey of Computer Methods*”, uma pesquisa de processamento de métodos em ampla gama de aplicações” (PARK; LEYDESDORFF, 2013, p. 756, tradução nossa). Na década de 1990, o termo foi associado ao desenvolvimento de software para grandes conjuntos de dados. Com o lançamento das revistas *Data Science Journal* (2002), *Journal of Data Science* (2003), *EPJ Data Science* (2012), e *Journal of Big Data* (2013), com foco em grande quantidade de dados, concentrou-se os estudos do termo “*Big Data*” a todo o espectro da vida e ciências (PARK; LEYDESDORFF, 2013).

Em conjunto com os dados científicos publicados para que a ciência avance, exige-se entendimento de todo o processo de produção e comunicação da informação (COSTA; CUNHA, 2014). Essa informação científica e tecnológica, para Costa e Cunha (2014, p. 194), “culminou com o surgimento da *e-science*”, que “altera fundamentalmente a maneira com que os cientistas realizam seu trabalho, as ferramentas que usam, os tipos de problemas que abordam, a natureza da documentação e a publicação que resulta da sua pesquisa” (MEDEIROS, CAREGNATO, 2012, p. 315) .

E-science, para Medeiros e Caregnato (2012, p. 312), “pode ser entendida como a infraestrutura que visa permitir que cientistas e pesquisadores possam ter acesso a dados científicos primários distribuídos, utilizando acesso remoto a esses

conteúdos”. Hoje, a *e-science*, “mais do que nunca, é uma atividade colaborativa” (CORDEIRO; *et.al*, 2013, p. 76), com projetos que integram grades computacionais, mas

[...] ainda longe da realidade de muitas instituições de pesquisa do mundo todo. A justificativa mais frequente para esse fato é a falta de recursos físicos, humanos e financeiros para criar e manter o ambiente computacional mínimo que é um requisito para a prática da *e-science*. Entretanto, o surgimento de uma nova tecnologia renovou as esperanças da comunidade científica em tornar a *e-science* mais “acessível”. Essa tecnologia é a computação em nuvem (CORDEIRO; *et.al*, 2013, p. 78).

Com o acesso aos dados científicos, seja pelas bases de dados ou repositórios institucionais, cabe analisar se “os artigos de alto impacto não são encontrados, necessariamente, nos periódicos de alto impacto” (SEGLEN, 1997; SUN & GILES, 2007 *apud* HEIDORN, 2008). Assim, o “prestígio do periódico, a reputação dos autores, das instituições, a relevância da pesquisa de campo, [...] taxas de citação e de impacto da revista são fatores para a avaliação na pesquisa relacionados à ciência” (SEGLEN, 1997, p. 498, tradução nossa). Dada a crescente utilização do fator de impacto da revista, cabe ao indicador a avaliação de pesquisa, a saber:

- a) o uso do fator de impacto da revista esconde a diferença nas taxas de citação dos artigos,
- b) o fator de impacto da revista não está relacionado à qualidade científica dos seus artigos;
- c) o fator de impacto depende do campo de pesquisa: fatores de alto impacto são prováveis em revistas que abrangem grandes áreas de pesquisa básica com uma rápida expansão, mas de curta duração na literatura que utilizam muitas referências por artigo; e
- d) a taxa de citação de artigo determina o fator de impacto da revista, e não vice versa (SEGLEN, 1997, p. 498, tradução nossa).

Os autores nem sempre publicam os resultados da pesquisa em periódicos de maior impacto. “Alguns autores, além do fator de impacto do periódico, verificam os fatores: área do periódico, a equidade e a agilidade no processo editorial, a probabilidade de aceitação e os custos de publicação” (SEGLEN, 1997, p. 499, tradução nossa).

Com isso, “recuperar informações das bases de dados de documentos continua sendo uma tarefa desafiadora, em que não é ponderada a contagem de citação com precisão” (SUN; GILES, 2007, p. 605, tradução nossa). Apesar de ser

utilizado na avaliação acadêmica, de acordo com Sun e Giles (2007, p. 606), “a contagem de citações de um trabalho individual não reflete a sua estrutura, diferente em cada disciplina, tendo os artigos de alto impacto e aqueles de baixo impacto tratado da mesma maneira na contagem”.

Além da contagem de citação, a qualidade do local onde o documento é publicado é importante para os usuários avaliarem o valor desse documento, uma vez que, nesses locais, há uma revisão, por pares, para a publicação (SUN; GILES, 2007, p. 606, tradução nossa). As contagens de citações são indicadores sólidos de desempenho da pesquisa (GARFIELD, 1972) e “o local da publicação remete ao acesso aberto, que pode ser definido pelo autor, quanto ao seu artigo” (HARNAD, et al., 2009, p. 38).

Assim, o crescente volume, de cientistas e dados gerados, contribui para o processo de análise na identificação do caminho percorrido (seja pelo pesquisador, periódico ou acontecimento social), cujos objetivos, ao serem atingidos, estarão disponíveis à sociedade em algum ponto da curva de itens publicados (quer na cabeça ou na parte longa da cauda de distribuição dos itens).

3 PROCEDIMENTOS METODOLÓGICOS

Neste capítulo, é tratada a caracterização da pesquisa e descrevem-se os procedimentos e recursos que são utilizados para o seu desenvolvimento; os instrumentos e técnicas de coleta; e o processo de tratamento e análise dos dados; ou seja, evidencia-se o caminho percorrido para a construção do estudo.

O objetivo desta análise é explorar as características das manifestações de “cauda longa”, derivadas de distribuições bibliométricas no domínio de auditoria da informação. Essa pesquisa evidencia-se quanto aos objetivos em exploratória: é conduzida quanto ao procedimento em documental, e tem abordagem quantitativa.

3.1 COLETA DE PUBLICAÇÕES NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO

As bases de dados para a coleta dos materiais foram escolhidas, considerando-se os seguintes critérios: disponibilidade de acesso através de acordo entre a Universidade Federal do Paraná (UFPR) e a Comissão de Aperfeiçoamento de Pessoal do Nível Superior (CAPES); facilidade de exportação dos dados bibliográficos dos artigos selecionados no formato de arquivo texto comum (txt); e, quantidade de metadados (autor, título, periódico, idioma, tipo de publicação, palavra(s)-chave e ano de publicação) disponíveis para a exportação dos dados.

A partir desses critérios, o universo de bases consultadas foi assim composto: EBSCOhost (via portal de periódicos da CAPES), Google Scholar, ISI/Web of Science^(C), SciELO, e SCOPUS (via portal de periódicos da CAPES). Cada base apresenta peculiaridades no processo de coleta de informações, descritas no quadro 7.

A seleção dos itens apresenta, como arco temporal, o período compreendido entre 1970 – indicado por Pestana (2014), como sendo um dos marcos do crescente interesse pela auditoria de informação – e 1º de setembro de 2016, data da realização da pesquisa.

QUADRO 7 – PECULIARIDADES NA DEFINIÇÃO DE BUSCA NAS BASES DE DADOS SELECIONADAS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 – 2016/SETEMBRO

Termos de busca	Base de Dados	Campo de busca	Especificidades na busca	Itens Recuperados (número de itens)
"Information Audit" OR	EBSCOhost	Pesquisa avançada ⁵	Limitação de resultado (texto completo e referências disponíveis)	26
"Auditoría de la Información" OR	Google Scholar	Pesquisa avançada ⁶	Excluídas patentes e citações	162
"Auditoria de Informação"	ISI/Web of Science ^(C)	Pesquisa básica ⁷	Principal coleção do <i>Web of Science</i> TM	59
	SciELO	Método integrado ⁸	Todos os itens ⁹	7
	SCOPUS	Pesquisa avançada	filtros: <i>article title, Abstract, Keywords</i> ¹⁰ .	92
TOTAL				346

FONTE: O autor (2016)

Quando a base de dados não apresentou uma funcionalidade de exportação de todos os itens recuperados, em formato .txt, cada item foi exportado, individualmente, garantindo-se a coleta dos metadados de título, autor, periódico e ano de publicação. Este procedimento foi utilizado nas bases EBSCOhost e Google Scholar.

A partir dos 346 itens recuperados (dados brutos), procedeu-se a eliminação dos itens duplicados. Este processo ocorreu, inicialmente, em cada base de dados, com a união dos arquivos¹⁰, no programa MS-Excel(T) – tabela 1.

Após a eliminação dos itens duplicados na própria base (EBSCOhost e Google Scholar com onze itens, cada; e SCOPUS com dois itens) foi realizada a

⁵ Selecionado "Academic Search Premier – ASP" e mais 28 bases (*Business Source Complete, Historical Abstracts with Full Text, Academic Search Premier, CINAHL with Full Text, Computers & Applied Sciences Complete, Dentistry & Oral Sciences Source, eBook Collection (EBSCOhost), Educational Administration Abstracts, ERIC, Human Resources Abstracts, Information Science & Technology Abstracts with Full Text, MEDLINE, MEDLINE Complete, Newswires, Public Administration Abstracts, Regional Business News, RILM Abstracts of Music Literature (1967 to Present only), RIPM – Retrospective Index to Music Periodicals, SocINDEX with Full Text, SPORTDiscus with Full Text, MathSciNet via EBSCOhost, FSTA – Food Science and Technology Abstracts, American Doctoral Dissertations, 1933-1955, E-Journals, eBook Academic Collection (EBSCOhost), International Pharmaceutical Abstracts e Education Source*).

⁶ Incluído no campo de busca o vocábulo "tudonotítulo" antes das palavras designadas como termo de busca.

⁷ Inserido no campo de busca os termos da pesquisa.

⁸ Com base no filtro de pesquisa, a procura é realizada nas bases de dados e fontes de informação que apresentam um resultado único (integrado).

⁹ Composto por: ano, autor, financiador, periódico, resumo e título.

¹⁰ As bases ISI/Web of Science^(C), SciELO e SCOPUS têm o recurso de eliminação de artigos em duplicidade no momento da exportação da busca. Já as bases EBSCOhost e Google Scholar necessitam da inserção de cada artigo, com os respectivos metadados, individualmente.

união de todos os itens, de cada base, no editor de planilhas da *Microsoft*, composto de 322 itens (Tabela 1).

TABELA 1 – QUANTIDADE DE ITENS QUE COMPÕEM O UNIVERSO DA PESQUISA APÓS A ELIMINAÇÃO DOS DUPLICADOS NA PRÓPRIA BASE NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 – 2016/SETEMBRO

Base de Dados	Recuperados (a)	Duplicados na própria base (b)	Universo inicial da pesquisa (a - b) = c
EBSCOhost	26	11	15
Google Scholar ¹¹	162	11	151
ISI/Web of Science ^(c) ¹²	59	-	59
SciELO	7		7
SCOPUS	92	2	90
Soma	346	24	322

FONTE: O autor (2016)

Com a constituição da nova base de dados (excetuadas as repetições), existiu a necessidade de excluir os itens duplicados na população (90 itens), sendo definido como critério a permanência do item que está na base de dados com maior quantitativo de informação (metadados), conforme cada base, e as informações disponíveis, a saber:

- a) EBSCOhost: título, autor(es), nome do periódico, idioma, ano e tipo de publicação;
- b) Google Scholar: título, autor(es), nome do periódico, idioma, ano de publicação;
- c) ISI/Web of Science ^(c): título, autor(es), nome do periódico, idioma, ano, tipo de publicação, palavra(s)-chave(s), país de publicação, e-mail de contato dos pesquisadores, referências utilizadas, volume e número da publicação.
- d) SciELO: título, autor(es), nome do periódico, idioma, ano, tipo de publicação, país de publicação e endereço eletrônico para localização do item;
- e) SCOPUS: título, autor(es), nome do periódico, idioma, ano, tipo, país de publicação e endereço eletrônico para localização do item.

¹¹ Indexa apenas os nomes dos primeiros e últimos autores (AGARWAL, 2016, p. 304).

¹² Inclui apenas as iniciais do primeiro nome do autor (Jane Smith, John Smith e Jorge Smith são identificados como Smith-J) (AGARWAL, 2016, p. 304).

As bases com o maior quantitativo de detalhes de metadados são: ISI/Web of Science^(C), SCOPUS, SciELO, EBSCOhost e Google Scholar, nesta ordem.

Assim, observa-se que, dos 26 itens recuperados da base EBSCOhost, onze deles foram excluídos, em virtude da duplicidade na própria base. A mesma quantidade foi excluída da base Google Scholar e dois itens foram eliminados da SCOPUS, totalizando 24 itens descartados. Os duplicados na união de todas as bases pesquisadas, somam 90 itens e o resultado final (universo final da pesquisa), compôs-se de 232 itens (três itens EBSCOhost, 120 itens Google Scholar, 59 itens ISI/Web of Science^(C), um item da SciELO e 49 itens da SCOPUS), elencados na tabela 2.

TABELA 2 – QUANTIDADE DE ITENS QUE COMPÕEM O UNIVERSO DA PESQUISA, POPULAÇÃO DOS ITENS COMPLETOS RECUPERADOS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 – 2016/SETEMBRO

Base de Dados	Universo inicial da pesquisa (c)	Duplicados na população (d)	Universo final da pesquisa (c – d)	Itens completos recuperados
EBSCOhost	15	12	3	2
Google Scholar	151	31	120	67
ISI/Web of Science ^(C)	59		59	57
SciELO	7	6	1	1
SCOPUS	90	41	49	46
Soma	322	90	232	173

FONTE: O autor (2016)

Em um segundo momento, devido às características do estudo, a obtenção do texto completo (incluídas as referências) foi um elemento crítico e, na impossibilidade de se coletarem todos os itens do universo (limites de acesso na instituição pelo pesquisador), aplicou-se um cálculo para a determinação da amostra (o grau de confiança foi de 95% e a margem de erro de cinco pontos percentuais), conforme a equação 9 (BRUNI, 2010), em que:

$$n = \frac{Z^2 \times p \times q \times N}{e^2 \times (N - 1) + Z^2 \times p \times q} \quad (9)$$

N = número de elementos da população

n = número de elementos da amostra

Z² = nível de confiança escolhido, expresso em número de desvio-padrão

p = porcentagem em que o fenômeno se verifica

q = porcentagem complementar

e^2 = erro máximo permitido

Com o cálculo da amostra, verificou-se que 144,9 itens (mínimo) deveriam ser objeto de análise. Foi possível recuperar 173 itens em texto completo (incluindo-se as referências) e a análise foi efetivada com 19,39% a mais de itens exigidos com o cálculo da amostra.

Os metadados extraídos das bases de dados dos itens foram: ano de publicação, veículo de publicação, local, idioma, periódico e autor; com os quais se constituiu, em um único arquivo carregado no programa MS-Excel(T), que foram submetidos às técnicas bibliométricas selecionadas.

3.2 TÉCNICAS BIBLIOMÉTRICAS SELECIONADAS PARA O ESTUDO

Para a submissão do material às técnicas bibliométricas, procedeu-se uma análise de potenciais ferramentas, entre as quais, o Bib Excel, o Cite Space, o In-Spire, entre outros (APÊNDICE 1).

Aplicados os critérios de facilidade na importação, a facilidade de manuseio do sistema e a geração de arquivo com as características e os resultados, definiu-se que os programas Stikis, Unicet e VOSviewer atendiam às necessidades da pesquisa. Para o uso do Stikis, é necessário gerar um único arquivo contendo os dados a serem estudados.

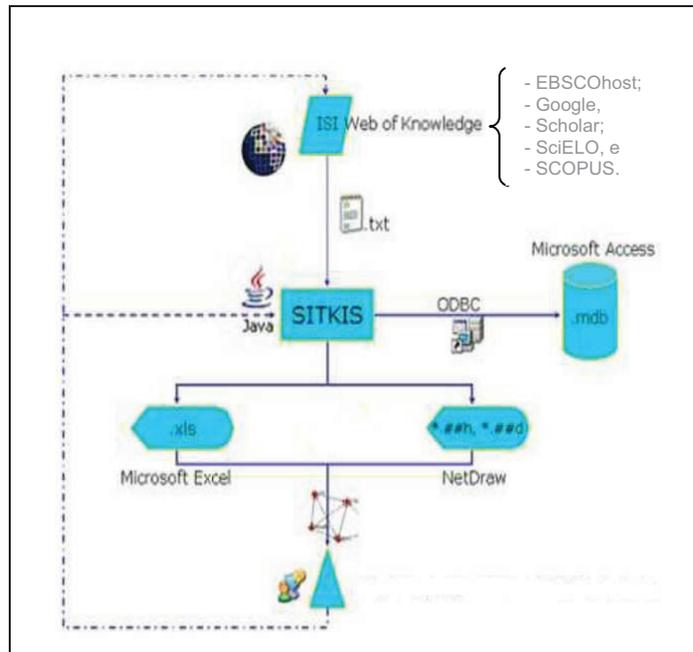
As funcionalidades disponíveis no Stikis são: *file*, *options*, *citations*, *authors*, *articles*, *help*, *draw*, *algorithms*, assim descritas:

- ✓ *file*: importação do arquivo .txt, gerado na base de dados consultada;
- ✓ *options*: escolha da base de dados cadastrada no *Microsoft Access* pelo usuário;
- ✓ *citations*: referência por artigo, frequência de referências por ano e rede de cocitação;
- ✓ *authors*: estatística, rede de citação e rede de palavras mais citadas;
- ✓ *articles*: instituições, países e autor por artigo, relação entre as palavras chave;
- ✓ *help*: versão do programa e contato;

- ✓ *draw*: geração de rede (função não cadastrada¹³); e
- ✓ *algorithms*: densidade da rede do grupo, relação da rede, estatística do grupo.

