

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL – UFRGS
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO – FABICO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO E INFORMAÇÃO –
PPGCOM

Rodrigo Aquino de Carvalho

ANÁLISE DE COCITAÇÃO DE AUTORES:
aspectos metodológicos e comparação com a
análise de cocitação de documentos

Porto Alegre
2019

Rodrigo Aquino de Carvalho

ANÁLISE DE COCITAÇÃO DE AUTORES:

aspectos metodológicos e comparação com a
análise de cocitação de documentos

Tese apresentada ao Programa de Pós-Graduação em Comunicação e Informação – PPGCOM/UFRGS, sob a orientação da Prof^a Dra. Sônia E. Caregnato.

Linha 1: Informação, redes sociais e tecnologias

Porto Alegre
2019

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL – UFRGS

Reitor: Prof. Dr. Rui Vicente Oppermann

Vice-Reitor: Prof^ª. Dr^ª. Jane Fraga Tutikian

FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO – FABICO

Diretor: Prof^ª. Dr^ª. Karla Maria Müller

Vice-diretor: Prof^ª. Dr^ª. Ilza Maria Tourinho Girardi

PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO E INFORMAÇÃO –
PPGCOM

Coordenador: Prof. Dra. Ana Taís Martins Portanova Barros

Coordenadora substituta: Prof^ª. Dr^ª. Nísia Martins do Rosário

C331a Carvalho, Rodrigo Aquino de.

Análise de cocitação de autores: aspectos metodológicos e
comparação com a análise de cocitação de documentos / Rodrigo
Aquino de Carvalho. – 2019.

200 f.

Tese (Doutorado) – Universidade Federal do Rio Grande do Sul,
Programa de Pós-Graduação em Comunicação e Informação, 2019.
Orientadora: Sônia Elisa Caregnato.

1. Cientometria. 2. Análise de cocitação de autores – ACA. 3.
Análise de cocitação de documentos – DCA. 4. Cocitação. 5.
Estudos de citação. I. Título.

CDD: t020

Catálogo na publicação: Rodrigo Aquino de Carvalho – CRB 10/1995.

PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO E INFORMAÇÃO

Faculdade de Biblioteconomia e Comunicação – Campus Saúde

R. Ramiro Barcelos, 2705 | Prédio: 22201

Porto Alegre/RS | CEP: 90035007

Telefone: +55 (51) 33085116

Fax: +55 (51) 33085379

Site: <http://www.ufrgs.br/ppgcom>

E-mail: ppgcom@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL – UFRGS
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO – FABICO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO E INFORMAÇÃO –
PPGCOM

Discente: Rodrigo Aquino de Carvalho (matrícula: 00250901)

Título: Análise de cocitação de autores: aspectos metodológicos e comparação com a análise de cocitação de documentos

BANCA EXAMINADORA DE TESE

Profª Dra. Ely Francina Tannuri de Oliveira – PPGCI/UNESP
(Membro externo)

Prof. Dr. Rene Faustino Gabriel Junior – PPGCIN/UFRGS
(Membro externo)

Profª Dra. Samile Andréa de Souza Vanz – PPGCOM/UFRGS
(Membro interno)

Profª Dra. Ana Maria Mielniczuk de Moura – PPGCOM/UFRGS
(Membro interno)

Profª Dra. Sônia E. Caregnato – PPGCOM/UFRGS
(Orientadora)

Dedico

Aos meus pais: Seu Paulo e Dona Izaura

Ao resto dos “parenti” e aos que torceram por mim

À irmã Rute (em memória)

Em especial aos meus amores: Catarina e Francisco

AGRADECIMENTOS

Agradeço ao Senhor! “Meus fardos são tão leves sobre os ombros Teus”.

À Professora Sônia. Obrigado pela paciência com as minhas limitações e frustrações; pela liberdade na escolha do tema e por me guiar no desenvolvimento da pesquisa; por colocar os meus pés no chão e me oferecer esperança em momentos difíceis; por me tratar como colega “professor” de tão forma respeitosa quando eu não merecia. Obrigado para sempre!

Às professoras Ana, Ely e Samile e ao professor Rene pela avaliação do trabalho. Agradeço a pronta disponibilidade em participar da banca, as considerações e a generosidade de vocês.

Aos demais professores que fizeram parte da minha formação: Alex, Bruno, Célia, João Becker, Mauren e Moises. Gratidão!

Aos amigos de curso no PPGCOM e em outros programas da UFRGS. Foi ótimo voltar a ser aluno e viver momentos que só a leveza de quem faz parte da “turma” pode curtir: cafés, fofocas, apoios, desabafos e muito aprendizado. Foi incrível!

Aos amigos professores do curso de Biblioteconomia da FURG. Agradeço a confiança na liberação para dedicação integral ao doutorado. Tenho muito orgulho de fazer parte dessa equipe e espero corresponder à confiança.

Nesse momento tão difícil da Educação Superior no Brasil, principalmente pelo que se vislumbra, fica um agradecimento especial à UFRGS e à FURG. Não são universidades perfeitas, mas existem e zelam pela excelência e mudam vidas; e eu sou uma dessas vidas. Obrigado às pessoas que fazem essas instituições funcionarem!

Agradeço ao Guilherme pela ajuda na parte da estatística, à Francieli e a Sabrina pela ajuda extração de parte dos dados da pesquisa. Eu não conseguia chegar aqui sem esse apoio!

Por fim, agradeço à minha família. Aos que torceram por mim de longe, principalmente meus pais e meus irmãos. Ao apoio da Catarina nas ausências e nas presenças “ausentes”, nas viagens, nos horários loucos, nas paranoias de prazos e pela revisão do texto. Ao Francisco pelas orações simples e sinceras para o bem do “papai” e para terminar o “trabalho”, e principalmente por me motivar quando eu não achava significado em quase nada. Amo muito vocês! Obrigado por me suportarem e me cuidarem.

Melhor é serem dois [cocitação] do que um [citação] – Eclesiastes 4:9a

RESUMO

CARVALHO, R. A. de. **Análise de cocitação de autores**: aspectos metodológicos e comparação com a análise de cocitação de documentos. Porto Alegre, 2019. 200 f. Tese (Doutorado em Comunicação e Informação) – Programa de Pós-Graduação em Comunicação e Informação, Universidade Federal do Rio Grande do Sul. 2019.

A presente pesquisa tem foco na análise de cocitação de autores (ACA) e na análise de cocitação de documentos (DCA). O objetivo geral é avaliar as implicações das escolhas metodológicas para efetuar cada fase de uma ACA e sua comparação com uma DCA. Os objetivos específicos são: i) propor uma forma de estabelecer um ponto de corte para a criação das matrizes; ii) identificar, quantificar, caracterizar e comparar os dados desconsiderados na fase de seleção de autores/documentos; iii) comparar as frequências de cocitação de autores, considerando os primeiros e todos os autores das referências; iv) comparar as frequências da ACA com a DCA; v) comparar os resultados dos agrupamentos considerando o valor da diagonal, matrizes normalizadas e absolutas e técnicas estatísticas de agrupamento; e vi) caracterizar e comparar um agrupamento da ACA. A pesquisa é descritiva, metodológica e cientométrica. Dois conjuntos de dados de referências foram utilizados: corpus principal (17.992 referências de 421 artigos da área da Organização do Conhecimento e Recuperação da Informação – 2015-2016); e corpus secundário (5.771 referências de 151 artigos da área da Organização do Conhecimento – 2011-2015). Três rankings foram criados com os dados do corpus principal: todos os autores das referências (19.590 autores); primeiros autores (9.338) e documentos (1.715 documentos com duas ou mais citações). Dois rankings foram criados com os dados do corpus secundário: todos os autores (5.333); e primeiros autores (2.938). Três matrizes de cocitação foram criadas com os dados do corpus principal: todos os autores (180x180); primeiro autor (72x72) e documentos (109x109). Um agrupamento com 17 autores desenvolvido com dados do corpus secundário foi analisado. Principais resultados: i) indica-se que o ponto de corte para a criação de matrizes de cocitação considere aproximadamente 20% da soma das citações oriundas de documento citantes, desconsiderando dados absolutos de dispersão (autores e documentos citados em apenas um documento citante); ii) as análises de cocitação a partir dos documentos geram maior dispersão do que as dos autores, pois, para dados do corpus principal, a ACA de todos e dos primeiros autores cobrem 79,33% e 61,76% dos citantes, respectivamente, enquanto a DCA cobre 47,74%; iii) quanto à comparação de uso de todos ou dos primeiros autores das referências, os resultados evidenciam que as abordagens lidam com dados de grandezas diferentes, ainda que oriundos da mesma fonte, ou seja, o uso de todos os autores traz ruídos, mas um ranking de citações considerando apenas as posições secundárias demonstra a força das coautorias no corpus; iv) a análise do ranking de documentos demonstra duas situações: a presença de documentos de autoria não pessoal, e a possibilidade de analisar a cocitação de documentos do mesmo autor; v) a análise fatorial das matrizes da ACA (dados absolutos, dados normalizados pelo r de Pearson, Cosseno de Salton e Índice de Jaccard, com zero, citação total, documento citante e cocitação do autor com ele mesmo na diagonal) indicam resultados muito distintos, mas utilizar citação total na diagonal não é uma boa solução, principalmente para normalizar os dados pelo Cosseno e Jaccard, o excesso de zeros nas matrizes interferem no uso do r de Pearson; e vi) a caracterização de um agrupamento de uma ACA (corpus secundário) demonstrou que agregar informações dos citantes e dos documentos dos autores melhora a capacidade informacional e que a identificação dos pares de cocitação nos parágrafos valida os dados das referências. A pesquisa conclui que estudos de cocitação são estudos de contexto e que os dados descartados são significativos para entender o alcance das propostas. Sugere-se o aprofundamento da análise de cluster, o uso do escalonamento multidimensional e a comparação das matrizes de cocorrência com as matrizes de ocorrência em pesquisas metodológicas futuras.

Palavras-chave: Análise de cocitação de autores – ACA. Análise de cocitação de Documentos – DCA. Estudos de citação. Estudos de cocitação.

ABSTRACT

CARVALHO, R. A. de. **Author cocitation analysis**: methodological aspects and its comparison with document cocitation analysis. Porto Alegre, 2019. 200 f. Tese (Doutorado em Comunicação e Informação) – Programa de Pós-Graduação em Comunicação e Informação, Universidade Federal do Rio Grande do Sul. 2019.

This study focuses on author cocitation analysis (ACA) and document cocitation analysis (DCA). Its main objective is to evaluate the methodological choices made at each stage of ACA and its comparison with a DCA. The specific objectives are: i) to establish a cut off point for the creation of the matrices; ii) to identify, quantify, characterize and compare data discharged in the selection phase of authors/documents; iii) to compare the frequencies of authors' cocitation, considering the first and all the authors in the references; iv) to compare the frequencies of ACA and DCA; v) to compare clustering results considering the value of the diagonal, normalized and absolute matrices and statistical techniques of clustering; and vi) characterize and compare a ACA cluster. The research is descriptive, methodological and scientometric in nature. Two sets of citation data were used: the main corpus (17,992 citations from 421 articles from the Knowledge Organization and Information Retrieval area - 2015-2016); and the secondary corpus (5,771 citations from 151 articles in the Knowledge Organization area - 2011-2015). Three rankings were created with data from the main corpus: all authors of the citations (19,590 authors); first authors (9,338) and documents (1,715 documents with two or more citations). Two rankings were created with data from the secondary corpus: all authors (5,333); and first authors (2,938). Three cocitation matrices were created with data from the main corpus: all authors (180x180); first author (72x72) and documents (109x109). A cluster of 17 authors developed with data from the secondary corpus was analyzed. Main results are: i) the cut off point for the creation of cocitation matrices should be approximately 20% of the sum of citations originating from citing documents, disregarding absolute dispersion data (authors and documents cited in only one citing document); ii) the DCA generate a greater dispersion than ACA, since, for data of the main corpus, the ACA of all and the first authors cover 79.33% and 61.76% of the citing documents, respectively, while the DCA covers 47.74%; iii) regarding the comparison of using all or just the first authors, results showed that both strategies deal with data from different magnitudes, although coming from the same source, that is, the use of all the authors brings noise, but a ranking of citations considering only the secondary positions demonstrates the strength of co-authoring in the corpus; iv) the analysis of document ranking demonstrates two situations: the presence of documents of non-personal authorship, and the possibility of analyzing cocitation of documents by the same author; v) the factorial analysis of ACA matrices (absolute data, data normalized by Pearson's r, Salton's cosine and Jaccard index, with zero, total citation, citing document and cocitation of the author with himself on the diagonal) indicate very different results, but using full diagonal citation is not a good solution, mainly to normalize data by Cosine and Jaccard, the excess of zeros in the matrices interfere in the use of the Pearson's r, as indicated in the literature; and vi) characterization of a cluster of ACA (secondary corpus) showed that adding information from citing documents and authors' documents amplifies the informational capacity and that the identification of the cocitation pairs in the paragraphs validates the data of the references. The research concludes that cocitation studies are context studies and that the data discharged are meaningful to understand the scope of the proposals. We suggest the deepening of the cluster analysis, the use of multidimensional scaling and the comparison of co-occurrence matrices with occurrence matrices in future methodological researche.

Key-words: Author cocitation analysis – ACA. Document cocitation analysis – DCA. Citation studies. Cocitation studies.

LISTA DE ABREVIATURAS E SIGLAS

AC	Análise de <i>Cluster</i> (agrupamentos)
ACA	<i>Authors cocitation analysis</i> (Análise de cocitação de autores)
AF	Análise fatorial
CI	Ciência da Informação
Cs	Cosseno de Salton
DCA	<i>Document cocitation analysis</i> (Análise de cocitação de documentos)
EMD	Escalonamento multidimensional
EC	Estudos de citação
IJ	Índice de Jaccard
OC	Organização do Conhecimento
RI	Recuperação da informação
WoS	<i>Web of Science</i>

LISTA DE QUADROS

Quadro 01	Matriz Epistemológica dos Estudos de citação: enfoques normativo e construtivista	27
Quadro 02	Matriz assimétrica representando o relacionamento entre documentos citantes e documentos citados.....	30
Quadro 03	Matriz quadrada de cocitação de documentos, com número de citação na diagonal	30
Quadro 04	Tipos de contagens para inclusão de autores em ACA	40
Quadro 05	Comparando fatores com ACA e MFTACA (BU et al, 2017)	47
Quadro 06	Fraguemento do índice geral do <i>corpus</i> principal da análise	65
Quadro 07	Distribuição das 12 primeiras posições da matriz de cocitação de autores comparando três valores distintos na diagonal	78
Quadro 08	Relação de coautoria dos autores “White, RW” e “Ruthven, I” em um documento citante	82
Quadro 09	Relação de coautoria dos autores “Boyack, KW” e “Klavans, R” em três documentos citantes.....	83
Quadro 10	Medidas de correlação do <i>ranking</i> de autores do <i>corpus</i> secundário considerando “totais de citações” e “citações por documentos citantes”	89
Quadro 11	Comparação da aplicação do Índice de Jaccard com dados de totais de citação e número de documentos citantes (<i>corpus</i> secundário).....	90
Quadro 12	Medidas de correlação dos rankings de autores do <i>corpus</i> principal considerando documentos citantes, citações totais e cocitação do autor com ele mesmo	99
Quadro 13	Relação aglomerativa do autor <i>Hjørland, B</i> (dados absolutos).....	134
Quadro 14	Artigos que citaram seis ou mais autores do agrupamento analisado (dados do <i>corpus</i> secundário de análise).....	136

LISTA DE FIGURAS

Figura 01	Cocitação e acoplamento bibliográfico	29
Figura 02	Relação das principais unidades de contagem em estudos de cocitação.....	33
Figura 03	Esquema de proximidade de cocitação em quatro níveis.....	33
Figura 04	Etapas de operacionalização de uma ACA	35
Figura 05	Diagrama de fluxo do algoritmo do método da MFTACA (Bu et al, 2017)	46
Figura 06	Distribuição do número de autores dos artigos que formam o corpus principal da análise.....	59
Figura 07	Distribuição do número de autores dos artigos que formam o corpus secundário da análise.....	62
Figura 08	Ilustração das etapas de extração e estruturação dos dados das Referências	64
Figura 09	Estrutura de construção das matrizes com base no valor da diagonal e no uso dos coeficientes de similaridade	80
Figura 10	Proporção dos pares de coautoria com os pares de cocitação (N=26623).....	84
Figura 11	Triângulos isósceles invertidos da produção de informações (Latour) com suas respectivas bases: redução e ampliação	85
Figura 12	Distribuição de valores de documentos citantes como categoria para definição de ponte de corte (todos os autores das referências – corpus secundário).....	92
Figura 13	Distribuição de valores de documentos citantes como categoria para definição de ponte de corte (primeiros autores – corpus secundário)	92
Figura 14	Distribuição de valores de documentos citantes (dois ou mais) como categoria para definição de ponte de corte (todos os autores – corpus secundário).....	94
Figura 15	Distribuição de valores de documentos citantes (dois ou mais) como categoria para definição de ponte de corte (todos os autores – corpus principal)	102
Figura 16	Distribuição de valores de documentos citantes (dois ou mais) como categoria para definição de ponte de corte (primeiros autores – corpus principal)	102
Figura 17	Distribuição de valores de citações dos documentos (duas ou mais) como categoria para definição do ponto de corte em DCA (corpus principal)	103
Figura 18	Triângulos isósceles da redução da informação em estudos de cocitação (duas primeiras fases).....	105
Figura 19	Distribuição da diferença entre o número total de autorias pessoais e autores citados apenas uma vez no corpus	107

Figura 20	Distribuição da diferença entre o número total de referências e documentos citados uma única vez no corpus	107
Figura 21	Triângulos isósceles da redução da informação em estudos de cocitação (dados não analisados)	115
Figura 22	Proporção dos tipos de autorias do <i>ranking</i> de autores (N=39058)	118
Figura 23	Triângulos isósceles da redução da informação em estudos de cocitação (primeiro <i>versus</i> todos os autores)	120
Figura 24	Triângulos isósceles da redução da informação em estudos de cocitação (análise de cocitação de documentos).....	124
Figura 25	Rede e distribuição dos principais documentos citados na análise do agrupamento (04 ou mais citações – corpus secundário da análise)	142

LISTA DE TABELAS

Tabela 01	Distribuição dos artigos do corpus principal de análise por ano de publicação e periódico.....	57
Tabela 02	Distribuição das primeiras 18 posições dos autores dos artigos que formam o corpus principal de análise.....	58
Tabela 03	Distribuição dos artigos do corpus secundário de análise por ano de publicação e periódico.....	60
Tabela 04	Distribuição das primeiras 15 posições dos autores dos artigos que formam o corpus secundário da análise	61
Tabela 05	Estatísticas descritivas dos indicadores das referências (<i>corpus</i> principal – 421 artigos)	68
Tabela 06	Distribuição das 23 primeiras posições (30 ou mais documentos citantes) do ranking de autores que inclui todas as posições das referências	69
Tabela 07	Distribuição das 25 primeiras posições (20 ou mais documentos citantes) do ranking de autores que considera a primeira posição e autoria única	71
Tabela 08	Distribuição das 26 primeiras posições (10 ou mais citações) do ranking de documentos	72
Tabela 09	Estatísticas descritivas dos indicadores das referências (<i>corpus</i> secundário – 151 artigos)	73
Tabela 10	Distribuição das 25 primeiras posições (13 ou mais documentos citantes) do ranking de autores que inclui todas as posições das referências (<i>corpus</i> secundário).....	74
Tabela 11	Distribuição das 24 primeiras posições (11 ou mais documentos citantes) do ranking de autores (primeira posição e autoria única – <i>corpus</i> secundário)	75
Tabela 12	<i>Ranking</i> de autores comparando citações totais com documentos citantes do <i>corpus</i> secundário (organizado por citações totais e com 153 posições).....	88
Tabela 13	Ranking de todos os autores das referências comparando valores de documentos citantes, citações totais e cocitação do autor com ele mesmo (<i>corpus</i> principal – 107 posições – 15 ou mais documentos citantes)	97
Tabela 14	Distribuição dos valores de cocitação dos autores com eles mesmos (<i>corpus</i> principal)	98
Tabela 15	Valores de citação e de cocitação dos autores e dos documentos utilizados na ACA e na DCA diante dos artigos citantes (N=421)....	109
Tabela 16	Distribuição dos periódicos, por ano, dos artigos citantes que não cocitam nenhum dos autores e documentos das análises	111
Tabela 17	Comparação dos valores de cocitação dos autores <i>Hjørland, B Croft, WB</i> com dados da matriz e com dados de dispersão	

	intermediária (todos os autores)	113
Tabela 18	Distribuição do <i>ranking</i> de autores a partir da posição secundária na referência (110 posições organizadas a partir de 10 documentos citantes)	119
Tabela 19	Distribuição dos autores que possuem documentos incluídos na DCA, considerando a primeira posição de autoria (nº de documentos e citações totais)	123
Tabela 20	Dados gerais da análise fatorial das matrizes de cocitação exclusiva de todos autores (dados absolutos e normalizados)	126
Tabela 21	Dados gerais da análise fatorial das matrizes de cocitação pura de primeiro autor (dados absolutos e normalizados)	129
Tabela 22	Distribuição dos autores de uma componente (cocitação pura de primeiro autor; zero na diagonal, normalizada pelo r de Pearson)	130
Tabela 23	Dados gerais da análise fatorial das matrizes de cocitação de documentos (dados absolutos e normalizados)	131
Tabela 24	Dados gerais da análise de cluster de todas as matrizes do estudo (dados absolutos e normalizados)	133
Tabela 25	Distribuição de dados relativos aos autores do agrupamento analisado (dados do <i>corpus</i> secundário de análise)	138
Tabela 26	Distribuição de frequências de citações dos documentos citados no agrupamento analisado (<i>corpus</i> secundário da análise)	141

SUMÁRIO

1 INTRODUÇÃO	16
1.1 <i>Problema de pesquisa</i>	20
1.2 <i>Objetivos</i>	21
1.3 <i>Justificativas</i>	22
1.4 <i>Hipótese norteadora</i>	23
2 ASPECTOS CONCEITUAIS E ESTUDOS RELACIONADOS	25
2.1 <i>Os estudos de citação: da citação/referência à cocitação</i>	25
2.1.1 <u>Tipos de cocitação e proximidade da ocorrência</u>	31
2.1.2 <u>Análise de cocitação de autores e documentos: técnica</u>	34
2.1.3 <u>Outros indicadores de relacionamento</u>	37
2.2 <i>Análise de cocitação de autores: estudos metodológicos</i>	39
2.3 <i>Análise de cocitação de autores: estudos aplicados</i>	48
2.4 <i>Análise de cocitação de autores: panorama no Brasil</i>	51
2.5 <i>Análise de cocitação de documentos: estudos recentes</i>	53
3 PROCEDIMENTOS METODOLÓGICOS	55
3.1 <i>Estratégia de busca para coleta de dados dos dois conjuntos de análise</i>	55
3.2 <i>Caracterização do corpus principal da pesquisa: periódicos e autores</i>	56
3.3 <i>Caracterização do corpus secundário da análise: periódicos e autores</i>	59
3.4 <i>Estruturação, limpeza e construção do conjunto de dados extraídos das referências</i>	62
3.5 <i>Caracterização das referências, dos rankings de autores e do ranking de documentos do corpus principal</i>	67
3.6 <i>Caracterização das referências e dos rankings de autores do corpus secundário</i>	73
3.7 <i>Parâmetros de construção e análise das matrizes de cocitação</i>	76
3.8 <i>Diferenciação das matrizes de cocitação com dados de coautoria</i>	82
4 RESULTADOS E DISCUSSÃO	85
4.1 <i>Uma proposta para definir o ponto de corte para estudos de cocitação</i>	86
4.1.1 <u>Discussão com os dados do corpus secundário</u>	87
4.1.2 <u>Discussão com os dados do corpus principal e apresentação das matrizes com valores absolutos</u>	95
4.2 <i>Sobre os dados não analisados na ACA e na DCA: mensuração,</i>	

<i>comparação e discussão sobre a perspectiva informacional</i>	106
<i>4.3 A utilização dos primeiros e de todos os autores das referências em ACA: uma análise empírico-comparativa sobre as vantagens e as desvantagens</i>	116
<i>4.4 DCA: o alcance, a relação com a ACA e as especificidades</i>	121
<i>4.5 Análise comparativa considerando valores da diagonal, matrizes normalizadas e técnicas de estatística multivariada</i>	125
<i>4.6 Uma proposta do agrupamento como unidade de análise em estudos de cocitação</i>	135
5 CONSIDERAÇÕES FINAIS	145
REFERÊNCIAS	149
APÊNDICE 1. Expressões extraídas dos títulos e das palavras-chaves dos 151 artigos que formam o <i>corpus</i> secundário de dados (ocorrência igual ou maior que dois).....	161
APÊNDICE 2. <i>Unitermos</i> extraídos dos títulos e das palavras-chaves dos 151 artigos que formam o <i>corpus</i> secundário de dados (ocorrência igual ou maior que dois).....	162
APÊNDICE 3. Distribuição dos periódicos com até quatro artigos publicados que formam o <i>corpus</i> principal de análise	163
APÊNDICE 4. Distribuição dos periódicos com até dois artigos publicados que formam o <i>corpus</i> secundário de análise.....	164
APÊNDICE 5. Artigos que formam o <i>corpus</i> principal de análise	165
APÊNDICE 6. Artigos que formam o <i>corpus</i> secundário de análise	177
APÊNDICE 7. Autores e documentos das matrizes de cocitação	182
APÊNDICE 8. Recorte da matriz de cocitação de autores (todos – sem coautoria)	191
APÊNDICE 9. Recorte da matriz de cocitação de autores (primeiros).....	192
APÊNDICE 10. Recorte da matriz de cocitação de documentos	193
APÊNDICE 11. Componentes da análise fatorial de duas matrizes de cocitação de primeiro autor.....	194
APÊNDICE 12. Componentes da análise fatorial de uma matriz de cocitação de documentos	196
APÊNDICE 13. Dendrograma de análise de cluster (primeiros autores, dados absolutos e diagonal com citação total)	199
APÊNDICE 14. Dendrograma de análise de cluster (primeiros autores, dados absolutos e diagonal com citação total)	200

1 INTRODUÇÃO

Em *As aventuras do sr. Pickwick*, [Charles Dickens diz] que um homem teria lido a respeito da metafísica chinesa na *Britannica*. No entanto, não havia tal verbete na enciclopédia: “Ele lera o texto sobre metafísica na letra M, e o texto sobre a China na letra C, e **combinara a informação de ambos.**” (GLEICK, 2001, p.396, grifo nosso).

A comunicação é uma etapa fundamental que caracteriza a ciência enquanto prática social. Uma série de disposições permeia a comunicação científica, que é inclusive considerada uma disciplina autônoma. As práticas editoriais (submissão, revisão, publicação etc.), a revisão por pares, a estrutura dos textos, o ato de citar e referenciar são alguns exemplos de processos que caracterizam a comunicação formal no campo científico, que se materializa principalmente pela produção de registros, com destaque especial para os periódicos científicos.

Popper (1975) chama esses registros de conhecimento objetivo, o que permite a inteligibilidade dos três mundos de sua teoria (mundo físico, mundo subjetivo ou das ideias e mundo do conhecimento objetivo ou dos registros da percepção do homem sobre o mundo físico e subjetivo: a patente, o artigo que relata a teoria da relatividade etc.). Ou seja, é possível fazer ciência a partir dos documentos, que são autônomos e passíveis de interpretações e inferências.

Nas diversas áreas do conhecimento há métodos e técnicas que utilizam os documentos na realização de pesquisas, desde as revisões sistemáticas e metanálises na área das Ciências da Saúde, às pesquisas documentais na História. A esses, pode-se incorporar também a Bibliometria, que é um conjunto de técnicas e também uma área que, utilizando o conhecimento objetivo publicado, busca, a partir da aplicação de modelos matemáticos e estatísticos (URBIZAGÁSTEGUI ALVARADO, 2007), mensurar a produção de informação diante de algum objetivo pré-estabelecido.

Segundo Glänzel (2003), a bibliometria objetiva atender três grandes vertentes: a) a bibliometria para pesquisadores ou profissionais da Bibliometria, com ênfase em questões metodológicas e aprimoramento de técnicas; b) a bibliometria para disciplinas científicas, com ênfase na necessidade que cada área possui para lidar com seu conhecimento publicado; e c) a bibliometria para política e gestão científica, com ênfase na avaliação da ciência a partir das estruturas que a institucionaliza.

Os estudos de citação (EC) se apresentam como uma ferramenta da Bibliometria para a compreensão da ciência, que neste contexto é chamada de Cientometria, e “[...] é

definida como o estudo da mensuração do progresso científico e tecnológico e que consiste na avaliação quantitativa e na análise das intercomparações da atividade, produtividade e progresso científico” (SILVA; BIANCHI, 2001, p. 6). Assim, os EC podem assumir uma das três vertentes bibliométricas, ou seja, ser do tipo metodológico, discutindo a eficácia e uso de algum índice ou dos motivos que levam a citar; aplicado a alguma área do conhecimento, quando usado na identificação dos seus autores mais relevantes; e voltado para política e gestão científica, quando utilizado na definição de quais periódicos devem ser assinados com os recursos financeiros disponíveis em uma unidade de informação, por exemplo.

Nesse sentido, a presente pesquisa tem foco na análise de cocitação de autores (ACA – *Author Cocitation Analysis*) e na análise de cocitação de documentos (DCA – *Document Cocitation Analysis*), aplicações dos EC que identificam o relacionamento bibliográfico entre entidades citadas e/ou referenciadas. Assim, a cocitação é um indicador de relacionamento derivado da citação, que Glänzel (2003) chama de indicador analítico de dados juntamente com o acoplamento bibliográfico, que estabelece a ligação entre duas entidades (autores, documentos, periódicos, instituições etc.) a partir de um documento citante, ou seja, a entidade citada ‘A’ e a entidade citada ‘B’ estão relacionadas pelo documento citante “Z”. Os estudos de acoplamento relacionam dois documentos citantes a partir de uma mesma entidade citada (GRÁCIO, 2016) e, portanto, tem um olhar para o futuro (*corpus* de análise) a partir das referências.

Assim, pode-se definir a análise de cocitação como a contagem das citações de dois documentos ou autores em um mesmo texto, e essa contagem determina, a partir do contexto do documento citante, uma potencial relação bibliográfica entre essas entidades (documentos/autores) (SPINAK, 1996; CO-CITATION ANALYSIS, c2013; REITZ, c2014), que pode se confirmar pela quantidade da ocorrência em um grande conjunto de documentos.

Estudos de cocitação são geralmente utilizados como técnica de análise de domínio (HJØRLAND, 2002; GRÁCIO; OLIVEIRA, 2013b; LIMA, 2015) ou para visualização e mapeamento de uma área específica do conhecimento (EOM, 2009; SCHNEIDER; LARSEN; INGWERSEN, 2009) produzido e publicado em teses e dissertações, artigos de periódicos, comunicações em eventos etc., mas olhando para o passado, para as referências.

A ACA consiste em seis etapas: a) seleção dos autores; b) recuperação das frequências de cocitação; c) compilação da matriz simétrica com os valores absolutos; d) conversão da matriz anterior em uma matriz de correlação ou normalizada; e) análise multivariada da matriz normalizada; e f) interpretação e validação dos agrupamentos identificados (McCAIN, 1990). A DCA segue os mesmos procedimentos, exceto pela unidade de contagem (EOM, 2009) que, nesse caso, é o documento. Essas etapas serão discutidas mais detalhadamente na revisão de literatura.

Assim, uma série de questões se apresenta para a realização de um estudo tradicional de cocitação (ou seja, aquele que estuda a cocitação que ocorre na lista de referências), a partir dos objetivos que se pretende alcançar. Cada escolha, desde a unidade de contagem (documento, autor e periódico), que define o tipo de análise de cocitação que será realizada, até a técnica de análise multivariada utilizada e a forma empregada para rotular os agrupamentos, implica em perda, ambiguidade, omissão e/ou ruído informacional, que interfere diretamente na análise do contexto estudado. Assim, o problema de pesquisa se desenha a partir dessa perspectiva, como será demonstrado adiante.

No caso da ACA, a seleção dos sujeitos que entrarão em um estudo incide em definir se todos os autores das referências serão considerados ou apenas os primeiros (SCHNEIDER; LARSEN; INGWERSEN, 2009; PERSON, 2011; CARVALHO; CAREGNATO, 2016). Esta decisão pode, por exemplo, tirar da análise autores muito produtivos ou reconhecidos (citados) que publicam em coautoria, mas não estão na primeira posição das referências citadas.

A recuperação da frequência de cocitação entre os autores ou documentos implica em definir a proximidade da ocorrência, que pode ser compilada a partir da lista de referência (proximidade por artigo) ou do texto propriamente dito (proximidade por seção do documento, parágrafo ou sentenças) (LIU; CHEN, 2012). Isso demonstra a necessidade de observar como os documentos são mencionados no corpo do texto, pois a lista de referências iguala documentos mencionados mais de uma vez com aqueles mencionados apenas uma vez (BU *et al*, 2017) ou mesmo nenhuma (CARVALHO *et al*, 2018).

Essas zonas de permanente discussão nos estudos de cocitação motivaram o problema desta pesquisa. A partir de algumas “[...] questões práticas, isto é, aquelas que pedem soluções concretas [...]” (BRAGA, 2005, p. 292), relacionadas às fases desse

tipo estudo de cocitação, enquanto método cientométrico (proximidade da cocitação, contar todos os autores etc.), surgiram duas inquietações ligadas ao que é efetivamente analisado, considerando inclusive os dados não utilizados, e também à necessidade de comparação com uma DCA que emprega o mesmo *corpus* de trabalho.

Além disso, o conceito de informação proposto por Latour (2006) também ampara a problematização proposta, quando o autor afirma ser a informação uma relação estabelecida entre dois lugares, sendo que o centro negocia o que retira da periferia, que pode ser exemplificado por um pesquisador (centro) e o seu trabalho em extrair informações de seu objeto de estudo (periferia). Assim, completa o autor que “[...] a produção de informações permite, pois, **resolver de modo prático, por operações de seleção, extração, redução**, a contradição entre a presença num lugar e a ausência desse lugar” (LATOUR, 2006, p. 24, grifo nosso). O autor ilustra essa noção a partir de um artigo da revista *Nature* que apresenta a sequência do DNA (*deoxyribonucleic acid*) humano, afirmando ser um absurdo considerar o desenho a expressão da realidade, pois a capacidade informativa da representação de todo o trabalho laboratorial é um caminho indireto para o gene (redução/ampliação). Essa é uma condição que pode ser estendida aos EC, sendo um caminho indireto para revelar padrões, tendências e particularidades de conjunto específico de documentos.

Latour (2006) ilustra essa condição e contradição de produção de informações por meio de triângulos isósceles invertidos e sobrepostos, pois se no vértice da redução há compatibilidade, padronização, texto, cálculo, circulação e universalidade relativa (exemplos apresentados na figura pelo autor para ilustrar o trabalho de produção de informações), há no vértice da ampliação localidade, particularidade, matérias, multiplicidade e continuidade (exemplos que ilustram a relação da produção da informação com o que se pretende informar).

Entre outros vários conceitos de informação aceitos na CI, o apresentado por Silva (2006) complementa a explicação de Latour e colabora na problematização do tema da informação apresentada em estudos de cocitação. O autor diz que a informação não é uma evolução de dado nem o que evolui para conhecimento, mas se apresenta como um “processo no qual o dado e o conhecimento unem-se em representação [...]”, não sendo (no contexto do ciberespaço) uma estrutura estabelecida ou um produto (SILVA, 2006, p.199).

Assim, a partir de cada escolha parece emergir uma relação entre os dados que serão utilizados em uma análise de cocitação e os que não serão ou que deixam de ser utilizados a partir de cada etapa, uma vez que as técnicas buscam apresentar as ligações mais frequentes entre os autores (generalizações) em um determinado conjunto de documentos sobre uma área do conhecimento.

Uma discussão sobre a informação que um estudo dessa natureza gera como resultados (analisados ou não) mostra-se necessária. Assim, a inicialmente contraditória relação entre generalização e verificação (BACHELARD, 2004) pode revelar novos problemas ou ao menos, diante um trabalho exaustivo de descrição dos dados (LATOUR, 2006), demonstrar efetivamente o que está sendo medido diante de um contexto mais amplo.

A dialética do conhecimento (GADOTTI, 1984) também pode oferecer bases, a partir de alguns princípios, que amparam um entendimento mais adequado do que ocorre com os resultados de uma pesquisa científica. O princípio da “passagem da quantidade à qualidade” do método dialético (KONDER, 1998) parece aplicável para os estudos de cocitação, ou seja, definir, identificar ou classificar a extensão qualitativa que ocorre em cada escolha nas fases de operacionalização da técnica, determinando por consequência quais seriam as diferenças meramente quantitativas (que não geram resultados amplamente significativos), pode colaborar na contextualização mais próxima de adequada para um estudo. Espera-se, portanto, que os dados empíricos dessa pesquisa possam ser base para uma crítica aos objetivos dos estudos de cocitação e para escolhas mais conscientes por parte dos pesquisadores.

A seguir apresenta-se o problema da pesquisa com base nos elementos discutidos.

1.1 Problema de pesquisa

Tendo em vista as escolhas necessárias para efetuar cada etapa de uma ACA e a sua comparação com uma DCA, surge a questão principal da pesquisa: o que é

efetivamente medido na análise de cocitação, levando em conta os dados utilizados, desconsiderados, perdidos e omitidos nas análises?

A questão secundária da pesquisa é: como objetivar a fase de validação e interpretação de uma ACA, na medida em que os agrupamentos gerados parecem pouco informativos e apresentam grande dificuldade de interpretação?

No próximo tópico seguem os objetivos da pesquisa.

1.2 Objetivos

O objetivo geral do trabalho é avaliar as implicações decorrentes das escolhas metodológicas necessárias para efetuar cada fase de uma ACA e sua comparação com uma DCA.

Os objetivos específicos são:

- a) Propor uma forma para estabelecer um ponto de corte para a criação das matrizes que serão analisadas, com base em dois indicadores e considerando a totalidade dos dados;
- b) Identificar, quantificar, caracterizar e comparar os dados desconsiderados na fase de seleção de autores e documentos da ACA e DCA, a partir da proposta de ponto de corte apresentada;
- c) Desenvolver e comparar as frequências de cocitação para ACA, considerando os primeiros e todos os autores das referências (excluindo dados de coautoria);
- d) Comparar as frequências de cocitação de autores (primeiros e todos os autores das referências) com as frequências de cocitação dos documentos, levando em conta principalmente o tamanho das matrizes e os valores de cocitação, bem como a relação entre os autores e seus documentos;
- e) Comparar os resultados, em termos de agrupamento das entidades, da análise multivariada dos dados (análise fatorial (AF) e análise de *cluster* (AC)), considerando os valores das matrizes (dados absolutos e convertidos pelos índices ‘*r* de Person’, ‘Cosseno de Salton’ e ‘Índice de Jaccard’); e

- f) Caracterizar e comparar um agrupamento da ACA (primeiro autores ou todos os autores), entre si, e com dados gerais do *corpus* (secundário), identificando o número de documentos citantes no *corpus* e no agrupamento, autores que citam, documentos citados e proximidade dos pares de cocitação no parágrafo.

1.3 Justificativas

O primeiro contato e o interesse em desenvolver um projeto sobre o tema da cocitação surgiram pela leitura do artigo “*Content-based author co-citation analysis*” (JEONG; SONG; DING, 2014). Parecia inicialmente, por preconceito a trabalhos mais descritivos de citação e falta de contato com a literatura qualificada da área da Cientometria, uma opção “absurdamente pioneira” e “analítica”, algo muito maior do que “apenas” medir. A leitura do artigo foi ao encontro da realização de um pequeno trabalho sobre menções (uso das referências no corpo do texto) em artigos (CARVALHO *et al*, 2018), com resultados publicados em um evento da área da Ciência da Informação em 2015. A intenção era, portanto, aprofundar-se nos textos completos e propor um trabalho “qualitativo” relacionado aos EC, já que a finalidade era realizar um trabalho que fosse além de “estruturar dados”.

Esse motivo se mostrou equivocado e a ACA se apresentou como um tema muito amplo para continuar mensurando, e com qualidade, apesar da literatura extensa publicada em periódicos qualificados. O projeto se desenhou como metodológico e com muitas opções de manipulação de dados e aprendizado sobre técnicas estatísticas e “significados” das citações. Dois trabalhos relacionados a ACA foram publicados (CARVALHO; CAREGNATO, 2016; 2017), respectivamente sobre a utilização dos autores das referências e a proximidade da ocorrência da cocitação.

Lidar com dados refinou as motivações e incluiu a DCA como parte do projeto, pois o trabalho de 2017 e o de 2019¹, em fase de publicação e com resultados dessa pesquisa, demonstraram que os dados de citação, oriundos de muitas publicações, eram omitidos pela ocorrência da cocitação na lista de referências e que autores com muitas

¹ Artigo com resultados dessa pesquisa intitulado: “Interpretação e validação de agrupamentos em análise de cocitação de autores: estudo exploratório”.

publicações tinham alguns documentos de destaque, cuja visualização se perdia parcialmente nos resultados das análises.

As justificativas pessoais amadureceram para justificativas científicas com a geração de dados para os trabalhos publicados e com a revisão de literatura. O contato com a literatura brasileira (MATTOS; DIAS, 2010; GRÁCIO; OLIVEIRA, 2013a, 2013b; 2015; LIMA, 2015; GRÁCIO, 2016; MACHADO, 2016; OLIVEIRA; ALVES, 2017) foi fundamental para entender o tema principal e os relacionados, como o acoplamento bibliográfico. Os trabalhos de Chen, Ibekwe-Sanjuan e Hou (2010), Bu, Liu e Huang (2016), Bu *et al* (2017) evidenciaram que há questões metodológicas que ainda permeiam o tema, principalmente na (i) fase de validação e interpretação dos agrupamentos, (ii) que os resultados finais de estudos tradicionais são pouco informativos, (iii) que a DCA pode colaborar na interpretação dos dados da ACA, (iv) que utilizar apenas o primeiro autor das referências é limitante, e (v) que as informações referentes aos citantes são importantes, pois caracterizam a formação dos agrupamentos.

Portanto, da necessidade de ampliação dos estudos, evidenciada tanto na literatura especializada como durante a realização de estudos empíricos, espera-se que os resultados dessa pesquisa possam colaborar na produção do conhecimento sobre os estudos de cocitação, principalmente no que se refere à descrição dos dados não analisados a partir de cada escolha técnica, para contextualizar o que é efetivamente medido, assim como apresentar dados empíricos da comparação de uma ACA com uma DCA e objetivar, ainda que não no nível do conteúdo, a fase de interpretação e validação de agrupamentos, ampliando as ações já propostas em artigo aceito para publicação com dados dessa pesquisa.

1.4 Hipótese norteadora

Partindo dos antecedentes que indicam que (i) há limitações informativas na ACA tradicional, principalmente pelo foco de análise nas relações dos autores, (ii) que a cocitação de documentos traz elementos comparativos e esclarecedores sobre as relações de cocitação dos autores e que, (iii) por ser uma técnica complexa, cada fase de sua operacionalização exige escolhas que acarretam em perdas ou omissões informativas, a pesquisa apresenta a seguinte hipótese norteadora: a descrição e a

mensuração adequada dos dados desconsiderados em cada fase da operacionalização de uma ACA, desde a seleção dos autores e do contexto de aplicação até o uso das técnicas de estatísticas multivariadas e a forma de rotulagem dos agrupamentos, têm influência semântica/explicativa na análise, pois esclarece e contextualiza o desenvolvimento das unidades chamadas de resultados: os agrupamentos; e assim, se apresentam também como parte desses resultados ou como resultados autônomos, dependendo dos objetivos de uma pesquisa e precisam fazer parte de estudos dessa natureza.

Na próxima seção são discutidos os aspectos conceituais dos estudos de cocitação e apresentadas pesquisas relacionadas, metodológicas e aplicadas.

2 ASPECTOS CONCEITUAIS E ESTUDOS RELACIONADOS _____

A citação é um dos mecanismos mais consagrados e importantes na comunicação formal dos resultados de pesquisa. Citar não é apenas mencionar a quem se refere uma opinião ou argumento sobre um assunto (REITZ, 2014; CUNHA; CAVALCANTI, 2008), pois essa é a condição básica e fundamental da citação, a relação citante/citado (ROMANCINI, 2010), que é evidenciada por um “caminho” ou “endereço”: a referência bibliográfica. A citação carrega no seu escopo elementos sociais e epistemológicos, que necessitam ser problematizados e aprofundados teoricamente (SILVEIRA, 2016). Assim, já não se busca unicamente quantificar quem ou qual documento é mais citado, mas objetiva-se também evidenciar os motivos e as condições que promovem esses resultados.

Está, portanto, no objetivo dos EC entender a citação e tudo que deriva desse fenômeno, inclusive a cocitação. Na revisão, em um primeiro momento, se discute os EC, principalmente seus objetos de estudos, a citação e a referência, até o estabelecimento de uma relação com a cocitação, e depois o foco da revisão é exclusivamente no tema específico da tese.

2.1 Os estudos de citação: da citação/referência à cocitação _____

Os EC se baseiam na relação entre os registros do conhecimento científico, e essa relação se materializa essencialmente, mas não exclusivamente, em duas unidades de análise: a citação e a referência (SILVEIRA; BAZI, 2009). A *indissociabilidade* dessas unidades, a que se referem Silveira e Bazi (2009), não é absoluta empiricamente, pois há documentos referenciados e não citados e citações não referenciadas (CARVALHO *et al*, 2018). De qualquer forma, é a partir desses recursos que a ciência se desenvolve, pois cada documento científico publicado – e a ciência só existe se for comunicada – precisa indicar ao final do texto a lista das fontes informacionais utilizadas, que mostra não apenas o caminho dos vínculos intelectuais entre o citante e o citado no corpo do texto, mas antes de tudo atribui reconhecimento ao que já foi produzido anteriormente, em termos de teorias, métodos e resultados, interligando, segundo Silveira e Caregnato (2017a), a comunidade e a literatura científica, por meio

de elementos objetivos e subjetivos, advindos da apropriação e uso do conhecimento disponível. Portanto, o documento científico deve evidenciar o que existe sobre um determinado assunto (citações/referências) e o que propõem de “novo” diante do todo ou de parte desse material utilizado.

Silveira e Caregnato (2017a, p. 148, grifo nosso) afirmam que as citações e as referências são moldadas pelas práticas dos pesquisadores, e juntas proveem elementos necessários “[...] para a preservação e conservação da memória científica, não somente por promover o acesso, o armazenamento, a organização e a disseminação de conhecimentos, mas [...] **por permitir a visualização da apropriação e uso desses [...]**”, colaborando enquanto recursos nos fluxos de informação da produção na ciência, remontando a linhagem histórico-epistemológica e a institucionalização das ciências.

Nesse sentido, a citação e as referências se apresentam como elementos objetivos e subjetivos dos EC, no primeiro caso há os registros e seus rastros (autorias, títulos, ano etc.) e no segundo caso há as percepções e as intenções de quem cita, a partir dos objetivos do seu texto e do contexto que o cerca (SILVEIRA; CAREGNATO, 2017a).

Os EC se desenvolvem, portanto, a partir dessas duas perspectivas. Os estudos teóricos, que buscam evidenciar o caráter científico e epistemológico desse campo (CRONIN, 1984; SILVEIRA; CAREGNATO, 2017b; TAHAMTAN; BORNMANN, 2018), para entender e identificar as intenções e razões genérico-classificadoras por trás das citações, a discussão da pertinência e uso de algum índice ou do resultado e construção de um *ranking* e suas relações com outras áreas. Os estudos aplicados em áreas específicas (VANZ; CAREGNATO, 2003) servem para medir o impacto e a visibilidade de autores e de periódicos, assim como outras fontes utilizadas, idioma etc., e isso permite identificar pesquisadores pioneiros e emergentes em uma área e as temáticas de pesquisa consolidadas e as novas, como as razões e as influências para citar etc.

Dois enfoques teóricos epistemológicos se destacam para compreender a dimensão dos EC, o normativo e o construtivista (SILVEIRA; CAREGNATO, 2017b; TAHAMTAN; BORNMANN, 2018), que podem ser estendidos aos estudos de cocitação. Os trabalhos de Silveira e Caregnato (2017b) e de Tahamtan e Bornmann (2018) sistematizam esses enfoques a partir de diversos autores que estudam a citação como

fenômeno científico e sociológico desde a década de 1960, com destaque para os trabalhos pioneiros de Merton e Gilbert, publicados em 1973 e 1977, respectivamente.

O Quadro 1 apresenta essas dimensões dos EC, comparando os enfoques teóricos apresentados, a partir de seis categorias: dimensão, objetos, variáveis, objetivos, métodos e técnicas e produtos.

Quadro 01. Matriz epistemológica dos estudos de citação: enfoques normativo e construtivista.

	Enfoque Normativo	Enfoque Construtivista
Dimensão	Produtiva	Discursiva
Objetos	Referências e seus elementos: a) constitutivos (registros referenciados); b) relacionados (registros que referenciam); e c) sociogeográficos (registros institucionais, espaciais e geográficos de quem referencia).	Razões das citações e suas relações com: a) os parágrafos; b) o texto; c) o contexto da publicação; e d) a especialidade temática.
Variáveis	Fatores relacionados: a) às indicações de produção e citação; b) às variações de tempo; c) às dinâmicas de produção do conhecimento; e d) às demarcações sociais e geográficas.	Fatores relacionados: - às formas narrativas e retóricas; - aos gêneros e tipologias textuais; - às filiações e marcas teóricas e metodológicas; e - aos contextos de produção e citação.
Objetivos	- Mapear e organizar as referências, seus elementos e suas relações; - Elaborar e comparar metodologias e índices quantitativos; e - Diagnosticar e analisar os campos e domínios por meio de representações.	- Levantar e analisar os motivos de citação dos cientistas, seus recursos discursivos e as formas e modalidades de apreensão do conhecimento; e - Elaborar e aplicar estruturas classificatórias dirigidas aos efeitos de persuasão dos cientistas e seu entorno social.
Métodos e Técnicas	Método bibliométrico baseado nas técnicas de: a) análise de citação; e b) análise de redes sociais.	- Análise de conteúdo; e - Análise de assunto.
Produtos	- Construção de instrumentos e procedimentos metodológicos e modelos de análise e avaliação; - Produção de rankings e índices normalizados, absolutos e relativos; e - Geração de indicadores e representações gráficas de comportamentos da produção científica de campos e domínios.	- Elaboração de instrumentos metodológicos voltados para a análise de estruturas textuais científicas; e - Produção de sínteses gerais e analíticas dos efeitos persuasivos dos cientistas e dos grupos sociais.

Fonte: Adaptado de Silveira e Caregnato (2017b, p. 259 e p. 267-268).

O enfoque normativo é mais simples, essencialmente atribui as citações recebidas ao reconhecimento do valor do texto pelos pares do(s) autor(es). Com base nele, originou-se uma geração de indicadores quantitativos, produzidos a partir das listas de referências e usados para avaliação da atividade científica. O enfoque construtivista, por outro lado, é analítico e procura evidenciar mais significado dos mecanismos de citar e referenciar, relacionando o conteúdo citado com o contexto de citação e os atores envolvidos. Vale salientar que os enfoques não são excludentes, ainda que a abordagem construtivista faça uma crítica ao pressuposto do enfoque

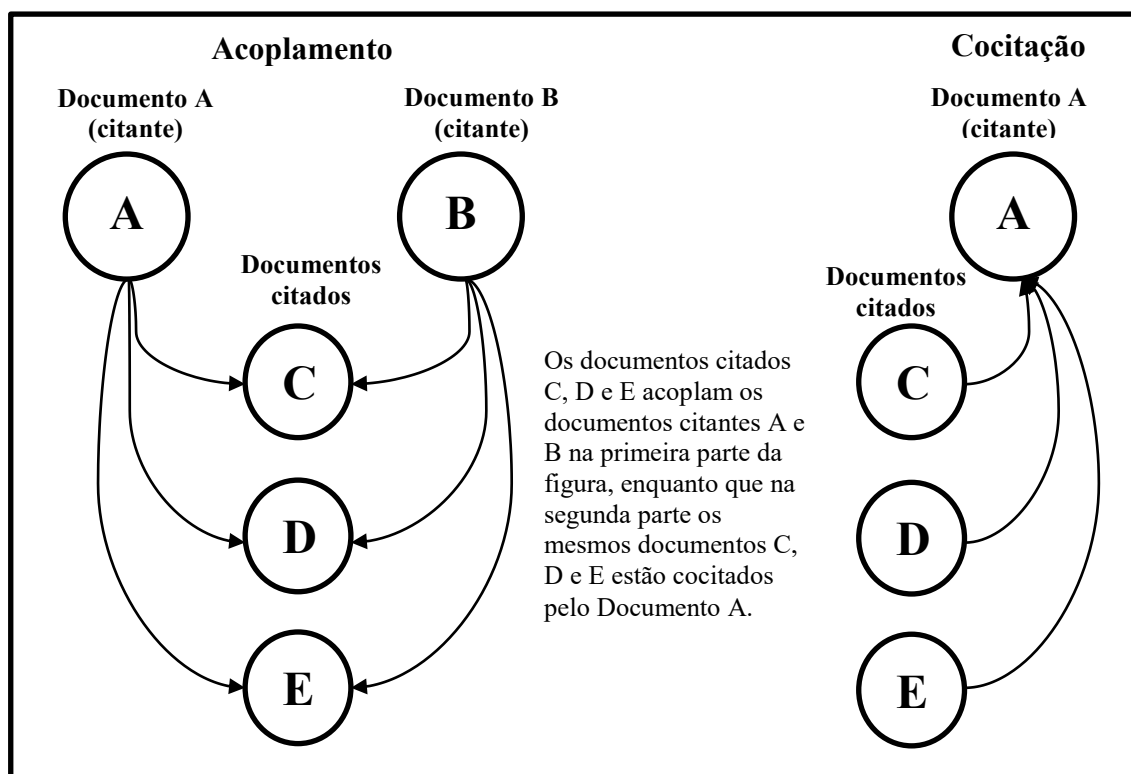
normativo, que trata o uso de citações como recompensa ou reconhecimento (TAHAMTAN; BORNMANN, 2018) e surja como instrumental analítico para responder demandas que os EC necessitavam e não tinham no enfoque normativo.

Portanto, os EC assumem diversas facetas teóricas e aplicadas que se evidenciam na literatura da área, inclusive com periódicos especializados dedicados ao tema, mas a Ciência da Informação (CI) não tem o monopólio de estudos dessa natureza, que são encontrados em todas as áreas do conhecimento. Desde os motivos apresentados por Garfield (1962) para citar, em texto pioneiro sobre automatizar a indexação de artigos, passando por diversos artigos de revisão (LIU, 1993; VANZ; CAREGNATO, 2003; BORNMANN; DANIEL, 2008; MARX; BORNMANN, 2016; TAHAMTAN; BORNMANN, 2018), trabalhos com abordagens teóricas e críticas (CRONIN, 1981, 1984; LEYDESDORFF *et al*, 2016; SILVEIRA; CAREGNATO, 2017a, 2017b), relações com outras áreas do conhecimento (HJØRLAND, 2013), estudos aplicados etc., os EC se consolidam, mas também se apresentam com muitas possibilidades de novas pesquisas.

A cocitação é um exemplo dessas possibilidades de pesquisa em EC, especialmente para mapeamento da ciência em domínios específicos. Trata-se de um indicador de relacionamento derivado da citação, mas efetivado por estudos realizados a partir das referências. Small (1973) publica as bases conceituais desse indicador de relacionamento bibliográfico, diferenciando-o de um antecessor, o acoplamento bibliográfico (KESSLER, 1963a, 1963b).

Cocitação e acoplamento bibliográfico se fundamentam na relação entre citante e citante (acoplamento), através de um citado, e entre citado e citado (cocitação), através de um citante. Ou seja, o acoplamento bibliográfico relaciona entidades citantes (documentos ou autores, por exemplo), a partir de uma entidade citada (KESSLER, 1963a; GLÄNZEL, 2003), portanto, tem um olhar para o ‘futuro’, enquanto a cocitação relaciona duas entidades citadas a partir de uma entidade citante (SMALL, 1973), portanto, tem um olhar para o passado. A Figura 1 ilustra a relação desses indicadores de relacionamento.

Figura 1. Cocitação e acoplamento bibliográfico.



Fonte: Adaptado de Garfield (2001, p.3).

A matriz apresentada no Quadro 2, que é assimétrica e de ocorrência (LEYDESDORFF; VAUGHAN, 2006), representa o relacionamento de publicações e suas referências, e mostra, com informações reais de um conjunto de dados utilizados nessa pesquisa e em trabalho derivado dela (CARVALHO, R. A.; CAREGNATO, 2017), que tanto a cocitação como o acoplamento partem de uma mesma base relacional, como já ilustrado na Figura 2. O documento citado “*What is knowledge organization?*” acopla os documentos citantes representados pelos rótulos “201102”, “201108”, “201208”, “201210” e “201302”. Da mesma forma, os documentos “Ref. 1”, “Ref. 4”, “Ref. 5”, “Ref. 7” e “Ref. 9” estão cocitados a partir do documento citante “201102”. Vale salientar que essa matriz é binária (utilização de zero e um), portanto de ocorrência, pois os documentos não podem ser citados mais de uma vez pelo citante, diferente da citação de autores, que podem ter difentes trabalhos citados.

Quadro 2. Matriz assimétrica representando o relacionamento entre documentos citantes e documentos citados.

Documentos citados (referências)	Documentos citantes								
	201102	201108	201208	201210	201302	201307	201308	201311	201325
<i>Ref. 1 What is knowledge organization?</i>	01	01	01	01	01				
Ref. 2 Domain analysis in information science		01			01				01
Ref. 3 Toward a new horizon in information science		01		01	01				
Ref. 4 Knowledge organization: a new science?	01						01		01
Ref. 5 Systems of knowledge organization for digital libraries	01	01	01		01	01		01	01
Ref. 6 Concept theory			01			01	01	01	01
Ref. 7 The intellectual foundation of information organization	01		01			01	01		
Ref. 8 Fundamentals of knowledge organization				01	01				
Ref. 9 Semantic and knowledge organization	01		01			01		01	

Fonte: dados da pesquisa.

Da matriz de relacionamento apresentada no Quadro 2 podem derivar duas matrizes simétricas de co-ocorrência, que indicam com mais clareza as relações de acoplamento ou cocitação em um contexto de análise, pois consideram a força da relação a partir de pares das entidades, ou seja, citante com citante ou citado com citado. O Quadro 3, desenvolvido a partir de dados do Quadro 2, apresenta um exemplo de uma matriz de cocitação de documentos, ou seja, uma das matrizes de coocorrência possíveis.

Quadro 3. Matriz quadrada de cocitação de documentos, com número de citação na diagonal.

	Ref. 1	Ref. 2	Ref. 3	Ref. 4	Ref. 5	Ref. 6	Ref. 7	Ref. 8	Ref. 9
Ref. 1	05	02	03	01	04	01	02	02	02
Ref. 2		03	02	01	03	01	00	01	00
Ref. 3			03	00	02	00	00	02	00
Ref. 4				03	02	02	02	00	01
Ref. 5					07	04	03	01	04
Ref. 6						05	03	00	03
Ref. 7							04	00	03
Ref. 8								02	00
Ref. 9									04

Fonte: dados da pesquisa.

Assim, uma matriz de ocorrência de documentos citados com documentos citantes gera uma matriz de frequência de cocitação, que evidencia a força dos pares das entidades analisadas. Vale salientar que no exemplo apresentado (Quadro 3), o valor da diagonal da matriz quadrada é o número de citações de cada documento, também retirados dos dados apresentados no Quadro 2.

Portanto, pode-se definir, a partir do enfoque normativo dos EC, a cocitação como a relação entre duas entidades (autor, documento, periódico, etc.) citadas em um documento (SMALL, 1973). Essa relação, dependendo do contexto e do volume de dados analisados, pode se mostrar significativa se a frequência de cocitação entre as entidades for alta. Assim, as análises de cocitação se concentram em identificar, medir e interpretar um grupo de entidades cocitadas em um contexto específico, como já destacado.