Posterior à inserção do arquivo no sistema, e do processamento, gera-se um arquivo, em formato .txt, compatível com o *software* Microsoft Excel. O criador do programa indica a necessidade de utilizar o Sitkis, em conjunto com o *Microsoft Access*, o *Microsoft Excel*, o *Unicinet* e o *NetDraw* (figura 5).

FIGURA 5 – FLUXO DE PROCESSAMENTO DE DADOS NO PROGRAMA SITKIS



FONTE: Adaptado de PONTES (2012, p. 1)

O processo de tratamento dos dados, portanto, compreende a(s) escolha(s) da(s) base(s), a inserção dos termos da pesquisa, bem como os filtros, e então, a extração do(s) arquivo(s) da(s) base(s) de dados, sendo a geração do arquivo em formato .txt. Este arquivo é importado no Sitkis (agregando-se o programa Java Enterprise 1.3 ou superior) e uma base de dados no *Microsoft Access* cadastrada (esta é fornecida pelo próprio programa, necessitando a configuração na máquina de execução). Após a escolha da função desejada, no programa Sitkis, este gera um arquivo, em formato .txt, que pode ser importado para o programa *Excel* (para a

¹³ Função não disponível, necessitando da instalação de outro programa para a realização da geração de rede.

criação de gráficos e tabelas), e pelo *NetDraw* (para a geração da rede de colaboração), com a extensão *##h* e/ou *##d*¹⁴.

Após a importação do arquivo gerado na base de dados, no programa *Stikis*, e o processamento das informações, foram escolhidas as funções no *software*:

- a) campo *citations*: referência por artigo, frequência de referência por ano;
- b) *authors*: estatística; e
- c) *articles*: países e autor por artigo.

Cada uma destas funções gerou arquivos em formato *.txt*, para serem submetidos e formatados em planilhas e gráficos, no programa MS-Excel^(T), relativos à contagem de frequência para os seguintes campos: distribuição temporal, veículo de publicação, local de publicação, idioma e autoria.

Inicialmente, definiu-se a análise da evolução quantitativa da produção; onde foram publicados os eventos; quais os veículos de publicação; o idioma de maior predomínio; a produtividade dos autores (Lei de Lotka), a ocorrência (e posicionamento) abrangendo palavras-chave (Leis de Zipf e o ponto de transição de Goffman). Neste último caso, agregaram-se, na contagem de frequências, as variações plural/singular (p. ex: *developing countries* e *developing country*) e as traduções (p. ex: auditoria do conhecimento, gestão da informação, gestão do conhecimento e gestão da informação).

Para Lotka, o número de autores que fazem “n” contribuições sobre um determinado campo científico é, aproximadamente, $1/n^2$ daqueles que fazem uma só contribuição; sendo a proporção deste de, aproximadamente, 60% do conjunto de autores (Lei do quadrado inverso) (LOTKA, 1926). Zipf explorou a premissa na qual a frequência com que determinada palavra ocorre no texto define sua posição e ordenação/classificação de modo decrescente (ZIPF, 1935). Portanto, há uma relação entre a frequência com que uma determinada palavra ocorre e a sua posição, ou seja: o produto da ordem de série de uma palavra (“r”) – sua ordem decrescente de ocorrência –, pela frequência de ocorrência (“f”) é, aproximadamente, uma constante (“c”) – o que se aplica apenas em palavras de alta frequência (GUEDES, BORSCHIVER, 2005), e é conhecido como primeira lei de Zipf.

¹⁴ Extensão do arquivo para a importação no *software UNICET*, na geração de gráficos utilizados na visualização de redes sociais.

De acordo com Pao (1978), Goffman – ao observar a primeira lei de Zipf – admitiu como hipótese que, nessa primeira zona de ocorrências de palavras, na região de transição, estariam as palavras de maior conteúdo semântico, denominando-a de Ponto de Transição (T).

Ao verificarem-se caudas longas, em especial a quantidade de itens publicados pela quantidade de veículo de publicações, dimensionou-se as relações: pelo tipo de publicação, idioma e palavras-chave. Após as relações, obteve a comparação dos resultados encontrados na cauda, com os apresentados na cabeça.

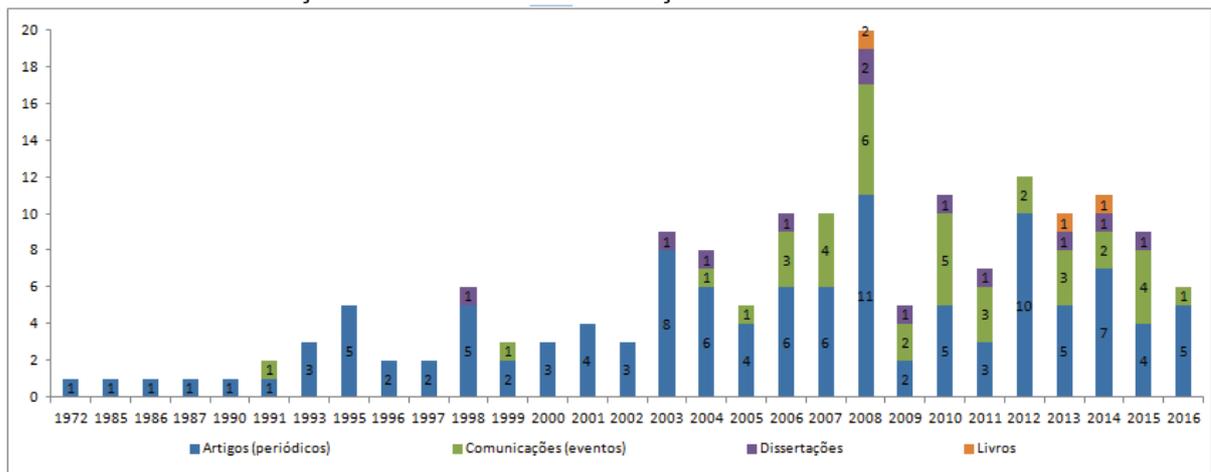
Além da produtividade dos autores, o posicionamento das palavras-chave, e o indicativo de cauda longa, verificou-se a formação de redes entre os pesquisadores (autores e coautores), investigando-se o quantitativo de citações e cocitações, levando-se em consideração o calculado. Para a composição dos itens de análise, na realização dos cálculos (Pearson, Salton e Jaccard), foram considerados, dos itens recuperados nas bases, aqueles referentes à artigos e comunicações em eventos (117 e 39, respectivamente), totalizando 156 itens, o que representa 7,66% a mais de itens exigidos com o cálculo da amostra. Alguns materiais recuperados nas bases de dados SCOPUS e Google Scholar (capítulo de livro dissertação, documentos de trabalho e livros) não disponibilizam o texto completo ou as referências, tendo sido, então, excluídos do cálculo. Na identificação dos autores mais citados, elaborou-se uma lista única, com todas as referências utilizadas nas publicações, mesmo quando estas eram autocitações. Utilizou-se ordem decrescente de citações e o corte realizado para os pesquisadores que foram citados em, pelo menos, vinte e cinco publicações (artigos e comunicações em eventos), totalizando o grupo-alvo de dez pesquisadores. A partir disso, construiu-se a matriz de dados absolutos (10 X 10). Para os cálculos (Pearson, Salton e Jaccard), utilizou-se o *Microsoft Excel* e, na realização das redes, *Unicinet* e *NetDraw*.

Desta forma, justifica-se abordar tanto as questões elencadas por Heidorn (2008) no item Cauda Longa, quanto a identificação de quais os dados que estão presentes na cabeça (dominado por poucos itens) e na cauda (pulverizados, representados por vários itens).

4 ANÁLISE DOS DADOS E RESULTADOS

A distribuição dos 173 itens referentes ao recorte temporal entre 1970 a 2016, estão apresentados no gráfico 3. Em termos quantitativos, de 1972 a 2002, verifica-se um comportamento estável, variando entre um a seis itens, condizente com o estágio de crescimento: lento no início (CRANE, 1971). Em 2008, há o ápice de publicações no período, com 21 itens (onze em periódicos, seis em eventos – dois nos Estados Unidos (América do Norte), dois na Malásia (Ásia), um na Argentina (América do Sul) e um na Alemanha (Europa) – duas dissertações e dois livros).

GRÁFICO 3 – EVOLUÇÃO DA PRODUÇÃO CIENTÍFICA NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO POR TIPO DE PUBLICAÇÃO – 1970 - 2016/SETEMBRO



FONTE: O autor (2016)

A maioria dos itens publicados no ano de 2008 descreve o processo de exploração, identificação e apresentação da auditoria de informação em companhias de eletricidade, utilidade pública, clínicas de saúde, tecnologia da informação (TI) e bibliotecas em instituições de ensino, abrangendo discussões sobre a compreensão, usabilidade, conceito e metodologia da auditoria de informação.

O número de itens publicados em eventos, no período de 1991 a 2016, foi de 39, o que representa, aproximadamente, 1,56 (itens/ano). A comparação das comunicações (eventos), com a localidade (continente) de publicação, demonstra na temática auditoria de informação, a publicação em todos os continentes – tabela 3. A partir do ano de 1998, há a publicação, além de em artigos de periódicos e comunicações em eventos, de dissertações e livros. De 2004 em diante, as

comunicações em eventos, mais especificamente, são publicadas nos continentes da África (um item), América do Norte (doze itens), América do Sul (dois itens), Ásia (oito itens), Europa (treze itens) e Oceania (um item), não apresentando ausência de publicação (comunicações) no período de 2004 a 2016. Os continentes que obtiveram maior número de itens publicados foram a América do Norte e a Europa, ambas com treze, o que representa 33,33% do total.

TABELA 3 – PUBLICAÇÃO POR ANO E CONTINENTE, EM COMUNICAÇÃO EM EVENTOS, NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 A 2016/SETEMBRO

Ano	África	América do Norte	América do Sul	Ásia	Europa	Oceania	Total
1991		1					1
1999						1	1
2004					1		1
2005					1		1
2006		1	1		1		3
2007		1		3			4
2008		2	1	2	1		6
2009		1			1		2
2010	1	1		1	2		5
2011		1		1	1		3
2012		1			1		2
2013		1			1	1	3
2014		1			1		2
2015		2		1	1		4
2016					1		1
Total	2,56%	33,33%	5,13%	20,51%	33,33%	5,13%	100,00%

FONTE: O autor (2016).

Em relação à distribuição geográfica dos 173 itens, registra-se que tal informação estava disponível, nas bases de dados somente para 57,80% da amostra (100 itens). A tabela 4 mostra o volume de produção por país. Nota-se que a maior produção foi originada nos Estados Unidos (13,87%), Inglaterra (8,67%), China (5,20%) e menor volume, os países que publicaram apenas um item, com representação de 0,58%, dos quais se pode citar: Argentina, Canadá e outros. O Brasil apresentou-se com 1,16% (dois itens) de participação. Os estudos brasileiros, identificados nas bases de dados são: Carvalho, A. V.; Navarro, M. E. (2016), publicado no *Journal of Librarianship and Information Science*; e Ferreira, R. de M. (2006), publicado no VII Encontro de Bibliotecários e Documentalistas da Justiça do Trabalho.

TABELA 4 – LOCAL DE PUBLICAÇÃO DA PRODUÇÃO CIENTÍFICA NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Local de Publicação	Ocorrências	%	Local de Publicação	Ocorrências	%
Não especificado na base de dados	73	42,20%	Antígua e Barbuda	1	0,58%
Estados Unidos	24	13,87%	Argentina	1	0,58%
Inglaterra	15	8,67%	Canadá	1	0,58%
China	9	5,20%	Colômbia	1	0,58%
Alemanha	6	3,47%	Equador	1	0,58%
Cuba	6	3,47%	Grécia	1	0,58%
Romênia	5	2,89%	Holanda	1	0,58%
Austrália	4	2,31%	Índia	1	0,58%
Escócia	4	2,31%	Japão	1	0,58%
Espanha	4	2,31%	Lituânia	1	0,58%
África do Sul	3	1,73%	Nova Zelândia	1	0,58%
Malásia	3	1,73%	Polônia	1	0,58%
Brasil	2	1,16%	Portugal	1	0,58%
Suíça	2	1,16%			
			Total	173	100,00%

FONTE: O autor (2016).

Pesquisou-se, também, o veículo de divulgação mais frequente, identificando-se a primazia de publicações na forma de artigos de periódicos (117 itens, 67,63%), seguido de comunicações em eventos (39 itens, 22,54%), dissertações (12 itens, 6,94%), livros (quatro itens, 2,31%) e documentos de trabalho¹⁵ (um item, 0,58%). O artigo é “um tipo de publicação que permite maior visibilidade” (SANTAREM, 2010, p. 66).

Quanto ao predomínio do idioma, o inglês é o idioma presente em 136 itens, seguido do espanhol, com dezesseis itens, sendo “a pluralidade de idiomas no campo da produção científica diminuído, tendo em conta a proeminência do Inglês” (ROSA; ALVES, 2010, p. 3). Justifica-se a predominância do idioma quando “a atividade de pesquisa é internacional [...] a maior parte dos periódicos de grande prestígio é redigido em inglês” (MEADOWS, 1999, p. 168). A hegemonia do idioma inglês, conforme Price (1971, p. 257, tradução nossa), “constitui pouco mais da metade da produção filosófica e científica do mundo”. Os idiomas, inglês e espanhol, representam 22,22% do total de idiomas (nove), que publicaram 87,86% dos itens analisados. Os idiomas, em conjunto com os itens publicados, demonstram o princípio de Pareto (87,86% dos itens analisados foram publicados em 22,22% dos idiomas, e os 77,78% dos idiomas representam 12,14% dos itens analisados).

¹⁵ São documentos ligados à elaboração de relatórios ou pareceres (suporte de trabalho desenvolvido pelo auditor).

TABELA 5 – IDIOMA DA PRODUÇÃO CIENTÍFICA NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

	Idioma	Frequência	Frequência %	
22,22%	Inglês	136	78,61%	} 87,86%
	Espanhol	16	9,25%	
	Chinês	5	2,89%	
	Coreano	5	2,89%	
	Português	5	2,89%	
	Alemão	2	1,16%	
	Ucraniano	2	1,16%	
	Filandês	1	0,58%	
	Polonês	1	0,58%	
	Total	173	100,00%	

FONTE: O autor (2016)

No que se refere ao número de autores, por item, os dados empíricos evidenciam uma variação entre um e sete. Consideram-se, para fins deste estudo, como “autores principais”, aqueles que aparecem como primeiro autor e, “coautores”, aqueles que aparecem da segunda (2ª) posição de autoria em diante. Na tabela 6, observa-se que, 43,93% dos itens foram produzidos por um único autor, significando baixa colaboração.

TABELA 6 – NÚMERO DE ITENS PUBLICADOS SEGUNDO O NÚMERO DE AUTORES PRINCIPAIS E COAUTORES NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

nº de autores por item	nº de itens publicados	%	% acum.
1	76	43,93%	43,93%
2	50	28,90%	72,83%
3	24	13,87%	86,71%
4	10	5,78%	92,49%
5	9	5,20%	97,69%
6	2	1,16%	98,84%
7	2	1,16%	100,00%
Total	173		

FONTE: O autor (2016).

Na tabela 7, mostra-se a quantidade de itens produzidos em relação ao número de autores e coautores, organizado pelo ano de publicação. A autoria da primeira publicação, dentro o período analisado (1970 a setembro de 2016), é do ano de 1972 e a última, no ano de 2016. Observa-se o número de diferentes documentos produzidos apenas pelo autor principal, em colaboração com um, dois, três autores e, assim, sucessivamente. Até o ano de 1993, as publicações foram realizadas individualmente ou em conjunto com mais um autor. No ano mencionado,

há publicações com dois e quatro autores. E, no período posterior, compreendido de 1995 a 2002, com publicações de até dois autores. Entre 2003 e 2008, começa o processo de colaboração entre os autores (principal e coautor) que, no ano de 2008, atinge o ápice de itens publicados – onze periódicos (artigos); seis comunicações (eventos); duas dissertações e dois livros - sendo: uma publicação com sete autores, outra com cinco, cinco publicações com três autores, oito com dois e seis apenas com um autor, que conforme Meadows (1999, p. 110), “a tendência geral é no sentido de crescente colaboração em todas as áreas”.

TABELA 7 – FREQUENCIA DE PUBLICAÇÃO DOS AUTORES PRINCIPAIS E COAUTORES, POR ANO, NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Ano	Número de autores							Total
	1	2	3	4	5	6	7	
1972	1							1
1985	1							1
1986	1							1
1987	1							1
1990	1							1
1991	1	1						2
1993		2		1				3
1995	4	1						5
1996	2							2
1997	2							2
1998	4	2						6
1999	3							3
2000	1	2						3
2001	4	1						5
2002		3						3
2003	5	2	1	1				9
2004	4	2	1		1			8
2005	1	1	2		1			5
2006	2	3	1	2	2			10
2007		3	2	2	1	1	1	10
2008	6	8	5		1		1	21
2009	3	1				1		5
2010	8	3						11
2011	6		1					7
2012	4	2	3	3				12
2013	4	2	1	1	2			10
2014	4	3	3		1			11
2015	1	4	4					9
2016	2	4						6
Total	76	50	24	10	9	2	2	173

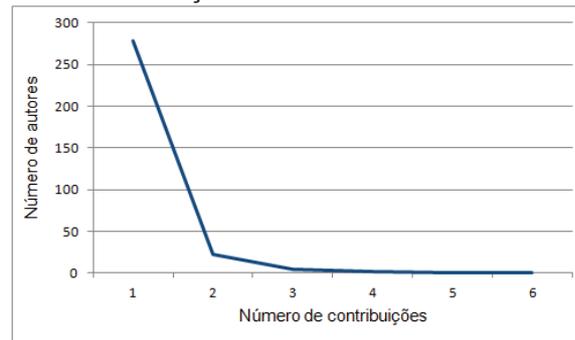
36,42% (para os anos 2003-2008)

19,08% (para os anos 2012-2014)

FONTE: O autor (2016).

A Lei de Lotka apresenta sua distribuição gráfica em forma de J inverso, com uma cauda de pequenos produtores (URBIZAGASTEGUI, 2008). O gráfico 4, que ilustra a cauda em forma de J inverso, é condizente com o padrão de queda acentuada dos autores que produzem um artigo para os que produzem dois, e tende à linearidade, conforme aumenta o número de publicações por autor, descrito por Urbizagastegui (2008).

GRÁFICO 4 – CAUDA LONGA DE AUTORES *VERSUS* ARTIGOS PRODUZIDOS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO



FONTE: O autor (2016).

A literatura, na forma de determinar a autoria, apresenta três possibilidades de contagem: a direta (crédito apenas ao autor principal), a completa (atribui crédito igual a todos os autores), e a fracionada (crédito fracionado entre os autores). Optou-se, inicialmente, pela análise das três formas, na tabela 8, e obteve-se o total de 152 autores que, em conjunto, produziram 173 itens. Desse total, 91,45% participaram na produção de um só item, mas em conjunto, elaboraram 80,35% dos itens, com média de um artigo por autor. No entanto, a produtividade média, do total de autores, é de 1,14 itens por autor.

TABELA 8 – PRODUTIVIDADE DE AUTORES PRINCIPAIS (CONTAGEM DIRETA) NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Nº. de autores principais	Nº. de itens por autores principais	% de autores principais	Nº. de itens	% de itens
139	1	91,45	139	80,35
9	2	5,92	18	10,40
2	3	1,32	6	3,47
1	4	0,66	4	2,31
1	6	0,66	6	3,47
152		100,00	173	100,00

FONTE: O autor (2016).

Os autores que se destacam, em relação aos 152, foram: Steven Buchanan (com seis publicações como autor principal), seguido de Susan Henczel (com quatro), e Sulfieza Mohd Drus e Xu Fe (ambos com três), que representam o primeiro *cluster*. O segundo *cluster*, de nove pesquisadores, contribui com duas publicações, cada um (tabela 9). O primeiro e segundo *clusters* representam 19,65% das publicações. A “cauda longa” é composta de 139 pesquisadores com uma única publicação.

TABELA 9 – AUTORES PRINCIPAIS MAIS PRODUTIVOS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Autor Principal	Quant. de Autor Principal	Quant.	%	Acumulado
Steven Buchanan	1	6	3,47%	3,47%
Susan Henczel	1	4	2,31%	5,78%
Sulfeeza Mohd Drus	1	3	1,73%	7,51%
Xu Fe	1	3	1,73%	9,25%
Cristina Soy i Aumatell	1	2	1,16%	10,40%
Fábio Dacêncio Pereira	1	2	1,16%	11,56%
Gao, P.	1	2	1,16%	12,72%
Gloria Ponjuán Dante	1	2	1,16%	13,87%
González Guitián, M. V.	1	2	1,16%	15,03%
Griffiths, P.	1	2	1,16%	16,18%
Mihalache D. Arsenie-Samoil	1	2	1,16%	17,34%
Redman, C.	1	2	1,16%	18,50%
Xavier Giné	1	2	1,16%	19,65%
Diversos Autores	139	139	80,35%	100,00%
Total	152	173	100,00%	

FONTE: O autor (2016).

Quanto à forma de contagem completa, em que é atribuído crédito a cada um dos autores envolvidos nos itens (tabela 10), foram observados 311 autores, sendo 89,71% participantes na produção de um só artigo e, em conjunto com o grupo de produtores únicos, foi responsável por 78,03% dos trabalhos produzidos. Na faixa de autores com um só artigo, a média é de, apenas, 0,48 artigos por autor. Porém, a produtividade média geral encontrada foi de 0,56 por autor.

TABELA 10 – PRODUTIVIDADE DE AUTORES PRINCIPAIS E COAUTORES (CONTAGEM COMPLETA) NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Nº. de autores principais	Nº. de itens por autores principais	% de autores principais	Nº. de itens	% de itens
279	1	89,71	135	78,03
23	2	7,40	19	10,98
5	3	1,61	9	5,20
2	4	0,64	4	2,31
1	5	0,32	0	0,00
1	6	0,32	6	3,47
311		100,00	173	100,00

FONTE: O autor (2016).

Na contagem ajustada/fracionada (crédito fracionado entre os autores), tendo sido atribuído valor um para cada item e, sendo este, dividido pela quantidade de autores (por até sete autores) – tabela 11. Estes quatorze autores, principais e

coautores, publicaram dois itens, onde 4,50% dos autores principais e coautores produziram, aproximadamente, 1,16% dos itens.

TABELA 11 – PRODUTIVIDADE DE AUTORES PRINCIPAIS E COAUTORES (CONTAGEM AJUSTADA/FRACIONADA) NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

	Nº. de autores principais e coautores	Nº. de contribuições por autores principais e coautores	% de autores principais e coautores	Nº. de itens	% de itens
96,14%	14	0,143	4,50	2,00	1,16
	12	0,167	3,86	2,00	1,16
	33	0,20	10,61	6,60	3,82
	22	0,25	7,07	5,50	3,18
	70	0,33	22,51	23,31	13,47
	7	0,45	2,25	3,15	1,82
	70	0,50	22,51	35,00	20,23
	1	0,65	0,32	0,65	0,38
	1	0,70	0,32	0,70	0,40
	1	0,75	0,32	0,75	0,43
	2	0,83	0,64	1,67	0,96
	1	0,90	0,32	0,90	0,52
	65	1,00	20,90	65,00	37,57
	3	1,50	0,96	4,50	2,60
	1	1,75	0,32	1,75	1,01
	4	2,00	1,29	8,00	4,62
	2	2,50	0,64	5,00	2,89
	1	3,00	0,32	3,00	1,73
1	3,5	0,32	3,5	2,02	
	311		100,00	173	100,00

FONTE: O autor (2016).

A percentagem de autores únicos (91,45% para a contagem direta, 89,71% para a contagem completa e 96,14% para a contagem ajustada/ fracionada), para Urbizagastegui (2008, p. 97), “parece ser característica das disciplinas em expansão e desenvolvimento na procura de sua institucionalização”.

Para uma visualização, quanto à mudança do critério de cálculo, o quadro 8 representa os três critérios, com a respectiva colocação de cada um dos autores, dando destaque para Steven Buchanan, Susan Henczel e Cristina Soy i Aumatell, sempre presentes entre as cinco primeiras colocações, compondo os autores de maior relevância sobre o assunto estudado.

Depois de atribuída a contagem (direta, completa, ajustada/fracionada), os autores, que obtiveram o mesmo valor, foram classificados em ordem alfabética. Desta forma, a colocação atribuída tem como o limite máximo de autor principal: 152, e autor principal e coautor: 311, sendo demonstrados, apenas os sessenta primeiros.

QUADRO 8 – PRODUTIVIDADE DE AUTORES PRINCIPAIS E COAUTORES (CONTAGEM AJUSTADA/FRACIONADA) NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

CONTAGEM								
Direta		Completa		Ajustada / Fracionada				
Col.	Autores Principais	Valor	Col.	Autores Principais e Coautores	Valor	Col.	Autores Principais e Coautores	Valor
1	Steven Buchanan	6	1	Steven Buchanan	6	1	Steven Buchanan	3,500
2	Susan Henczel	4	2	Forbes Gibb	5	2	Susan Henczel	3,000
3	Sulfeeza Mohd Drus	3	3	Susan Henczel	4	3	Cristina Soy i Aumatell	2,500
4	Xu Fe	3	4	Xu, C.	4	4	Forbes Gibb	2,500
5	Cristina Soy i Aumatell	2	5	Cristina Soy i Aumatell	3	5	Gloria Ponjuán Dante	2,000
6	Fábio Dacêncio Pereira	2	6	Siti Salbiah Mohamed Shariff	3	6	Griffiths, P.	2,000
7	Gao, P.	2	7	Sulfeeza Mohd Drus	3	7	Mihalache D. Arsenie-Samoil	2,000
8	Gloria Ponjuán Dante	2	8	Xu Fe	3	8	Redman, C.	2,000
9	González Guitián, M. V.	2	9	Yu, F.	3	9	Siti Salbiah Mohamed Shariff	1,750
10	Griffiths, P.	2	10	Aliza Abdul Latif	2	10	González Guitián, M. V.	1,500
11	Mihalache D. Arsenie-Samoil	2	11	Dong Soo Kim	2	11	Peter Griffiths	1,500
12	Redman, C.	2	12	Edward David Moreno Ordonez	2	12	Sulfeeza Mohd Drus	1,500
13	Xavier Giné	2	13	Fábio Dacêncio Pereira	2	13	Amita S. Pradhan	1,000
14	Alexandre Pinheiro	1	14	Gao, P.	2	14	Anu Hakkarainen-Kiri	1,000
15	Alonso Perez-Soltero	1	15	Gloria Ponjuán Dante	2	15	Barton, J.	1,000
16	Amita S. Pradhan	1	16	González Guitián, M. V.	2	16	Bloomfield, R.	1,000
17	Anu Hakkarainen-Kiri	1	17	Graham Roberston	2	17	Budzak, D.	1,000
18	Bacigalupo, R.	1	18	Griffiths, P.	2	18	David R. Worlock	1,000
19	Barton, J.	1	19	Hee Wan Kim	2	19	De Stricker, U.	1,000
20	Bloomfield, R.	1	20	Izyana Ariffin	2	20	Dennis Steven Erthanegara	1,000
21	Booth, A.	1	21	J. A. Boon	2	21	Désirée Lamoral	1,000
22	Budzak, D.	1	22	Mawaddah Mohamad Nadzir	2	22	Dubois, C. P. R.	1,000
23	Carvalho, Andrea V.	1	23	Mihalache D. Arsenie-Samoil	2	23	Edward David Moreno Ordonez	1,000
24	Chandra Kant Sharma	1	24	Peter Griffiths	2	24	Edwin Michael Cortez	1,000
25	Cheung, N. W.	1	25	Rafael Keenan Mazer	2	25	Elena Cabal García	1,000
26	Correia, Z. P.	1	26	Redman, C.	2	26	Fábio Dacêncio Pereira	1,000
27	Craig Wright	1	27	Sang-jiin Lee	2	27	Fryc, L. M.	1,000
28	David R. Worlock	1	28	Shen Yue	2	28	Gabriella Signàs	1,000
29	De Stricker, U.	1	29	Xavier Giné	2	29	Gilman, P. L.	1,000
30	Dennis Steven Erthanegara	1	30	Yue, G.	2	30	Graham Roberston	1,000
31	Désirée Lamoral	1	31	Zhang, D.	2	31	Gretchen M. Ptacek	1,000
32	Ding, L.	1	32	Adolfo Alberto Vanti	1	32	Guenther, K.	1,000
33	Dubois, C. P. R.	1	33	Adriean Parlog	1	33	Huan, V. T	1,000
34	Edwin Michael Cortez	1	34	Akash Kumar Singh	1	34	Hugh Parkes	1,000
35	Elena Cabal García	1	35	Alexandre Pinheiro	1	35	Ibrarhim Ramjaun	1,000
36	Elisabeth Jokobsen	1	36	Alonso Perez-Soltero	1	36	Isabel García Soler	1,000
37	Ellis, D.	1	37	Amita S. Pradhan	1	37	J. A. Boon	1,000
38	Emily A. Langley	1	38	Anay Jiménez Rivas	1	38	Jesús Tramullas	1,000
39	Evangelos Alexopoulos	1	39	André Schneider	1	39	Jin, H.	1,000
40	Evan-Wong, S.	1	40	Andrew Fearn	1	40	John Niklasson	1,000
41	Fei Yu	1	41	Andrew J. Palmer	1	41	Jones, H.	1,000
42	Florin Dobre	1	42	Ann Kritin H. Johansen	1	42	Kilzer, R.	1,000
43	Franck, L. S.	1	43	Anu Hakkarainen-Kiri	1	43	Knott, David	1,000
44	Fraudzi, M. A	1	44	Arief Deswandi	1	44	Lian van Uden	1,000
45	Fryc, L. M.	1	45	Aziz, S. J. S	1	45	Marian Stoica	1,000
46	Fu, X.	1	46	Azizah Abd Rahman	1	46	McMurdo, G.	1,000
47	Gabriella Signàs	1	47	B. Parolin	1	47	Michael Fanning	1,000
48	Gallegos, F.	1	48	Babis Theodoulidis	1	48	Mitea, N.	1,000
49	Gilman, P. L.	1	49	Bacigalupo, R.	1	49	Munoz, JVR	1,000
50	Gretchen M. Ptacek	1	50	Barker, R.	1	50	Ntuen, Celestine A.	1,000
51	Guenther, K.	1	51	Barrulas, M. J.	1	51	Oleksandr Sherstiuks	1,000
52	Guimarães, F. S.	1	52	Barton, J.	1	52	Olivia Pestana	1,000
53	Hammersley, J. S.	1	53	Bastião, L.	1	53	Pete Dalton	1,000
54	Hanneri Botha	1	54	Bertrand, W. E.	1	54	Petra Petroff	1,000
55	Haro Acosta Mauricio Orlando	1	55	Bloomfield, R.	1	55	Petuchovaite, R.	1,000
56	Hepworth, Mark	1	56	Booth, A.	1	56	Rebecka Sheffield	1,000
57	Huan, V. T	1	57	Brian Freedman	1	57	Regina de Marco Ferreira	1,000
58	Hugh Parkes	1	58	Bridgart, R.	1	58	Richie, R. D.	1,000
59	Ibrarhim Ramjaun	1	59	Budzak, D.	1	59	Rittberger, M.	1,000
60	Isabel García Soler	1	60	Burwell, B.	1	60	Robinson, Charles B.	1,000

FONTE: O autor (2016).

Na análise foram considerados apenas os autores principais, pela contagem direta de publicações superiores a um item, a saber: Steven Buchanan com seis publicações, Susan Henczel com quatro e Sulfeeza Mohd Drus com três.

Steven Buchanan, em destaque na cor vermelha, obteve a primeira colocação entre os 152 autores principais. Considerando-se, na contagem completa, que ele não desenvolveu nenhum trabalho em coautoria (obtem o mesmo valor), continuando ainda na primeira colocação e, na contagem ajustada/ fracionada, como desenvolveu seus trabalhos apenas com mais um autor (cinco publicações) e uma, sozinho, obteve o valor de 3,5, ficando, também, na primeira posição.