Hjørland (2013) apresenta a análise de cocitação como uma técnica dos EC que servem como abordagem para a Organização do Conhecimento, enquanto Small (1973) introduz a nova medida com vistas a estudar a estrutura das especialidades científicas ou mapeá-las, como define o próprio Small (1977), Eom (2009) e Chen (2017).

A cocitação potencializa, portanto, a citação e as referências para um nível analítico, ainda que não discursivo, como no enfoque construtivista dos EC, pelo menos nos trabalhos mais tradicionais e normativos e na sua concepção básica.

Nos próximos tópicos serão apresentados: os tipos de cocitação, a partir das unidades de contagem; cada etapa da técnica da análise de cocitação realizada através de autores e documentos; e outros indicadores de relacionamentos e estudos de ACA e DCA metodológicos, aplicados e publicados no Brasil, para aprofundar o entendimento sobre esse indicador de relacionamento de citação.

2.1.1 Tipos de cocitação e proximidade da ocorrência

Os tipos de estudos de cocitação se desenvolvem a partir de duas perspectivas: a unidade de contagem dos pares, que advém dos elementos que formam uma referência bibliográfica ou outro metadado identificável a partir desses elementos; e a proximidade da relação, ou seja, a parte do texto citante onde ocorre a cocitação, pois dois

documentos cocitados na lista de referências podem não ser cocitados na mesma seção do documento citante, por exemplo.

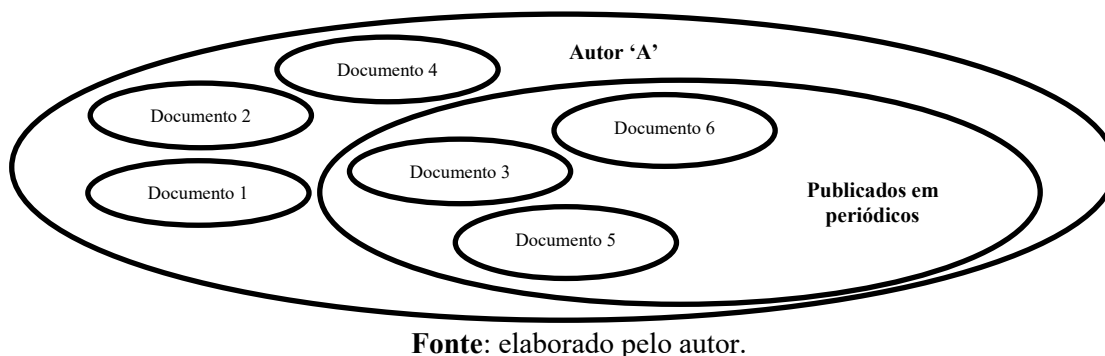
Os principais elementos de uma referência bibliográfica que servem para estudos de cocitação são os autores, os títulos (que representam o documento) e, no caso do contexto científico, os periódicos. A partir deles, emerge uma tipologia pela unidade de contagem: a análise de cocitação de autores (WHITE; GRIFFITH, 1981, McCAIN, 1990; ROUSSEAU; ZUCCALA, 2004; EOM, 2009, MATTOS; DIAS, 2010), a análise de cocitação de documentos (SMALL, 1973, 1977; CHEN *et al*, 2005; ÖZÇINAR, 2015; TRUJILLO; LONG, 2018) e a análise de cocitação de periódicos (McCAIN, 1991; TSAY; XU; WU, 2003; HU *et al*, 2011; MACHADO, 2015). Desenvolver estudos de cocitação a partir de outros elementos implícitos em uma referência bibliográfica é tecnicamente possível, como país ou região de origem ou as instituições dos autores, que são elementos amplamente utilizados em estudos de colaboração científica.

Cada tipo de análise de cocitação tem suas especificidades, que as diferenciam diante dos objetivos de aplicação de cada pesquisa. Mas é importante salientar que a unidade bibliográfica, que é representada no documento, é única e as relações de cocitação de autores e periódicos ocorrem a partir dos documentos, ainda que essa não seja a unidade de contagem escolhida para uma determinada aplicação.

Assim, quando se escolhe o autor, por exemplo, para realizar um estudo de cocitação, toda a sua produção é evidenciada, o que torna os dados mais robustos, mas também traz mais ruídos do que um estudo de cocitação pelo documento.

A Figura 2 ilustra a relação entre as unidades de contagem mais utilizadas em estudos de cocitação. A ACA utiliza todos os rastros de publicação de um autor citado em um conjunto específico de publicações e é por natureza mais extensiva que a DCA e que a cocitação de periódicos. A DCA pode incluir mais de uma publicação de um autor, desde que haja volume de dados para isso e como não existe restrição quanto ao tipo de documento tem um alcance diferente de um estudo de cocitação de periódicos. A análise de cocitação de periódicos se concentra na citação de artigos e dependendo da área de aplicação pode ter o alcance mais restrito.

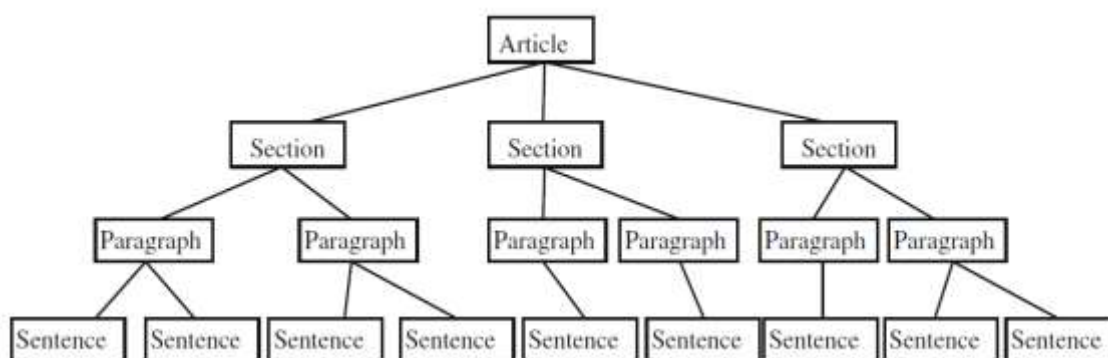
Figura 2. Relação das principais unidades de contagem em estudos de cocitação.



Fonte: elaborado pelo autor.

Escolhida a unidade de contagem, que caracteriza o tipo de análise que será realizada, é momento de escolher o que os autores Liu e Chen (2012) chamam de proximidade da ocorrência da cocitação, que é dividida em quatro níveis, da menor para a maior distância. São eles: sentença ou frase; parágrafo; seção do documento; e artigo/documento. Os autores definem que a cocitação por proximidade em artigo ocorre quando duas referências são citadas em seções distintas de um mesmo documento e não necessariamente na lista de referências. Essa distinção é apresentada porque os autores retiram as ocorrências de cocitação do texto completo do *corpus* analisado, sem considerar a lista de referências. A figura abaixo ilustra a relação da proximidade que os autores utilizam.

Figura 3. Esquema de proximidade de cocitação em quatro níveis.



Fonte: Liu e Chen (2012, p. 498).

O texto de Liu e Chen (2012) compara a extração das ocorrências nos quatro níveis simultaneamente, ou seja, a contagem começa pela identificação da ocorrência na sentença até chegar ao nível do artigo, sendo que se a cocitação ocorre no nível da

sentença, a contagem não é considerada para os outros níveis. Mas há a possibilidade de realizar estudos considerando, por exemplo, a ocorrência da cocitação na seção do documento e, assim, as ocorrências no parágrafo e na frase contam para esse tipo de proximidade. Carvalho e Caregnato (2017) apresentam um trabalho em que a proximidade da cocitação por sentença é considerada na proximidade por parágrafo.

Além do aspecto da proximidade é importante mencionar que grande parte dos estudos de cocitação é realizada com os dados das referências bibliográficas, o que faz com que sejam usualmente chamados de tradicionais (LIU; CHEN, 2012; JEONG; SONG; DING, 2014; BU *et al*, 2017). Teoricamente não há diferença entre a proximidade por artigo, como propõem Liu e Chen (2012), e o que se pode chamar de proximidade nas referências ou estudo tradicional, exceto que, extraíndo dados do texto completo há a possibilidade de inclusão de citações não referenciadas e a ausência de referências não citadas, o que empiricamente ocorre.

O presente projeto é focado na ACA e a DCA se apresenta como uma possibilidade de análise dos agrupamentos que serão formados, para atender à questão principal da pesquisa. O estudo é majoritariamente tradicional e a proximidade por parágrafo será utilizada como um elemento de análise das relações de cocitação (CARVALHO; CAREGNATO, 2017). No próximo tópico são apresentadas as etapas de operacionalização da ACA e da DCA.

2.1.2 Análise de cocitação de autores e documentos: técnica

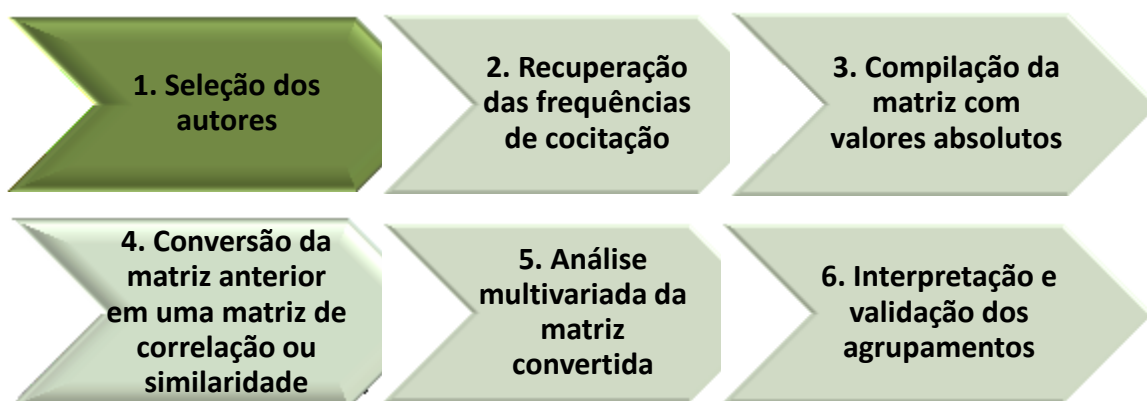
A operacionalização da ACA tradicionalmente se realiza a partir de seis passos, segundo McCain (1990), e exceto pelas necessidades de escolha que a contagem por autores exige, o procedimento é o mesmo para a DCA (EOM, 2009).

Apesar de o primeiro passo ser a seleção dos autores ou documentos, grande parte dos estudos aplicados e metodológicos de ACA e de DCA parte de um contexto específico, seja das disciplinas científicas, como a Ciência da Informação (WHITE; MCCAIN, 1998; ZHAO; STROTMANN, 2008b), a Engenharia Química (MILMAN; GAVRILOVA, 1993) e a Filosofia (KREUZMAN, 2001), seja de ramos específicos de pesquisa, como “Modelo de negócios” (LI; QIAO; WANG, 2017), “Violência envolvendo mulheres” (MUÑOZ-MUÑOZ; MIRÓN-VALDIVIESO, 2017) e pesquisa

sobre AIDS² (SMALL; GREENLEE, 1989), ou ainda, de periódicos de uma determinada área do conhecimento (PERSSON, 1994; JEONG; SONG; DING, 2014).

A Figura 4 ilustra os seis passos do procedimento que McCain (1990) e Eom (2009) aplicam para a ACA.

Figura 4. Etapas de operacionalização de uma ACA.



Fonte: Adaptado de McCain (1990).

A seleção dos autores parte quase sempre de um contexto, como indicado anteriormente, e os critérios para a seleção podem ser vários, como consultar pesquisadores ou listas, autores com mais citações totais (McCAIN, 1990), optar pelos autores mais citados a partir do número de documentos citantes (GRÁCIO; OLIVEIRA, 2013b) etc.

A recuperação das frequências de cocitação é o próximo passo. A definição de utilizar todos os autores das referências (PERSSON, 2001; ROUSSEAU; ZUCCALA, 2004; SCHNEIDER; LARSEN; INGWERSEN, 2009; CARVALHO; CAREGNATO, 2016) ou apenas os primeiros, como procedem alguns estudos (BU; LIU; HUANG, 2016; BU *et al*, 2017), pode ser definida nessa etapa, dependendo da natureza dos dados, ou ser estabelecida previamente, assim como a proximidade da relação de cocitação (LIU; CHEN, 2012), sendo que a proximidade pela referência, ou tradicional, é a opção mais utilizada. Tanto nessa, como na etapa anterior, a seleção de documentos para uma DCA é similar, exceto pelas implicações que a coautoria traz para a ACA.

² “[...] causada pela infecção do vírus da imunodeficiência humana (HIV é a sigla em inglês). Esse vírus ataca o sistema imunológico [...]”. Disponível em: <http://portalm.s.saude.gov.br/saude-de-a-z/aids>

A compilação da matriz de cocitação é a próxima fase e não consiste apenas em indicar os valores recuperados para cada par de autores/documentos cocitados, mas também em definir os valores da diagonal (McCAIN, 1990; GRÁCIO; OLIVEIRA, 2013b), já que a matriz (Quadro 3) é quadrada (autores vs autores ou documentos vs documentos) e esse dado influencia na conversão da matriz de correlação ou similaridade. Essa é uma discussão metodológica que será apresentada no tópico 2.2

A conversão da matriz principal em uma matriz de correlação ou de similaridade consiste no uso de índices de similaridade que normalizam os dados, sendo que os mais utilizados são o r de Pearson, o Cosseno de Salton (Cs) e o Índice de Jaccard (IJ) (LEYDESDORFF, 2008; GRÁCIO; OLIVEIRA, 2015), mas há também o uso de um coeficiente chamado *Ochiai*, que segundo Zhou e Leydesdorff (2016), em estudo comparativo com o Cosseno, pode trazer resultados similares para normalização de matrizes de cocitação desenvolvidas sem dados confiáveis para atribuir à diagonal. Esse tópico é um dos mais controversos na literatura da CI sobre manipulação de matrizes, mesmo porque os valores de uma matriz de cocitação são valores de mesma grandeza, que em tese não necessitam de normalização (ZHOU; LEYDESDORFF, 2016). Segundo Tan, Kumar e Srivastava (2004), em estudo comparativo de 21 medidas objetivas de análise de associação, Cs e IJ têm propriedades similares de aplicação.

A análise multivariada da matriz de correlação é a próxima etapa, que é normalmente realizada a partir das seguintes técnicas: análise fatorial (AF), análise de *cluster* (AC) e escalonamento multidimensional (EMD) (McCAIN, 1990; EOM, 2009). As técnicas evidenciam os agrupamentos mais pertinentes.

A interpretação e a validação dos agrupamentos compõem a última e mais livre fase da ACA e da DCA. McCain (1990) afirma que essa etapa pode ser realizada comparando os resultados com outros dados, consultando especialistas da área e/ou utilizando validação estatística. Chen, Ibekwe-Sanjuan e Hou (2010) utilizam dados dos citantes e dos documentos citados para interpretar os agrupamentos da ACA, enquanto Bu, Liu e Huang (2016) e Bu *et al* (2017) optam pela inclusão de metadados para objetivar essa fase e ampliar a capacidade informativa dos agrupamentos, que segundo os autores é um dos grandes problemas de estudos tradicionais de ACA.

Além das técnicas estatísticas multivariadas apresentadas, há pesquisas optando pela análise de redes sociais com os dados gerados em estudos de cocitação. Grácio e Oliveira (2013b) comparam as frequências absolutas de cocitação com dados relativos

em um estudo metodológico. Jeong, Song e Ding (2014) usam a técnica em um estudo metodológico de ACA baseada em conteúdo de sentenças. Também empregam a análise de redes sociais os seguintes trabalhos: Lima (2015), em uma pesquisa sobre a análise de domínio no periódico *Knowledge Organization*; Piovezan e Fujita (2015), que estudaram as relações de cocitação de autores na temática “avaliação de indexação”; e Muñoz-Muñoz e Mirón-Valdivieso (2017), que apresentaram as redes de cocitação na área de pesquisa “violência envolvendo mulheres”. O uso da técnica de redes sociais, além de uma opção autônoma de análise, utilizando os dados da matriz com valores absolutos, também pode servir para visualização dos agrupamentos, ainda que alguma técnica de estatística multivariada seja empregada.

Cada etapa apresentada envolve discussões e decisões metodológicas significativas sobre a ACA e sobre a DCA, mas também demonstra a complexidade em realizar os estudos. A seleção dos autores, por exemplo, implica em determinar a utilização de todas as autorias dos documentos (inclusive as omitidas por normas de citação e de referência bibliográfica, regras das revistas e descrição dos registros em bases de dados) ou apenas as primeiras, como já salientado. A recuperação das frequências considera a proximidade da ocorrência da cocitação (referência, artigo, seção do documento, parágrafo ou sentenças) (LIU; CHEN, 2012) e isso implica em observar como as referências são mencionadas no corpo texto (DING et al, 2013; BU et al, 2017; CARVALHO et al, 2018), já que a lista de referências bibliográficas iguala o uso dos documentos, mas há dados que demonstram que alguns documentos são mais utilizados do que outros, além da tendência dos autores mais citados terem mais de um documento utilizado em um mesmo documento citante, como demonstram Carvalho et al (2018). Alguns desses temas serão aprofundados no tópico 2.2.

2.1.3 Outros indicadores de relacionamento

Além dos tipos de cocitação já discutidos anteriormente e do acoplamento bibliográfico, que é um indicador de relacionamento antecedente à cocitação, pois ajuda na sua explicação e também se deriva da citação/referência, há outros tipos de indicadores de relacionamento utilizados na Bibliometria, que servem a propósitos específicos e que se relacionam de alguma forma ao tema desse projeto.

A análise de coocorrência de palavras ou *co-words* é uma técnica de análise de conteúdo, que seguindo os princípios do relacionamento de referências, busca identificar padrões de pares de itens textuais (palavras, termos, frases etc.) em um *corpus* de estudo (tantos nos metadados título, palavras-chaves ou resumo, como no texto completo), para mapear relações de assuntos dentro de alguma área a partir de um índice, que podem ser agrupados por proximidade e visualizados em rede, evidenciando, a partir de diferentes períodos de tempo, a dinâmica de evolução e mudança da ciência (HE, 1999), ou permitindo a comparação com textos tecnológicos (CALLON *et al*, 1983). Essa possibilidade de mapeamento ocorre também com os resultados de uma ACA ou de uma DCA, mas, segundo Ding e Chen (2014), as três técnicas apresentam limitações para identificação de tópicos emergentes de pesquisa.

De qualquer forma, as palavras possuem importante papel nos estudos de cocitação, seja como um elemento isolado que contribui para rotular agrupamentos (BRAAM; MOED; Van RAAN, 1991a, 1991b), ou como técnica própria de relacionamento para mapeamento de disciplinas científicas (KHASSEH *et al*, 2017), oferecendo um contraponto para as análises que os estudos de cocitação apresentam.

O acoplamento bibliográfico já foi discutido anteriormente, e apesar de ser postulado na década de 1960 (KESSLER, 1963a) ainda é utilizado (LUCAS, GARCIA-ZORITA, 2014; FERREIRA, 2018), inclusive com variações, como apresentam Liu (2017), que emprega os títulos das referências para qualificar o acoplamento entre dois documentos citantes. Zhao e Strotmann (2008a), ao invés de acoplar documentos, acoplam os autores citantes em uma variação da técnica chamada *Author Bibliographic-Coupling Analysis* (ABCA), para mapear os autores ativos em comparação com uma ACA tradicional. Ni, Sugimoto e Jiang (2013), por outro lado, apresentam uma variação da técnica chamada *Venue-Author-Coupling*, pela qual observam o acoplamento de periódicos na área da Biblioteconomia e Ciência da Informação, mas a partir dos autores citados e não dos documentos.

A intenção dessas variações do acoplamento bibliográfico é aproveitar as outras unidades de contagem oferecidas por uma referência bibliográfica, que acabam agregando de maneiras diferentes os documentos, oferecendo resultados distintos que enriquecem o mapeamento de uma disciplina científica.

Por fim, vale mencionar a coautoria, que é um dos muitos indicadores de colaboração científica. Esse indicador se assemelha à cocitação não somente pela

unidade de contagem, mas principalmente pela manipulação das matrizes quadradas (autor *versus* autor). Visto que atribuir um valor à diagonal da matriz quadrada, aplicar um índice de correlação/similaridade e utilizar técnicas de estatísticas multivariadas para evidenciar agrupamentos (VANZ, 2009) são práticas comuns e necessárias aos dois indicadores de relacionamento. Vale salientar que outros níveis de colaboração científica (instituições, departamentos, países etc.) lidam da mesma forma com as matrizes, mas a coautoria se evidencia com a particularidade de ter a mesma unidade de contagem da cocitação de autores, como já evidenciado.

2.2 Análise de cocitação de autores: estudos metodológicos

Desde sua concepção, os estudos de cocitação de autores (WHITE; GRIFFITH, 1981) são usados não apenas para o mapeamento de disciplinas científicas, mas também como tema de discussões metodológicas que criticam a técnica, propõem variações, novas formas de interpretar e agregar informações aos agrupamentos, de desenvolver e avaliar as matrizes etc.

Essa parte da revisão aborda especialmente esse aspecto das publicações sobre o tema, com quatro enfoques: a) a forma de contagem dos autores das referências; b) a forma de determinar o número de autores incluídos na análise c) os índices de similaridades utilizados e o valor da diagonal das matrizes; e d) as formas de validar e interpretar os agrupamentos. Esses elementos perpassam por quase todas as etapas da técnica de ACA e se apresentam mais como possibilidades para testar, entender e adequar os resultados do que como itens definitivos que engessam a operacionalização da técnica. De qualquer forma é bom salientar que esses elementos podem implicar não somente em mudanças quantitativas dos resultados, mas também em mudanças qualitativas (KONDER, 1998), que merecem mais atenção na discussão das análises.

Entender o efeito da contagem de todos os autores das referências em estudos de cocitação é uma preocupação necessária. Alguns estudos optam pelo uso apenas dos primeiros autores e entendem que essa escolha limita a pesquisa (BU *et al*, 2017). Rousseau e Zuccala (2004) propõem uma classificação para a contagem dos autores das referências e suas implicações na ACA. O Quadro 4 apresenta o esquema proposto pelos autores.

Quadro 4. Tipos de contagens para inclusão de autores em ACA.

Formas cocitação	Implicações de uso
1 Cocitação pura de primeiro autor O par de cocitação é contado apenas considerando a primeira posição ou autoria única das referências	<ul style="list-style-type: none"> • Adequado para identificar um quadro temático de um domínio ou assunto • As relações entre primeiros autores e únicos de referências são coerentes na demonstração de semelhanças intelectuais
2 Cocitação pura de todos os autores O par de cocitação é contado considerando todas as autorias das referências, independentemente da posição do autor e desconsiderando coautoria	<ul style="list-style-type: none"> • Os autores, independentemente da posição na referência, contribuem para sua área de pesquisa • Dados mais precisos para áreas que produzem muito em coautoria, evidenciando autores produtivos que não ocupam a primeira posição da referência • Mais adequado para avaliar o desempenho dos autores em um domínio
3 Cocitação geral de autor geral O par de cocitação é contado considerando todas as autorias das referências e não fazendo distinção de coautoria, ou seja, relações de coautoria são contados como relação de cocitação	<ul style="list-style-type: none"> • Apresenta dados mais precisos para áreas com altas taxas de coautoria, ainda que não haja distinção da relação entre os autores
4 Cocitação geral de autor indicando coautorias Mesma definição de “cocitação de autor geral”, mas com uma contagem especial, destacando a relação de coautoria	<ul style="list-style-type: none"> • A similaridade intelectual dos autores é mais precisa se apresentado os dados de coautoria • As contagens de coautoria podem ser incluídas nas contagens de cocitação ou contadas separadamente e sobrepostas em uma rede para melhor visualização

Fonte: Adaptado de Rousseau e Zuccala (2004, p. 517).

Zhao (2006) apresenta uma classificação com três categorias para a contagem dos autores em ACA, muito próxima à proposta anterior, mas sem uma categoria especial para contagem separada de coautoria. O autor as denomina como: cocitação de primeiro autor (*first-author co-citation*), que inclui também referências com autoria única; cocitação inclusiva de todos os autores (*inclusive all-author co-citation*), que inclui todos os autores das referências e conta coautoria como cocitação; e cocitação exclusiva de todos os autores (*exclusive all-author co-citation*), que considera todos os autores das referências, mas não inclui coautoria como cocitação.

Persson (2001) indica a necessidade da inclusão de todos os autores para análises de citação, mostrando dados que sustentam uma diferença considerável em um *ranking*, pois grande parte dos autores mais citados trabalha em colaboração. O mesmo estudo observa que o uso de todos os autores não altera significativamente a formação dos pares de cocitação, mas indica também que isso não é justificativa para o não uso da contagem de todos os autores. Já Schneider, Larsen e Ingwersen (2009), utilizando o método “*inclusive all-author cocitation*” (coautoria = cocitação), concluem, diferente

de Persson (2001), que o agrupamento dos autores é mais coerente quando utilizados todos os autores das referências.

Carvalho e Caregnato (2016) mostraram que 67,60% das referências analisadas em um recorte específico (n=3.295) são de autoria única, mas representam apenas 42,65% das autorias do estudo (n=4.807), sendo que 1.774 autorias não participariam de uma ACA que considerasse apenas a primeira posição da referência. Os autores elaboraram dois *rankings* com 37 posições, o primeiro considerando todos os autores e o segundo considerando apenas a primeira posição e autoria única e os dados mostraram uma diferença de oito autores entre as listas (quatro em cada *ranking*), sendo que o valor mínimo de citação considerando todas as autorias e primeira posição/autoria única foram 10 e oito, respectivamente. Essa diferença pode implicar na perda de um agrupamento mais significativo. Mas de qualquer forma a opção de utilizar apenas os primeiros autores é válida e precisa ser descrita nas pesquisas evidenciando suas vantagens e limitações.

O segundo tópico é sobre o número de autores que farão parte de uma matriz de cocitação, assim como o indicador utilizado para a inclusão dos autores. Parte dos estudos opta pela inclusão dos autores pelo número total de citações recebidas por eles e não há uma explicação sobre o número de autores que farão parte da matriz final com dados absolutos em relação a todos os autores que formam um *corpus* de análise. Bu, Liu e Huang (2016) utilizam os 100 autores mais citados e Bu *et al* (2017) utilizam os 500 autores mais citados para criarem as matrizes. White e McCain (1998) e Zhao e Strotmann (2014) analisam os 120 e 250 autores mais citados na área da CI, respectivamente, a partir de 12 periódicos.

McCain (1990) cita alguns critérios para essa situação identificados a partir das práticas de alguns pesquisadores de ACA³, como utilizar taxas médias de cocitação acima de nove em um período específico de tempo e cocitação com pelo menos um

³ Trecho: *ACA researchers, concerned with the possible instability of small cocitation counts, have selected for further analysis only those authors meeting certain ad hoc criteria, such as mean cocitation rates above nine (for 10 years of Social Scisearch data) and cocitation with at least one-third of the entire author set (B. C. Griffith, personal communication). Penan (1989) restricted the final author set to the 20% receiving the highest number of citations and cocitations in initial retrieval trials. I have experimented with mean cocitation rates as low as four across a 50-author set for four or five years of Scisearch and Social Scisearch data. The results have been satisfactory in terms of face validity and interpretability of the maps and clusters.* (McCain, 1990, p.435).

terço dos autores que fazem parte da análise. Penan⁴ (1989 citado por McCAIN, 1990, p. 435) estabeleceu em 20% os autores com mais citações e cocitações, a partir de alguns testes de recuperação de dados. McCain (1990) indica que fez testes com cortes menores e que os resultados foram satisfatórios em termos de interpretação dos agrupamentos gerados.

Essa discussão é relevante e diversos aspectos surgem para ponderar as propostas dos autores citados acima, inclusive a diferença de tempo que impacta em termos tecnológicos a capacidade de recuperação e manuseio dos dados. Um corte muito alto, por exemplo, pode acarretar na não identificação de uma área emergente com dados de autores que estejam um pouco abaixo do ponto de corte estabelecido, assim como um corte muito baixo pode apresentar relações não efetivamente temáticas entre os cocitados e atrapalhar a interpretação adequada de um determinado agrupamento.

Quanto ao indicador utilizado, os trabalhos metodológicos de Grácio e Oliveira (2013a, 2013b, 2015) optam pelo uso do número de documentos que citam os autores para a formação das matrizes. Já o trabalho de Carvalho *et al* (2018) traz dados empíricos que evidenciam que o número de documentos citantes de um autor é mais apropriado para ACA do que a contagem total de citações, pois os principais autores são mencionados várias vezes por documentos citantes, até pelo número de documentos citados de alguns autores específicos, como *Birger Hjørland*, que possui 22 citações em apenas oito documentos citantes nos dados apresentados. Dois pequenos *rankings* com 30 posições, organizados por citações totais e número de documentos citantes, mostram que 23 autores fazem parte das duas listas e 14 autores distintos diferenciam os *rankings*. Certamente isso interfere na melhor identificação das relações de cocitação. Eom e Farris (1996) utilizam a cocitação do autor com ele mesmo, o que facilita a definição dos dados inseridos na diagonal da matriz, mas eles alertam que utilizar essa informação pode implicar em limitações, principalmente em áreas multidisciplinares. Empiricamente é possível um autor ser citado por todos os documentos citantes de um *corpus* de análise e não ter nenhuma cocitação com ele mesmo, sem contar que estudos que utilizam a primeira posição das referências limitam ainda mais o uso dessa informação.

⁴ Penan, H. (1989). **Pour une gestion bibliometrique de l' information scientifique et technique des entreprises**. Application en theorie microeconomique et financiere. Doctoral dissertation, Universite des Sci-ences Sociales. Toulouse I, Toulouse, France. [Texto completo não localizado].

O uso dos índices de similaridade e o valor da diagonal aparecem como um terceiro elemento complicador na análise das matrizes e podem gerar resultados distintos dependendo das escolhas. Os principais índices discutidos na literatura são o r de Pearson, o Cs e o IJ.

Ahlgren, Jarneving e Rousseau (2003) fazem uma crítica ao uso do r de Pearson como medida de similaridade por não atender a dois requisitos que os autores entendem ser necessários para uso da medida em matrizes de ACA: 1) a medida de similaridade dos autores A e B, em uma situação hipotética, deve continuar a mesma se os autores forem incluídos em um conjunto com outros autores e não apresentarem cocitação com eles; 2) os autores A, B, C e D formam ou fazem parte de um grupo com medida de similaridade já estabelecida, se o grupo for ampliado e nenhum desses quatro autores apresentarem cocitação com os novos autores do grupo, exige-se que, se antes da expansão, a similaridade entre A e B fosse menor do que entre C e D, que continue igual após a expansão. Os autores afirmam que o r de Pearson não atende esses requisitos, mas mesmo assim é um índice bem utilizado em estudos de cocitação (OZÇINAR, 2015; BU *et al*, 2017) e recomendado por Oliveira e Alves (2017), em relação ao Cs e IJ, se o passo posterior for utilizar alguma técnica de estatística multivariada.

White (2003) faz uma crítica ao artigo de Ahlgren, Jarneving e Rousseau (2003), que não testaram os requisitos propostos de flutuação do valor de similaridade de Pearson com dados reais. O autor faz os testes e recomenda o uso dessa medida de correlação, pois as mudanças não são significativas nem para os requisitos propostos e nem para comparação com similaridade utilizando o Cs.

Leydesdorff (2005) salienta a sensibilidade da medida de Pearson para os zeros da matriz, algo que não ocorre com o Cs. Em outro estudo, Leydesdorff (2008) recomenda o uso do índice de Jaccard quando não há confiabilidade nos dados recuperados, uma vez que o índice foca apenas na interseção e soma de dois conjuntos, mas Oliveira e Alves (2017) apresentam dados que mostram ser esse índice menos robusto que o Cs, gerando dados proporcionalmente menores com os mesmos valores de cocitação, ou seja, as medidas possuem menos força.

A discussão do uso dessas medidas é desenvolvida na literatura brasileira sobre cocitação, com destaque para as autoras Grácio e Oliveira (2013a, 2013b, 2015) e para o

trabalho de Oliveira e Alves (2017), como já salientado. No tópico 2.5 desse projeto há outras informações sobre o tema.

Por fim, Zhou e Leydesdorff (2016) compararam o uso de um coeficiente chamado *Ochiai* com o Cs e concluem que os índices são similares, mas a diagonal principal não pode ser preenchida com zeros. Os autores indicam que a matriz de coocorrência não precisa ser normalizada e que há semelhança entre a normalização pelo Cs e pelo r de Pearson.

Os mesmos autores destacam que os valores que devem ser atribuídos à diagonal dependem do índice de similaridade que será posteriormente utilizado e que tecnicamente os valores em uma matriz de cocitação são de mesma grandeza e que não precisam ser normalizados (ZHOU; LEYDESDORFF, 2016). McCain (1990) alerta, porém, que utilizar um índice de similaridade transforma os valores em medidas que estabelecem a força da associação. A aplicação do Cs e do IJ, por exemplo, exige que o valor da diagonal não seja menor que o valor máximo de cocitação do autor. Isso inviabiliza o uso para matrizes com zeros na diagonal, ou mesmo com valores de cocitação do autor com ele mesmo. Esse tópico é recorrente nos estudos metodológicos de cocitação.

Ainda sobre o valor utilizado na diagonal das matrizes Ahlgren, Jarneving e Rousseau (2003) consideram que o uso de dados de citação é arbitrário, uma vez que a matriz é construída com valores de outra natureza. Calcular a cocitação de um autor com ele mesmo é possível, mas pouco usual. Há trabalhos utilizando zero (BU *et al*, 2017) e não atribuindo valores (AHLGREN; JARNEVING; ROUSSEAU, 2003). White (2003) cita o uso das médias de cocitação do autor como opção, mas recomenda que o valor máximo de cocitação de um autor com outro seja colocado na diagonal, o que possibilita a aplicação do Cs e do IJ. A pesquisa aplica na diagonal citação total, citação oriunda de documentos citantes, zero e cocitação do autor com ele mesmo, mas somente nos dois primeiros casos foi possível comparar resultados dos agrupamentos com os três índices de similaridades: Cs, IJ e r de Pearson.

No quarto tópico dos estudos metodológicos, ou seja, as formas de validar e interpretar os agrupamentos, seguem três destaques de pesquisas que buscam apresentar formas de validar e interpretar os agrupamentos em ACA.

O artigo de Chen, Ibekwe-Sanjuan e Hou (2010) teve como objetivo apresentar uma nova perspectiva aos estudos de cocitação, com múltiplas facetas, no sentido de analisar os padrões estruturais, temporais e semânticos, além de acrescentar informações dos citados e dos citantes para ampliar a interpretação e o nível de confiabilidade dos agrupamentos formados. Os dados foram recuperados de documentos publicados em 12 periódicos da área da Ciência da Informação (1996-2008), totalizando 10.853 registros, que citaram 129.060 referências únicas através de 206.180 citações. Essas referências apresentaram 58.711 autores únicos. Os agrupamentos da ACA foram definidos a partir de 663 autores e os agrupamentos da DCA a partir de 655 documentos. Um sistema automático rotulou os agrupamentos com dados dos documentos citantes (*citers*) e os autores concluíram que é recomendado realizar as duas técnicas juntas – ACA e DCA. Os autores pontuam ainda a dificuldade de interpretar os resultados obtidos em ACA e em DCA, que se concentram nas entidades citadas (autores/documentos) como fonte principal de evidência. Os autores afirmam que se concentrar nos documentos ou autores citantes pode melhorar a compreensão dos agrupamentos. (CHEN; IBEKWE-SANJUAN; HOU, 2010).

Um recorte menor da análise (2001-2005) de Chen, Ibekwe-Sanjuan e Hou (2010) validou tanto os agrupamentos, em relação ao trabalho de Zhao e Strotmann (2008), com uma sobreposição de 82% dos membros, como os rótulos gerados de forma automática com rótulos desenvolvidos manualmente, já que os trabalhos utilizaram o mesmo *corpus* e recorte temporal.

Bu, Liu e Huang (2016) propõem um método de ACA modificado, chamado de MACA (*Modified Author Co-citation Analysis*). A pesquisa explora o uso de quatro metadados descritivos para qualificar o relacionamento que os dados de cocitação apresentam: a) autores citados/referenciados (contagem de cocitação); b) o tempo (data) em que a citação/referência foi publicada; c) suporte ou fonte onde a citação/referência foi publicada; e e) palavras-chaves de uma citação/referência.

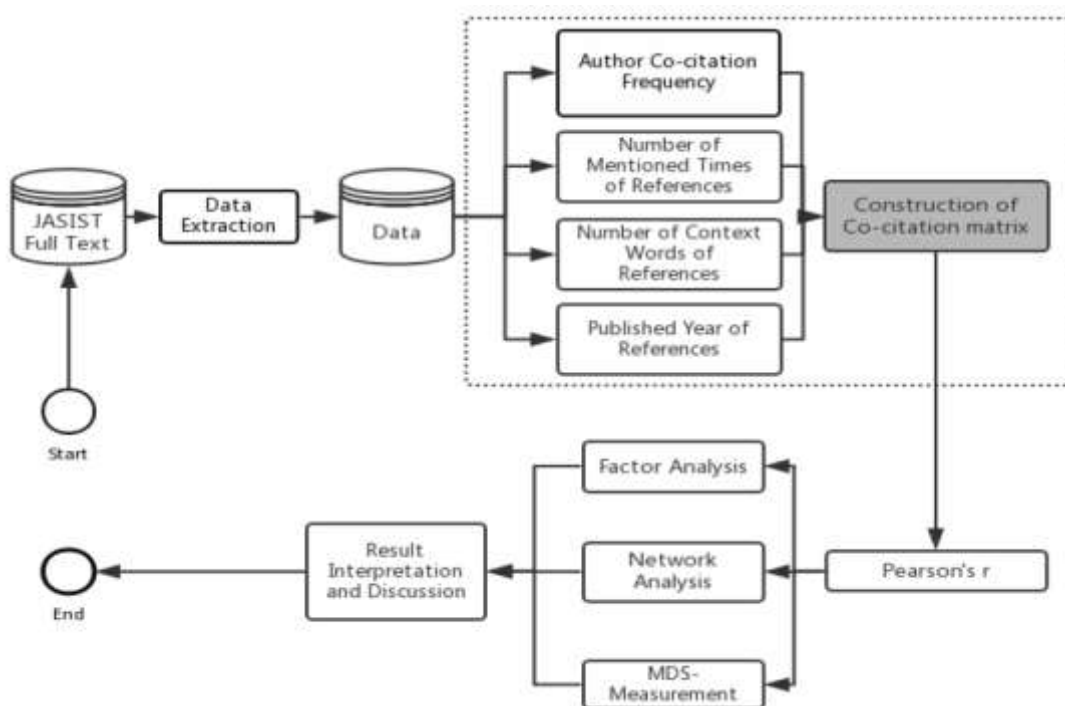
O conjunto de dados para a pesquisa foi recuperado no periódico *Journal of the American Society for Information Science and Technology* (JASIST) de 2003 a 2012. Vários metadados descritivos foram retirados de 2.038 artigos e 68.606 referências e no final os 100 autores mais citados foram utilizados para finalizar a pesquisa. A AF e o EMD foram as técnicas de estatística multivariada utilizadas e as matrizes foram construídas com diagonal igual a zero e considerando apenas os primeiros autores das

referências. Os autores concluíram que MACA oferece mais detalhes utilizando MDS como técnica estatística, ou seja, adiciona mais informações à ACA tradicional (BU; LIU; HUANG, 2016).

Por último, Bu *et al* (2017) apresentam um método de ACA combinado com metadados do texto completo, chamado de MFTACA (*Metadata-in-full-text-based ACA*). Os autores partem do pressuposto que as menções a uma referência no corpo do texto modificam a suposta “igualdade” apresentada na lista de referências e isso tem consequência nas análises de cocitação.

Os dados da pesquisa foram extraídos de um conjunto de 1.420 artigos do periódico *Journal of the American Society for Information Science and Technology* (JASIST). Os 500 autores mais citados foram utilizados para formar a matriz de cocitação, o número de menções de cada referência, o número de palavras do contexto de uso das referências e ano de publicação das referências foram utilizados para ponderar as relações de cocitação. A Figura 5 demonstra o fluxo do algoritmo proposto pelos autores.

Figura 5. Diagrama de fluxo do algoritmo do método da MFTACA (Bu *et al*, 2017).



Fonte: Bu *et al* (2017, p. 4).

Observa-se na figura que a matriz de correlação foi normalizada pelo r de Pearson e a análise fatorial, a análise de redes sociais e o MDS foram as técnicas utilizadas para gerar os agrupamentos.

Bu *et al* (2017) utilizaram o número de menções das referências no texto completo para definir se a relação de cocitação de dois autores deveria ser considerada mais relevante e, portanto, receber uma contagem distinta que valoriza essa relação: *number of mentioded times of references*. Quanto ao “*number of context words of references*” os autores contam as palavras de cada frase citante, pois entendem que esse número pode refletir a importância de uma referência utilizada e assim, se dois autores possuem muitas palavras associadas a sua menção a contagem dessa relação de cocitação é realizada de forma distinta. Sobre o “*published year of references*” os autores utilizam as diferenças das datas de publicação, conforme trabalho anterior (BU; LIU; HUANG, 2016), entendendo que datas próximas de publicação sugerem uma aproximação temática dos autores cocitados e grandes diferenças indicam distância temática (BU *et al*, 2017).

Esses parâmetros apresentados balizaram o uso de um algoritmo e a análise fatorial gerou 18 fatores com MFTACA e 11 com ACA tradicional. O Quadro 5 apresenta as diferenças.

Quadro 5. Comparando fatores com ACA e MFTACA (BU *et al*, 2017).

FATORES	ACA	MFTACA
Information retrieval	X	x
Information seeking behavior	X	x
Information usage, digital library		x
Language model, query, clustering	X	x
Classification algorithms, information organizations		x
Text mining, machine learning	X	x
User interface	X	x
User acceptance of information technology		x
Information systems		x
Data Mining, Data Analysis		x
Evaluation indicator, index	X	x
Webometrics, social network analysis	X	x
Visualization, mapping		x
Modeling		x
Scholarly communication	x	x
Journal citation analysis, interdisciplinarity, evaluation of algorithms	x	x
Network analysis	x	x
Bioinformatics	x	x

Fonte: adaptado de Bu et al (2017, p. 7).

Os autores concluem (BU *et al*, 2017) que apesar de não utilizarem informação no nível do conteúdo, como fazem Jeong, Song e Ding (2014), o método proposto apresenta melhor desempenho na formação dos agrupamentos do que a ACA tradicional, além de fornecer mais detalhes.

Os estudos, de maneira geral, têm buscado melhorar a qualidade dos agrupamentos para ajudar na interpretação do que realmente ocorre ou existe de relação entre os autores/documentos. Vale salientar, que essa revisão não é sistemática e buscou apresentar os principais trabalhos sobre o tema.

2.3 Análise de cocitação de autores: estudos aplicados _____

Apesar de ser uma técnica complexa que exige muita discussão metodológica, fazendo com que grande parte das pesquisas acabe se orientando nesse sentido, a ACA também é utilizada como ferramenta metodológica em estudos aplicados. Alguns desses estudos são destacados a seguir em ordem de data publicação e identificando, na medida do possível, os parâmetros de operacionalização da técnica.

Culnan, O'Reilly e Chatman (1990) utilizam a ACA para identificar a estrutura intelectual da área de Organização comportamental, entre os anos de 1972 e 1984. O destaque é para a forma de seleção dos autores, por meio de pesquisa com chefes de departamentos de organização comportamental e gestores de universidades. A matriz quadrada ficou com 52 autores, o valor da diagonal principal foi estabelecido como fez Ozçınar (2015), somando os três principais valores da coluna/linha do autor e dividindo por dois. A análise fatorial (análise de componentes principais) foi a técnica de estatística empregada.

Eom e Farris (1996) aplicaram uma ACA na área do campo da *Decision Support Systems* (DSS). A escolha dos autores para a análise se deu em dois processos, pois os dados foram retirados da WoS que indexa apenas o primeiro autor das referências e houve um trabalho para a identificação das coautorias. Assim, os 67 autores com pelo menos 25 cocitações consigo mesmo entraram na análise. Os autores, a partir de diversos critérios, principalmente propostos por McCain (1990), decidiram que essa era

a melhor forma de definir um ponto de corte de inclusão dos autores, mas salientam as limitações dessa forma, principalmente para áreas altamente multidisciplinares (EOM; FARRIS, 1996). O trabalho detalha ainda que a matriz com dados absolutos foi utilizada para análise de componentes principais (técnica fatorial de análise estatística), com solução inicial de nove fatores, cargas maiores que “0,4” e rotação oblíqua. Os nove fatores responderam por 84,66% da variância total. A matriz de correlação foi analisada pela técnica de escalonamento multidimensional para visualizar similaridade e dissimilaridade. Os autores concluíram, para o contexto da época de publicação do relato, que esse campo de estudo se dividia em quatro disciplinas e que os pesquisadores ainda estavam em processo de desenvolvimento do campo.

White e McCain (1998) realizaram um estudo de larga escala, englobando o período de 1972 a 1995, na área da Ciência da Informação. A matriz foi produzida a partir dos 120 principais autores recuperados de 12 periódicos. O estudo utilizou a correlação de Pearson como medida de similaridade e à diagonal principal não foi atribuída nenhum valor. As análises de fatores (componentes principais) e *clusters* foram empregadas no estudo.

Kreuzman (2001) desenvolve uma ACA com 62 autores para entender a relação entre filosofia da ciência e epistemologia, abrangendo um período de 13 anos e meio (1980 a julho de 1993), a partir de dados da versão eletrônica do *Arts and Humanities Citation Index*. A matriz com dados absolutos ficou sem valor atribuído na diagonal e o *r* de Pearson foi utilizado para converter as matrizes. O autor salienta que a correlação alta (mais próxima do um positivo) indica proximidade temática entre autores cocitados enquanto a correlação baixa (mais próximo do um negativo) indica uma distância intelectual. As técnicas de escalonamento multidimensional e a análise de clusters foram utilizadas na parte final do estudo. Os autores concluem o texto analisando a proximidade dos autores a partir das temáticas que representam e salientam que a análise de cocitação é apenas uma técnica de identificação de estruturas de disciplinas e autores.

Cassillas e Acedo (2007) estudaram o tema *Family Business* como uma subdisciplina da Gestão e indicaram aplicar uma ACA em um periódico chamado *Family Business Review* (1988-2005). O total de artigos analisados foi 339 que se desdobrou para 10.072 referências. Esse artigo chama atenção por utilizar a forma “autor/data” para representar documentos, realizar uma DCA, mas designar o estudo

como ACA. Isso ocorre também com o trabalho de García-Lillo, Úbeda-García e Marco-Lajara (2016). Vale salientar que utilizar “autor/data” não caracteriza o estudo como ACA, uma vez que designar o documento dessa forma facilita a limpeza de dados para realizar uma DCA, como fizeram Serra *et al* (2012).

Osareh e McCain (2008) realizaram um estudo sobre os autores mais citados em publicações iranianas na área da Química, entre os anos de 1990 e 2006. Uma análise fatorial (análise componentes principais) apresentou sete fatores a partir dos 43 autores mais influentes no campo. Os dados mostraram pouca relação dos autores iranianos com os autores internacionais mais citados. O artigo não ofereceu muitos parâmetros para a operacionalização da ACA.

Zhao e Strotmann (2014) apresentam um estudo de identificação da estrutura intelectual e frente de pesquisa na área da Ciência da Informação, entre 2006 e 2010, usando também o acoplamento bibliográfico de autores. Os resultados mostram a influência da Web na CI e o crescimento do tópico de pesquisa “mapeamento da ciência”. A análise fatorial foi a técnica de estatística multivariada utilizada e o valor da diagonal principal da matriz foi omitido e substituído pela média da rotina da análise estatística. Os fatores gerados foram 16, a partir dos 150 autores cocitados na análise. Não há no texto informação sobre o uso de algum índice de similaridade e a análise de redes foi utilizada para visualização das relações. Os autores concluem que combinar as técnicas (cocitação e acoplamento) ajuda na identificação de tendências de pesquisa.

Khasseh, Soheili e Chelak (2018) publicaram um trabalho que objetivou descrever o desenvolvimento de pesquisas sobre “responsabilidade social corporativa” em 12 periódicos internacionais de negócios, selecionando 43 autores com pelo menos 50 citações. A matriz foi gerada com diagonal principal igual zero. Cinco fatores foram gerados a partir de uma análise fatorial (análise de componentes principais), que aproveitou cargas fatoriais acima de “0,6”, que mostra bastante rigor na análise. Não foi identificado o emprego de nenhum índice de similaridade/correlação.

Zhao, Zhang e Kwon (2018) aplicaram uma ACA no campo das métricas da informação (*iMetrics*) para mapear a estrutura intelectual desse domínio. Os dados são formados por 5.944 registros recuperados na WoS entre 1978 e 2014. Cerca de 170 autores foram incluídos na matriz e a técnica multivariada empregada foi análise de *clusters*. Os autores mencionam a matriz de correlação, mas não são específicos quanto ao índice utilizado, e não há menção sobre o valor da diagonal principal.

No próximo tópico são apresentados alguns estudos brasileiros sobre o tema, que aprofundam questões metodológicas e relatam aplicações mais objetivas da técnica.

2.4 Análise de cocitação de autores: panorama no Brasil

No Brasil alguns estudos de cocitação se destacam a partir de 2010, especialmente na literatura da Ciência da Informação, com enfoque aplicado e também metodológico. Os principais estudos seguem em ordem cronológica.

Graeml, Macada e Rossoni (2010) apresentam um relato de pesquisa cientométrica na área da Administração da Informação, a partir de artigos publicados em um evento entre 1997 e 2006. Parte da pesquisa é a aplicação de um estudo de cocitação autores que utiliza a análise fatorial como técnica estatística, mas não há outras informações sobre os parâmetros de construção das matrizes.

Mattos e Dias (2010) apresentam uma revisão sobre questões metodológicas que envolvem a aplicação de ACA e concluem ser esse um campo de pesquisa com muito potencial para a Cientometria no Brasil, com vistas a um diálogo com a produção internacional. Por fim, salientam a dificuldade de padronização e consistência dos dados para os EC.

Santarem (2010) utiliza a cocitação como técnica de análise de domínio em dissertação de mestrado. A autora apresenta uma rede de cocitação de autores com 15 posições na área do “Tratamento temático da informação” no Brasil.

Oliveira e Grácio (2011, 2012) publicam um estudo sobre a visibilidade de pesquisadores em um grupo de trabalho da ANCIB⁵ e no periódico *Scientometrics*⁶, respectivamente. As autoras identificam uma rede coesa de cocitação e apresentam dados interessantes sobre 15 citantes do *corpus* de análise que não citam nenhum dos autores mais citados que formam a rede de cocitação no trabalho de 2011. No texto de 2012 as autoras identificam 53 artigos de pesquisadores brasileiros no periódico estudado e elaboraram uma matriz quadrada com os 19 autores mais citados. Elas concluem que há uma interlocução entre pesquisadores brasileiros e estrangeiros.

⁵ Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação. Disponível em: <https://www.ancib.org.br>

⁶ An International Journal for all Quantitative Aspects of the Science of Science, Communication in Science and Science Policy. Disponível em: <https://link.springer.com/journal/11192>

Custódio (2012) apresenta uma dissertação que estuda a cocitação na área da Educação, especificamente em teses e dissertações desenvolvidas na UNESP⁷, entre 2004 e 2009. A autora utiliza a análise de redes sociais como técnica para lidar com os dados de cocitação e faz uma divisão de análise por tipo de documento (dissertação/tese). Verifica que há apenas seis autores em comum cocitados nas teses e dissertações, e conclui sobre a diversificação da área da Educação na universidade.

Serra *et al* (2012) analisaram artigos publicados no periódico *Strategic Management Journal*, entre 2001 e 2007, e realizaram uma DCA com os 50 artigos mais citados no recorte proposto. Os dados foram levantados a partir de 31.481 referências de 489 artigos, sendo que 15.899 eram documentos distintos. A matriz de dados absolutos foi convertida utilizando o r de Pearson e analisada a partir do escalonamento multidimensional. Os dados apresentam dois agrupamentos, sendo que o principal foi dividido em três subgrupos.

Grácio e Oliveira (2013a, 2013b, 2015) apresentam três trabalhos que tratam respectivamente de um estudo comparativo sobre indicadores de proximidade em ACA (Cs e IJ), sobre a análise de indicadores absolutos e relativos em ACA, onde o Cs é utilizado, e por fim fazem outra comparação desses coeficientes, mas utilizando a correlação de Pearson e o Cs. Concluem, a partir de dados empíricos para os três estudos, que os índices apresentam as mesmas tendências de proximidade.

Lima (2015) apresenta uma dissertação de mestrado em que faz uma ACA dos artigos do periódico *Knowledge Organization*⁸, entre 2004 e 2013. A técnica utilizada foi a análise de redes sociais e autores com pelo menos cinco documentos citantes fizeram parte da rede.

Machado (2015) realiza um estudo de cocitação de periódicos na área de célula-tronco e separa os periódicos pelas categorias de que são vinculados na WoS, pois esse tema de pesquisa é muito interdisciplinar. O autor ainda apresenta uma dinâmica de mudança dos periódicos em um período de duas décadas.

Piovezan e Fujita (2015) apresentam um estudo de cocitação sobre o tema “Avaliação de indexação”. As autoras utilizam a técnica de redes sociais em uma matriz quadrada com 31 autores e identificam três correntes sobre o assunto.

⁷ Universidade Estadual Paulista. Disponível em: www.unesp.br

⁸ Knowledge Organization (ISSN 0943-7444) is the official bi-monthly journal of ISKO (International Society for Knowledge Organization). Disponível em: http://www.ergon-verlag.de/isko_ko/

Grácio (2016) apresenta uma revisão teórico-conceitual sobre estudos de cocitação e acoplamento bibliográfico. A autora destaca a relevância dos métodos e salienta o pouco uso do acoplamento, mesmo em nível internacional. A autora destaca a ACA como o método de cocitação mais aplicado nas últimas duas décadas e a *Author Bibliographic-Coupling Analysis* (ABCA) como método emergente, que evidencia as atividades científicas em curso.

Por fim, Oliveira e Alves (2017) apresentam dados comparativos de uso dos três principais indicadores normalizados utilizados nos estudos de cocitação: Cs, IJ e Correlação de Pearson. Os autores concluem que o Cs é o indicador mais adequado, mas não descartam a qualidade dos outros dois, principalmente o indicador de correlação de Pearson, se o objetivo da aplicação for análise estatística multivariada.

2.5 Análise de cocitação de documentos: estudos recentes _____

Ainda que ocorra a predominância de ACA em relação a outros tipos de estudos de cocitação, a DCA tem sido utilizada, inclusive em publicações recentes e em outras áreas do conhecimento. Abaixo seguem quatro destaques.

Jankovic, Kaufmann e Kindler (2008) aplicaram uma DCA na pesquisa sobre anestesia nas publicações de 23 revistas indexadas na WoS, em 2003. O estudo identificou 46 campos ativos de pesquisa nessa área, a partir de 8.243 referências mais citadas (cinco ou mais citações) e utilizando os títulos dos documentos. Os campos de pesquisa foram classificados de acordo com a proporção de documentos recentes nos agrupamentos.

Ozçinar (2015) aplicou uma DCA para identificar a estrutura da pesquisa sobre a formação de professores e suas possíveis mudanças, já que o recorte temporal era de 1992 a 2012. O domínio é caracterizado por diversas especialidades, segundo o autor, e nenhuma dessas especialidades se desenvolveu de forma a ser a principal tendência na pesquisa sobre formação de professores. O autor utilizou o r de Pearson como índice de correlação na matriz de valores absolutos, análise fatorial para processar a matriz de correlação e a análise de redes sociais para visualização das relações. O valor da diagonal foi definido somando os três maiores valores de cada linha/coluna e dividindo por dois.

González-Alcaide *et al* (2016) utilizaram a DCA no campo de pesquisa sobre “abuso de substâncias” e identificaram 56 agrupamentos de pesquisa em periódicos indexados na WoS, entre 2001 e 2012, a partir de 698 documentos. Os autores construíram a rede de cocitação com os valores absolutos de matriz, optando por um caminho que elimina a necessidade de normalizar matrizes e estabelecer valores para a diagonal.

Trujillo e Long (2018) realizam uma rede de cocitação de documentos como forma de evidenciar pesquisas transdisciplinares em artigos publicados em revistas indexadas na WoS. Os autores evidenciam que a técnica tem essa função e trabalham os dados empíricos do texto de forma a validar a DCA. O uso da correlação de *Spearman*⁹ validou estatisticamente as redes de cocitação.

Encerra-se a revisão que dará suporte para discutir os resultados da pesquisa. Na próxima seção são apresentados os procedimentos metodológicos.

⁹ “[...] medida padronizada da força do relacionamento entre duas variáveis que não depende das hipóteses de um teste paramétrico” (FIELD, 2009, p.643).

3 PROCEDIMENTOS METODOLÓGICOS

A pesquisa se caracteriza como descritiva, de caráter metodológico e cientométrico. O conjunto de dados principal da análise foi recuperado em uma base de dados multidisciplinar internacional (*Web of Science – WoS*) e é constituído de informações das referências de artigos de periódicos científicos da Ciência da Informação, que versam sobre o tema Organização do conhecimento e Recuperação da Informação e que tenham sido publicados entre 2015-2016, neste trabalho denominado “*corpus principal*” de análise.

Um segundo conjunto de dados de referências, apresentado em artigo publicado (CARVALHO; CAREGNATO, 2017), também foi utilizado para auxiliar na definição do ponto de corte de autores e documentos para análises de cocitação, para parte dos resultados sobre objetivar a fase de interpretação e validação de agrupamentos de ACA, para definir os termos de busca do conjunto principal de dados, além de amparar alguns testes do projeto que terminou nesta tese. As referências que o compuseram foram retiradas de artigos de pesquisa recuperados na WoS, da área da Organização do Conhecimento enquanto subdisciplina da Ciência da Informação, publicados entre 2011 e 2015, neste trabalho denominado “*corpus secundário*” de análise.

3.1 Estratégia de busca para coleta de dados dos dois conjuntos de análise

O *corpus principal* de análise é formado por 421 artigos de pesquisa coletados na base de dados WoS em 21 de fevereiro de 2018 (Apêndice 5). A estratégia de busca consistiu no uso de quatro termos compostos utilizados no campo ‘TÓPICO’, como segue: “TOPIC: ("knowledge organization" OR "information organization" OR "knowledge representation" OR "information retrieval").

Os dados recuperados foram refinados da seguinte forma: “WEB OF SCIENCE CATEGORIES: (INFORMATION SCIENCE LIBRARY SCIENCE) AND DOCUMENT TYPES: (ARTICLE); Timespan: 2015-2016. Indexes: SCI-EXPANDED, SSCI, A&HCI, ESCI”.

A busca inicial recuperou 461 documentos, sendo que 40 foram retirados por serem textos sem referências, sem seção de referências e textos em outros tipos de alfabeto, principalmente cirílico, de um periódico russo.

As expressões de busca foram definidas a partir dos títulos e das palavras-chaves dos 151 artigos de pesquisa que formam o *corpus* secundário (Apêndices 1 e 2 listam as expressões e termos únicos, respectivamente), também da área da Organização do Conhecimento, recuperados na WoS em maio de 2017 e publicados entre 2011 e 2015 (Apêndice 6). A busca foi realizada pelo termo “*Knowledge organization*” no campo “TÓPICO” da base e refinado pela categoria “INFORMATION SCIENCE LIBRARY SCIENCE”, foram considerados apenas artigos. Alguns dados desse conjunto tiveram resultados publicados por Carvalho e Caregnato (2017). Os resultados foram exportados para uma planilha de dados para posterior recuperação dos textos completos dos artigos que formam os dois conjuntos.