Susan Henczel, na cor laranja, na contagem direta, obteve a segunda colocação, com quatro publicações. Como não desenvolveu nenhum item em coautoria, na contagem completa cai uma posição, ficando na terceira colocação, com o valor de quatro. Na contagem ajustada/fracionada, como duas das suas publicações são desenvolvidas com mais um autor, e duas publicações individuais, o valor atribuído é três, ficando na segunda colocação.

Sulfeeza Mohd Drus, com três publicações como autor principal, obteve, na contagem direta, a terceira colocação e a sétima, na completa. Como desenvolveu seus itens apenas com mais um colaborador, obteve, na contagem fracionada/ ajustada, o valor de 0,5 para cada item, que multiplicado por três itens, ficou na décima segunda colocação.

No tocante às palavras-chave, aplicaram-se a Lei de Zipf e o Ponto de Transição de Goffman (T). Inicialmente, no registro na base de dados de cada item, foram coletadas as palavras-chave, às quais, após elencadas agregaram-se, na contagem de frequências, às palavras que possuíam o mesmo significado/traduições, mas estavam repetidas, no plural ou singular, de modo que não existissem significados repetidos que comprometessem as análises posteriores.

A palavra mais citada: *audit information*, foi excluída da análise, em virtude de ser a palavra-chave para a coleta dos itens, em que “[...] a lei de Zipf gerou o princípio do menor esforço, em que uma mesma palavra é utilizada muitas vezes e que as palavras mais usadas indicam o assunto do documento” (BEUREN; SILVA, 2014, p. 40).

Obteve-se como resultado, o mapeamento de 612 palavras-chave, com 425 delas sendo diferentes (quadro 9) e, 336 (54,90%) ocorrem, apenas, uma única vez. Assim, os 45,10% das palavras-chave utilizadas nos itens analisados foram citadas

duas ou mais vezes. As palavras *information management* e *knowledge management*, de frequência 24 (3,92%) e 15 (2,45%), respectivamente, “preconiza que um pequeno número de palavras é usado mais frequentemente” (BEUREN; SILVA, 2014, p. 40).

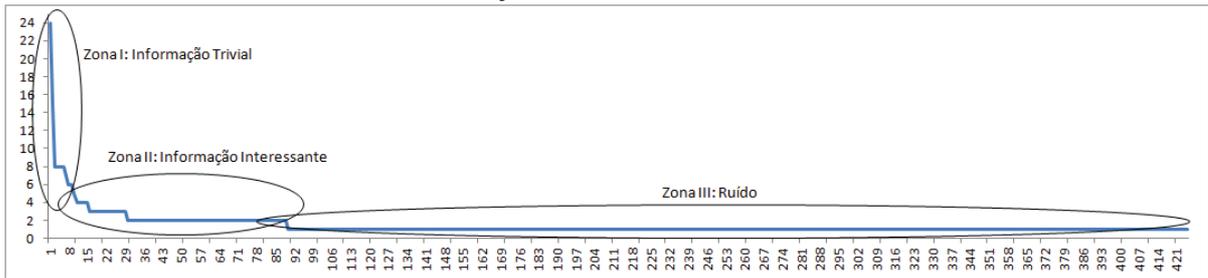
QUADRO 9 – FREQUÊNCIA DAS PALAVRAS-CHAVE NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Frequência (número de vezes)	Quantidade de palavras-chave diferentes	Quantidade de palavras-chave	Agregado %	Palavras-chave
1	336	336	54,90	...
2	60	120	19,61	accident prevention; accounting information system; action research; architectural design process; audit process; Australia; automated decisions; badness information; compliance; computer system; data audit; decision audit; decision process; decision quality; disease; education ; electricity industry; enterprise information management; female; game theory; health; health services research ; information asymmetry; information behaviour; informations hantering; Information policy; information process; information processing; information professionals; informations granskning; informations resurer; infrastructure; internal audit; job description; latin America; length of stay; libraries; management audit; management information system; medical audit; neonatal nursing ; newborn nursing; nurse; nurse clinicians; nutrition surveys; patient satisfaction; political factors; politics; postnatal care; pregnancy; questionnaire; research methodology; resource allocation; risk management; security of data; statistics; systemic approach; text character; work.
3	14	42	6,86	article; audit planning; auditing, finalcial audit; health service; health services needs and demand, infant, information assets; information audit; information security; intranets; organization; strategis planning; United Kingdom
4	5	20	3,27	audit; decision making; family health; information strategy; records management
5	1	5	0,82	methodology
6	2	12	1,96	patient education; information resource management
7	2	14	2,29	newborn; knowledge audit
8	3	24	3,92	information service; human, information system
15	1	15	2,45	knowledge management
24	1	24	3,92	information management
Total	425	612	100,00	

FONTE: O autor (2016).

As palavras quantificadas evidenciam predominância de temas centrados nas questões de gestão da informação, que totalizam 3,92% do total de palavras-chave. Há, também, ocorrência de outras palavras ligadas à gestão, ou ferramentas de controle, análise e método. Conforme Guedes, Borschiver (2005), o comportamento das palavras de alta frequência (informação trivial) e de baixa frequência (ruído) resulta em duas extremidades e o meio da frequência (informação interessante), denominada de “ponto de transição”, expressa no gráfico 5, o comportamento das palavras-chave analisadas.

GRÁFICO 5 – AS TRÊS ZONAS DE DISTRIBUIÇÃO NA CURVA DE ZIPF NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO



FONTE: O autor (2016).

Ao aplicar-se a fórmula de Zipf e o Ponto de Transição de Goffman (T), ordenaram as 612 palavras-chave de acordo com a frequência de ocorrência (“f”), classificando-as em ordem decrescente de frequência de ocorrência e, assim, obtendo-se uma ordem de série “r”. Executando-se a multiplicação da ordem de série (“r”) e a frequência de ocorrência (“f”), obteve-se a constante da primeira lei de Zipf (“c”) – quadro 10.

QUADRO 10 – FREQUÊNCIA DAS PALAVRAS-CHAVE PARA APLICAÇÃO DA PRIMEIRA LEI DE ZIPF NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Qtd. de palavras-chave classificadas na ordem de série	Ordem de série “r”	Frequência de Ocorrência “f”	Constante da Primeira Lei de Zipf (r.f=c)
1	1	23	23
1	2	14	28
3	3	8	24
2	4	7	28
2	5	6	30
1	6	5	30
5	7	4	28
14	8	3	24
60	9	2	18
336	10	1	10
425			

FONTE: O autor (2016).

As 425 palavras-chave foram classificadas nas áreas de conhecimentos, estabelecidas pela CAPES, que agrupa as ciências e especialidades científicas e tecnológicas em nove áreas, a saber: ciências da saúde; ciências agrárias; ciências biológicas; ciências exatas e da terra; ciências humanas; ciências sociais e aplicadas; engenharias; linguística, letras e artes; e multidisciplinares (APÊNDICE 2).

Assim, estabelece a quantidade de palavras-chave que constituem cada área do conhecimento (tabela 12), com destaque para a área do estudo, ciências sociais e aplicadas (70,35%) e multidisciplinar (12,47%).

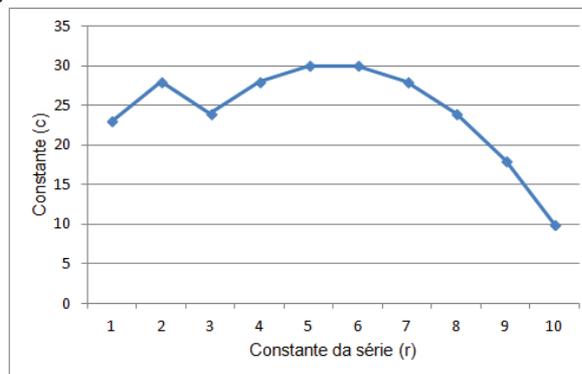
TABELA 12 – FREQUÊNCIA DAS PALAVRAS-CHAVES POR ÁREA DE CONHECIMENTO ESTABELECIDADA PELA CAPES NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Área de conhecimento	Freq.	%
Ciência da Saúde	38	8,94%
Ciências Agrárias	2	0,47%
Ciências Biológicas	3	0,71%
Ciências exatas e da terra	16	3,76%
Ciências Humanas	4	0,94%
Ciências Sociais e Aplicadas	299	70,35%
Engenharias	5	1,18%
Linguística, Letras e artes	5	1,18%
Multidisciplinar	53	12,47%
Total	425	100%

FONTE: O autor (2016).

Observa-se, no gráfico 6, o ponto máximo 30 (nas ordens de série, cinco e seis), que representa o produto da ordem de série pela frequência de ocorrência. À esquerda, no gráfico 6, situam-se as palavras com menor ordem de série (4, 3, 2, 1) e as de maior frequência (7, 8, 14, 23). À direita, situam-se as palavras-chave com maior ordem de série (7, 8, 9, 10), mas com menor frequência de ocorrência (4, 3, 2, 1). Verifica-se uma parábola e a tendência decrescente, resultante da ordem de série multiplicada pela frequência de ocorrência.

GRÁFICO 6 – APLICAÇÃO DA PRIMEIRA LEI DE ZIPF NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO



FONTE: O autor (2016).

Uma vez aplicada a primeira Lei de Zipf, calculou-se, na sequência, o Ponto de Transição de Goffman (T) – equação 10.

$$n = \frac{-1 \sqrt{1+8 \times (336)}}{2} \quad n = 50,85 \quad (9)$$

Assim, conclui-se que, sendo $n = 50,85$, a 51ª palavra e expressão mapeada, realizada pela contagem da lista apresentada no quadro 9, pela

frequência da maior para a menor, é o Ponto de Transição (T) de Goffman. Logo, a palavra e expressão mapeada na 51ª colocação, foi “*health service research*”, “indica, aproximadamente, as ordens de série nas quais devem encontrar as palavras significativas de um texto” (LIMA; MAIA, 1973, p. 99), representada pela nona ordem de série (quadro 10).

As palavras-chave que apresentam a mesma frequência de ocorrência foram classificadas utilizando-se as áreas de conhecimento estabelecida pela CAPES (tabela 13). As palavras-chave que não pertencem à área de maior frequência (Ciências Sociais e Aplicadas) (tabela 12) estão concentradas na região de maior conteúdo semântico (GUEDES, 2012).

TABELA 13 – PALAVRAS-CHAVE CLASSIFICADAS POR ÁREA DE CONHECIMENTO ESTABELECIDADA PELA CAPES NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Ciência da Saúde	Ciências Biológicas	Ciências exatas e da terra	Ciências Humanas	Ciências Sociais e Aplicadas	Engenharias	Linguística, Letras e artes	Multidisciplinar
accident prevention disease health health services research neonatal nursing newborn nursing nurse nurse clinicians nutrition surveys patient satisfaction postnatal care pregnancy	female	Australia computer system latin america statistics	developing country	accounting Information system architectural design process audit process badness information compliance data audit decision audit decision process decision quality enterprise information management game theory information asymmetry information behaviour Information policy information process information processing information professionals informations granskning informations resurer informationshantering infrastructure internal audit job description libraries management audit management Information system medical audit political factors politics resource allocation risk management security of data work	automated decisions electricity industry	education	action research length of stay questionnaire research methodology systemic approach text character

FONTE: O autor (2016).

Na tabela 13, observa-se o aparecimento de diversas palavras-chave na área de saúde, fruto da aplicabilidade de auditoria de informação em clínicas e regionais de saúde nas quais o objetivo se voltava à necessidade de se analisar em que condições a informação poderia contribuir para o objetivo do negócio.

Quando aos veículos de publicação, identificaram-se 141 títulos, com uma média de 1,23 (173/141) itens por periódico. As revistas *International Journal of Information Management* e *Revista Cubana de Informacion en Ciencias de la Salud*

publicaram, juntas, 8,67% do total. 124 periódicos (71,68%) publicaram somente um item, indicando que, na cabeça, existe o predomínio de poucos, e um alongamento diversificado na cauda (tabela 14).

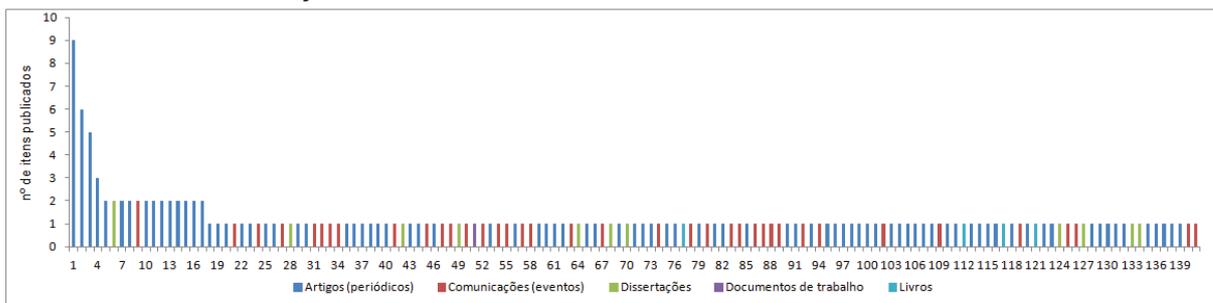
TABELA 14 – NÚMERO DE VEÍCULOS DE PUBLICAÇÕES, SEGUNDO A QUANTIDADE DE ITENS PUBLICADOS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Nº de veículo de publicações	nº de itens publicados	Quant. Publicada	%	% acum.
1	9	9	5,20	5,20
1	6	6	3,47	8,67
1	5	5	2,89	11,56
1	3	3	1,73	13,29
13	2	26	15,03	28,32
124	1	124	71,68	100,00
Total	141	173		

FONTE: O autor (2016).

As relações veículos de publicações e palavras-chave serão compostas pela cauda com 124 veículos de publicações. A cauda longa dos veículos de publicações (124 veículos) apresenta comportamento estável até o ano de 2003, com aumento no ano de 1995 e seu ápice, no ano de 2008, com 13 publicações (gráfico 7).

GRÁFICO 7 – EVOLUÇÃO DA PRODUÇÃO CIENTÍFICA NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO



FONTE: O autor (2016)

A predominância na publicação e nos eventos no continente da Europa e da América do Norte, com treze publicações cada um que, em conjunto, representam 70,27% do total, ou seja, 33,33% dos continentes correspondem a cerca de 70% dos itens e publicações compostas na cauda longa, e 29,73% dos itens representam 66,67% dos continentes, tendo a evidência do princípio de Pareto. O evento *Knowledge Management International Conference*, que ocorreu no ano de 2008, apresenta duas publicações, e está presente na cabeça da cauda.

Quanto à publicação do autor mais produtivo – Steven Buchanan –, que publicou todos os seus itens (seis artigos) no periódico *International Journal of Information Management*, que é o que apresenta maior frequência de publicações. O segundo, o terceiro e o quarto autores mais produtivos (respectivamente, Susan Henczel, Sulfeeza Mohd Drus e Xu Fe) publicaram os seus trabalhos em periódicos que não estão na cabeça dos veículos de publicações. Estes veículos tiveram apenas uma publicação.

Já os autores mais cocitados (Steven Buchanan; Forbes Gibb) como realizaram seus itens em conjunto, estão publicando no periódico de maior frequência (tabela 16). Os autores Susan Henczel; Xu, F e Cristina Soy publicaram em veículos pertencentes à cauda, com apenas uma publicação.

Observa-se, no quadro 11, a colocação, por ordem decrescente, quanto à frequência na quantidade de itens publicados em periódicos e eventos. Estes, presentes na cabeça.