3.2 Caracterização do *corpus* principal da pesquisa: periódicos e autores _____

Os 421 artigos que formam o *corpus* principal da análise foram publicados por 69 periódicos distintos, sendo que os títulos que se destacam são o “*Knowledge Organization*” (17,10%) e o “*Journal of The Association for Information Science and Technology*” (14,25%), que publicaram juntos 132 artigos. Alguns periódicos brasileiros se destacam: *TransInformação* (cinco artigos), *Informação & Sociedade* (quatro artigos), *Perspectivas em Ciência da Informação* (quatro artigos) e *Revista Ibero-Americana de Ciência da Informação* (quatro artigos).

A Tabela 1 apresenta a distribuição, por ano de publicação, dos 20 periódicos que publicaram cinco ou mais artigos presentes no *corpus* principal da análise. Esses dados foram retirados e estruturados a partir dos metadados dos arquivos de cada artigo e não dos metadados da WoS.

O Apêndice 3 apresenta a distribuição dos outros 49 periódicos (com até quatro artigos publicados) divididos por ano de publicação.

Tabela 1. Distribuição dos artigos do *corpus* principal de análise por ano de publicação e periódico.

Periódicos	2015	2016	Total	%
1. KNOWLEDGE ORGANIZATION	32	40	72	17,10
2. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY	23	37	60	14,25
3. INFORMATION PROCESSING & MANAGEMENT	16	16	32	7,60
4. JOURNAL OF INFORMATION SCIENCE	11	11	22	5,23
5. JOURNAL OF DOCUMENTATION	07	12	19	4,51
6. SCIENTOMETRICS	12	03	15	3,56
7. ELECTRONIC LIBRARY	05	09	14	3,33
8. SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO	07	06	13	3,09
9. ASLIB JOURNAL OF INFORMATION MANAGEMENT	05	04	09	2,14
10. INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES	02	06	08	1,90
11. LIBRARY HI TECH	04	04	08	1,90
12. ONLINE INFORMATION REVIEW	05	03	08	1,90
13. QUALITATIVE & QUANTITATIVE METHODS IN LIBRARIES	07	01	08	1,90
14. JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION	03	04	07	1,66
15. DESIDOC JOURNAL OF LIBRARY & INFORMATION TECHNOLOGY	04	02	06	1,43
16. INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL	03	03	06	1,43
17. INVESTIGACION BIBLIOTECOLOGICA	02	03	05	1,19
18. JOURNAL OF INFORMETRICS	02	03	05	1,19
19. JOURNAL OF KNOWLEDGE MANAGEMENT	05	--	05	1,19
20. TRANSINFORMACAO	02	03	05	1,19
Quatro artigos (08 periódicos)	18	14	32	7,60
Três artigos (06 periódicos)	11	07	18	4,28
Dois artigos (09 periódicos)	08	10	18	4,28
Um artigo (26 periódicos)	13	13	26	6,18
Total	207	214	421	100

Fonte: dados da pesquisa.

Os artigos utilizados na análise principal dessa pesquisa possuem 1.097 autorias para 976 autores distintos. A Tabela 2 destaca as 18 primeiras posições dos autores com mais artigos na análise (a partir de três artigos) e distribui os dados pelo tipo de autoria dos artigos: única e coautoria (subdividida em principal e secundária). Os dados foram retirados dos arquivos do texto completo de cada artigo e a limpeza dos nomes dos autores foi realizada manualmente, considerando os sobrenomes compostos.

Destacam-se os autores “Guimarães, JAC”, “Hjørland, B”, “Lee, JH” e “Martínez-Ávila, D” com cinco ou mais artigos presentes na análise. Vale ressaltar a

presença de autores brasileiros entre os mais produtivos, pois além do autor “Guimarães, JAC” há também a participação dos autores “Fujita, MSL”, “Meireles, MRG”, “Sales, R” e “Tognoli, NB”.

Tabela 2. Distribuição das primeiras 18 posições dos autores dos artigos que formam o *corpus* principal de análise.

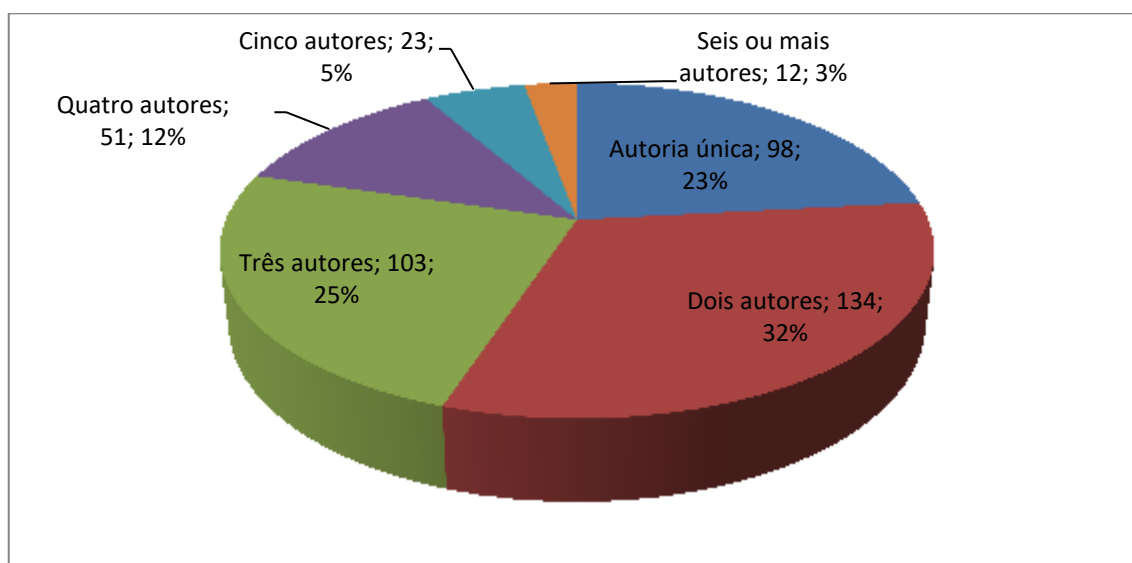
Autores	Autoria única	Autoria principal	Autoria secundária	Total	%
1. Guimarães, JAC	--	03	03	06	0,55
2. Hjørland, B	05	--	--	05	0,46
3. Lee, JH	--	03	02	05	0,46
4. Martínez-Ávila, D	01	04	--	05	0,46
5. Shah, C	--	01	03	04	0,36
6. Tudhope, D	--	01	03	04	0,36
7. Fujita, MSL	--	02	01	03	0,27
8. Madalli, DP	--	02	01	03	0,27
9. Meireles, MRG	--	03	--	03	0,27
10. Sales, R	01	01	01	03	0,27
11. Sánchez-Alonso, S	--	--	03	03	0,27
12. Smiraglia, RP	02	--	01	03	0,27
13. Soergel, D	01	--	02	03	0,27
14. Tamine, L	--	01	02	03	0,27
15. Tognoli, NB	01	--	02	03	0,27
16. Wolfram, D	02	--	01	03	0,27
17. Yan, E	--	01	02	03	0,27
18. Zhu, Y	--	02	01	03	0,27
Autoria em até dois documentos*	17	37	94	148	13,49
Autoria de um documento	260	554	70	884	80,58
Total	290	615	192	1097	100,00

* 74 autores distintos

Fonte: dados da pesquisa.

A distribuição de autores mostra que a maioria (80%) dos artigos possui entre um e três autores, sendo que dois autores por artigo apresenta o maior escore (32%), seguindo por três autores (25%) e autoria única (23%). A Figura 6 ilustra esses dados.

Figura 6. Distribuição do número de autores dos artigos que formam o *corpus* principal da análise.



Fonte: Dados da pesquisa.

A seguir é apresentada a caracterização do *corpus* secundário da análise.

3.3 Caracterização do *corpus* secundário da análise: periódicos e autores _____

Os 151 artigos que formam o *corpus* secundário de análise foram publicados em 23 periódicos distintos. Os títulos que se destacam são o “Knowledge Organization” (54, 97%), o “Journal of the Association for Information Science and Technology” (7,28%) e o “Journal of Documentation” (6,62), que publicaram juntos 104 artigos, sendo que o primeiro é responsável por mais da metade dos artigos da análise. Alguns periódicos brasileiros se destacam: Perspectivas em Ciência da Informação (sete artigos), TransInformação (três artigos) e Informação & Sociedade (dois artigos).

A Tabela 3 apresenta a distribuição dos oito periódicos com mais artigos utilizados nessa pesquisa divididos a partir do ano de publicação e o Apêndice 4 lista os demais títulos.

Os dados que caracterizam esse *corpus* secundário foram retirados dos arquivos do texto completo de cada artigo e não dos metadados da WoS, no sentido de seguir o padrão de estruturação dos dados das referências.

Tabela 3. Distribuição dos artigos do *corpus* secundário de análise por ano de publicação e periódico.

Periódicos	Ano de publicação					Total	%
	2011	2012	2013	2014	2015		
1. KNOWLEDGE ORGANIZATION	07	12	18	20	26	83	54,97
2. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY	2	2	4	01	02	11	7,28
3. JOURNAL OF DOCUMENTATION	02	02	03	01	02	10	6,62
4. JOURNAL OF INFORMATION SCIENCE	--	--	04	02	01	07	4,64
5. PERSPECTIVAS EM CIENCIA DA INFORMACAO	04	03	--	--	--	07	4,64
6. LIBRARY TRENDS	--	02	02	--	02	06	3,97
7. SCIENTOMETRICS	--	--	02	01	--	03	1,99
8. TRANSINFORMACAO	01	--	--	01	01	03	1,99
Dois artigos (seis periódicos)	00	06	02	02	02	12	7,95
Um artigo (nove periódicos)	01	02	02	01	03	09	5,96
Total	17	29	37	29	39	151	100,00

Fonte: dados da pesquisa.

Os artigos que compõem o *corpus* secundário de análise dessa pesquisa possuem 289 autorias para 227 autores distintos. A Tabela 4 lista os 15 autores mais produtivos (a partir de três artigos) e distribui os dados pelo tipo de autoria dos artigos: única e coautoria (subdividida em principal e secundária). Os dados foram retirados dos arquivos do texto completo de cada artigo e a limpeza foi realizada manualmente, considerando os sobrenomes compostos.

Destacam-se os autores “Hjorland, B”, “Guimaraes, JAC” e “Martinez-Avila, D”, presentes também como destaque no *corpus* principal e os autores “Smiraglia, RP”, “Szostak, R” e “Tennis, JT”, com quatro ou mais artigos presentes na análise.

Além de “Guimarães, JAC” outros autores brasileiros se destacam, a saber: “Bufrem, LS”, “Café, LMA” e “Tognoli, NB”.

A estratégia de busca dos dois conjuntos de dados acarreta essa diferença pelo número de termos utilizados na recuperação e pelo recorte temporal utilizado, pois enquanto o *corpus* principal é mais heterogêneo, no *corpus* secundário há a predominância de artigos oriundos de um periódico específico.

Tabela 4. Distribuição das primeiras 15 posições dos autores dos artigos que formam o *corpus* secundário da análise.

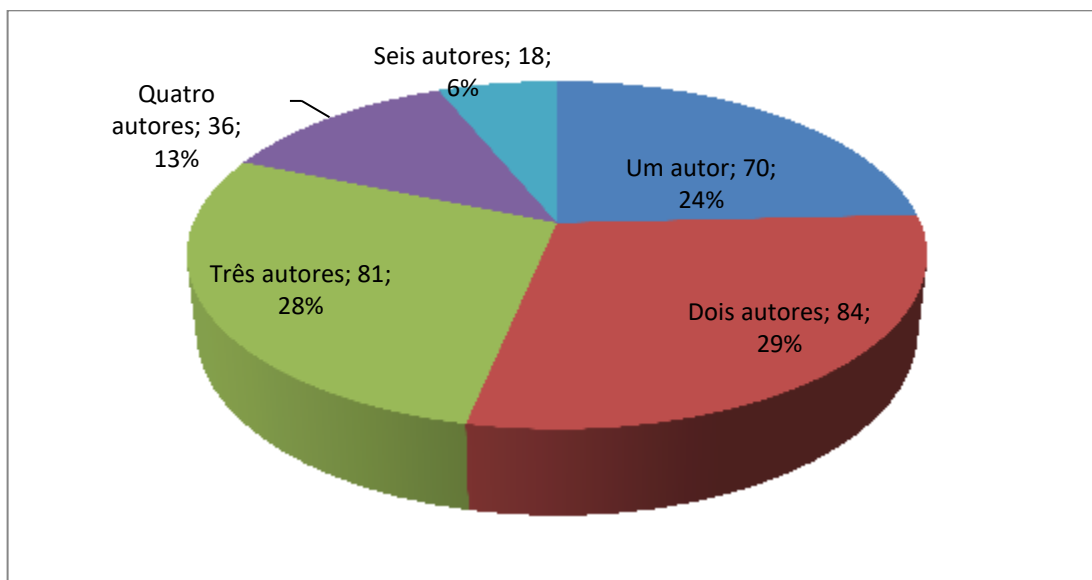
AUTORES	AUTORIA			Total	%
	Única	Principal	Secundária		
1. Hjorland, B	09	--	--	09	3,11
2. Guimaraes, JAC	--	01	05	06	2,08
3. Martinez-Avila, D	01	02	01	04	1,38
4. Smiraglia, RP	01	01	02	04	1,38
5. Szostak, R	04	--	--	04	1,38
6. Tennis, JT	03	--	01	04	1,38
7. Bufrem, LS	--	01	02	03	1,04
8. Cafe, LMA	--	--	03	03	1,04
9. Fox, MJ	--	02	01	03	1,04
10. Friedman, A	01	02	--	03	1,04
11. Gutierrez, AG	02	01	--	03	1,04
12. Lopez-Huertas, MJ	01	--	02	03	1,04
13. Thellefsen, M	--	01	02	03	1,04
14. Tognoli, NB	--	01	02	03	1,04
15. Tudhope, D	--	--	03	03	1,04
Autoria em até dois documentos	04	14	20	38	13,15
Autoria de um documento	44	55	94	193	66,78
Total	118	150	252	289	100,00

Fonte: dados da pesquisa.

A distribuição do número de autores dos artigos nesse conjunto de dados é similar aos dados do conjunto anterior, mas um pouco menos dispersa, com mínimo de um e o máximo de seis autores por artigo. A maioria dos artigos possui entre um e três autores, sendo que dois autores por artigo apresentam o maior escore (29%), comportamento similar aos dados do *corpus* principal, seguido por três autores por artigo (28%), autoria única (24%), quatro autores (13%) e seis autores por artigo (6%). A Figura 7 ilustra a distribuição desses dados com os valores absolutos e percentuais.

A coautoria (dois ou mais autores por artigo) é um comportamento presente na produção acadêmica do conjunto estudado, como evidencia os dois conjuntos de dados, mas a autoria única ainda se faz significativa.

Figura 7. Distribuição do número de autores dos artigos que formam o *corpus* secundário da análise.



Fonte: dados da pesquisa.

A seguir são apresentados aspectos relativos à estruturação e limpeza dos dados das referências.

3.4 Estruturação, limpeza e construção do conjunto de dados extraídos das referências

Os dados das referências foram retirados dos arquivos “pdf” de cada artigo que forma o conjunto de dados e não dos metadados da WoS. A estruturação e a limpeza dos dados foram realizadas manualmente. Esse procedimento se mostrou necessário na manipulação do *corpus* secundário, que foi realizada anteriormente em caráter de teste, principalmente pelas referências com autorias múltiplas, que é um elemento significativo para ACA, e por oferecer segurança quanto ao domínio dos dados e marcação adequada das informações dos artigos citantes, que se mostrou necessária para auxiliar na interpretação dos agrupamentos e nos dados que não serão utilizados, mas que podem ser analisados.

Com a coleta dos dados das referências foram gerados um arquivo de texto e uma planilha de dados para cada artigo citante, dos 421 que formam o *corpus* principal de análise, com o(s) autor(es), data de publicação e título do documento.

Na planilha de dados foram acrescentados alguns rótulos aos dados das referências, como o código do documento citante (que ajuda na construção das matrizes assimétricas “citante *vs* documentos” e “citantes *vs* autores”), código de referência (que permite retirar dados de coautoria da matriz de cocitação que inclui todos os autores das referências) e posição da autoria (para a geração do índice de autores apenas com as primeiras posições ou autoria única e também destacar os documentos, não permitir dupla contagem na DCA, pois todo o autor do índice teve um documento atribuído no índice, mas para documentos elaborados em coautoria foi necessário estabelecer uma forma para não duplicar a contagem).

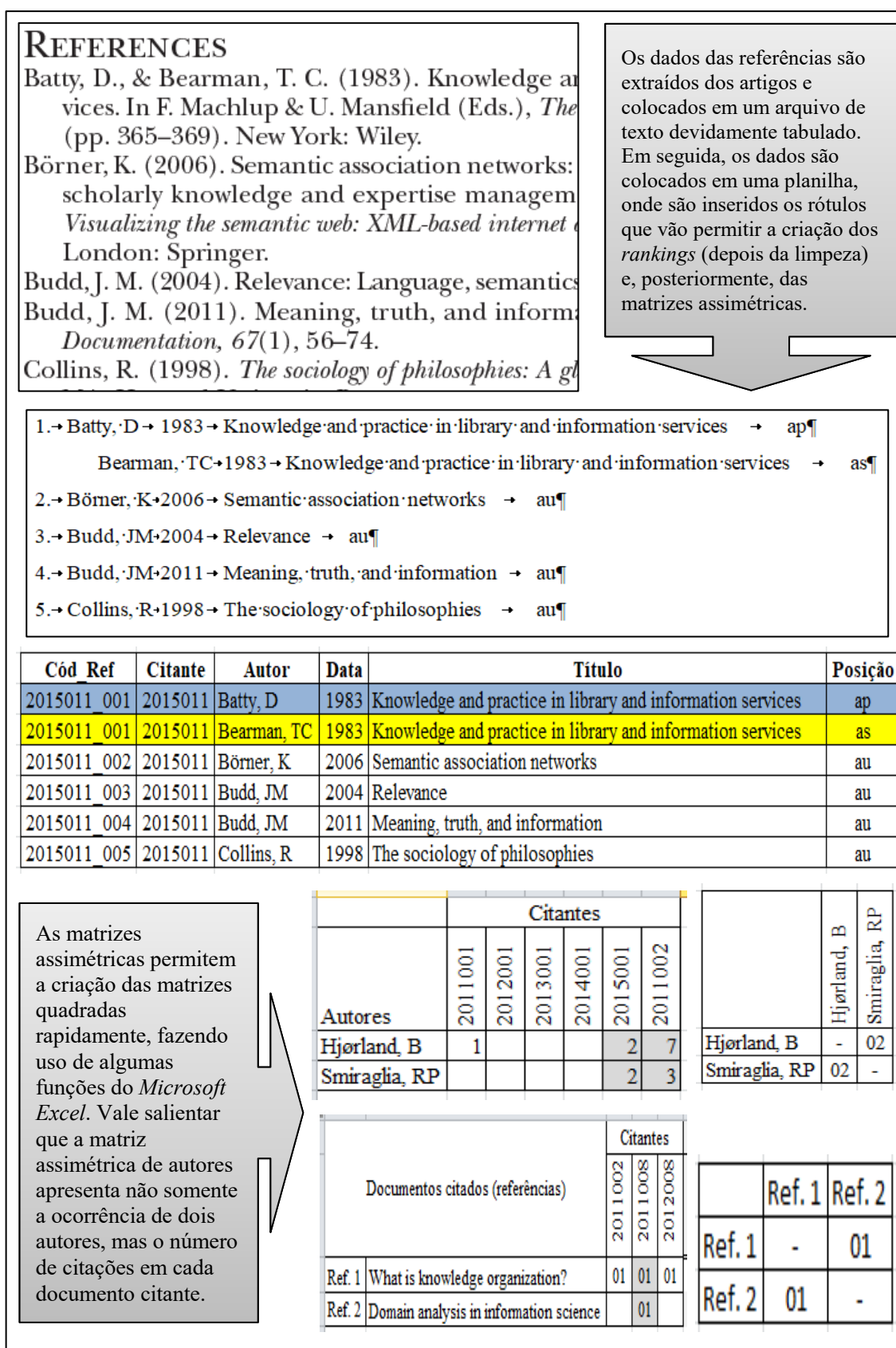
Após a criação das planilhas de cada artigo citante, os dados foram colocados em uma única planilha e foi iniciado o processo de limpeza dos nomes dos autores e dos títulos dos documentos, respectivamente.

Depois dessa etapa foram gerados três índices com os devidos rótulos: a) índice de autor incluindo todas as autorias das referências e não incluindo os documentos com autoria não pessoal (entradas por título, jurisdição, entidades coletiva etc.); b) índice de autor de primeira posição ou autoria única (não incluindo os documentos com autoria não pessoal); c) e índice de documentos (título), incluindo documentos de autoria não pessoal.

Em seguida foram geradas as matrizes assimétricas de ocorrência e depois as matrizes quadradas de coocorrência: todos os autores excluindo coautoria, todos os autores incluindo coautoria (mas não utilizados na pesquisa), primeiros autores e documentos. A Figura 8 apresenta e ilustra as etapas adotadas para os procedimentos descritos.

Quando da estruturação dos dados das referências do *corpus* secundário da análise, não foram incluídos os títulos dos documentos, pois não era um objetivo naquele momento, além do rótulo de “Código da referência”. Assim a matriz utilizada no tópico 4.6 (Tabela 25) apresenta dados de coautoria para a parte superior (dados de cocitação retirados das referências), dados esses que foram retirados manualmente para a contagem da cocitação na proximidade por parágrafo (parte inferior da matriz).

Figura 8. Ilustração das etapas de extração e estruturação dos dados das referências.



Fonte: elaborado pelo autor.

A Figura 8 apresenta um exemplo de documento elaborado em coautoria e a colocação dos rótulos permite o entendimento da distribuição dos dados. O código da referência (“Cód. Ref.”) possibilitou a não inclusão de dados de coautoria na matriz de citação que possui todos os autores. O rótulo de posição permitiu que o autor “Bearman, TC” não fosse incluído no índice de autores que considera apenas a primeira posição ou autoria única, como também permitiu a contagem correta dos documentos para que a DCA não fosse realizada com dados duplicados, pois documentos com a posição “AS” (autoria secundária) não foram considerados. Foi necessário atribuir o título do documento para cada autor para facilitar a recuperação dos dados na fase de interpretação e validação dos agrupamentos da ACA.

O Quadro 6 apresenta uma parte do índice geral elaborado com dados das referências, que tem 39.947 posições ou linha, independente da repetição dos autores ou títulos. Vale salientar que a coluna do “Código de referência” foi retirada para efeito de apresentação do Quadro 6 e essa informação faz parte do rótulo do código de referência (“Cód. Ref.”).

Quadro 6. Fragmento do índice geral do *corpus* principal da análise.

Cód. Ref.	Autor	Ano	Título	Posição
2016120_034	Nam-Park, O	2015	Development of Linked Data for Archives in Korea	AU*
2016162_023	Nan, X	2010	A comparison of a graph database and a relational database	AS*
2016056_016	Nana, E	2013	Quantum associative memory for the diagnosis of some tropical diseases	AS
2015075_034	Nanas, N	2004	Beyond TREC’s filtering track	AP*
2015139_001	Nanas, N	2011	Exploring ant colony optimisation for adaptive interactive search	AS
2016037_044	Nanba, H	2009	Classification of research papers into a patent classification system using two translation models	AP
2016037_043	Nanba, H	2008	Overview of the patent mining task at the NTCIR-7 workshop	AP
2016010_002	Nandy, P	2010	The YouTube video recommendation system	AS
2016142_074	Nannini, S	2011	L’anima e il corpo	AU
2016102_003	Nanny, M	2003	Turning techno-savvy into info-savvy	AS
2016011_028	Nanopoulos, A	2005	Content-Based Music Information Retrieval in Wireless Ad-Hoc Networks	AS
2016200_006	Nantajeewarawat, E	2016	Anontology-based framework for semantic reconciliation in humanitarian aid inEmergency information systems	AS
2015038_023	Naor, M	2001	Rank aggregation methods for the web	AS
2015076_011	Naor, M	2001	Rank aggregation methods for the web	AS
2015084_015	Naor, M	2001	Rank aggregation methods for the Web	AS
2016024_020	NAP*	2005	Libro blanco de los Títulos de Grados en Ciencias de la Comunicación	AO*
2015133_015	NAP	2013	#Saudi Arabic world’s 2nd most Twitter-happy nation	AO
2016030_041	NAP	2014	(Wikipedia) List of languages by number of native speakers	AO
2016031_113	NAP	2012	“Use” versus “usage.”	AO

*Legendas: NAP: autoria não pessoal; AU: autoria única; AS: autoria secundária; AP: autoria principal; e AO: autoria outra, utilizada para autoria não pessoal.

Fonte: dados da pesquisa.

O fragmento do índice geral apresentado no Quadro 6 requer algumas explicações: (i) a limpeza foi realizada apenas nos nomes dos autores nessa fase, a limpeza dos títulos para realizar a DCA foi realizada posteriormente e sem a inclusão dos “títulos” indicados com “AS” (autoria secundária) na coluna “Posição” e apenas com os autores que tinham a partir de dois documentos citados no *corpus* principal da análise; (ii) na coluna “Autor”, a sigla “NAP” significa “autoria não pessoal” (documentos com entrada na referência por título, jurisdição, entidade coletiva etc.) e implica que esses dados não foram utilizados para os índices dos autores, apenas para o índice de documentos, por isso na coluna “Posição” é atribuído o valor “AO” (autoria outra), além disso, todos os casos com autor “NAP” participaram da limpeza dos dados do índice de documentos; (iii) a partir da lista geral foram criadas outras três listas: todos os autores das referências (com 39.058 posições), primeiros autores e autoria única (com 17.109 posições) e lista de documentos (com 17.972 posições).

Outro aspecto importante a salientar é que todos os autores das referências foram identificados, ainda que omitidos nas listas no final dos artigos. Mesmo que a intenção fosse trabalhar quase que estritamente com os rastros dos artigos citantes, quando houve uma indicação de omissão de autores na referência, na medida do possível, o texto do documento referenciado ou outra fonte *on-line* disponível foi consultada. Há nos dados seis referências com mais de 30 autores, e uma obra com 81 autores.

Para realizar a desambiguação dos nomes dos autores foram consultados os títulos dos documentos e as respectivas datas de publicação presentes nas listas, assim como fontes externas *on-line* quando necessário, mas sempre a partir dos rastros dos artigos citantes. Além disso, todos os sobrenomes compostos foram verificados. A lista tinha inicialmente 23.283 “autores”, após a limpeza ficou com 19.590 autores distintos (diferença de 3.693).

A limpeza da lista de documentos ocorreu primeiro retirando os autores que tinham apenas uma citação, em seguida, foram retirados dos autores que tinham no mínimo duas citações, os documentos citados apenas uma vez, ou seja, primeiro saiu da lista autores com uma citação e depois documentos com uma citação dos autores que restaram. O total de documentos distintos foi de 14.901, sendo que 1.715 possuem duas ou mais citações e foram listados no *ranking* de documentos, que será detalhado no próximo tópico. Traduções foram identificadas e consideradas como mesma obra, assim como edições. Uma das dificuldades presentes na limpeza dessa etapa foram os

documentos citados em partes, pois foram publicados dessa forma e as partes foram tratadas de forma distinta apenas quando citadas pelo mesmo artigo, pois o impacto final foi pequeno e a intenção, como já salientado, era trabalhar com os rastros dos artigos e não fazer verificações muito detalhadas em fontes externas, o que impossibilitaria finalizar o trabalho.

No próximo tópico são caracterizados os dados das referências e os *rankings* produzidos com esses dados.

3.5 Caracterização das referências, dos rankings de autores e do ranking de documentos do corpus principal

Os dados extraídos das referências dos 421 artigos do *corpus* principal de análise serviram ao desenvolvimento de quatro conjuntos de dados, a saber: dados de caracterização das referências; *ranking* de autores incluindo todas as posições das referências; *ranking* de autores considerando a primeira posição e autoria única; e *ranking* de documentos, que inclui documentos de autoria não pessoal.

O conjunto de dados de caracterização das referências é composto de 11 indicadores relacionados a cada artigo citante, como seguem:

- i) Número de referências totais;
- ii) Número de referências de autoria pessoal;
- iii) Número de referências com um autor (autoria única);
- iv) Número de referências com dois ou mais autores;
- v) Número de referências com entrada por título, jurisdição, verbete, entidade coletiva etc.;
- vi) Número de autorias pessoais;
- vii) Número de referências desconsideradas, se necessário;
- viii) Número de autores citados apenas uma vez por artigo citante diante dos dados do *corpus*;
- ix) Número de documentos citados apenas uma vez por artigo citante diante dos dados do *corpus*;
- x) Fonte de publicação (periódico); e
- xi) Ano de publicação.

Os sete primeiros indicadores já foram desenvolvidos em trabalho anterior (CARVALHO; CAREGNATO, 2017), que também constituem o *corpus* secundário de análise.

Os indicadores oito e nove tiveram os valores retirados dos *rankings*, e servem para identificar a distribuição dos artigos citantes que mais possuem autores e documentos citados apenas uma vez no *corpus*, indo ao encontro do segundo objetivo específico da tese.

Os dois últimos indicadores já foram abordados no tópico 3.2 dessa seção, sendo que o periódico publicado pode ser necessário para a interpretação dos agrupamentos, como evidenciam Bu, Liu e Huang (2016), ou mesmo para caracterizar e entender os dados não analisados.

Na Tabela 5 são apresentadas as estatísticas descritivas dos nove primeiros indicadores retirados das referências, acrescentado o indicador “Nº de coautorias não pessoais”, que surgiu com a manipulação dos dados.

Tabela 5. Estatísticas descritivas dos indicadores das referências
(*corpus* principal – 421 artigos).

Indicadores	Total	Medidas de centralidade			Medidas de dispersão		
		Média	Mediana	Moda	Desvio Padrão	Mínimo	Máximo
Nº de referências	17992	42,74	38	38	27,37	02	236
Nº de referências de autoria pessoal	17112	40,65	36	46	26,89	02	221
Nº de referências com um autor	6822	16,20	13	14	15,18	00	164
Nº de referências com dois ou mais autores	10292	24,45	22	31	20,56	00	187
Nº de referências de autoria NÃO pessoal	862	2,05	00	00	3,95	00	34
Nº de referências desconsideradas	18	0,04	00	00	0,25	00	03
Nº de autorias pessoais	39060	92,78	81	33	70,30	02	621
Nº de coautorias NÃO pessoais	07	0,02	--	--	0,13	00	01
Nº de autores com 01 citação	13934	33,10	24	08	32,62	00	364
Nº de documentos com 01 citação	13124	31,17	26	23	23,33	00	226

Fonte: dados da pesquisa.

O segundo conjunto de dados é o *ranking* de todos os autores das referências. Os indicadores relacionados para cada autor são os seguintes:

- i) Número total de citações no *corpus*;
- ii) Número de artigos do *corpus* que citam o autor (documento citante);
- iii) Número de citações oriundas de documentos publicados em autoria única; e
- iv) Número de citações oriundas de documentos publicados em coautoria, desdobrado para autoria na posição principal e na posição secundária;

A Tabela 6 apresenta as 23 primeiras posições desse *ranking*, organizadas a partir do número de documentos citantes (30 ou mais). Vale salientar que havia sido previsto a identificação do número de documentos distintos de cada autor, mas esse trabalho se mostrou mais relevante para os autores que fazem parte da matriz final de cocitação. O *ranking* ficou com 19.590 autores distintos, a partir da limpeza realizada.

Tabela 6. Distribuição das 23 primeiras posições (30 ou mais documentos citantes) do *ranking* de autores que inclui todas as posições das referências.

Autores	Citações totais no corpus	Documentos citantes	Autoria única	Coautoria	Autoria principal	Autoria secundária
1. Hjørland, B	191	68	155	36	31	05
2. Croft, WB	136	60	03	133	09	124
3. Järvelin, K	77	45	03	74	15	59
4. Baeza-Yates, R	60	45	00	60	38	22
5. Salton, G	62	44	12	50	40	10
6. Robertson, SE	69	42	11	58	34	24
7. Spink, A	74	41	05	69	31	38
8. Saracevic, T	68	40	48	20	07	13
9. Belkin, NJ	95	38	20	75	22	53
10. Jansen, BJ	66	37	08	58	38	20
11. Manning, CD	49	36	00	49	35	14
12. Voorhees, EM	59	34	21	38	11	27
13. Olson, HA	106	33	75	31	16	15
14. Dumais, ST	65	33	01	64	04	60
15. Smiraglia, RP	73	32	54	19	08	11
16. Ingwersen, P	53	32	13	40	19	21
17. Buckley, C	49	32	02	47	23	24
18. Marchionini, G	49	32	26	23	04	19
19. Zhang, J	46	32	06	40	21	19
20. Ribeiro-Neto, B	37	32	00	37	01	36
21. Bates, MJ	47	30	46	01	01	00
22. Schutze, H	35	30	01	34	01	33
23. Raghavan, P	30	30	00	30	00	30

Fonte: dados da pesquisa.

Esses indicadores servem para contextualizar a participação de cada autor citado na pesquisa e, principalmente, para o desenvolvimento das matrizes de cocitação. A posição de cada autor nas referências (PERSSON, 2001; ROUSSEAU; ZUCCALA, 2004; SCHNEIDER; LARSEN; INGWER-SEN, 2009) permitiu a criação da matriz utilizando apenas os primeiros autores das referências, incluindo as referências de autoria única, e a matriz utilizando todos os autores das referências, que atende ao terceiro objetivo da pesquisa.

Assim, o terceiro conjunto de dados é o *ranking* de autores considerando apenas a primeira posição nas referências e a autoria única. Os indicadores são:

- i) Número total de citações no *corpus*;
- ii) Número de artigos do *corpus* que citam o autor (documento citante);
- iii) Número de citações oriundas de documentos publicados em autoria única; e
- iv) Número de citações oriundas de documentos publicados em coautoria, mas considerando apenas a primeira posição.

A Tabela 7 apresenta as 25 primeiras posições desse *ranking* organizadas, assim como no *ranking* anterior, a partir do número de artigos citantes (20 ou mais). O *ranking* ficou com 9.338 posições ou autores distintos, a partir da limpeza realizada.

Os dados se mostram menos robustos do que os apresentados na Tabela 6, mas é necessário contextualizar a diferença. Por exemplo, os autores “Baeza-Yates, R” e “Ribeiro-Neto, B” (Tabela 6) possuem uma relação de coautoria que os colocam nas primeiras posições, mas um único documento gera grande parte das citações dos autores (*Modern information retrieval* – várias datas), principalmente as citações de “Ribeiro-Neto, B”, ou seja, não há um ganho efetivo de relação bibliográfica com a participação dos dois autores em uma rede de cocitação.

Outro exemplo da diferença que a contagem de todos os autores pode proporcionar é o autor “Croft, WB”, que com 136 citações totais no *corpus* (Tabela 6) possui apenas 12 citações oriundas de documentos produzidos em autoria única ou autoria principal.

De qualquer forma, esses resultados são ilustrativos para evidenciar a aplicação dos indicadores propostos. As diferenças serão exploradas na seção de resultados desta pesquisa.

Tabela 7. Distribuição das 25 primeiras posições (20 ou mais documentos citantes) do *ranking* de autores que considera a primeira posição e autoria única.

Autores	Citações totais no corpus	Documentos citantes	Autoria principal	Autoria única
1. Hjørland, B	186	66	31	155
2. Salton, G	52	40	40	12
3. Robertson, SE	45	34	34	11
4. Olson, HA	91	33	16	75
5. Baeza-Yates, R	38	33	38	00
6. Saracevic, T	55	31	07	48
7. Manning, CD	35	31	35	00
8. Smiraglia, RP	62	30	08	54
9. Bates, MJ	47	30	01	46
10. Jansen, BJ	46	29	38	08
11. Belkin, NJ	42	27	22	20
12. Blei, DM	39	27	33	06
13. Spink, A	36	27	31	05
14. Ingwersen, P	32	25	19	13
15. Marchionini, G	30	25	04	26
16. Tennis, JT	42	24	04	38
17. Voorhees, EM	32	23	11	21
18. Lancaster, FW	32	22	06	26
19. Berners-Lee, T	31	22	16	15
20. White, RW	46	21	46	00
21. Kuhlthau, CC	36	21	03	33
22. Beghtol, C	31	21	00	31
23. Dahlberg, I	30	21	00	30
24. White, HD	45	20	16	29
25. Mai, JE	32	20	00	32

Fonte: dados da pesquisa.

O quarto conjunto de dados é constituído do *ranking* dos documentos. Esse conjunto apresenta como indicador o número de citações totais no *corpus* de análise e o tipo de autoria relacionada, que são: autoria única pessoal, coautoria pessoal e autoria não pessoal. O número de citações tem a mesma função do número de documentos citantes utilizado para os autores, pois um documento não pode ser referenciado duas vezes por um mesmo citante.

A Tabela 8 apresenta as 26 primeiras posições desse *ranking* (10 ou mais citações) que ficou com 1.715 documentos distintos, contabilizando os que possuíam a partir de duas citações no *corpus* de análise.

Tabela 8. Distribuição das 26 primeiras posições
(10 ou mais citações) do *ranking* de documentos.

Títulos – primeiro autor (ano)	Citações totais	Autoria Única	Coautoria	Autoria não pessoal
1. Modern information retrieval – Baeza-Yates, R (várias datas)	27	--	27	--
2. Latent Dirichlet allocation – Blei, DM (2003)	23	--	23	--
3. Introduction to information retrieval – Manning, CD (2008)	18	--	18	--
4. Relevance – Saracevic, T (várias datas)	17	17	--	--
5. The turn – Ingwersen, P (2005)	17	--	17	--
6. Domain analysis in information science – Hjørland, B (2002)	15	15	--	--
7. Toward a New Horizon in Information Science - Hjørland, B (1995)	15	01	14	--
8. Information Retrieval – Van Rijsbergen, CJ (1979)	14	14	--	--
9. The semantic web – Berners-Lee, T (2001)	14	--	14	--
10. The Power to Name – Olson, HA (2002)	13	13	--	--
11. What is Knowledge Organization (KO)? – Hjørland, B (2008)	13	13	--	--
12. Indexing and abstracting – Lancaster, FW (várias datas)	12	12	--	--
13. A vector space model for automatic indexing – Salton, G (1975)	11	--	11	--
14. An algorithm for suffix stripping – Porter, MF (1980)	11	11	--	--
15. Exploratory search – Marchionini, G (2006)	11	11	--	--
16. Inside the search process – Kuhlthau, CC (1991)	11	11	--	--
17. Perceived usefulness, perceived ease of use and user acceptance of information technology – Davis, FD (1989)	11	11	--	--
18. The design of browsing and berrypicking techniques for the online search interface – Bates, MJ (1989)	11	11	--	--
19. A language modeling approach to information retrieval – Ponte, JM (1998)	10	--	10	--
20. A survey of automatic query expansion in information retrieval – Carpineto, C (2012)	10	--	10	--
21. How are we searching the World Wide Web? – Jansen, BJ (2006)	10	--	10	--
22. ISO 25964-2 (2013)	10	--	--	10
23. Knowledge organization: a new science? – Dahlberg, I (2006)	10	10	--	--
24. Looking for information – Case, DO (2012)	10	10	--	--
25. Seeking meaning – Kuhlthau, CC (várias datas)	10	10	--	--
26. The PageRank citation ranking – Page, L (1999)	10	--	10	--

Fonte: dados da pesquisa.

Os dados se mostram ainda menos robustos pela especificidade da unidade de contagem, mas são mais informativos que os anteriores, no primeiro contato. Ocorre a inclusão de documento de autoria não pessoal, além da possibilidade de relacionar documentos do mesmo autor, por exemplo, os autores “Hjørland, B” e “Kuhlthau, CC” possuem mais de um documento com pelo menos 10 citações.

A seguir é apresentada a caracterização dos dados de referências e dos *rankings* de autores do *corpus* secundário de análise.

3.6 Caracterização das referências e dos rankings de autores do corpus secundário ____

As referências dos 151 artigos de pesquisa que formam o *corpus* secundário de análise geraram três conjuntos de dados distintos. Esses dados serviram de testes para as propostas apresentadas e, especificamente, para estabelecer o ponto de corte dos autores para as matrizes de cocitação e para parte dos dados que atendem o sexto e último objetivo específico proposto.

A Tabela 9 apresenta os dados descritivos de caracterização das referências que compõem o primeiro conjunto de dados desse *corpus* secundário. Os indicadores são:

- i) Número de referências;
- ii) Número de referências de autoria pessoal;
- iii) Número de referências com um autor (autoria única);
- iv) Número de referências com dois ou mais autores;
- v) Número de referências com entrada por título, jurisdição, verbete, entidade coletiva, etc.;
- vi) Número de autorias pessoais; e
- vii) Número de referências desconsideradas (se necessário).

Tabela 9. Estatísticas descritivas dos indicadores das referências
(*corpus* secundário – 151 artigos).

Indicadores	Total	Medidas de centralidade			Medidas de dispersão		
		Média	Moda	Mediana	Desvio padrão	Mínimo	Máximo
Nº de referências	5771	38,22	28	34	22,50	05	120
Nº de referências de autoria pessoal	5449	36,06	21	32	21,81	04	111
Nº de referências com um autor pessoal	3469	23,08	07	20	17,09	02	88
Nº de referências com dois ou mais autores pessoais	1980	13,11	02	09	11,72	00	59
Nº de referências de autoria não pessoal	318	2,11	00	01	5,29	00	55
Nº de autorias pessoais (incluindo todas as posições)	9311	61,66	33	49	43,00	06	213
Nº de referências desconsideradas	04	0,03	00	00	0,26	00	03

Fonte: dados da pesquisa.

O segundo conjunto de dados é um *ranking* que inclui todos os autores das referências e que possui os seguintes indicadores: total de citações no *corpus*, número de documentos citantes, citações oriundas de documentos publicados em autoria única, autoria principal e autoria secundária. O *ranking* ficou 5.333 posições ou autores distintos e a Tabela 10 apresenta as 25 primeiras posições organizadas a partir do número de documentos citantes.

Tabela 10. Distribuição das 25 primeiras posições (13 ou mais documentos citantes) do *ranking* de autores que inclui todas as posições das referências (*corpus* secundário)

Autores	Citações no <i>corpus</i>	Documentos citantes	Autoria única	Autoria principal	Autoria secundária
1. Hjørland, B	228	70	180	38	10
2. Olson, HA	61	31	51	07	03
3. Smiraglia, RP	68	29	51	08	09
4. Dahlberg, I	44	28	44	--	--
5. Mai, J-E	47	25	46	--	01
6. Tennis, JT	34	23	27	02	05
7. Hodge, Gail	25	23	23	01	01
8. Zeng, ML	30	21	07	14	09
9. Ranganathan, SR	31	19	30	01	--
10. Albrechtsen, H	23	19	02	01	20
11. Gnoli, C	35	18	22	12	01
12. Berners-Lee, T	32	18	20	10	02
13. Svenonius, E	25	18	25	--	--
14. Broughton, V	22	17	13	06	03
15. Soergel, D	26	16	19	04	03
16. Chan, LM	21	16	02	03	16
17. López-Huertas, MJ	21	16	12	01	08
18. Vickery, BC	20	15	20	--	--
19. Beghtol, C	20	14	20	--	--
20. Gruber, TR	17	14	17	--	--
21. McGuinness, DL	17	14	04	03	10
22. Szostak, R	59	13	54	01	04
23. Guimarães, JAC	32	13	08	16	08
24. Star, SL	17	13	02	02	13
25. Lancaster, FW	16	13	13	02	01

Fonte: dados da pesquisa.

O terceiro conjunto de dados é um *ranking* de autores que considera apenas autoria única ou autoria principal e que ficou com 2.938 autores distintos. A Tabela 11 apresenta as 24 primeiras posições desse *ranking* organizadas a partir do número de documento citantes. Além do número de documentos citantes, a tabela apresenta o

número total de citações no *corpus* e o número total de citações por posição (autoria única ou principal).

Tabela 11. Distribuição das 24 primeiras posições (11 ou mais documentos citantes) do *ranking* de autores (primeira posição e autoria única – *corpus* secundário).

Autores	Citações totais no <i>corpus</i>	Documentos citantes	Autoria única	Autoria principal
1. Hjørland, B	218	69	180	38
2. Olson, HA	58	30	51	07
3. Dahlberg, I	44	28	44	--
4. Smiraglia, RP	59	27	51	08
5. Mai, J-E	46	24	46	--
6. Hodge, Gail	24	22	23	01
7. Tennis, JT	29	21	27	02
8. Ranganathan, SR	31	19	30	01
9. Gnoli, C	34	18	22	12
10. Svenonius, E	25	18	25	--
11. Berners-Lee, T	30	17	20	10
12. Soergel, D	23	16	19	04
13. Zeng, ML	21	16	07	14
14. Vickery, BC	20	15	20	--
15. Broughton, V	19	15	13	06
16. Beghtol, C	20	14	20	--
17. Gruber, TR	17	14	17	--
18. Szostak, R	55	12	54	01
19. Frohmann, B	21	12	21	--
20. Lancaster, FW	15	12	13	02
21. Buckland, MK	14	12	12	02
22. Bowker, GC	12	12	--	12
23. Guimarães, JAC	24	11	08	16
24. Bates, MJ	11	11	10	01

Fonte: dados da pesquisa.

Esses dados auxiliares foram fundamentais para o entendimento do tema, para definir os objetivos da tese, e resultaram em uma publicação (CARVALHO, CAREGNATO, 2017) e um artigo aprovado aguardando publicação, com resultados presentes nessa pesquisa. Os títulos dos artigos que formam esse *corpus* foram utilizados para definir a estratégia de busca que resultou no *corpus* principal de análise.

No próximo tópico serão abordados os parâmetros de construção das matrizes de cocitação.

3.7 Parâmetros de construção e análise das matrizes de cocitação _____

Os parâmetros de construção das matrizes de cocitação de autores ou documentos com valores absolutos pressupõem três principais aspectos:

- i) O indicador utilizado para a inclusão dos autores/documentos: que pode ser o número de citações que cada um recebeu, por exemplo;
- ii) A definição de um ponto de corte para essa inclusão, ou seja, a definição do número de autores ou documentos que farão parte da análise de cocitação; e
- iii) O valor da diagonal da matriz quadrada, que é o tipo de matriz utilizada para fazer as análises estatísticas em estudos tradicionais de cocitação.

Em relação aos autores, a presente pesquisa utilizou o número de documentos citantes (GRÁCIO; OLIVEIRA, 2013b; CARVALHO; CAREGNATO, 2017) ao invés do número de citações (BU; LIU; HUANG, 2016; BU *et al*, 2017) ou a cocitação do autor com ele mesmo (McCAIN, 1990) para definir a ordem e a inclusão dos autores nas matrizes. O tópico 4.1 dos resultados vai justificar essa posição com base nos dados empíricos dos dois conjuntos de dados de referências apresentados anteriormente, comparando as distribuições de citações totais de cada autor com o número de documentos que os citam (documentos citantes) e os valores da cocitação do autor com ele mesmo. Já em relação aos documentos, o número de citações e o número documentos citantes são empiricamente idênticos.

Quanto ao ponto de corte para a inclusão dos autores/documentos, foi adotado o princípio de Pareto. Assim, conta-se a citação de cada entidade (autor/documento) a partir dos valores de artigos citantes e incluem-se os autores/documentos até aproximadamente 20% desses dados de citação (esse percentual foi calculado depois de testes, desconsiderando autores e documentos citados uma única vez, o que foi chamado de dispersão absoluta). Por exemplo, se 20% das citações oriundas de documentos citantes apresentarem 100 autores, mas outros autores apresentarem número de documentos citantes igual ao autor da centésima posição, eles serão incluídos. Essa proposta é parte do resultado da pesquisa. Dados empíricos para sustentar a proposta são apresentados no tópico 4.1 da seção dos resultados.

Quanto ao valor da diagonal, como já mencionado, não há um consenso na literatura sobre o indicador a ser utilizado. Entretanto, há concordância quanto à utilização do valor das citações parecer arbitrário (AHLGREN; JARNEVING;

ROUSSEAU, 2003), uma vez que a matriz é construída com valores de outra natureza. Zhou e Leydesdorff (2016) enfatizam que o valor da diagonal impacta no uso de um coeficiente de similaridade e o trabalho dos autores compara dois desses índices com base em definir a diagonal com zeros, pois a forma que se recupera os dados de cocitação não permite reconhecer essa informação.

Na presente pesquisa foi possível definir alguns valores a partir da forma de extração dos dados, inclusive a cocitação do autor com ele mesmo, pois a matriz assimétrica “citantes vs autores” carrega os valores de citação e não apenas a ocorrência dos autores. Sendo assim, para a ACA as matrizes com dados brutos possuem os seguintes valores na diagonal:

- i) Citação total no *corpus*. Entende-se que esse valor não é tão arbitrário como sugerem Ahlgren, Jarneving, Rousseau (2003) porque são dados que contextualizam os autores no *corpus* representando a influência de cada um, e como o estudo é metodológico, parece válida uma comparação com outros tipos de valores atribuídos na diagonal;
- ii) Número de documentos citantes no *corpus*. O indicador foi utilizado para definir a inclusão dos autores nas matrizes e há trabalhos que o utilizaram para esse propósito;
- iii) Cocitação do autor com ele mesmo. A forma de extração dos dados dessa pesquisa permite recuperar essa informação, mas há autores que não são cocitados dessa forma e há a presença de zeros ocasionais na diagonal. Outro aspecto importante é a limitação para os estudos que utilizam apenas os primeiros autores das referências; e
- iv) Zero. O zero foi atribuído na diagonal das matrizes para acompanhar os estudos metodológicos sobre tema.

No caso da DCA apenas o valor total de citações no *corpus* e o zero foram utilizados na diagonal das matrizes com valores absolutos.

O Quadro 7 apresenta as 12 primeiras posições da matriz quadrada de cocitação de autores (inclui todos autores das referências sem dados de coautoria) comparando os valores da diagonal (exceto com zero) com citação, número de documentos citantes e cocitação do autor consigo mesmo, respectivamente. Observa-se que a cocitação do autor com ele mesmo apresenta valores mais próximos aos dados de cocitação entre os autores presentes no quadro.

Quadro 7. Distribuição das 12 primeiras posições da matriz de cocitação de autores comparando três valores distintos na diagonal.

	Hjørland, B	Croft, WB	Baeza-Yates, R	Järvelin, K	Salton, G	Robertson, SE	Spink, A	Saracevic, T	Belkin, NJ	Jansen, BJ	Manning, CD
Diagonal com total de citações no corpus											
Hjørland, B	191	07	03	06	08	05	01	06	05	02	05
Croft, WB		136	25	13	22	24	10	11	09	11	19
Baeza-Yates, R			60	11	16	14	12	13	07	10	15
Järvelin, K				77	08	09	15	11	14	12	05
Salton, G					62	14	04	05	05	04	14
Robertson, SE						69	04	05	02	04	14
Spink, A							74	20	17	15	05
Saracevic, T								68	15	15	09
Belkin, NJ									95	11	03
Jansen, BJ										66	05
Manning, CD											49
Diagonal com nº de documentos de citantes											
Hjørland, B	68	07	03	06	08	05	01	06	05	02	05
Croft, WB		60	25	13	22	24	10	11	09	11	19
Baeza-Yates, R			45	11	16	14	12	13	07	10	15
Järvelin, K				45	08	09	15	11	14	12	05
Salton, G					44	14	04	05	05	04	14
Robertson, SE						42	04	05	02	04	14
Spink, A							41	20	17	15	05
Saracevic, T								40	15	15	09
Belkin, NJ									38	11	03
Jansen, BJ										37	05
Manning, CD											36
Diagonal com cocitação do autor com ele mesmo											
Hjørland, B	34	07	03	06	08	05	01	06	05	02	05
Croft, WB		28	25	13	22	24	10	11	09	11	19
Baeza-Yates, R			09	11	16	14	12	13	07	10	15
Järvelin, K				14	08	09	15	11	14	12	05
Salton, G					11	14	04	05	05	04	14
Robertson, SE						16	04	05	02	04	14
Spink, A							21	20	17	15	05
Saracevic, T								15	15	15	09
Belkin, NJ									20	11	03
Jansen, BJ										15	05
Manning, CD											08

Fonte: dados da pesquisa.

Quanto aos índices ou coeficientes de similaridade/correlação, três foram utilizados na pesquisa. O Cosseno de Salton (GRÁCIO; OLIVEIRA, 2013a, 2015; ZHOU; LEYDESDORFF, 2016), que é definido pela divisão da frequência de cocitação de dois elementos (autores/documentos) pela raiz quadrada de número de citações (ou número de documentos citantes) recebidas pelo elemento A multiplicado pelo número de citações do elemento B (GRÁCIO; OLIVEIRA, 2013a, p. 200).

$$\text{Fórmula: CS} = \frac{\text{cocitação (A,B)}}{\sqrt{\text{citação}^{10} \text{ (A)} \cdot \text{citação (B)}}$$

O Índice de Jaccard (LEYDESDORFF, 2007; GRÁCIO; OLIVEIRA, 2013a), que é determinado pela frequência de cocitação de dois elementos (autores/documentos) dividida pela soma das citações dos elementos, menos a cocitação entre eles (GRÁCIO; OLIVEIRA, 2013a, p. 200).

$$\text{Fórmula: IJ} = \frac{\text{cocitação (A, B)}}{\text{citação}^{11} \text{ A} + \text{citação B} - \text{cocitação}^{12} \text{ (A, B)}}$$

O coeficiente de correlação r de Pearson (FIELD, 2009; BECKER, 2015; GRÁCIO; OLIVEIRA, 2015).

Coeficiente que indica a força de associação entre quaisquer duas variáveis métricas. O sinal (+ ou -) indica a direção da relação. O valor pode variar de -1 a +1, em que +1 indica uma perfeita relação positiva, 0 indica relação nenhuma e -1, uma perfeita relação negativa ou reversa (quando uma variável se torna maior, a outra fica menor). (HAIR *et al*, 2009, p. 150).

A aplicação do coeficiente de correlação do r de Pearson foi feita diretamente no *software SPSS*¹³ e os demais foram aplicados manualmente, a partir das fórmulas apresentadas. A proposta é comparar o uso dos coeficientes em termos de agrupamentos das entidades depois das análises multivariadas.

¹⁰ Essa “citação” refere-se ao valor atribuído na diagonal da matriz de coocorrência.

¹¹ Idem anterior.

¹² Fórmula adaptada para o entendimento de aplicação.

¹³ IBM® SPSS® Statistics (Version 20).

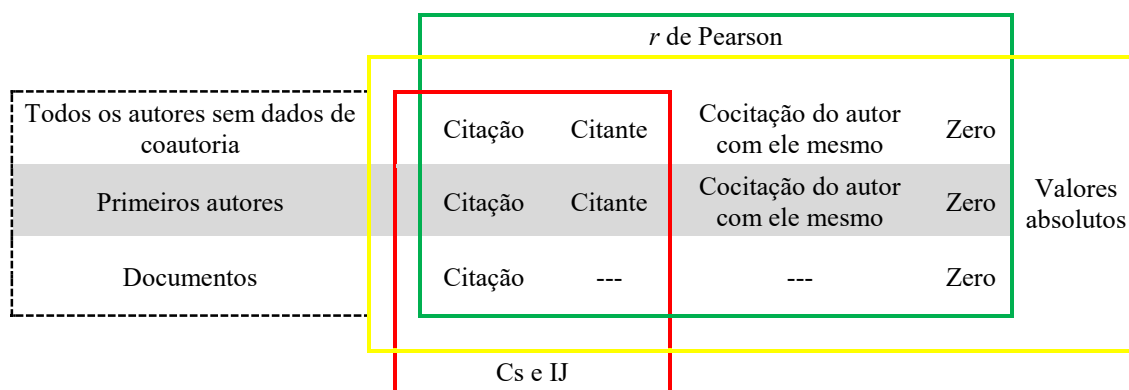
O uso do “*Ochiai Coefficient*” (ZHOU; LEYDESDORFF, 2016) foi proposto no projeto, mas a discussão que os autores propõem lida com a manipulação de matrizes de ocorrência, coocorrência e binárias, além da ausência de um valor confiável para ser atribuído na diagonal. Zhou e Leydesdorff (2016) concluem que o Cs e o Ochiai têm resultados similares e a equação três do Ochiai apresentada pelos autores é idêntica à fórmula do Cs apresentada por Grácio e Oliveira (2013a, p. 220).

Quanto às matrizes, as dos autores são divididas a partir de duas perspectivas: i) todos os autores excluindo os dados de coautoria; e ii) primeiros autores.

As quatro matrizes com valores absolutos de cada item anterior são diferenciadas pelo valor da diagonal (citação, documentos citantes, cocitação do autor com ele mesmo e zero) e duas dessas matrizes (citação e documentos citantes) foram normalizadas pelo Cs pelo IJ e todas as matrizes foram normalizadas pelo r de Pearson.

Assim, 24 matrizes foram analisadas quanto ao agrupamento dos autores. As matrizes dos documentos são seis, sendo duas com dados absolutos e diferenciadas pelo valor da diagonal (citação e zero), uma normalizada pelo Cs, uma normalizada pelo IJ e duas normalizadas pelo r de Pearson. A Figura 9 apresenta graficamente a estrutura de construção das matrizes com base no valor da diagonal e no uso dos coeficientes.

Figura 9. Estrutura de construção das matrizes com base no valor da diagonal e no uso dos coeficientes de similaridade.



Fonte: dados da pesquisa.

As técnicas de estatísticas multivariadas utilizadas, a partir do *software SPSS*, foram duas das mais utilizadas na literatura sobre cocitação: análise fatorial e análise de

clusters.

A análise fatorial (técnica de componentes principais) busca identificar a correlação (inter-relação) de um grupo de variáveis para definir um conjunto com dimensões comuns, os fatores ou componentes (HAIR *et al*, 2009; FIELD, 2009).

A rotina para aplicação dessa técnica foi estabelecida a partir do método de “componentes principais”, analisando a matriz de correlação, exibindo diagrama de sedimentação, sem número fixo de fatores e com no máximo de 25 interações por convergências, inicialmente. O método de rotação é o “Varimax”, com exibição rotacionada da solução e do diagrama, com interações máximas por convergência igual a 25, inicialmente (alguns dados exigiram a modificação desse valor para a geração da matriz rotacionada). Quanto às opções finais: classificação dos coeficientes de cada variável por tamanho dentro de cada componente; e coeficientes abaixo de “0,30” foram suprimidos.

Foram observados na aplicação dessa técnica os valores de comunalidades, que evidenciam a variância compartilhada de cada variável com todas as demais (HAIR *et al*, 2009), a variância total das componentes e a variância do principal agrupamento, número de componentes e interações, além da indicação observações sobre os resultados.

A análise de *clusters* ou agrupamento foi a segunda técnica utilizada. Ela tem a finalidade de agregar ou agrupar objetos a partir de suas características de similaridade (HAIR *et al*, 2009), ou seja, agregam-se as variáveis pela homogeneidade e separam-se os grupos ou partições pela heterogeneidade. A rotina para a aplicação da técnica foi o “Cluster hierárquico” das variáveis, exibindo estatísticas e diagramas, com “programação de aglomeração” e associação do *Cluster* com faixas de solução entre “02” e “06”, inicialmente, com exibição de dendrograma do cluster orientado verticalmente. O método utilizado foi o “Vizinho mais distante”, com intervalo de “Correlação de Pearson”, sem transformação de medidas e valores. A análise do dendrograma deu-se na medida de evitar o isolamento de algum autor/documento, já que há subjetividade na escolha dos grupos e foi indicada a presença “clara” de algum subgrupo. Foi testado o método de “Ward”, mas optou-se pela forma como Vanz (2009) aplicou a técnica em um estudo de coautoria institucional (vizinho mais distante). O EMD não foi utilizado na pesquisa.