QUADRO 11 – CLASSIFICAÇÃO DOS VEÍCULOS DE PUBLICAÇÃO PELA QUANTIDADE DE PUBLICAÇÕES, FATOR DE IMPACTO E ÍNDICE H - ITENS PRESENTES NA CABEÇA NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO¹⁶

Col.	ITENS	TIPO DE VEÍCULO	QUALIS*	FATOR DE IMPACTO**	ÍNDICE H**	QUANT.	FREQ.	ACUM. FREQ.
1º	International Journal of Information Management	Periódico	A1	0,148	6	9	5,20%	5,20%
2º	Revista Cubana de Informacion en Ciencias de la Salud	Periódico		0,253	7	6	3,47%	8,67%
3º	Business Information Review	Periódico		0,223	9	5	2,89%	11,56%
4º	Brazilian Journal of Information Science: Research Trends	Periódico	B1	?	?	3	1,73%	13,29%
5º	Advances in Culture, Tourism and Hospitality Research	Periódico		0,103	4	2	1,16%	14,45%
6º	Aslib Journal of Information Management	Periódico	B2	0,65	29	2	1,16%	15,61%
7º	British Food Journal	Periódico	A2	0,329	35	2	1,16%	16,76%
8º	International Journal of Digital Content Technology and its Applications	Periódico		0,118	22	2	1,16%	17,92%
9º	Journal of Librarianship and Information Science	Periódico		1,062	20	2	1,16%	19,08%
10º	KMICE 2008 - Knowledge Management International Conference 2008	Evento				2	1,16%	20,23%
11º	Library Journal	Periódico		0,161	17	2	1,16%	21,39%
12º	Ovidius University Annals, Series Physical Education and Sport/Science, movement and health	Evento		?	?	2	1,16%	22,54%
13º	Patient Education and Counseling	Periódico		1,279	96	2	1,16%	23,70%
14º	Profesional de la información	Periódico	A1	0,422	15	2	1,16%	24,86%
15º	Swedish School of Library and Information Science	Dissertação		?	?	2	1,16%	26,01%
16º	Universidade Oberta de Catalunya	Periódico		?	?	2	1,16%	27,17%
17º	World Bank Policy Research Working Paper	Periódico		?	?	2	1,16%	28,32%

FONTE: O autor (2016) – * CAPES (2015) – ** Scimago (2015)

Em relação aos veículos de publicação, um periódico que encontra no alongamento da cauda é o que tem o maior fator de impacto. Porém, não há um artigo de alto impacto publicado nele. Percebe-se, ainda que Buchanan, Gibb e Orna, são os autores mais citados (ver na tabela 16) cujas publicações situam-se na cabeça da cauda, o que confirma – para o universo analisado – a afirmativa de

¹⁶ Qualis = Site da Capes (2015) – Fonte: <http://www.periodicos.capes.gov.br/>;
Fator de Impacto e índice H = site da Scimago 2015 – Fonte: <http://www.scimagojr.com/>.

Seglen (1997) e Sun; Giles (2007) (*apud* HEIDORN, 2008) na qual “os artigos de alto impacto não são encontrados, necessariamente, nos períodos de alto impacto”.

O periódico *Patient Education and Counseling*, apesar de possuir o maior fator de impacto, não representa o elevado quantitativo de produção na temática de auditoria de informação.

Quanto à relação entre os veículos de publicações e a cauda longa das palavras-chaves, estas formam um conjunto de 461 palavras, sendo 338 distintas (tabela 15). Há o predomínio da expressão *information management* (quatorze menções em material publicado entre os anos de 2004 a 2014, com exceção em 2006).

TABELA 15 – FREQUÊNCIA DAS PALAVRAS-CHAVE NA CAUDA LONGA DOS VEÍCULOS DE PUBLICAÇÕES NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Frequência (número de vezes) (a)	Quantidade de palavras-chave (b)	Quantidade Geral (a x b)	% Acumulado	Palavra(s)- chave
14	1	14	3,04	information management
8	1	8	1,74	human
7	2	14	3,04	newborn; information service
6	1	6	1,30	patient education
5	3	15	3,25	knowledge management; methodology; knowledge audit
4	3	12	2,60	decision making, family health, information system
3	7	21	4,56	...
2	51	102	22,13	...
1	269	269	58,35	...
Total	338	461	100,00	

FONTE: O autor (2016)

Observa-se que, independente da relação estudada, seja quantidade da produção, veículos de publicação, idioma, produtividade de autores, ou ainda, a frequência de palavras-chave, há a existência de cauda longa e de mini-caudas, com diferentes características, tais como distintas distribuições em distintos tipos de materiais. A relação entre itens se refere à identificação de citações entre documentos, o que identifica possível rede de colaboração.

4.1 CITAÇÃO E COCITAÇÃO

Para a formação das redes de colaboração, que são formadas a partir dos trabalhos realizados em coautoria e as cocitações, com ênfase na importância da

parceria, foram analisados 156 itens (117 artigos e 39 comunicações). Não foi possível a coleta total dos 173 itens em virtude da indisponibilidade da obtenção do texto completo de todos eles e, conseqüentemente, da listagem de referências.

Foi, inicialmente, levantado o rol de referências e realizado o tratamento: as autorias múltiplas foram desdobradas para que se contemplasse o nome de todos os autores citados; e as substituições dos traços pelo nome do próprio autor.

No período analisado, foram identificados 4.605 pesquisadores citados, para um total de 2.685 referências, destacando-se que, 2.170 (47,12%) foram citados apenas uma vez. Tendo em vista o significativo volume de dados, os pesquisadores foram ordenados de maneira decrescente de citações, e o corte foi definido com o seguinte critério: que os de maior impacto e visibilidade obtivessem, pelo menos, vinte e cinco publicações (artigos e comunicações em eventos), o que constituiu um grupo de dez pesquisadores – tabela 16.

TABELA 16 – PESQUISADORES DE MAIOR IMPACTO E VISIBILIDADE IGUAL E SUPERIOR A VINTE E CINCO PUBLICAÇÕES NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Pesquisador	Nº de artigos em que foi citado
Buchanan, S. J.	106
Gibb, F.	101
Orna, E.	82
Henczel, S.	69
Horton, F.	50
Boon, J. A.	40
Burk, C. F..	40
Botha, H.	36
Barker, R. L.	28
Griffiths, P.	25
Total	577

FONTE: O autor (2016).

A partir da matriz de ocorrência, gerada pelo *software Sitkis*, e trabalhada no programa *Excel*, utilizando a função matemática “SOMARPRODUTO”, criou-se a matriz quadrada e simétrica, de tamanho 10 x 10, de autores mais citados (quadro 12). A formação deste grupo enfatiza o descrito por Meadows (1999, p. 10), de que “cada área temática contém um grupo relativamente pequeno de pessoas que dominam suas áreas preferidas”.

O grupo dos dez principais autores corresponde a 12,52% do percentual total, com 577 citações, referindo-se ao número total de 4.605 citações e, 0,21% dos autores, são responsáveis por 12,52% das citações totais encontradas na pesquisa.

O processo realizado consiste, conforme descrito por Ahkgren *et al* (2006), numa análise de cocitação que apresenta quatro passos. Primeiro, a matriz de dados brutos é compilada; depois, é feita uma conversão dessa para uma matriz de proximidade, associação ou similaridade. O terceiro passo é a análise das multivariadas relações entre os autores presentes na matriz. Após as análises, ocorre a última etapa do processo, que é a interpretação dos dados.

Os autores Buchanan, S. J. e Gibb, F. têm o maior número de citações, o que é explicado pelo local que exercem suas atividades produtivas e linhas de pesquisa. Ambos os autores “são ingleses, trabalham na mesma instituição (*Universidad de Strathclyde*, na Grã Bretanha) e investigam a mesma linha” (GUITIÁN, M. V. G., 2009, p. 10, tradução nossa). Estes, “investigam a efetividade e eficiência dos serviços e sistemas de informação em nível individual e organizacional, evolução de serviços em bibliotecas digitais” (GUITIÁN, M. V. G.; PÉREZ, M. R. de Z.; PORRA, J. L, 2015, p. 46, tradução nossa).

QUADRO 12 – MATRIZ DE VALORES ABSOLUTOS (10 X 10) NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

	Barker, R. L.	Boon, J. A.	Botha, H.	Buchanan, S. J.	Burk, C. F..	Gibb, F.	Griffiths, P.	Henczel, S.	Horton, F.	Orna, E.
PESQUISADORES										
Barker, R. L.	28	6	5	7	2	9	1	6	10	7
Boon, J. A.		40	23	10	0	10	2	8	8	9
Botha, H.			36	9	0	9	4	10	8	9
Buchanan, S. J.				106	0	51	2	11	9	18
Burk, C. F..					40	0	0	0	4	0
Gibb, F.						101	2	12	11	23
Griffiths, P.							25	1	3	0
Henczel, S.								69	9	19
Horton, F.									50	8
Orna, E.										82

FONTE: O autor (2016).

Observa-se que, esse conjunto de autores, é formado por pesquisadores que desenvolveram um método de auditoria de informação, indicando que o grupo de autores selecionados para a análise de cocitação apresenta consistência teórica. Conforme Buchanan, Gibb (2008, p. 3, tradução nossa), “a metodologia de auditoria de informação têm sido revisada por Barker (1990), Buchanan e Gibb (1998), e Botha and Boon (2003), e as mais citadas são: Burk and Horton (1988), Orna (1990, 1999), Buchanan and Gibb (1998) e Henczel (2001)”.

Para a análise dos resultados das matrizes de proximidade foram evidenciadas as diferenças entre os valores presentes na matriz de valores

absolutos, cosseno de Salton, índice de Jaccard e Correlação de Pearson, e os valores mais altos, com intuito de observar as associações e as proximidades entre os autores.

Os cinco maiores valores absolutos (quadro 12), identificados com a cor amarela, destacam-se as cocitações entre Boon, J. A. e Botha, J. A.; entre Buchanan, S. J. e Gibb, F.; entre Gibb, F. e Henczel, S.; entre Gibb, F. e Orna, E.; e, entre Henczel, S. e Orna, E.. Os cinco menores valores de cocitação, identificados em verde, entre os autores, na matriz absoluta, a saber: entre Barker, R. L. e Burk, C. F.; entre Barker, R. L. e Giffiths, P.; entre Boon, J. A. e Giffiths, P.; entre Buchanan, S. J. e Giffiths, P.; e, entre Giffiths, P. e Henczel, S., significam que há pouca aproximação entre os pesquisadores.

Constituída a matriz de valores absolutos, foram constituídas as matrizes de medidas de similaridade bilateral (cosseno de Salton – quadro 13 – e o índice de Jaccard – quadro 14), “que são usadas para comparar relações entre pares” (VANZ; STUMPF, 2010, p. 72), a partir da aplicação das fórmulas elencadas no referencial teórico.

QUADRO 13 – INDICADOR NORMATIVO - COSSENO DE SALTON - NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

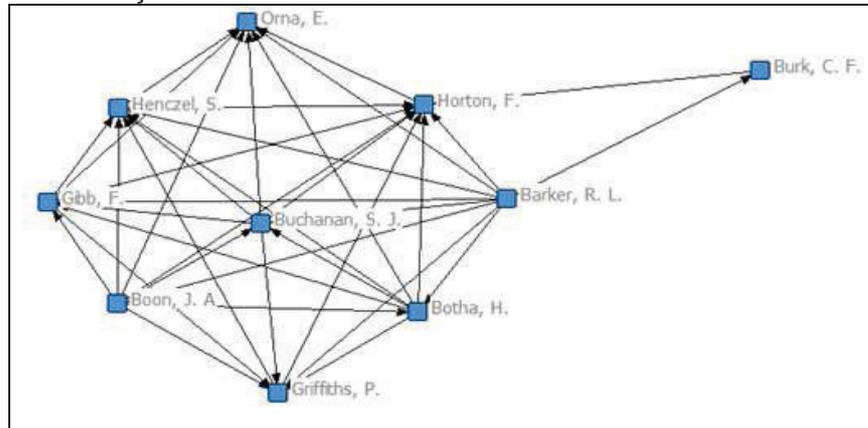
PESQUISADORES	Barker, R. L.	Boon, J. A.	Botha, H.	Buchanan, S. J.	Burk, C. F.	Gibb, F.	Giffiths, P.	Henczel, S.	Horton, F.	Orna, E.
Barker, R. L.	1,00	0,18	0,16	0,13	0,06	0,17	0,04	0,14	0,27	0,15
Boon, J. A.		1,00	0,61	0,15	0,00	0,16	0,06	0,15	0,18	0,16
Botha, H.			1,00	0,15	0,00	0,15	0,06	0,20	0,19	0,17
Buchanan, S. J.				1,00	0,00	0,49	0,04	0,13	0,12	0,19
Burk, C. F.					1,00	0,00	0,00	0,00	0,09	0,00
Gibb, F.						1,00	0,04	0,14	0,15	0,25
Giffiths, P.							1,00	0,02	0,08	0,00
Henczel, S.								1,00	0,15	0,25
Horton, F.									1,00	0,12
Orna, E.										1,00

FONTE: O autor (2016).

Observam-se os valores significativos de índice de cosseno de Salton – CS – advindos da proximidade temática identificada pelos citantes, que explicitam a concepção de cocitação, ou podem indicar coautoria entre os pares de autores cocitados. No gráfico 8, observa-se a relação dos indivíduos dentro do grupo de autores ou da rede de relacionamentos, conforme resultado da matriz do cosseno de Salton, com a utilização do *software Unicet*, utilizado como ferramenta de apoio para análise e interpretação dos dados, pois para Spinak (1996, p. 16), “as redes de

cocitação podem traduzir-se em mapas ou nós, onde os pontos denotam documentos e as linhas de união representam as relações de cocitações”.

GRÁFICO 8 – GRÁFICO DE ANÁLISE DE REDES SOCIAIS: AUTORES E COAUTORES, NA MATRIZ DO COSSENO DE SALTON NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO



FONTE: O autor (2016).

O gráfico 8, mostra a relação que cada autor apresenta sobre outros autores. Aqueles que estão no centro (Buchanan, S. J.) apresentam um maior número de citações, recebendo ou realizando a citação. Isso pode ser comprovado, através da comparação com a tabela 16, sendo os autores que contribuem com o maior número de citações, a posicionados no centro da rede.

QUADRO 14 – INDICADOR NORMATIVO – ÍNDICE DE JACCARD - NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

PESQUISADORES	Barker, R. L.	Boon, J. A.	Botha, H.	Buchanan, S. J.	Burk, C. F.	Gibb, F.	Griffiths, P.	Henczel, S.	Horton, F.	Orna, E.
Barker, R. L.	1,00	0,10	0,08	0,06	0,03	0,08	0,02	0,07	0,15	0,07
Boon, J. A.		1,00	0,43	0,07	0,00	0,08	0,03	0,08	0,10	0,08
Botha, H.			1,00	0,07	0,00	0,07	0,07	0,11	0,10	0,08
Buchanan, S. J.				1,00	0,00	0,33	0,02	0,07	0,06	0,11
Burk, C. F.					1,00	0,00	0,00	0,00	0,05	0,00
Gibb, F.						1,00	0,02	0,08	0,08	0,14
Griffiths, P.							1,00	0,01	0,04	0,00
Henczel, S.								1,00	0,08	0,14
Horton, F.									1,00	0,06
Orna, E.										1,00

FONTE: O autor (2016).

Nota-se, no quadro 14, que os valores do índice de Jaccard são menores que os índices para o cosseno de Salton, com exceção do Botha, H. e Griffiths, P., que apresentam o valor normalizado de 0,06 e 0,07, respectivamente. Comparando-

se os indicadores de medida de similaridade bilateral (cosseno de Salton e índice de Jaccard), afirma o exposto, “que o cosseno de Salton tende a produzir uma medida de similaridade relativa que se aproxima, em valor, a duas vezes o número daquele obtido pelo índice de Jaccard” (HAMERS, 1989, p.316).

A constituição da medida de similaridade multivariada – correlação de Pearson, “relaciona coautoria entre um par de autores e com todos os outros autores envolvidos” (VANZ; STUMPF2010, p. 72) – quadro 15.

QUADRO 15 – INDICADOR NORMATIVO – CORRELAÇÃO DE PEARSON - NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

PESQUISADORES	Barker, R. L.	Boon, J. A	Botha, H.	Buchanan, S. J.	Burk, C. F..	Gibb, F.	Griffiths, P.	Henczel, S.	Horton, F.	Orna, E.
Barker, R. L.	1,00	0,20	0,17	0,18	0,06	0,23	0,04	0,19	0,34	0,20
Boon, J. A		1,00	0,62	0,20	-0,01	0,19	0,06	0,19	0,20	0,20
Botha, H.			1,00	0,19	-0,01	0,19	0,14	0,26	0,21	0,21
Buchanan, S. J.				1,00	-0,01	0,51	0,05	0,14	0,15	0,20
Burk, C. F..					1,00	-0,01	-0,01	-0,01	0,09	-0,01
Gibb, F.						1,00	0,05	0,15	0,18	0,26
Griffiths, P.							1,00	0,03	0,10	-0,01
Henczel, S.								1,00	0,17	0,26
Horton, F.									1,00	0,14
Orna, E.										1,00

FONTE: O autor (2016).

Constituída a correlação de Pearson, o resultado elevado ao quadrado proporciona a medida da variabilidade em uma variável que é explicada pela variabilidade da outra, conforme o quadro 16.