A seguir é apresentado como as matrizes de cocitação com todos os autores foram criadas, levando em conta os dados de coautoria.

3.8 Diferenciação das matrizes de cocitação com dados de coautoria

A forma de recuperação das relações de cocitação que inclui todas as posições das referências (*corpus* principal) ocasionou a necessidade de incluir um rótulo de identificação de cada referência (“Cód. Ref.” – ver Figura 8 e Quadro 6) para gerar uma matriz quadrada de coautoria dos autores selecionados para essa parte do estudo, já que a matriz quadrada incluía as relações de coautoria como cocitação, sem apresentar uma distinção entre elas. Parecia simples fazer a subtração da matriz com os valores de cocitação pela matriz com os valores de coautoria, mas a forma de extração incluía na contagem todas as coautorias nos documentos citantes e não apenas a ocorrência (contar “um” quando houvesse e “zero” quando não houvesse coautoria). O Quadro 8 demonstra a relação de coautoria de dois autores presentes em um único documento citante e indica que a contagem não era simples.

Quadro 8. Relação de coautoria dos autores “White, RW” e “Ruthven, I” em um documento citante.

	2016016_004	2016016_021	2016016_049	2016016_053	2016016_054	2016016_064	2016016_065	2016016_068	2016016_069	2016016_070	2016016_071	2016016_072	2016016_073	2016016_074	2016016_075	2016016_076	2016016_077	2016016_078	2016016_079	2016016_080
White, RW	--	01	--	--	--	--	--	--	01	01	01	01	01	01	01	01	01	01	01	01
Ruthven, I	01	--	01	01	01	01	01	01	--	--	01	01	--	--	01	01	01	01	01	01

Fonte: dados da pesquisa.

Todos os dados apresentados no Quadro 8 contam como uma única relação de cocitação para esses autores, e apenas a relação em destaque, em vermelho, já validaria essa condição. Entretanto, surgiram casos dos mesmos autores e de outros em que a condição era como a que se apresenta em destaque em verde, onde apenas uma citação

dos autores não era em coautoria entre eles no documento citante, e ficou decidido que essa situação validaria a contagem de cocitação. Seriam desconsiderados apenas os casos como no destaque em azul, ou seja, todas as citações do par de autores (ou apenas uma) tinham que ser em coautoria. O Quadro 9 complementa a explicação com dados reais do *corpus* principal de análise dos autores “Boyack, KW” e “Klavans, R”.

Quadro 9. Relação de coautoria dos autores “Boyack, KW” e “Klavans, R” em três documentos citantes.

Referências citadas	2015043_004	2015043_005	2015043_006	2015043_008	2015043_011	2015043_012	2015043_014	2015043_017	2015043_021	1ª condição DESCONSIDERAR a cocitação. Pois no documento citante “2015043” há um único documento citado dos autores e elaborado em coautoria.
Boyack, KW								01		
Klavans, R								01		
Referências citadas	2015125_011	2015125_012	2015125_014	2015125_015	2015125_016	2015125_017	2015125_018	2015125_024	2015125_026	2ª condição CONSIDERAR a cocitação. Pois no documento citante “2015125” pelo menos um documento citado desses autores não é elaborado em coautoria (2015125_011)
Boyack, KW	01		01	01	01					
Klavans, R			01	01	01					
Referências citadas	2016212_007	2016212_008	2016212_009	2016212_010	2016212_011	2016212_012	2016212_013	2016212_015	2016212_017	3ª condição DESCONSIDERAR a cocitação. No documento citante “2016212” há seis documentos citados desses autores e TODOS são elaborados em coautoria.
Boyack, KW	01	01	01	01	01	01				
Klavans, R	01	01	01	01	01	01				

Fonte: dados da pesquisa.

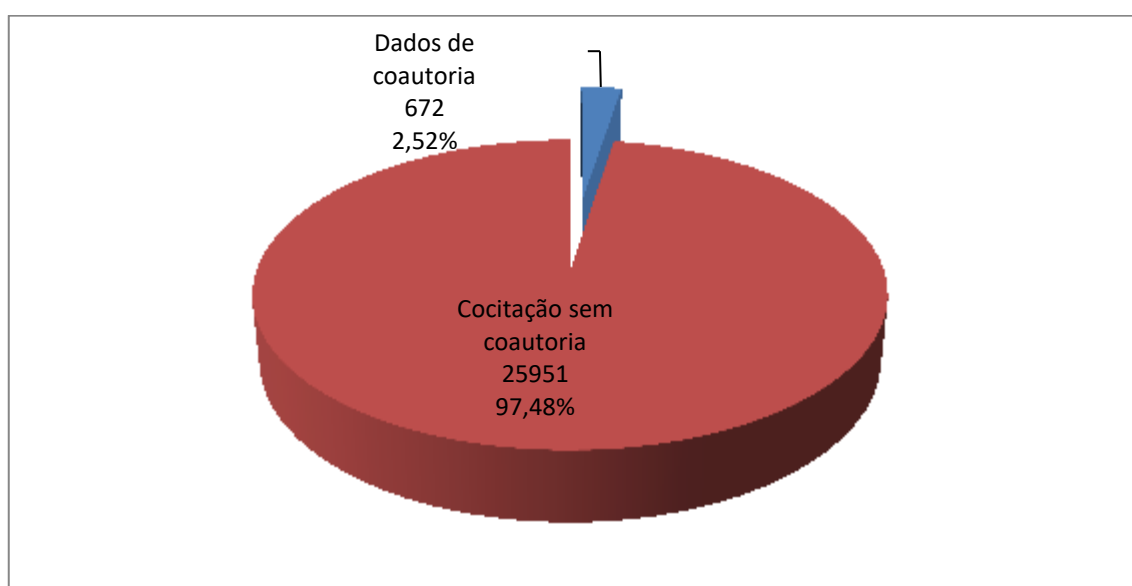
Para uma matriz quadrada de cocitação criada a partir de uma matriz de ocorrência (autores vs citantes), com a inclusão de todos os autores das referências, as três condições apresentadas no Quadro 9 são consideradas relações de cocitação, mas para efeito desse trabalho apenas a segunda condição se manteve como tal. Os autores “Boyack, KW” e “Klavans, R” não possuem no *corpus* uma condição de cocitação “pura”, onde são citados em um mesmo documento a partir de referências distintas.

Vale salientar que foi considerada a possibilidade de contar a 3ª condição apresentada no Quadro 9 como uma relação de cocitação válida, pois há mais de um

documento envolvido e assim a não contagem omite uma relação bibliográfica em ACA, no caso entre os seis documentos citados, mas essa condição foi descartada.

A figura 10 ilustra a diferença do número de pares de cocitação após a retirada dos dados de coautoria da matriz de cocitação que incluiu todos os autores das referências. Os dados são contabilizados a partir de 180 autores citados em pelo menos 12 artigos (documentos citantes).

Figura 10. Proporção dos pares de coautoria com os pares de cocitação (N=26623).



Fonte: dados da pesquisa.

As relações entre os autores que demandaram esse total de pares de coautoria foram da ordem de 329, sendo que depois de analisadas caso a caso, 140 relações foram definidas com “zero”, ou seja, existe coautoria, mas a forma de contagem estabelecida não interfere na contagem da ocorrência de cocitação dos autores.

Os resultados são apresentados e discutidos no próximo capítulo.

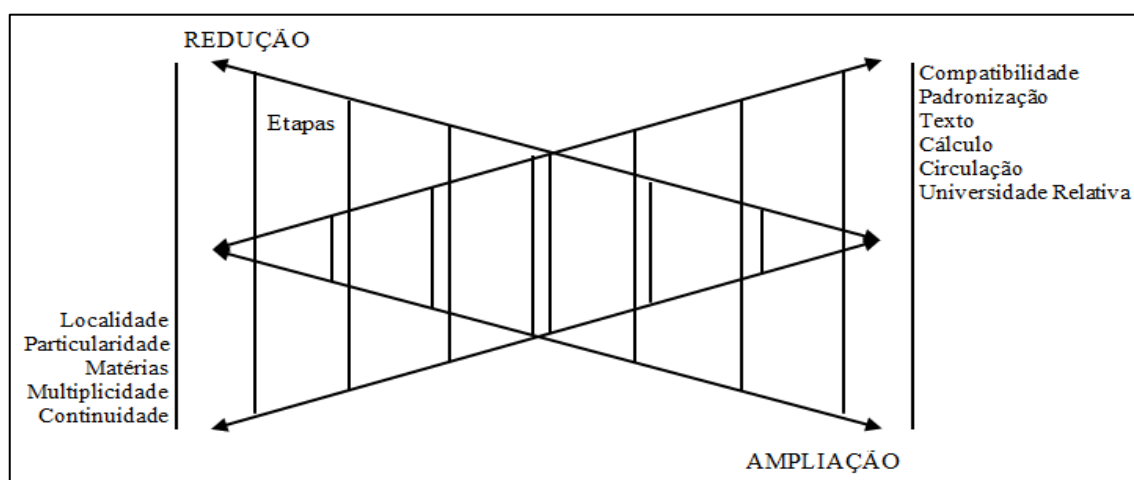
4 RESULTADOS E DISCUSSÃO

Os seis objetivos específicos propostos nessa pesquisa se dividem basicamente em dois grandes núcleos de análise: todos os dados que formam o *corpus* principal até a definição de um ponto de corte para a inclusão dos elementos (autores/documentos) que fazem parte da análise; e a análise da “elite”, que se destaca no conjunto de dados a partir dos critérios estabelecidos.

A problematização proposta se baseia, portanto, não apenas na aplicação de uma técnica de estatística multivariada em uma matriz quadrada de citações de autores ou documentos para identificar agrupamentos, mas em contextualizar e descrever o caminho para chegar nessa condição, detalhando, na medida das possibilidades oferecidas pela estruturação dos dados: as implicações das escolhas de cada fase de uma ACA/DCA; a relação dos citantes com essa elite; o alcance da elite analisada em relação ao todo; as especificidades das unidades de contagem, etc.

Nessa perspectiva, a conceito de informação de Bruno Latour foi adotado como norte para entender a informação que um estudo de citação transmite a partir do que o autor sugere como operações da produção de informação: seleção, extração e redução (LATOUR, 2006). A Figura 11 apresenta os triângulos isósceles invertidos que ilustram o jogo de redução e ampliação da produção de informações de que fala o autor.

Figura 11. Triângulos isósceles invertidos da produção de informações (Latour) com suas respectivas bases: redução e ampliação.



Fonte: Latour (2006, p.24).

Essa figura, devidamente adaptada, vai aparecer algumas vezes ainda nesse texto a partir de cada etapa da ACA e DCA operacionalizada.

Os dois grandes núcleos de análise mencionados no início desta seção têm como ponto de divisão, considerando uma área específica do conhecimento em um recorte temporal, o ponto de corte para inclusão dos autores e documentos. Essa definição diz o que é “importante” e o que é dispersão na análise, e será o tema da próxima parte dos resultados, onde os dois conjuntos de dados são utilizados, a fim de apresentar mais dados empíricos para justificar a proposta apresentada.

4.1 Uma proposta para definir o ponto de corte para estudos de cocitação _____

Para atender ao primeiro objetivo proposto é necessário entender a relação entre os indicadores “citações”, “documentos citantes” (quantidade de citações oriundas do número de documentos que cita um autor) e “cocitação do autor consigo mesmo”. Qualquer autor pode ser referenciado/citado mais de uma vez por um documento citante, desde que sua produção permita essa condição. Assim, contar o número de documentos que citam um autor permite entender melhor a distribuição de suas citações e o seu real alcance em relação a um *corpus* de análise, bem como se o autor é citado mais de uma vez pelo mesmo documento citante, quando então ocorre uma “cocitação” do autor com ele mesmo.

Esses três indicadores são utilizados na literatura sobre cocitação para ordenar os autores e para a atribuição do valor da diagonal da matriz quadrada (McCain, 1990; EOM; Farris, 1996; White; McCain, 1998; Grácio; Oliveira, 2013b; BU; Liu; Huang, 2016; Carvalho; Caregnato, 2017; BU *et al*, 2017).

No próximo tópico, utilizando dados do *corpus* secundário da análise, serão apresentados dados que evidenciam a relação entre as citações totais e as citações contadas por documentos citantes, além de uma proposta de ponto de corte para inclusão dos autores, utilizando as citações a partir de documentos citantes como parâmetro. Espera-se que essas evidências empíricas favoreçam escolhas para futuras pesquisas.

4.1.1 Discussão com os dados do *corpus* secundário

A Tabela 12 (próxima página) apresenta um *ranking* com 153 autores, que relaciona citações e documentos citantes. É possível observar que poucos possuem diferença igual a zero (apenas onze – onde a citação total é igual a distribuição dos autores pelos documentos citantes do *corpus*). O autor “Hjørland, B” (destaque em vermelho nas bordas), que é o primeiro do *ranking*, tem suas 228 citações distribuídas por 70 citantes dos 151 artigos que formam o *corpus* secundário da pesquisa.

Os autores em destaque na Tabela 12 (**negrito**, *itálico* e sublinhado) evidenciam que há uma discrepância entre as citações e o número de documentos citantes. Isso pode interferir na inclusão desses autores em uma matriz de cocitação, uma vez que a distribuição dos autores pelos documentos citantes pode oferecer mais possibilidades de ocorrências de cocitação com os outros autores, e isso é uma preocupação nos estudos, inclusive como critério de inclusão dos autores (McCain, 1990). Ou seja, um autor muito citado e pouco cocitado tem pouca relevância para um estudo tradicional de cocitação, que tenha como objetivo mapear uma disciplina.

O autor “Ostwald, W” (destaque em verde nas bordas) possui 26 citações no *corpus* secundário e certamente seria incluso em uma análise de cocitação que consideraria o número de citações totais como indicador de inclusão, mas esse dado é oriundo de apenas três citantes e isso interfere nas possíveis relações de cocitação. O autor “Fahlman, SE” (destaque em azul nas bordas) é um caso extremo, pois possui nove citações oriundas de apenas um documento e é outro exemplo de que a distribuição dos autores pelo *corpus* é mais significativa para estudos de cocitação do que o número total de citações.

Mas é importante salientar que esses indicadores são próximos e sua distinção só é possível com um domínio adequado dos dados dos documentos citantes e que dependendo da forma de recuperação nem sempre é possível estabelecer essa diferença.

Para identificar a proximidade dos indicadores foi utilizada a medida de correlação r de Pearson em alguns arranjos do *ranking* de autores que inclui todas as posições das referências. Esses dados evidenciam que os resultados apresentam forte correlação mesmo com as suas especificidades. O Quadro 10 apresenta os valores de correlação para cinco arranjos desses dados.

Tabela 12. Ranking de autores comparando citações totais com documentos citantes do corpus secundário (organizado por citações totais e com 153 posições).

AUTORES	CITAÇÕES	CITANTES	DIFERENÇA	AUTORES	CITAÇÕES	CITANTES	DIFERENÇA	AUTORES	CITAÇÕES	CITANTES	DIFERENÇA
Hjørland, B	228	70	158	La Barre, Kathryn	12	09	03	Miles, Alistair	08	06	02
Smiraglia, RP	68	29	39	Ingwersen, Peter	12	08	04	Moya-Anegón, F	08	06	02
Olson, HÁ	61	31	30	<i>Lykke Nielsen, M</i>	12	05	07	Saracevic, Tefko	08	06	02
Szostak, R	59	13	46	<i>Pinho, FA</i>	12	05	07	Andersen, Hanne	08	05	03
Mai, J-E	47	25	22	Bates, MJ	11	11	-	Wilson, TD	08	05	03
Dahlberg, I	44	28	16	Hendler, James	11	11	-	Bizer, Christian	08	04	04
Gnoli, Claudio	35	18	17	Lassila, Ora	11	11	-	Klein, JT	08	04	04
Tennis, JT	34	23	11	Jacob, EK	11	10	01	Maltese, Vincenzo	08	04	04
Berners-Lee, Tim	32	18	14	Kuhn, TS	11	10	01	Milani, SO	08	04	04
Guimarães, JAC	32	13	19	Mazzocchi, F	11	10	01	Welty, CA	08	04	04
Ranganathan, SR	31	19	12	Spiteri, LF	11	10	01	<i>Cahier, JP</i>	08	03	05
Zeng, ML	30	21	09	Furner, Jonathan	11	09	02	<i>Rosch, E</i>	08	03	05
García Gutiérrez, A	30	08	22	Taylor, AG	11	09	02	<i>Sorensen, Bent</i>	08	02	06
<i>Thellefsen, TL</i>	27	09	18	Keizer, Johannes	11	07	04	Belkin, NJ	07	07	-
Soergel, Dagobert	26	16	10	Pejtersen, AM	11	05	06	Börner, Katy	07	07	-
<i>Ostwald, W</i>	26	03	23	Poli, Roberto	11	05	06	Cronin, Blaise	07	07	-
Hodge, Gail	25	23	02	<i>Ding, Y</i>	11	04	07	Foskett, AC	07	07	-
Svenonius, Elaine	25	18	07	<i>Pepper, Steve</i>	11	04	07	Foskett, DJ	07	07	-
Thellefsen, MM	24	11	13	Berman, Sanford	10	09	01	Rowley, JE	07	07	-
Albrechtsen, H	23	19	04	Gilchrist, Alan	10	08	02	Sowa, JF	07	07	-
Broughton, Vanda	22	17	05	Hartel, Jenna	10	08	02	Day, RE	07	06	01
Chan, LM	21	16	05	Miksa, FL	10	08	02	Doerr, Martin	07	06	01
López-Huertas, MJ	21	16	05	Losee, RM	10	06	04	Frické, Martin	07	06	01
Frohmann, Bernd	21	12	09	Van Harmelen, F	10	06	04	Golder, SA	07	06	01
Vickery, BC	20	15	05	Giunchiglia, F	10	04	06	Lauser, Boris	07	06	01
Beghtol, Clare	20	14	06	Tognoli, NB	10	04	06	Rorty, Richard	07	06	01
Gruber, TR	17	14	03	van den Heuvel, C	10	04	06	Vizine-Goetz, D	07	06	01
McGuinness, DL	17	14	03	<i>Tudhope, D</i>	10	02	08	Willis, Craig	07	06	01
Star, SL	17	13	04	Bawden, David	09	09	-	Bufrem, LS	07	05	02
Smith, Barry	17	11	06	Liang, AC	09	08	01	Chen, Chaomei	07	05	02
<i>Peirce, CS</i>	17	08	09	Dousa, TM	09	07	02	Dumais, ST	07	05	02
<i>Rayward, WB</i>	17	08	09	Järvelin, K	09	07	02	Farrow, JF	07	05	02
<i>Brier, Soren</i>	17	05	12	McCulloch, Emma	09	07	02	Fox, MJ	07	05	02
Lancaster, FW	16	13	03	Panzer, Michael	09	07	02	Jiménez-Contreras, E	07	05	02
Feinberg, M	16	08	08	Salton, G	09	07	02	Miller, GA	07	05	02
Martínez-Ávila, D	16	06	10	Chowdhury, GG	09	06	03	Brickley, Dan	07	04	03
<i>Fujita, MSL</i>	16	03	13	Garshol, LM	09	06	03	Cabré, MT	07	04	03
Friedman, Alon	15	10	05	Lakoff, G	09	05	04	Campos, MLA	07	04	03
Green, Rebecca	15	10	05	<i>Bourdieu, Pierre</i>	09	03	06	Dutta, Biswanath	07	04	03
Buckland, MK	14	12	02	<i>Golub, K</i>	09	03	06	Fernández-Molina, JC	07	04	03
<i>Budd, JM</i>	14	08	06	<i>Witten, JH</i>	09	03	06	Hapke, T	07	04	03
Small, HG	14	06	08	<i>Herre, Heinrich</i>	09	02	07	Marteleteo, RM	07	04	03
Wittgenstein, L	13	08	05	<i>Fahlman, SE</i>	09	01	08	Tudhope, Douglas	07	04	03
Guarino, Nicola	13	06	07	Aitchison, Jean	08	07	01	White, HD	07	04	03
Otlet, P	13	06	07	Eco, Umberto	08	07	01	<i>Bakhtin, Mikhail</i>	07	03	04
Bowker, GC	12	12	-	Ibekwe-SanJuan, F	08	07	01	<i>Blei, DM</i>	07	03	04
Noy, NF	12	11	01	Shiri, AA	08	07	01	<i>Tillett, BB</i>	07	03	04
Andersen, Jack	12	10	02	Stock, WG	08	07	01	<i>Weller, K</i>	07	03	04
Greenberg, Jane	12	10	02	Bechhofer, Sean	08	06	02	<i>Beak, Jihee</i>	07	02	05
Hansson, Joacim	12	10	02	Ellis, David	08	06	02	<i>Sigel, Alexander</i>	07	02	05
Foucault, Michel	12	09	03	Lambe, Patrick	08	06	02	<i>Zacklad, Manuel</i>	07	02	05

Fonte: dados da pesquisa.

Quadro 10. Medidas de correlação do *ranking* de autores do *corpus* secundário considerando “totais de citações” e “citações por documentos citantes”.

ARRANJO		Valor da correlação (r de Pearson)
1	5333 posições <i>Ranking</i> total considerando todos os autores das referências	,903
2	933 posições Considerando os autores a partir de dois documentos citantes	,905
3	1249 posições Considerando os autores a partir de duas citações	,893
4	153 posições Recorte do <i>ranking</i> apresentado na Tabela 12 desse texto	,889
5	156 posições Recorte do <i>ranking</i> com número de autores próximos ao da Tabela 12 organizados pelo número de documentos citantes (quatro ou mais)	,913

Fonte: dados da pesquisa.

A força de correlação se mostra próxima para todos os arranjos, mas os maiores valores estão ligados à organização do *ranking* pelo número de documentos citantes (itens 2 e 5 do Quadro 10), que são valores mais consistentes e regulares que os valores de citações totais. Em DCA, por exemplo, não há essa especificidade que existe na ACA.

Além da questão da classificação dos autores para a inclusão em estudos, é preciso salientar que a escolha do indicador tem relação com o valor atribuído na diagonal da matriz quadrada e isso influencia diretamente nos resultados da aplicação de algum índice de similaridade. As fórmulas do Cs e do IJ apresentadas por Grácio e Oliveira (2013b) consideram o valor atribuído na diagonal para a operacionalização dos índices e as autoras utilizaram, como já relatado, o número de documentos citantes.

O Quadro 11 apresenta a aplicação do IJ, conforme a fórmula descrita por Grácio e Oliveira (2013b), para uma pequena matriz quadrada com 10 posições dos autores com mais documentos citantes do *corpus* secundário da análise. Os dados evidenciam a diferença nas medidas a partir do uso dos valores totais de citação e dos valores de citação definidos a partir de documentos citantes. Ou seja, os valores de distribuição dos autores no *corpus* (documentos citantes) são mais próximos da realidade das relações de cocitação e, portanto, destacam essas relações que o número total de citações diminui.

Quadro 11. Comparação da aplicação do Índice de Jaccard com dados totais de citação e número de documentos citantes (*corpus secundário*).

AUTORES	Hjørland, B	Olson, HA	Smiraglia, RP	Dahlberg, I	Mai, J-E	Tennis, JT	Hodge, Gail	Zeng, ML	Ranganathan, SR	Albrechtsen, H
Hjørland, B	17	19	23	20	16	14	8	14	19	
Olson, HA		11	06	12	13	03	01	01	05	
Smiraglia, RP			09	08	11	06	01	02	06	
Dahlberg, I				08	06	07	04	05	09	
Mai, J-E					06	06	03	03	08	
Tennis, JT						01	02	05	09	
Hodge, Gail							07	02	04	
Zeng, ML								02	01	
Ranganathan, SR									03	
Albrechtsen, H										
Citações totais										
228	Hjørland, B	0,063	0,069	0,092	0,078	0,065	0,059	0,032	0,057	0,082
61	Olson, HA		0,093	0,061	0,125	0,159	0,036	0,011	0,011	0,063
68	Smiraglia, RP			0,087	0,075	0,121	0,069	0,010	0,021	0,071
44	Dahlberg, I				0,096	0,083	0,113	0,057	0,071	0,155
47	Mai, J-E					0,080	0,091	0,041	0,040	0,129
34	Tennis, JT						0,017	0,032	0,083	0,188
25	Hodge, Gail							0,146	0,037	0,091
30	Zeng, ML								0,034	0,019
31	Ranganathan, SR									0,059
23	Albrechtsen, H									
Documentos citante										
70	Hjørland, B	0,202	0,238	0,307	0,267	0,208	0,177	0,096	0,187	0,271
31	Olson, HA		0,224	0,113	0,273	0,317	0,059	0,020	0,020	0,111
29	Smiraglia, RP			0,280	0,249	0,354	0,183	0,020	0,052	0,187
28	Dahlberg, I				0,257	0,188	0,224	0,120	0,157	0,299
25	Mai, J-E					0,208	0,208	0,094	0,096	0,292
23	Tennis, JT						0,022	0,048	0,135	0,273
23	Hodge, Gail							0,189	0,050	0,105
21	Zeng, ML								0,053	0,026
19	Ranganathan, SR									0,086
19	Albrechtsen, H									

Fonte: dados da pesquisa.

Assim, o valor de citação por documentos citantes pode ser considerado também um bom indicador para definir o número de autores que farão parte de um estudo de cocitação. Do mesmo modo, o Princípio de Pareto (BECKER, 2015) se apresenta como uma alternativa, pois determina uma proporção entre dados triviais e dados importantes para explicar algum fenômeno. Embora essa proporção não seja absoluta, o mais comum é a utilização da relação 80-20, onde aproximadamente 80% dos dados são considerados triviais ou de dispersão e 20% são dados mais significativos. Vale salientar que mais do que determinar um número de autores para a inclusão em uma matriz de cocitação, parece necessário relacionar de alguma forma essa definição com os dados que não serão utilizados ou considerados como dispersão. Penan¹⁴ (1989 citado por McCAIN, 1990, p. 435) utilizou essa proporção de autores, mas usando dados de citação e cocitação e não número de documentos citantes.

Assim, utilizando os valores de documentos citantes de cada autor foi identificado um número específico de citados (considerando o número de documentos citantes e não o total de autores) que mais ficasse próximo dos 20%, ou que não excedesse muito esse percentual. Essa definição dos dados do *corpus* secundário serve para comparar com os dados do *corpus* principal, no próximo tópico, que geraram efetivamente as matrizes utilizadas nessa pesquisa.

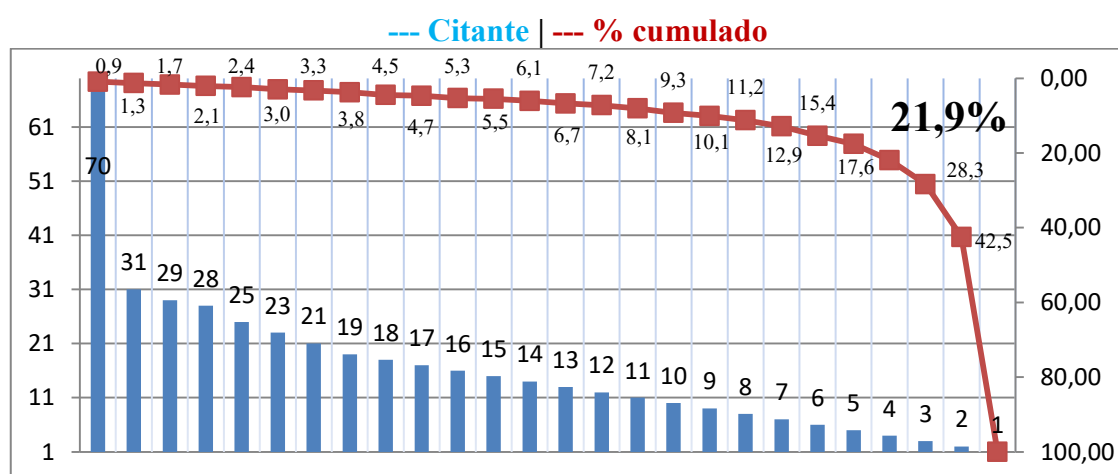
As Figuras 12 e 13 ilustram o primeiro teste para entender a distribuição em percentual dos valores de documentos citantes dos dois *rankings* gerados com os dados do *corpus* secundário: todos os autores das referências, com 5.333 posições; e dos primeiros autores, com 2.938 posições.

A elaboração das figuras utilizou os valores absolutos de documentos citantes como categorias (azul), onde o mínimo era um citante por autor e o máximo era 70 para todos os autores e 69 para os primeiros autores somente respectivamente. Foi calculado o percentual acumulado dessa distribuição, que é apresentado em uma linha vermelha invertida que realça a relação com as categorias em azul. A partir dessa perspectiva, os autores citados em quatro documentos ou mais do *ranking* com todos os autores das referências entram na análise de cocitação, totalizando 227 autores, que correspondem a 21,9% (Figura 12) das citações contadas a partir dos documentos citantes. No caso do

¹⁴ Penan, H. (1989). **Pour une gestion bibliometrique de l' information scientifique et technique des entreprises**. Application en theorie microeconomique et financiere. Doctoral dissertation, Universite des Sciences Sociales. Toulouse I, Toulouse, France. [Texto completo não localizado].