QUADRO 16– COEFICIENTE DE DETERMINAÇÃO R² - NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

PESQUISADORES	Barker, R. L.	Boon, J. A	Botha, H.	Buchanan, S. J.	Burk, C. F..	Gibb, F.	Griffiths, P.	Henczel, S.	Horton, F.	Orna, E.
Barker, R. L.	1,00	0,04	0,03	0,03	0,00	0,05	0,00	0,04	0,11	0,04
Boon, J. A		1,00	0,38	0,04	0,00	0,04	0,00	0,04	0,04	0,04
Botha, H.			1,00	0,04	0,00	0,03	0,02	0,07	0,04	0,04
Buchanan, S. J.				1,00	0,00	0,26	0,00	0,02	0,02	0,04
Burk, C. F..					1,00	0,00	0,00	0,00	0,01	0,00
Gibb, F.						1,00	0,00	0,02	0,03	0,07
Griffiths, P.							1,00	0,00	0,01	0,00
Henczel, S.								1,00	0,03	0,07
Horton, F.									1,00	0,02
Orna, E.										1,00

FONTE: O autor (2016).

O valor de 0,38 entre Boon, J. A e Botha, H. quanto multiplicado por cem demonstra que 38% das citações recebidas por Boon são explicadas por Botha, ou seja, a variação de Boon explica 38% da variação de Botha. No quadro 16, o resultado de cada um dos citantes demonstra o quanto uma citação esta atrelada a outra citação.

Analisando as normalizações, tanto pelo cosseno de Salton, índice de Jaccard e correlação de Pearson, os maiores valores são entre Boon, J. A e Botha, H.. Embora o valor absoluto para cocitação não tenha sido o maior entre os pesquisadores, Boon, J. A. e Botha, H. apresentam os maiores valores normalizados, em virtude da diferença da intensidade de proximidade, que não é significada pela frequência absoluta (tabela 17).

TABELA 17 – AUTORES E VALORES EXTREMADOS DOS INDICADORES ABSOLUTOS E NORMALIZADOS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Autores cocitados	Índice absoluto	Indicadores Normalizados		
		Cosseno de Salton	Índice de Jaccard	Correlação de Pearson
Boon, J. A. x Botha, H.	23 cocitações	0,31	0,43	0,62
Buchanan, S. J. x Gibb, F.	51 cocitações	0,49	0,33	0,51
Gibb, F. x Henczel, S.	12 cocitações	0,14	0,08	0,15
Gibb, F. x Orna, E.	23 cocitações	0,25	0,14	0,26
Henczel, S. x Orna, E.	19 cocitações	0,25	0,14	0,26
Barker, R. L x Burk, C. F.	2 cocitações	0,06	0,03	0,03
Barker, R. L x Griffiths, P.	1 cocitação	0,04	0,02	0,04
Boon, J. A. x Griffiths, P.	2 cocitações	0,06	0,03	0,06
Buchanan, S. J. x Griffiths, P.	2 cocitações	0,04	0,02	0,05
Griffiths, P. x Henczel, S.	1 cocitação	0,02	0,01	0,03

FONTE: O autor (2016).

O valor de 0,62 para r de Pearson significa maior correlação entre os pesquisados que foram analisados. O resultado é fruto de 23 cocitações relativizadas, para quarenta citações de Boon, J. A. e 36 e Botha, H.

Na comparação entre autores mais produtivos, autores e coautores mais citados, e autores mais cocitados, observa-se a presença dos pesquisadores Buchanan e Henczel, nos mais produtivos, mais citados e mais cocitados, enfatizando o descrito por Meadows (1999, p. 109), que “os trabalhos mais citados, [...], escrito em colaboração, envolvem os pesquisadores mais produtivos e eminentes”.

TABELA 18 – AUTORES MAIS PRODUTIVOS, AUTORES E COAUTORES MAIS CITADOS E AUTORES MAIS COCITADOS NA TEMÁTICA DE AUDITORIA DE INFORMAÇÃO – 1970 - 2016/SETEMBRO

Col.	Autores mais produtivos	Col.	Autores e coautores mais citados	Col.	Autores mais cocitados
1º	Buchanan, S.	1º	Buchanan, S.	1º	Boon, J. A. x Botha, H.
2º	Henczel, S.	2º	Gibb, F.	2º	Buchanan, S. J. x Gibb, F.
3º	Drus, S. M.	3º	Orna, E.	3º	Gibb, F. x Henczel, S.
4º	Xu Fe	4º	Henczel, S.	4º	Gibb, F. x Orna, E.
5º	Dante, G. P.	5º	Horton, F.	5º	Henczel, S. x Orna, E.
		6º	Boon, J. A.		
		7º	Burk, C. F..		
		8º	Botha, H.		
		9º	Barker, R. L.		
		10º	Griffiths, P.		

FONTE: O autor (2016).

No entanto, o predomínio de um pequeno número de pesquisadores (Buchanan, Gibb, Orna, Henczel), na temática auditoria de informação, demonstra o domínio de poucos (cabeça), mas a presença de muitos, formando a cauda longa.

5 CONSIDERAÇÕES FINAIS

O estudo apresentou como questão de pesquisa a seguinte indagação: “como se configura o domínio de auditoria de informação ao se ampliar não apenas o universo de bases de dados, mas igualmente as técnicas bibliométricas utilizadas acompanhadas da análise do “efeito de cauda longa?””, tendo como objetivo geral explorar as características das manifestações de “cauda longa”, derivadas de distribuições bibliométricas, no domínio de auditoria da informação. A discussão com base em técnicas bibliométricas, fundamentada nas Leis de Lotka, Zipf, Correlação de Pearson, Cosseno de Salton, Índice de Jaccard, identificação de cauda longa e indicadores de produtividade dos autores, foi realizada por meio de análise de 173 itens publicados indexados ISI/Web of Science^(C), SCOPUS, SciELO, EBSCOhost e Google Scholar.

Verificou-se que os itens analisados tiveram aderência em periódicos da área da saúde, contabilidade, financeiro, inovação, informação e auditoria externa. Constatou-se que, no ano de 2008, a publicação na temática auditoria de informação atingiu o ápice de publicações, com itens referentes à aplicação em empresas de eletricidade, utilidade pública, saúde, além de discussões sobre a compreensão, usabilidade, conceito e diversas metodologias dos pesquisadores. Os Estados Unidos produziram o maior volume de pesquisa e o idioma inglês apresenta hegemonia.

Verificou-se que 54,90% das palavras-chave (simples ou compostas) citadas nesses itens foram utilizadas apenas uma vez, enquanto 19,61% foram usadas duas vezes, o que revela uma dispersão dos termos utilizados para indexação e sugere forte heterogeneidade dos temas abordados nos itens.

Na aplicação da Lei de Lotka os testes revelaram que a produção em auditoria da informação segue as características da lei, que são: muitos autores produzem pouco e poucos autores produzem muitos. A dimensão dos autores que publicam apenas uma vez alcançou 80,35%. Uma explicação para o comportamento pode ser o fato de auditoria de informação ser uma área de aplicação prática.

Na aplicação empírica da Lei de Lotka (prevê a relação entre autores e artigos) e os pressupostos de produtividade dos autores, os achados indicam baixa consolidação do tema, dada a ocorrência de um elevado número de autores que

publicaram apenas uma vez (89,71%), e um número significativo de publicações dispersas em vários periódicos.

Na relação da análise de cocitação evidenciaram-se as diferenças entre os valores presentes nas matrizes absolutas (índices Salton, Jaccard, Pearson). Quanto aos valores mais altos na matriz absoluta (total de citação ou presença de cada autor no universo analisado, derivados dos cálculos referentes ao conceito de Salton) os autores Buchanan x Gibb e Boon x Botha são os de maior destaque. Estes mesmos autores, ao se calcular o índice de Jaccard, relativo a intensidade de cocitação, e a correlação de Pearson (que indica maior correlação entre os autores pesquisados) - permanecem em destaque. Buchanan, Boon, Botha e Gibb representam a “frente de pesquisa”, confirmando à que “cada área temática contém um grupo relativamente pequeno de pessoas que dominam suas áreas” (MEADOWS, 1999, p. 101). Estes autores publicaram estudos do uso da auditoria de informação em diferentes áreas, tais como tecnologia da informação (TI), bibliotecas em instituição de ensino, explorando abordagens voltadas a compreensão, uso, conceito e metodologias da AI.

Quanto a discussão trazida Seglen (1997) e Sun; Giles (2007) (*apud* HEIDORN,2008) de que “artigos de alto impacto não são encontrados, necessariamente, nos periódicos de alto impacto”, esta se confirma com os dados coletados, nos quais o periódico *International Journal of Infomation Management* e a Revista Cubana de *Infomation em Ciencias de la Salud* não apresentam um alto fator de impacto, mas publicam artigos dos autores mais citados (QUADRO 11). O periódico com o maior fator de impacto (*Patient Education and Couseling*) não agrega artigos que receberam número significativo de citações. Porém, cabe lembrar que o cálculo do fator de impacto dos periódicos pode não ter uma relação direta com a publicação de artigos na temática de auditoria de informação.

Ao se explorar o fenômeno da cauda longa consideraram-se as seguintes relações: local de publicação dos eventos e dos periódicos (por continente) em relação com idioma, produtividade de autores e palavras-chave.

Quanto aos itens que compõem a cauda longa visualizada no Gráfico 7, percebe-se o domínio de eventos no continente Europeu e da América do Norte. Destes, o *Knowledge Management International Conference* aparece com destaque e a palavra-chave de maior frequência (*information management*) é a que identifica tais itens.

Espera-se que este estudo possa contribuir para futuras investigações da literatura sobre a temática auditoria da informação, especialmente aquelas que ampliem o escopo das bases ora analisadas (EBSCOhost, Google Scholar, ISI/Web of Science^(C), SciELO, e SCOPUS), auxiliando a consolidar investigações sobre o conhecimento científico que vem sendo construído neste campo do conhecimento.

Como limitações da pesquisa podem-se destacar o fato de que nem todos os periódicos que compuseram o universo da pesquisa foram analisados; assim como dificuldades na captura dos metadados devido à homonímia relativa ao nome dos autores, os quais sofreram tratamento manual por parte do pesquisador. Tal condição pode se tornar inviável em universos maiores.

Futuras pesquisas podem replicar o estudo em outras bases de dados possibilitando a comparação de resultados. A análise da produção em auditoria de informação igualmente pode-se beneficiar da aplicação da lei de Bradford a qual não foi utilizada neste estudo. As técnicas bibliométricas selecionadas podem igualmente ser utilizadas em outras temáticas, reforçando a importância dos estudos métricos da informação na área interdisciplinar.

REFERÊNCIAS

AHKGREN, P.; *et al.*. Requirements for a cocitação similarity measure, with special reference to Pearson's Correlation Coefficient. **Journal of the American Society of Information Science & Technology**, New York, v. 54, n. 6, p. 1616-1628, 2006.

AGUNE, R.M.; CARLOS, J. A. A cauda longa e a gestão pública. In: I Congresso Cosad de Gestão Pública, 2008, Brasília. **Anais...** Brasília, 2008. Disponível em: <http://repositorio.fjp.mg.gov.br/consad/bitstream/123456789/97/4/C1_PP_GESTAO%20DO%20CONHECIMENTO%20NO%20SETOR%20PUBLICO.pdf>. Acesso em: 24 out. 2016.

ALELUIA, L. R. de. **Revista Baiana de saúde pública: produção do conhecimento científico de 1974 a 2008**. 180 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal da Bahia, Salvador, 2009.

ALVES, B. H. **Aportes bibliométricos à produção científica nos principais periódicos da área de ciência da informação do Brasil, no período de 2006-2010**. 113f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciência, Marília, 2013.

ANDERSON, C. **A cauda longa**: do mercado de massa para o mercado de nicho. Rio de Janeiro: Elsevier, 2006.

ANDERSON, C.. **The long tail**: why the future of business is selling less of more. 2006: BRYNJOLFSSON, E.; HU, Y. J.; SMITH, M. D. From niches to riches: anatomy of the long tail. **Sloan Management Review**, 2006. v. 47, n. 4, p. 67-71 *apud* WU, L-L.; LUESUKPRASERT, L.; LEE, L. Research and the long tail: a large-scale citation analysis In: INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 42nd. **Proceedings**. Hawaii, 2009, p. 1-10. Disponível em: <<https://www.computer.org/csdl/proceedings/hicss/2009/3450/00/09-13-02.pdf>>. Acesso em: 4 abril, 2016.

ARAÚJO, C. A. Bibliometria: evolução histórica e questões atuais. **Em questão**, Porto Alegre, v.12, n.1, p.11-32, 2006. Disponível em: <<http://www.revistas.univerciencia.org/index.php/revistaemquestao/article/view/3707/3495>>. Acesso em: 09 jun. 2016.

BELOCHIO, V. C. **Jornalismo colaborativo em redes digitais: estratégia comunicacional no ciberespaço o caso de zero hora.com**. 214 f. Dissertação (Mestre em Comunicação) – Universidade Federal de Santa Maria, Santa Maria, 2009. Disponível em: <http://cascavel.cpd.ufsm.br/tede/tde_arquivos/29/TDE-2009-04-03T131156Z-1962/Publico/VIVIANDECARVALHOBELOCHIO.pdf>. Acesso em: 18 jun. 2016.

BEUREN, I. M.; SILVA, M. Z.. Característica bibliométricas dos artigos sobre gestão hospitalar publicados em periódicos de alto impacto. **Revista Cubana de Información en Ciencias de la Salud**, v.25, n.1, p.36-65, 2014. Disponível em: <<http://www.acimed.sld.cu/index.php/acimed/article/view/274/380>>. Acesso em: 13 jun. 2016.

BORGMAN, C. L.; *et. al.* Data Management in the Long Tail: Science, Software and Service. **International Journal of Digital Curation**, v.11, n.1, p.128-149, 2016. Disponível em: <<http://www.ijdc.net/index.php/ijdc/article/view/11.1.128/432>>. Acesso em: 20 out. 2016.

BORTOLOSSI, H. J.; QUEIROZ, J. J. D. B.; SILVA, M. M. da. A lei de Zipf e outras leis de potência em dados empíricos. **Instituto de matemática e estatística**. Universidade Federal Fluminense. [s.n]. Disponível em: <<http://klein.sbm.org.br/wp-content/uploads/sites/17/2016/02/Zipt-bortolossi-queiroz-dasilva-lpp-projeto-klein.pdf>>. Acesso em: 01 jul. 2016.

BOTHA, H.; BOON, J. A. The information audit: Principles and guidelines. **Libri**, v. 53, p. 23-38, 2003.

BUCHANAN, S.; GIBB, F. The Information audit: an integrated strategic approach. **The International Journal of Information Management**, v.18, n.1, p. 29-48, 1998.

_____. The information audit: role and scope. **The International Journal of Information Management**, v.27, p. 159-172, 2007.

BRUNI, A. L. **Estatística aplicada à gestão empresarial**. 2. ed. São Paulo: Atlas, 2010.

CARNEIRO, A. **Auditoria e controlo de sistemas de informação**. Lisboa: FCA, 2009.

CRANE, D. **Information needs and uses**. **Annual Review of Information Science and Technology**, Chicago, v. 6, p. 3-38, 1971.

CORDEIRO, D.; *et. al.* Da ciência à e-ciência: paradigmas da descoberta do conhecimento. **Revista USP**, São Paulo, n.97, p. 71-80, março/abril/maio 2013. Disponível em: <<http://www.revistas.usp.br/revusp/article/view/61867>>. Acesso em: 29 out. 2016.

CÔRTEZ, M. R; *et al.* Cauda Longa e a mudança no modelo de negócio no mercado fonográfico: reflexões acerca do impacto das novas tecnologias. In: XXVIII Encontro Nacional de Engenharia de Produção, 2008, Rio de Janeiro. **Anais...** Rio de Janeiro, 2008. p. 1-13. Disponível em: <http://abepro.org.br/biblioteca/enegep2008_TN_STP_069_490_11917.pdf>. Acesso em: 18 jun. 2016.

COSTA, J. F. da R. **A Cauda Longa**: a sua existência no mercado de retalho online português. 95f. Dissertação (Mestre em Economia e Administração de Empresas) – Universidade do Porto, 2015. Disponível em:<<https://repositorio-aberto.up.pt/bitstream/10216/81128/2/124857.pdf>>. Acesso em: 18 jun. 2016.

COSTA, M. M.; CUNHA, M. B.. O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais. **Perspectivas em Ciência da Informação**, v. 19,

n.3, p. 189-206, jul/set/2014. Disponível em:
<<http://www.scielo.br/pdf/pci/v19n3/a10v19n3.pdf>>. Acesso em 29 out. 2016.

CUSTÓDIO, P. A. G. da R. **Dissertações e teses da pós-graduação em educação da Unesp/Campus de Marília: um estudo das citações e cocitações (2004 a 2009)**. 145 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2012. Disponível em:
<https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/Custodio_P_A_G_da_R_mestrado_C I_2012.pdf>. Acesso em: 15 jun. 2016.

DANUELLO, J. C. **Produção científica docente em tratamento temático da informação no Brasil: uma abordagem métrica como subsídio para a análise de domínio**. 110 f. Dissertação (Mestre em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2007. Disponível em:
<<http://repositorio.unesp.br/handle/11449/95532>>. Acesso em: 18 jun. 2016.

DANTE, P. G. Auditoria da informação e do conhecimento organizacional: gênese de uma integração. **Brazilian journal of information science**. São Paulo, v. 2, n. 2, p. 3-16, jul./dez. 2008.

DIAS, C. da C. A análise de domínio, as comunidades discursivas, a garantia da literatura e outras garantias. **Informação & Sociedade**, João Pessoa, v. 25, n. 2, p. 7-17, maio/ago. 2015. Disponível em:
<<http://www.ies.ufpb.br/ojs/index.php/ies/article/viewFile/7/13741>>. Acesso em: 18 jun. 2016.

FRANCO, S. F. **Os nichos da informação: um blog sobre seriados televisivos**. 2012. 60 f. Trabalho de conclusão de curso (Graduação em Comunicação Social) – Centro Universitário de Brasília, Brasília, 2012. Disponível em:<<http://repositorio.uniceub.br/bitstream/235/7218/1/20961700.pdf>>. Acesso em: 19 jun. 2016.

GARFIELD, E. Citation Analysis as a Tool in Journal Evaluation. **Essays of an Information Scientist**, v.1, p. 527-544, 1972. Disponível em:
<<http://www.garfield.library.upenn.edu/essays/V1p527y1962-73.pdf>>. Acesso em: 17 out. 2016.

_____. Is Citation Analysis a Legitimate Evaluation Tool?. **Scientometrics**, Amsterdam, v. 1, n.4, p. 359-375, 1979. Disponível em:<<http://link.springer.com/article/10.1007/BF02019306>>. Acesso em: 18 jun. 2016.

GINN, L. K. Citation analysis of authored articles. In: Library & Science research. **Mississippi Libraries**, v.67, p. 106-109, 2003.

GOUVEIA, F. C.; LANG, P. Da webometria à altmetria: uma jornada por uma ciência emergente. In: ALBAGLI, S. (Org.). **Fronteiras da ciência da informação**. Rio de Janeiro: IBICT, 2013. p. 172-195.

GRACIO, M. C. C.; OLIVEIRA, E. F. T. Análise de cocitações de autores: um estudo teórico-metodológico dos indicadores de proximidade, aplicados ao GT 7 da ANCIB. **Liinc em revista**, Rio de Janeiro, v.9, n.1, p. 196-213, maio 2013. Disponível em: <<http://liinc.revista.ibict.br/index.php/liinc/article/viewFile/527/401>>. Acesso em: 16 jun. 2016.

GRACIO, M. C. C.; OLIVEIRA, E. F. T. Indicadores de proximidades em análise de cocitação de autores: um estudo comparativo entre coeficiente de correlação de Pearson e Cosseno de Salton. **Informação & Sociedade**, João Pessoa, v.25, n.2, p. 105-116, maio/ago 2015. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/105>>. Acesso em: 14 jun, 2016.

GREGOLIN, J. A. R. (coord.). Análise de produção científica a partir de indicadores bibliométricos In: LANDI, R. L.; GUSMÃO, R. **Indicadores de ciência, tecnologia e inovação em São Paulo**. São Paulo: Fapesp, 2005, p. 5-43. Disponível em: <http://www.fapesp.br/indicadores2004/volume1/cap05_vol1.pdf>. Acesso em: 08 jun. 2016.

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: CIFORM – Encontro Nacional de Ciência da Informação, 6., 2005, Salvador. **Anais...** Salvador: ICI/UFBA, 2005. Disponível em <<http://www.feg.unesp.br/~fmarins/seminarios/Material%20de%20Leitura/Bibliometria/Artigo%20Bibliometria%20-%20Ferramenta%20estat%edstica%20VaniaLSGuedes.pdf>>. Acesso em: 13 jun. 2016.

GUEDES, V. L. S. A bibliometria e a gestão e do conhecimento científico e tecnológico: uma revisão da literatura. **Ponto de Acesso**, Salvador, v.6, n.2, p. 74-109, ago, 2012. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/view/5695/4591>>. Acesso em: 16 jun. 2016.

GUIMARÃES, F.; QUARESMA, P. Auditoria de informação baseada em ontologia. **Inf**, Portugal, 2015. Disponível em:<<http://dspace.uevora.pt/rdpc/handle/10174/17209>>. Acesso em: 19 jun. 2016.

GUITIÁN, M. V. G. Auditorías de información: análisis de dominio en la base de datos LISA. **Acimed**, v.19, n.4, p. 1-13, 2009.

GUITIÁN, M. V. G.; PÉREZ, M. R. de Z.; PORRA, J. L. Auditoría de información y auditoría de conocimiento: acercamiento a su visualización como dominios científicos. **Revista Cubana de Información em Ciências de la Salud**, v.26, n.1, p. 34-52, 2015.

HAMERS, L. et al. Similarity measures in scientometric research: the Jaccard index versus Salton's cosine formula. **Information Processing & Management**, v.25, n.3, p. 315-318, 1989. Disponível em:

<<http://www.sciencedirect.com/science/article/pii/0306457389900484>>. Acesso em: 14 jun. 2016.

HARNAD, S. et al. Maximizing and Measuring Research Impact Through University and Research-Funder Open-Access Self-Archiving Mandates.

Wissenschaftsmanagement, v.5, n.4, 2009, p. 36-41. Disponível em: <<http://eprints.soton.ac.uk/266616>>. Acesso em: 17 out. 2016.

HEIDORN, P. B. Shedding light on the dark data in the long tail of Science. **Library Trends**, v.57, n. 2, 2008, p. 280-299. Disponível em:

<http://muse.jhu.edu/journals/library_trends/v057/57.2.heidorn.html>. Acesso em: 1 jul. 2016.

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

HJØRLAND, B. Epistemology and the socio-cognitive perspective um Information Science. **Journal of the American Society for Information Science and Technology**, v. 53, n. 4, p. 257-270, 2002.

HJØRLAND, B. Domain analysis: a socio-cognitive orientation for information science research. **Bulletin of the American Society for Information Science and Technology**, v.30, n.3, p. 17-21, 2004.

JOB, I. **Educação Física no PPGCMH/UFRGS**: uma visão a partir da análise de citações e perfil dos pesquisadores. 145 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

Disponível em:

<<http://www.lume.ufrgs.br/bitstream/handle/10183/7771/000556580.pdf;sequence=1>>. Acesso em: 09 jun 2016.

KROEFF, M. S. et al. Análise de citações dos artigos publicados em periódicos da área da Ciência da informação que versam sobre gestão da informação. **Revista Brasileira de Biblioteconomia e Documentos**, São Paulo, v.11, n.1, p. 41-65,

jan/jun. 2015. Disponível em: <<https://rbbd.febab.org.br/rbbd/article/view/301>>.

Acesso em: 14 jun. 2016.

KURAMOTO, H. Informação científica: proposta de um novo modelo para o Brasil.

Ciência da Informação, v. 35, n. 2, mai./ago. 2006, p. 91-102. Disponível em:

<<http://www.scielo.br/pdf/ci/v35n2/a10v35n2.pdf>>. Acesso em: 13 nov. 2016.

LEITE, F. C. L.; COSTA, S. Repositórios institucionais como ferramentas de gestão do conhecimento científico no ambiente acadêmico. **Perspectiva Ciência e Informação**, v.11, n.2, p. 206-219, mai./ago. 2006. Disponível em:

<<http://www.scielo.br/pdf/%0D/pci/v11n2/v11n2a05.pdf>>. Acesso em: 29 jun. 2016.

<<http://www.scielo.br/pdf/%0D/pci/v11n2/v11n2a05.pdf>>. Acesso em: 29 jun. 2016.

LEYDESDORFF, L.; VAUGHAN, L. Co-occurrence matrices and their applications in information science: extending ACA to the web environment. **Journal of the American**

Society for Information Science and Technology, v.57, n. 12, p. 1616-1628, 2006. Disponível em: <<http://www.leydesdorff.net/aca/aca.pdf>>. Acesso em: 19 jun. 2016.

LIMA, E.; MAIA, S. Comportamento bibliométrico da língua portuguesa, como veículo de representação da informação. **Ciência da Informação**, v. 2, n. 2, 1973. Disponível em: < <http://revista.ibict.br/ciinf/article/view/31>>. Acesso em: 02 nov. 2016.

LOPES, E. C; et al. A construção do conhecimento sobre governança corporativa: uma análise bibliométrica. **Data grama Zero**, v.14, n.2, abr 2013. Disponível em: <<http://hdl.handle.net/11449/115026>>. Acesso em: 13 jun. 2016.

LOTKA, A. J. The frequency distribution of scientific productivity. **Journal of the Washington Academy of Sciences**, v.16, n.12, p.317-323, June 1926.

LUUKKONEN, T.; PERSSON, O.; SIVERTSEN, G. Understanding patterns of international scientific collaboration. **Science, Technology & Human Values**, Thousand Oaks, v.17, n.1, Winter, 1992, p.101-126. Disponível em: < https://www.researchgate.net/publication/248204625_Understanding_Patterns_of_International_Scientific_Collaboration>. Acesso em: 07 jul. 2016.

MACIAS-CHAPULA, C. A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciências da Informação**, Brasília, v.27, n.2, p. 134-140, maio/ago 1998. Disponível em: <<http://www.scielo.br/pdf/ci/v27n2/macias.pdf>>. Acesso em: 18 jun. 2016.

MACHADO, J.R., C. As leis da bibliometria em diferentes bases de dados científicos. **Ciências da Administração**, Santa Catarina, v.18, n.44, p.111-123, 2016. Disponível em: <<https://periodicos.ufsc.br/index.php/adm/article/view/2175-8077.2016v18n44p111>>. Acesso em: 09 jul. 2016.

MAI, J. E. Analysis in indexing: document and domain centered approaches. **Information Processing and Management**, v.41,n.3, p. 599-611, 2005. Disponível em: <http://ac.els-cdn.com/S030645730300116X/1-s2.0-S030645730300116X-main.pdf?_tid=636213f4-3565-11e6-86b3-00000aab0f26&acdnat=1466262205_10177bd7a5c011edf04f49b1b17ee415>. Acesso em: 18 jun. 2016.

MARCELO, C.; CORDEIRO, R. **Auditoria interna e operacional: fundamentos, conceitos e aplicações práticas**. São Paulo: Atlas, 2013.

MATTAR, F. N.; AUAD, M. Nicho de mercado: um conceito ainda indefinido. In: Seminários em Administração. In: SEMEAD, 2, 1997, São Paulo. **Anais...** São Paulo: FEA/USP, 1997.

MARSHALL JUNIOR, I. et al. **Gestão da qualidade**. 9. ed. Rio de Janeiro: FGV, 2008.

McAFEE, A.; BRYNJOLFSSON, E. Big Data: The Management Revolution. **Harvard Business Review**. p. 59-68; 2012. Disponível em:

<http://www.rosebt.com/uploads/8/1/8/1/8181762/big_data_the_management_revolution.pdf>. Acesso em: 24 out. 2016.

MEADOWS, A. J. **A comunicação científica**. Tradução de: LEMOS, A. A. B. de. Brasília, 1999. Original em inglês.

MEDEIROS, J. da S.; CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc. Em revista**, Rio de Janeiro, v.8, n.2, p. 311-322, set. 2012. Disponível em: <https://chasqueweb.ufrgs.br/~jacksonmedeiros/pubs/2012_escience_liinc.pdf>. Acesso em: 29 out. 2016.

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO. **Aumenta o investimento em C&T no Brasil**. Disponível em: < http://www.mcti.gov.br/noticia/-/asset_publisher/epbV0pr6eIS0/content/aumenta-o-investimento-em-c-t-no-brasil>. Acesso em: 9 jul. 2016.

MUELLER, S. P. M. O crescimento da ciência, o comportamento científico e a comunicação científica: algumas reflexões. **Escola da Biblioteconomia UFMG**, v.24, n.1, p. 63-84, jan/jun. 1995. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/index.php/article/download/13022>>. Acesso em 17 out. 2016.

NEIGHBORS. **Software Construction Usin Components**. Doutorado, 1991 *apud* KERR, E. S. **Ketib: um processo de representação de informações para textos complexos**. .82 f. Dissertação (Mestrado Profissional em Ciência da Computação) - Universidade Estadual de Campinas, Campinas, 2003. Disponível em: <<http://www.bibliotecadigital.unicamp.br/document/?code=vtls000311557&fd=y>>. Acesso em: 08 jun de 2016.

OLIVEIRA, E. F. T.; GRACIO, M. C. C. A produção científica e organização e representação do conhecimento no Brasil: uma análise bibliométrica do GT-2 da ANCIB. In: Encontro Nacional de Pesquisa em Ciência da Informação, 10, 2009. João Pessoa. **Anais...** João Pessoa: ANCIB, p. 2037-2056, 2009.

OLIVEIRA, E. F. T. de. **Análise de domínio em “estudos métricos” no Brasil: produção, impacto e visibilidade em âmbito nacional e internacional**. 193 f. Tese (Doutorado Livre docência) – Universidade Estadual Paulista, Marília, 2013. Disponível em: <<http://docplayer.com.br/2498209-Analise-de-dominio-em-estudos-metricos-no-brasil-producao-impacto-e-visibilidade-em-ambito-nacional-e-internacional.html>>. Acesso em: 18 jun.2016.

PAO, M. L. Automatic text analysis based on transition phenomena of word occurrences. **Journal of the American Society for Information Science**, New York, v. 29, n.3, p. 121-124, May 1978. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.4630290303/pdf> >. Acesso em: 14 jun. 2016.

PARK, H.; LEYDESDORFF, L. Decomposing Social and Semantic Networks In Emerging “Big Data Research. **Journal of Informetrics**, v.7, n.3, p. 756-765, 2013.

Disponível em:<

<http://www.sciencedirect.com/science/article/pii/S1751157713000473>>. Acesso em: 20 out. 2016.

PECK, S. Science suffers when getting a Grant becomes the goal. **The Chronicle of Higher Education**, v.55, n.7, p. 42, 2008. Disponível em:

<<https://www.ideals.illinois.edu/bitstream/handle/2142/10672/heidorn.pdf?sequence=2&isAllowed=y>>. Acesso em: 17 out. 2016.

PEREIRA, T. F. et al. Gestão do conhecimento em projetos de simulação: um estudo bibliométrico. **Perspectiva em Ciência da Informação**, v.20, n.1, p. 138-155, jan./mar. 2015. Disponível em:

<<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2038>>. Acesso em: 13 jun. 2016.

PESTANA, O. Auditoria de informação: definição e evolução da atividade no contexto da gestão da informação e das organizações. **Páginas A&B**. v.2, p. 49-64, 2014. Disponível em:

<https://sigarra.up.pt/flup/pt/pub_geral.show_file?pi_gdoc_id=502411>. Acesso em: 03 abr. 2016.

PIRES, E. A. de N. Auditoria de informação no contexto biblioteconômico. In Encontro Nacional de Estudantes de Biblioteconomia, Documentação, Ciência da Informação e Gestão da Informação, XXXV, 2012, Minas Gerais. **Anais...** Minas Gerais, UFMG, 2012. Disponível em: <<http://docplayer.com.br/4091902-Auditoria-da-informacao-no-contexto-biblioteconomico-gt-5-abordagem-contemporanea.html>>. Acesso em: 19 jun. 2016.

POBLACIÓN, D. A.. Literatura cinzenta ou não convencional: um desafio a ser enfrentado. **Ciência da informação**, v. 21, n.3, p. 243-246, set./dez. 1992.

Disponível em:< <http://revista.ibict.br/ciinf/article/view/438/438>>. Acesso em: 13 out. 2016.

PONTES, E. N. **A Cauda Longa**: uma aplicação aos fundos de investimentos. 55f. Dissertação (Mestre em Matemática Financeira) – Lisboa Scholl of Economics & Management, 2014. Disponível em:

<<http://www.repository.utl.pt/handle/10400.5/8116>>. Acesso em: 19 jun. 2016.

PONTS, E. T. Sitkis: uma herramienta bibliométrica para el desarrollo del estado de la cuestión. **Textos univerritaris de biblioteconomia i documentació**. n. 28, p. 1-8, jun. 2012. Disponível em: <<http://http://bid.ub.edu/28/tarrats2.htm>>. Acesso em: 27 jun, 2016.

PRICE, J. D. de S. The expansion of scientific knowledge. **Annals of the New York Academy of Sciences**, v. 184, p. 257-259, jun. 1971. Disponível em:

<<http://onlinelibrary.wiley.com/doi/10.1111/j.1749-6632.1971.tb41329.x/pdf>>. Acesso em:13 jun. 2016.

QUONIAN, I.et al. Inteligência obtida pela aplicação de data minig em base de teses francesas sobre o Brasil. **Ciência da Informação**, v.30, n.2, p. 20-28, maio/ago.

2001. Disponível em: <<http://www.scielo.br/pdf/ci/v30n2/6208.pdf>>. Acesso em: 13 jun. 2016.

RAJNEESH; RANA, M.S. Content Analysis and application of Xipf's Law in computer science literatura. In: International Symposium on Emerging Trend and Technologies in Libraries and Information Services, 4, 2015, Noida. **Anais...**Noida: IEEE, 2015. Disponível em: <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7048202&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D7048202>. Acesso em 16 jun. 2016.

RIBEIRO, C. J. S. Big Data: os novos desafios para o profissional da informação. **Informação & Tecnologia**. v.1, p. 96-105, 2014. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/itec/article/view/19380/11156>>. Acesso em 24 out. 2016.

ROSA, A. R.; ALVES, M. A. "For the English To Read": Subalternidade e a Hegemonia da Língua Inglesa nos estudos organizacionais Brasileiros. In: EnANPAD, 2010. **Anais...**Rio de Janeiro, 2010. Disponível em: <http://www.anpad.org.br/diversos/trabalhos/EnANPAD/enanpad_2010/EPQ/2010_EPQ2725.pdf>. Acesso em: 30 jun. 2016.

SAINT CLAIR, G. Ask the customers. **The One-Person Library: a newsletter for librarians & management**, v.11, n.9, p. 1-5, 1995.

SANTAREM, L. G. S. **Caracterização dos pesquisadores em tratamento temático da informação**: um estudo da produção científica por meio da análise de domínio. 112 f. Dissertação (Mestre em Ciência da Informação) – Universidade Estadual Paulista, 2010. Disponível em: <https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/santarem_lgs_me_mar.pdf>. Acesso em: 02 nov. 2016.

SANTOS, R. N. M. dos; KOBASHI, N. Y. Bibliometria, cientometria, infometria: conceitos e aplicações. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, Brasília, v. 2, n.1, p. 155-172, jan./dez. 2009. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/view/21/43>>. Acesso em: 13 jun, 2016.

SEGLIN, P. Why the impact factor of journals should not be used for evaluating research. **British Medical Journal**, v. 314, 1997, p. 498-502. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2126010/pdf/9056804.pdf>>. Acesso em: 10 abr. 2016.

SILVA, E. L.; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. Florianópolis: UFSC, 2005. Disponível em: <http://www.convibra.com.br/upload/paper/adm/adm_3439.pdf>. Acesso em: 19 dez. 2016.

SILVA, M. A. da. **Aplicação de leis de potência para tratamento e classificação de tamanho de empresas**: uma proposta metodológica para pesquisas contábeis.

118 f. Dissertação (Mestre em Controladoria e Contabilidade) – Universidade de São Paulo, Ribeirão Preto, 2008. Disponível em:
<www.teses.usp.br/teses/disponiveis/96/96133/tde.../MarliAuxiliadoraSilva.pdf>. Acesso em: 10 jul. 2016.

SMALL, H. On the shoulders of Robert Merton: towards a normative theory of citation. **Scientometrics**. v.60, n.1, p. 71-79, mai. 2004. Disponível em:
<<http://link.springer.com/article/10.1023/B%3ASCIE.0000027310.68393.bc>>. Acesso em: 18 jun. 2016.

SOY, C. **Auditoría de la información identificar y explotar la información en las organizaciones**. Barcelona: UOC, 2013.

SPINACK, E. **Dicionário enciclopédico de Bibliometria, cienciometria e infometria**. Venezuela: UNESCO, 1996.

SUN, Y.; GILES, C. L. Popularity weighted rankind for academic digital libraries. In: Proceedings of the 29th European conference on IR research, 2007, Berlin. **Proceedings...Roma,2007**, p.605-61. Disponível em:
<https://devaka.ru/files/Popularity_Weighted_Ranking_for_Academic.pdf>. Acesso em 17 out. 2016.

TARGINO, M. das G.; *et. al.* Repositórios institucionais Brasileiros: entre o sonho e a realidade. **Revista FSA**, Teresina, v.11, n.1, p. 117-133, jan./mar. 2014. Disponível em:< <http://www4.fsnet.com.br/revista/index.php/fsa/article/view/320/143>>. Acesso em: 29 out. 2016.

TENNIS, J. T. Com o que uma análise de domínio se parece no tocante a sua forma, função e gênero? **Bjus**, Marília, v.6, n.1, p. 3-15, jan./jun. 2012. Disponível em:
<<https://dialnet.unirioja.es/descarga/articulo/4328123/1.pdf>>. Acesso em: 18 jun. 2016.

URBIZAGASTEGUI, R. A produtividade dos autores sobre a Lei de Lotka. **Ciência da Informação**, Brasília, v.37, n.2, p. 81-102, mai/ago. 2008.

VALENTE, S. B. M. **Análise das estratégias comunicacionais de grifes de luxo na internet**. 169 f. Dissertação (Mestre em Comunicação) – Universidade Estadual Paulista, Bauru, 2009. Disponível em:<http://repositorio.unesp.br/bitstream/handle/11449/89462/valente_sbm_me_bauru.pdf;sequence=1>. Acesso em: 18 jun.2016.

VALÉRIO, E. D.; *et al.* Produção científica em ciência da informação: análise de domínio das dissertações do PPGCI/UFPE. In: Conferência sobre tecnologia, cultura e memória, 2013, Pernambuco. **Anais...Pernambuco**, 2013. Disponível em:
<<http://docplayer.com.br/2497599-Producao-cientifica-em-ciencia-da-informacao-analise-de-dominio-das-dissertacoes-do-ppgci-ufpe.html>>. Acesso em: 18 jun. 2016.

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

2002. Disponível em:

<http://www.cin.ufpe.br/~ajhol/futuro/references/03%23_Da%20bibliometria%20%E0%20webometria_12918.pdf>. Acesso em: 15 jun. 2016.

VANZ, S. A. de S.; STUMPF, I. R.C. Procedimentos e ferramentas aplicados aos estudos bibliométricos. **Informação & Sociedade**, João Pessoa, v.20, n.2, p. 65-75, maio/ago. 2010. Disponível em:

<<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/4817/4358>>. Acesso em: 16 jun. 2016.

VARELA, A.; et al. Rumos da Comunicação Científica diante do Acesso Aberto à Informação. In: XXIV Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação, 2011, Alagoas. **Anais...Alagoas**, 2011. Disponível em:

<<http://repositorio.ufba.br/ri/bitstream/ri/3827/3/639-2600-1-PB.pdf>>. Acesso em: 17 out. 2016.

VO-TRAN, H. Adding Action to the Information Audit. **Journal of Information Systems Evaluation**, Australia, v.14, n.2, p. 271 – 281, 2011. Disponível em: <www.ejise.com/issue/download.html?idArticle=777>. Acesso em 02 jul. 2016.

ZIPF, G. K. **The psycho-biology of language**: na introduction to dynamic philology. Boston: Houghton Mifflin Co., 1935.

APÊNDICE 1 – PROGRAMAS DE ANÁLISE BIBLIOMÉTRICA DE COLABORAÇÃO CIENTÍFICA E DE VISUALIZAÇÃO DE DADOS

NOME DO PROGRAMA	CARACTERÍSTICAS
Bib Excel	auxilia na análise de dados bibliográficos, gerando arquivos de dados que podem ser importados pelo <i>Microsoft Excel</i> ou outro programa que utiliza dados tabelados para processamento. URL: http://bibliothek.univie.ac.at/bibliometrie/
Cite Space	aplicativo em Java para a visualização e análise das tendências, foca em encontrar pontos críticos no desenvolvimento de um campo. URL: http://cluster.cis.drexel.edu/~cchen/citespace/
In-Spire	<i>software</i> de visualização de dados. URL: http://in-spire.pnnl.gov/
Leydesdorff	<i>software</i> de análise de mapeamento de periódicos URL: http://www.leydesdorff.net/
NetDraw	programa de visualização de redes sociais, instalado automaticamente com o Unicet. URL: http://pt.freownloadmanager.org/Windows-PC/NetDraw-GRATUITO.html
Network workbench	análise de rede em larga escala. URL: http://nwb.cns.iu.edu/
Publish or Perish	recupera e analisa citações acadêmicas por meio do <i>Google Scholar</i> . URL: http://www.harzing.com/resources/publish-or-perish
Revis	apoia a seleção e avaliação de qualidade de estudos primários em revisões sistemáticas, pontuando os artigos na qualidade do conteúdo. URL: http://infoserver.lcad.icmc.usp.br/infovis2/Tools
SCImago	periódico e <i>ranking</i> de países, um portal que inclui periódicos e os indicadores científicos dos países desenvolvidos extraídos da base de dados Scopus®. URL: http://www.scimagojr.com/
Sci2 Tool:	ferramenta para o estudo da ciência, suporta análises temporais, geo espaciais, de tópicos. URL: https://sci2.cns.iu.edu/user/index.php
Stikis	aplicativo em <i>Microsoft Access</i> auxilia na análise de dados bibliométricos, gerando arquivos que podem ser utilizados no <i>Microsoft Excel</i> . URL: http://users.tkk.fi/hschildt/sitkis/downloads.html
Vantage Point	ferramenta para a descoberta de conhecimento, auxilia na rápida navegação em resultados de buscas extensas. URL: https://www.thevantagepoint.com/
VisPipeline-Graph	ferramenta para análise de redes sociais, permite a construção e visualização de grafos. URL: http://infoserver.lcad.icmc.usp.br/infovis2/VisPipelineGraphs

(conclusão)

NOME DO PROGRAMA	CARACTERÍSTICAS
<i>VOS viewer</i>	constrói mapas nos dados da rede utilizando técnicas de mapeamento e agrupamento do VOS URL: http://www.vosviewer.com/
Unicet	programa em Windows, que auxilia na geração de redes sociais. URL: http://www.arschile.cl/ucinet/

FONTE: O autor (2016).

**APÊNDICE 2 – ÁREAS DO CONHECIMENTO DA CAPES E PALAVRAS-CHAVE
PRESENTES NA CAUDA LONGA DO TEMA AUDITORIA DA
INFORMAÇÃO**

ÁREA DO CONHECIMENTO	PALAVRAS-CHAVE	
Ciências da Saúde	accident prevention diabetes mellitus disease family health healthcare delivery health health care planning health knowledge health service Health services needs and demand health services research health visitors human infant infant newborn infomationsbeteende midwife midwifery national health service	neonatal nursing network joint guarantee newborn disease newborn nursing newborn nurse nurse clinicians nutrition nutrition surveys Ontologia paediatric anaesthesia patient education patient satisfaction postnatal care pregnancy public health puerperal disorder state medicine vaccine
Ciências Agrárias	environmental protection	veterinary
Ciências Biológicas	female U. S. department of agriculture	Poultry
Ciências exatas e da terra	Australia BM algorithms caribbean region central america computer system database email factores cualitativos	Great Britain intranets latin America petroleum industry pretroleum industry south and central america statistics user interfaces
Ciências Humanas	developing country London 2012	social network analysis United Kingdom
Ciências Sociais e Aplicadas	3D GIS academic libraries accounting Information system accounting policies agency theory android security monitoring anxiety architectural design process article attitude to health attitudes audit audit check items audit committee audit committee effectiveness audit execution measures audit fees audit framework audit information	information strategy Information system Information system audit Information system development Information systems audit Information systems threats information technology governance technology security information transfer information visualization informations behov informations granskning informations resurer informationshantering informationsresuruser infrastructure integral audit of information knowledge integrated Information systems integrated water resource management

		(continuação)
	<p>audit information source audit methodology audit of financial information audit of Information systems audit planning audit procedures audit process audit seniors auditability auditing auditor size auditoria integral de informação- conhecimento auditors badness information behavior Benchmarking business classification business processes change management</p> <p>character library client's information communication communication technologies community development compliance computer audit computer science consultancy consultants consumers content audit contingency planning core banking system corporate intelligence crisis financiera internacional critical information customer satisfaction data audit data collection Data Discovery Data Governance data interpretation Data Lineage data model</p> <p>data privacy Data security data storage equipment</p> <p>data transfer standard data/information stewardship database systems decision audit decision making</p> <p>decision process</p>	<p>intellectual property</p> <p>intelligence audit intelligence systems internal audit IT Governance IT infrastructure library ITC Project job description joint audits knowledge management Knowledge</p> <p>knowledge audit knowledge economy knowledge management knowledge sharing knowledge strategy process knowledge worker Law Leadership learning environment legacy data management legislation Libraries library science library system loan repayment rate local area networks local government location based service Location information Maintenance Management management audit management information management Information system Manager managerial objectives matching allowing errors medical audit medical library Metadados MIL-Q-9858A/AQAP-1 mobile security audit models of care Multi-Pattern matching multiple approximate string match needs analysis network processo normas internacionais de auditoria object audit OLAP Olympic Games open source organisational information modellig</p>

		(continuação)
decision quality	decision theory	organization
digital forensics	document control standard	organization and management
document delivery	document analysis	organization al objectives
domain analysis	Domanin analysis	organization al theory
earnings management	e-commerce	outcome and process
economic factors	economic Information system	assessment
electronic commerce	email audit	outsourcing
email standard file analysis	emergency	Parent
enterprise information management	ESI	Pattern matching
evaluation	event log	Performance measures
evidence	executive involvement	personal information audit
experiment design	financial activities	political factors
financial audit	financial information	politics
financial performance	financial state machine	practice
flexible manufacturing systems	fraud detection	Private information
fraud risk factors	fraud risk factors	private location information
fuzzy neural network	game theory	problem representation
GIS	government performance review act	process control
Hash	Henczel's information audit	process modelling
methodology	high-risk subjects selection system	production control
Housing	importancia relativa	professional judgment of the
industrial management	industrial plants	auditor
industrial plants	infoCage series	protocol analysis
informatics audit	informatics audit	public housing
informatics audit	information	public relations
information	information analysis	Public Sector
information analysis	information and knowledge	public utility company
information architecture	information architecture	quality
information asset registration	information assets	Quality assurance
information assets	information asymmetry	quality control
information asymmetry	information audit methodology	Records management
information audit methodology	information audit system	regional health planning
information audit system	information audits	regulatory compliance
information audits	information behaviour	resource allocation
information behaviour		return on investiments
		Review
		risk assessments
		risk assessmentss
		risk management
		safety policy
		scientific workflows
		security
		Security Governance
		security information
		management
		security of data
		selection of high-risk subjects
		serious game
		service evaluation
		service life cycle
		skills accreditation
		SME
		social networks
		social psychology
		societies and institutions
		software audit
		spatial decision support system
		Standard
		standards for the information audit

	<p>information culture information gain information industry information leakage prevention information leakage prevention system Information literacy information management information management standards information needs information organization s Information policy information process information processing information professionals</p> <p>information provision information resource information resource entities information resource management information resources inventory information retrieval information security information service information sharing</p>	<p>(conclusão)</p> <p>strategic management strategic planning statistical activiries subsystem CLC number</p> <p>Supply chain management systems theory Technology Tetrotechnology texture character the audit process Tiedonhallinta treatment outcome university libraries vector space model web-based Information system audit Websites Work Information systems Vulnerabilities information technology</p>
Engenharias	<p>automated decisions electricity companies electricity industry</p>	<p>method engineering military engineering</p>
Linguística, Letras e Artes	<p>education Multi-language Mixed Processamento Língua Nacional</p>	<p>unlisted words Web Semântica</p>
Multidisciplinar	<p>action research agent assessment bioterrorims case study Cisa certifications CISM certifications cloud offloading colaboratory community renewal data analysis discretionary accruals emerging energy company security forskare group INFOTUR innovation interoperability investments length of stay licensing load balancing methodology MIL-I-45208A/AQAP-4</p>	<p>Models Outcomes PPGIS Questionnaire Rationality Regulatory Research research methodology romanian market Salary Satisfaction Scheduling systemic approach text character TP311 Trie Uinfant use studies user tools Valuation Victoria Visualisation 13F239 X3D XES</p>

FONTE: Elaborado com base na tabela de áreas de conhecimento/ avaliação da CAPES (2012)