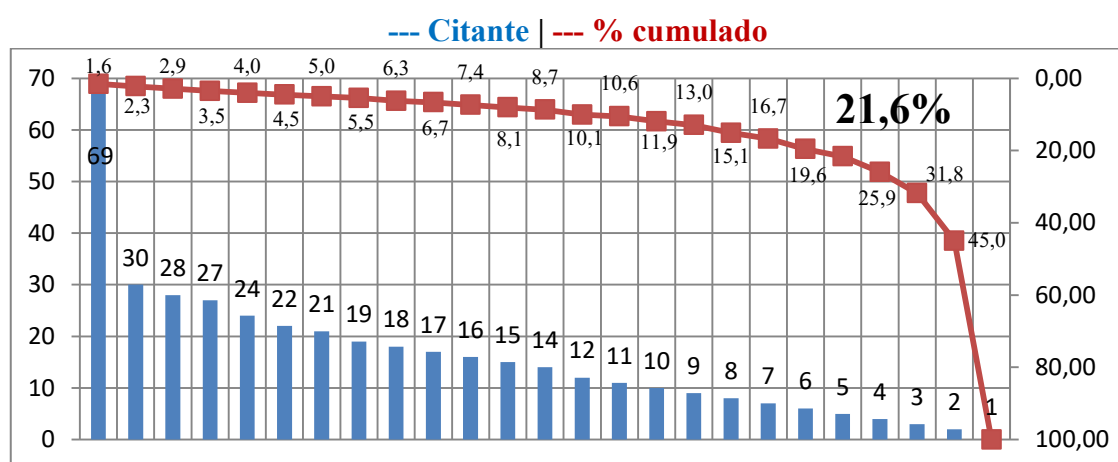
ranking com os primeiros autores, o ponto de corte é estabelecido a partir de cinco documentos citantes, no total de 96 autores, que corresponde a 21,6% (Figura 13) das citações contadas a partir dos documentos citantes. A intenção, portanto, foi escolher os autores pelo rastro dos documentos onde são citados e o número total de autores não precisa ser redondo, como ocorre em outros estudos (WHITE; MCCAIN; 1998; ZHAO; STROTMANN, 2014; LIU; HUANG, 2016; BU *et al*, 2017).

Figura 12. Distribuição de valores de documentos citantes como categoria para definição de ponte de corte (todos os autores das referências – *corpus* secundário).



Fonte: dados da pesquisa.

Figura 13. Distribuição de valores de documentos citantes como categoria para definição de ponte de corte (primeiros autores – *corpus* secundário).



Fonte: dados da pesquisa.

Esse primeiro teste se mostrou parcialmente equivocado por dois motivos: o número de documentos citantes por autor era baixo e a quantidade de autores parecia muito grande em decorrência desse corte. Ainda que McCain (1990) tenha tido bons resultados com valores baixos, estudos com grandes quantidades de dados (WHITE; McCAIN, 1998; KREUZMAN, 2001; ZHAO; STROTMANN, 2014; KHASSEH; SOHEILI; CHELAK, 2018; ZHAO; ZHANG; KWON, 2018) analisaram menos autores. Mas a preocupação maior foi o número limite de documentos citantes por autor.

Nesse sentido, um segundo teste com os mesmos dados descartou todos os autores citados em apenas um documento citante nos dois *rankings*. Esses autores foram considerados sem um alcance mínimo de relacionamento no *corpus*, ou seja, sem ser citado em pelo menos dois documentos. Essa condição de ser citado em apenas um documento, no contexto do *corpus*, caracteriza mais esses autores pela relação com o documento citante do que pela relação com os outros autores citados. Para os estudos de cocitação, no entendimento dessa pesquisa e no contexto estrito do *corpus* analisado, são dados de natureza dispersiva absoluta.

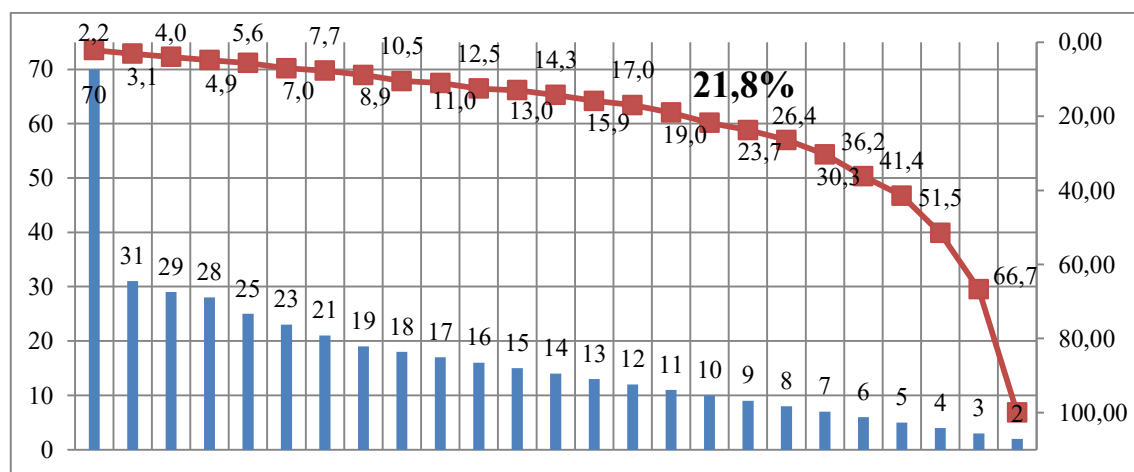
A partir dessa decisão foram, inclusive, definidos os indicadores “Número de autores citados apenas uma vez por artigo citante diante dos dados do *corpus*” e “Número de documentos citados apenas uma vez por artigo citante diante dos dados do *corpus*”, que caracterizam os dados das referências no tópico 3.5 dos procedimentos metodológicos desta pesquisa.

A Figura 14 ilustra os resultados do segundo teste com o *ranking* que inclui todas as posições das referências e exclui os autores citados em apenas um documento citante. Nessas condições o *ranking* ficou com 933 posições, sendo que 4.400 autores foram retirados dessa fase da análise. Assim, a partir das condições estabelecidas, os autores com citações em pelo menos 10 documentos citantes seriam selecionados para um estudo de cocitação, ou seja, 43 autores (com 21,8% das citações oriundas de documentos citantes).

A aplicação no *ranking* com os primeiros autores das referências definiu o ponto de corte em autores citados em pelo menos 12 documentos citantes (22,4% das citações oriundas de documentos citantes). O *ranking* nas condições estabelecidas para o teste ficou com 519 posições, sendo que 2.419 autores foram retirados dessa fase da análise. Considerando as posições analisadas, 22 autores fariam parte de um estudo de cocitação. A inclusão de mais autores para esse caso não seria arbitrária, pois o que

parece efetivamente relevante na proposta apresentada é fazer a relação adequada dos dados analisados com os dados não analisados. Nesse caso específico do *ranking*, considerando os primeiros autores, a inclusão a partir de 10 documentos citantes incluiria apenas mais oito autores, sendo eles “Guimarães, JAC”, “Bates, MJ”, “Friedman, A”, “Smith, B”, “Kuhn, TS”, “Spiteri, LF”, “Andersen, J” e “Noy, NF”; e subiria o percentual para 26,53% das citações oriundas de documentos citantes.

Figura 14. Distribuição de valores de documentos citantes (dois ou mais) como categoria para definição de ponto de corte (todos os autores – corpus secundário).



Fonte: dados da pesquisa.

Propõem-se, portanto, uma breve taxonomia para os dados a partir da proposta estabelecida para definição do ponto de corte para a inclusão dos autores em ACA, a saber:

- i) Dados analisados: incluindo autores que, somados os valores de citação de documentos citantes, correspondam a aproximadamente 20% dos dados do *corpus*, excluindo valores de dispersão absoluta;
- ii) Dados intermediários de dispersão: inclui autores citados em pelo menos dois documentos até o limite estabelecido para a análise e pode ser dividido em dois subconjuntos:
 - a. metade próxima aos dados analisados: autores com alta potencialidade de uso em estudo de cocitação. Nos dados ilustrados na Figura 14, seriam autores entre seis e nove citações oriundas de documentos citantes; e

- b. metade próxima aos dados de dispersão absoluta: autores com baixa potencialidade de uso em estudos de cocitação. Nos dados apresentados na Figura 14, seriam autores citados entre dois e cinco documentos citantes;
- iii) Dispersão absoluta: autores citados em apenas um documento.

Essa forma para definir os autores em um estudo de cocitação não tem a pretensão de ser definitiva ou a mais correta, mas ressalta-se que os estudos dessa natureza deveriam apresentar dados descritivos que mostram a proporcionalidade dos autores utilizados e não utilizados no contexto de um estudo. Para esse fim, ao menos, essa abordagem parece válida.

No próximo tópico serão apresentados os resultados relativos aos dados do *corpus* principal da análise, já desconsiderando a dispersão absoluta, utilizando citações a partir de documentos citantes e apresentação e caracterizando as matrizes quadradas de cocitação que respondem aos demais objetivos dessa pesquisa.

4.1.2 Discussão com os dados do *corpus* principal e apresentação das matrizes com valores absolutos

A definição dos autores ou documentos que farão efetivamente parte de um estudo de cocitação é uma questão crucial para a problematização apresentada nessa pesquisa, como já enfatizado algumas vezes.

Sendo assim, o uso dos dados utilizados para testes da proposta se deu pela necessidade de analisar os resultados (nessa perspectiva do ponto de corte) em dois conjuntos de dados distintos, ainda que de natureza temática similar, mas principalmente por permitir inferências em dois conjuntos de tamanhos diferentes, pois o *corpus* principal é quase três vezes maior que o *corpus* secundário em número de artigos citantes (421 *versus* 151), mais de três vezes em números de referências (17.992 *versus* 5.771), mais de quatro vezes em número de autorias pessoais (39.060 *versus* 9.311), só para citar alguns exemplos das diferenças dos conjuntos de dados.

Mas para além da necessidade de comparar os dados, os resultados destacados do *corpus* principal incluem a DCA, além dos dados de cocitação do autor com ele mesmo.

A Tabela 13 apresenta, na próxima página, um *ranking* com 107 posições que compara valores dos três indicadores considerando todos os autores das referências: documentos citantes, citações totais e cocitação do autor com ele mesmo. Os dados são organizados a partir de autores citados em pelo menos 15 artigos do *corpus*.

Os dados dos autores “Raghavan, P”, “Ng, AY”, “Hendler, JA” e “Park, JR” (em destaque com bordas em vermelho) demonstram que a “cocitação do autor com ele mesmo” tem pouco impacto na distribuição dos autores pelos documentos citantes e mesmo sendo uma informação de natureza relacional, assim como a cocitação entre autores distintos, parece um dado arbitrário, assim como o número total de citação, como indicam Ahlgren, Jarneving e Rousseau (2003) para definir a inclusão de autores em ACA. Os autores citados não apresentam nenhuma cocitação consigo mesmo, mas possuem um alcance relativo em termos de artigos citantes, com valores iguais a 30, 26, 19 e 17, respectivamente.

Vale salientar também que utilizar esse indicador na diagonal da matriz quadrada, mesmo com autores citados em muitos documentos e com muitas citações totais, pode implicar, como nos destaques mencionados acima, em zeros que impedem a aplicação adequada dos índices de similaridade Cs e IJ, pois é necessário que o valor da diagonal seja no mínimo igual ao maior valor de cocitação do autor com outro autor.

Outros sujeitos em destaque (**negrito**, *itálico* e sublinhado), com cocitação consigo mesmo igual ou menor a cinco, evidenciam a distância de dados de cocitação do autor com ele mesmo para dados de citações totais e de documentos citantes. Salienta-se a relevância dessa informação, mas não se recomenda seu uso para definir a inclusão de autores em ACA, a não ser com um objetivo estritamente relacionado com a natureza dessa informação.

Os autores com destaque em verde nas bordas evidenciam a diferença dos indicadores “documentos citantes” e “citações totais”, como na tabela 12, que era organizada pelas citações totais. Esses dados reforçam que o uso dos valores de citação a partir de documentos citantes é mais adequado para definir a inclusão dos autores em estudos de cocitação.

Tabela 13. Ranking de todos os autores das referências comparando valores de documentos citantes, citações totais e cocitação do autor com ele mesmo (*corpus* principal – 107 posições – 15 ou mais documentos citantes).

AUTORES	DC	CT	CoA	AUTORES	DC	CT	CoA
Hjørland, B	68	191	34	<i>Bawden, D</i>	20	26	03
Croft, WB	60	136	28	<i>Li, Y</i>	20	25	04
Järvelin, K	45	77	14	<i>Harman, DK</i>	20	23	02
Baeza-Yates, R	45	60	09	Guimarães, JAC	19	40	08
Salton, G	44	62	11	Wilson, TD	19	38	08
Robertson, SE	42	69	16	Callan, JP	19	38	07
Spink, A	41	74	21	<i>Ruthven, I</i>	19	34	02
Saracevic, T	40	68	15	<i>Hu, X</i>	19	32	04
Belkin, NJ	38	95	20	Garfield, E	19	30	06
Jansen, BJ	37	66	15	<i>Sparck Jones, K</i>	19	26	05
Manning, CD	36	49	08	<i>Ellis, D</i>	19	25	05
Voorhees, EM	34	59	16	Hendler, JA	19	19	00
Olson, HA	33	106	17	<i>Li, H</i>	18	36	05
Dumais, ST	33	65	16	<i>Fox, EA</i>	18	35	04
Smiraglia, RP	32	73	18	Lafferty, JD	18	34	10
Ingwersen, P	32	53	11	Nie, JY	18	33	08
Marchionini, G	32	49	13	<i>Zobel, J</i>	18	29	05
Buckley, C	32	49	09	<i>Moffat, A</i>	18	27	05
Zhang, J	32	46	06	<i>Jones, Rosie</i>	18	23	05
<i>Ribeiro-Neto, B</i>	32	37	03	<i>MacFarlane, A</i>	18	22	04
Bates, MJ	30	47	10	Soboroff, I	17	30	07
<i>Schutz, H</i>	30	35	05	Macdonald, C	17	29	06
Raghavan, P	30	30	00	<i>Wildemuth, BM</i>	17	28	05
Tennis, JT	29	54	12	<i>Bar-Ilan, J</i>	17	24	04
Blei, DM	28	51	10	<i>Svenonius, E</i>	17	21	04
Metzler, D	28	40	08	<i>Chen, H</i>	17	21	01
Sanderson, M	28	37	06	<i>Page, L</i>	17	19	02
Vakkari, P	27	47	08	<i>Jones, SA</i>	17	18	01
White, RW	26	70	12	Park, JR	17	17	00
Ng, AY	26	26	00	Leydesdorff, L	16	51	09
Lancaster, FW	25	36	07	Glänzel, W	16	44	10
<i>Van Rijsbergen, CJ</i>	25	29	03	<i>Small, H</i>	16	33	05
Ding, Y	24	46	09	<i>Wolfram, D</i>	16	28	03
Berners-Lee, T	24	39	10	Xu, J	16	27	06
Ounis, I	24	39	07	<i>Han, J</i>	16	25	03
Furner, J	24	32	07	<i>McCallum, A</i>	16	25	02
Zeng, ML	24	32	07	<i>Singhal, A</i>	16	19	03
<i>Jordan, MI</i>	24	31	03	<i>Buckland, MK</i>	16	18	02
Witten, IH	23	43	07	<i>Chan, LM</i>	16	16	00
De Rijke, M	23	41	09	Yan, E	15	29	07
Allan, J	23	39	06	<i>Dervin, B</i>	15	21	04
<i>Li, X</i>	23	28	05	<i>Chowdhury, GG</i>	15	21	02
Zhai, CX	22	56	10	<i>Toms, EG</i>	15	20	02
Kelly, D	22	40	08	<i>Fidel, R</i>	15	19	02
Beghtol, C	22	32	06	<i>Kamps, J</i>	15	19	02
<i>Teevan, J</i>	22	31	05	<i>Van Raan, AFJ</i>	15	18	02
<i>Walker, S</i>	22	24	02	<i>Brin, S</i>	15	17	02
White, HD	21	50	11	<i>Zhang, M</i>	15	17	02
<i>Kuhlthau, CC</i>	21	36	05	<i>Broder, AZ</i>	15	16	01
Joachims, T	21	34	06	<i>Cronin, B</i>	15	16	01
Dahlberg, I	21	30	07	<i>Yu, Y</i>	15	16	01
Borlund, P	20	35	06	<i>Nejdl, W</i>	15	15	00
Mai, JE	20	32	08	<i>Strohman, T</i>	15	15	00
<i>Chen, C</i>	20	28	04				

Legendas: DC: documentos citantes; CT: citações totais; CoA: cocitação do autor com ele mesmo.

Fonte: dados da pesquisa.

Não pareceu necessário aprofundar alguns aspectos já abordados em relação ao *corpus* secundário, mas apresentar alguns dados que demonstrem com mais qualidade a relação da cocitação do autor com ele mesmo é necessário. A Tabela 14 apresenta a distribuição para os dois *rankings* (todos e primeiros autores) dos valores de cocitação do autor com ele mesmo e evidencia que mais 85% dos autores, nos dois *rankings*, não possuem cocitação desse tipo e que uma única ocorrência fica em 11,18% para todos os autores (N=19.590) e 10,44%, considerando as primeiras posições das referências (N=9.339).

Tabela 14. Distribuição dos valores de cocitação dos autores com eles mesmos (*corpus* principal).

	Todos os autores das referências		Primeiros autores das referências	
	Nº de autores	%	Nº de autores	%
Sem cocitação	16831	85,92	8160	87,38
01 cocitação	2191	11,18	975	10,44
02 cocitações	314	1,60	105	1,12
03 cocitações	107	0,55	34	0,36
04 cocitações	45	0,23	22	0,24
05 cocitações	31	0,16	14	0,15
06 cocitações	18	0,09	06	0,06
07 cocitações	15	0,08	09	0,10
08 cocitações	09	0,05	05	0,05
09 cocitações	05	0,03	02	0,02
10 cocitações	06	0,03	02	0,02
Acima de 10 cocitações	18	0,09	05	0,05
Total	19590	100,00	9339	100,00

Fonte: dados da pesquisa.

O Quadro 12 apresenta a medida de correlação (r de Pearson) dos três indicadores em alguns arranjos para os dois *rankings*. Os dados mostram uma relação próxima dos indicadores nos quatro cenários propostos: todas as posições dos *rankings*; considerando autores citados em dois ou mais documentos; considerando autores a partir de uma cocitação com ele mesmo; e autores a partir de duas citações totais.

Os destaques no quadro (fundo cinza) mostram valores de correlação abaixo de “0,8” em dois arranjos para os dois *rankings* entre os indicadores “documentos citantes”

e “cocitação do autor com ele mesmo”. A cocitação do autor com ele mesmo é uma situação que ocorre a partir do número de citações totais em menos documentos citantes e por isso a correlação é mais alta entre valores de cocitação desse tipo e citações totais.

Quadro 12. Medidas de correlação dos *rankings* de autores do *corpus* principal considerando documentos citantes, citações totais e cocitação do autor com ele mesmo.

TODOS OS AUTORES DAS REFERÊNCIAS				PRIMEIROS AUTORES DAS REFERÊNCIAS			
		DC	CT			DC	CT
19590 posições	DC	1		9339 posições	DC	1	
	CT	0,933	1		CT	0,922	1
	CoA	0,791	0,910		CoA	0,770	0,902
<hr/>				<hr/>			
4509 posições (dois ou mais citantes)	DC	1		1891 posições (dois ou mais citantes)	DC	1	
	CT	0,928	1		CT	0,920	1
	CoA	0,810	0,923		CoA	0,810	0,924
<hr/>				<hr/>			
2759 posições (uma ou mais cocitação)	DC	1		1179 posições (uma ou mais cocitação)	DC	1	
	CT	0,930	1		CT	0,920	1
	CoA	0,836	0,931		CoA	0,831	0,936
<hr/>				<hr/>			
5610 posições (duas ou mais citações totais)	DC	1		2426 posições (duas ou mais citações totais)	DC	1	
	CT	0,920	1		CT	0,911	1
	CoA	0,747	0,893		CoA	0,732	0,888

Legenda: DC - documentos citantes; CT - citações totais; CoA - Cocitação do autor com ele mesmo

Fonte: dados da pesquisa.

Todos esses dados se encaixam no enfoque normativo dos EC (SILVEIRA; CAREGNATO, 2017b), conforme demonstra o Quadro 1. Sendo os objetos os elementos constitutivos dos registros referenciados, foi possível, a partir de um único *ranking*, inferir sobre suas duas divisões (todos *versus* primeiros autores) e comparar as especificidades de três indicadores distintos (citações totais, documentos citantes e cocitação do autor com ele mesmo) que se relacionam e que possuem a mesma origem. Toda essa complexidade diminui quando o elemento de contagem é o documento, pois as citações totais dos documentos têm a mesma capacidade de mapear a distribuição das

citações como o “documento citante” dos autores e como um documento não pode ser citado duas vezes pelo mesmo citante não é possível contar a cocitação para essa condição.

Discutido as relações e especificidades dos três indicadores relacionados aos *rankings* de autores e considerando o número de artigos citantes, o valor mais apropriado para definir a inclusão dos autores em estudos de cocitação é o momento de formar as matrizes, ou seja, estabelecer os autores que continuarão no estudo.

Os testes iniciais para definir os autores partiram dos valores de documentos citantes e incluíram os dados de dispersão absoluta, ou seja, todos os autores dos dois *rankings*. Para o *ranking* com todos os autores das referências, o ponto de corte ficou estabelecido inicialmente em 606 autores citados em pelo menos seis artigos e os valores de citação a partir dos documentos citantes corresponderam a 20,90% do total das 31.849 citações oriundas de documentos citantes. Para o *ranking* com os primeiros autores das referências o ponto de corte ficou definido inicialmente em 331 autores citados em pelos menos cinco artigos do *corpus* e os valores de citações a partir de documentos citantes corresponderam a 21,55% do total das 14.326 citações contadas a partir dos documentos citantes.

Esses dados foram ao encontro dos resultados obtidos na análise do *corpus* secundário, ou seja, muitos autores para um número de documentos citantes baixo. Os dados de dispersão absoluta interferem de maneira significativa na análise e são retirados para realizar a segunda rodada de testes, que acabam por definir o número de autores para as análises.

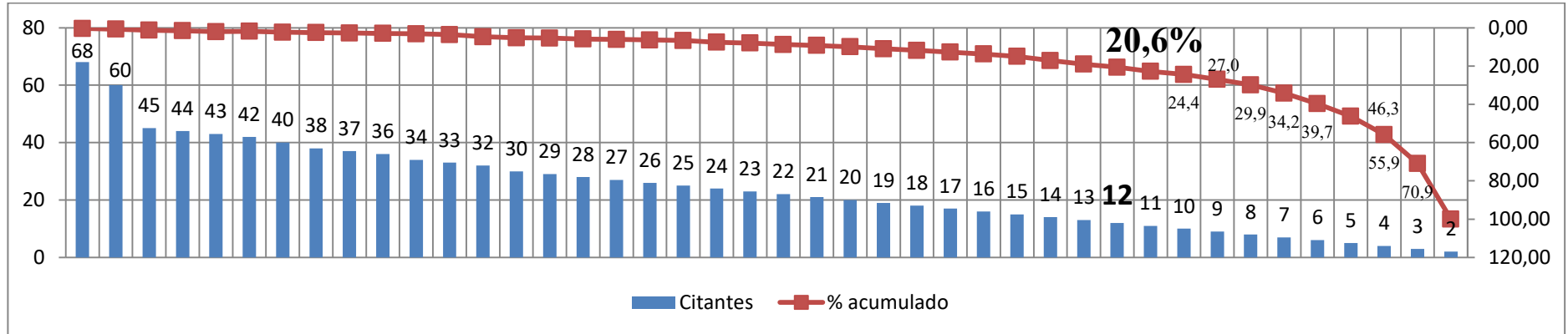
A Figura 15, na página 102, apresenta a distribuição dos valores de documentos citantes como categoria (em azul) e o percentual acumulado (em vermelho) da soma das citações oriundas de documentos citantes, considerando todos os autores das referências do *corpus* principal. O gráfico destaca que os autores citados em pelo menos 12 artigos, que correspondem a 20,6% dos valores de documentos citantes, excluindo os dados de dispersão absoluta, fazem parte da análise e esse total de autores é 180. A inclusão de autores que fazem parte dos dados intermediários de dispersão é possível, principalmente a metade próxima aos dados de análise. A inclusão de autores citados em 10 e 11 documentos, por exemplo, incluiria mais 27 e 33 autores, respectivamente. Cabe ao pesquisador definir essa necessidade e justificar a inclusão.

A Figura 16, na próxima página, segue o esquema da figura anterior e apresenta a distribuição dos valores de documentos citantes como categoria (em azul) e o percentual acumulado (em vermelho) da soma dos valores das citações oriundas de documentos citantes, considerando as primeiras posições das referências do *corpus* principal. A figura destaca que os autores citados em pelo menos 11 documentos, que representam 20,1% dos valores somados de citações oriundas de documentos citantes, excluindo os dados de dispersão absoluta, fazem parte da análise e somam 72 autores ao todo. Como já salientado, a inclusão de autores que fazem parte dos dados intermediários de dispersão é possível. A inclusão de autores citados em 10 documentos incluiria mais 15 indivíduos na análise.

Quanto aos documentos, a abordagem para estabelecer o ponto de corte foi similar, já que as citações totais para esse elemento têm a mesma função que os valores de documentos citantes para os autores. O teste inicial, incluindo os documentos citados uma única vez, evidenciou ainda mais a força da dispersão. Para o *ranking* com todos os documentos, o ponto de corte ficou estabelecido a partir de duas citações e simplesmente dividiu os dados em duas metades, incluindo na análise 1.715 documentos. A retirada dos documentos com uma citação (13.186) racionalizou os dados e reforçou a abordagem que desconsidera citações únicas para calcular o ponto de corte.

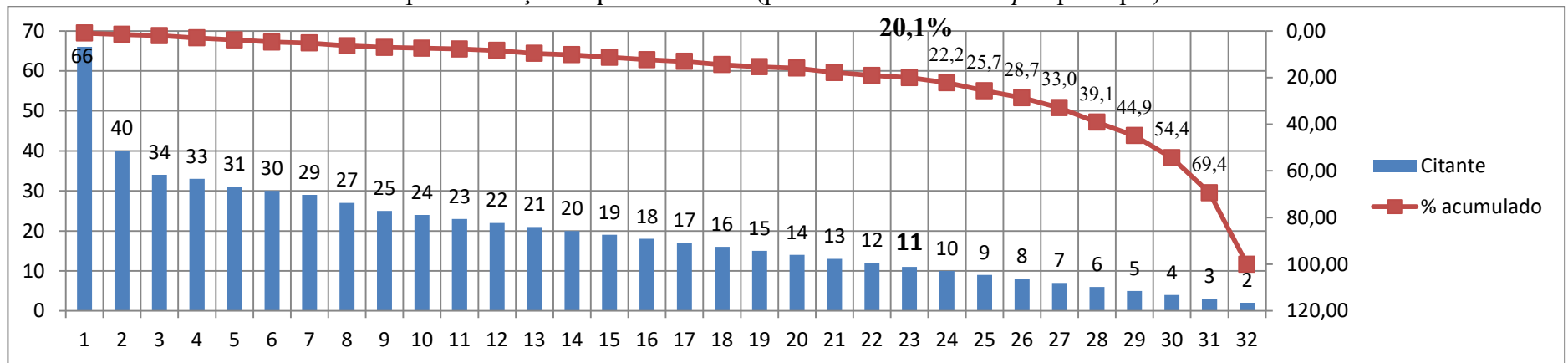
Assim, considerando os documentos a partir de duas citações, ficou estabelecido como ponto de corte seis citações, o que incluiu 109 documentos na análise, que equivalem a 19,4% das citações desconsiderando os dados de dispersão absoluta. A inclusão de documentos com cinco citações acarretaria no uso de mais 45 documentos e deixaria o percentual da soma das citações em 24,11%. Considerou-se, portanto, os dados intermediários de dispersão entre documentos com duas e cinco citações, sendo a metade mais próxima dos dados da análise a partir de quatro citações.

Figura 15. Distribuição de valores de documentos citantes (dois ou mais) como categoria para definição de ponte de corte (todos os autores – *corpus* principal).



Fonte: dados da pesquisa.

Figura 16. Distribuição de valores de documentos citantes (dois ou mais) como categoria para definição de ponte de corte (primeiros autores – *corpus* principal).

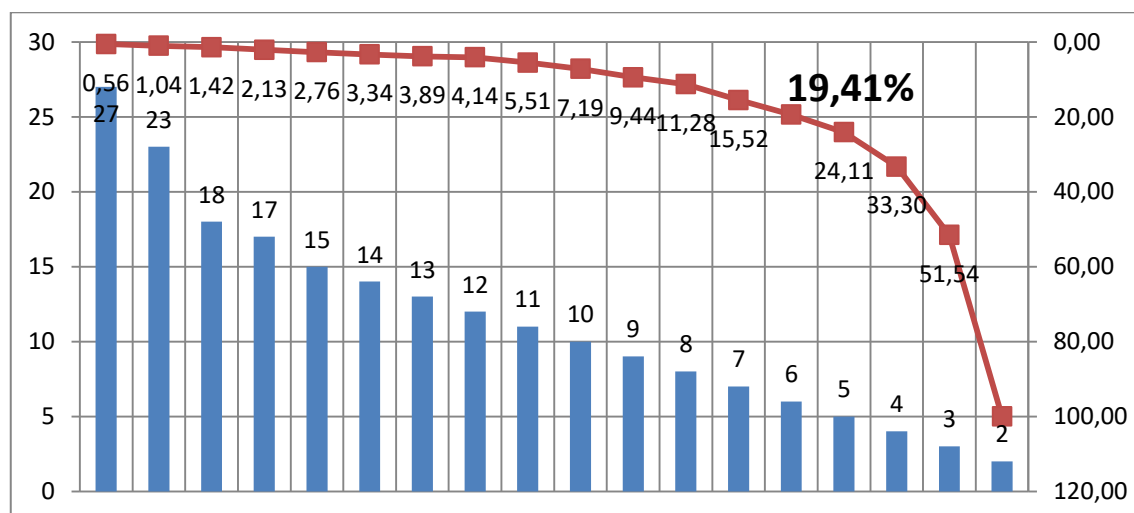


Fonte: dados da pesquisa.

A Figura 17 apresenta a distribuição dos valores de citações dos documentos, nos moldes dos dados dos autores, sendo o número de citações como categoria em azul e o percentual acumulado das somas das citações em vermelho, com uma curva invertida para deixar a relação mais destacada.

A proposta para definir o ponto de corte tem como objetivo principal descrever a relação entre os dados analisados e os dados considerados triviais, além de utilizar, no caso dos autores, um indicador que favoreça a melhor distribuição diante dos artigos que formam o *corpus* de análise.

Figura 17. Distribuição de valores de citações dos documentos (duas ou mais) como categoria para definição do ponto de corte em DCA (*corpus* principal).



Fonte: dados da pesquisa.

As matrizes de cocitação foram formadas como seguem e as listas completas dos autores e documentos estão disponíveis no Apêndice 7:

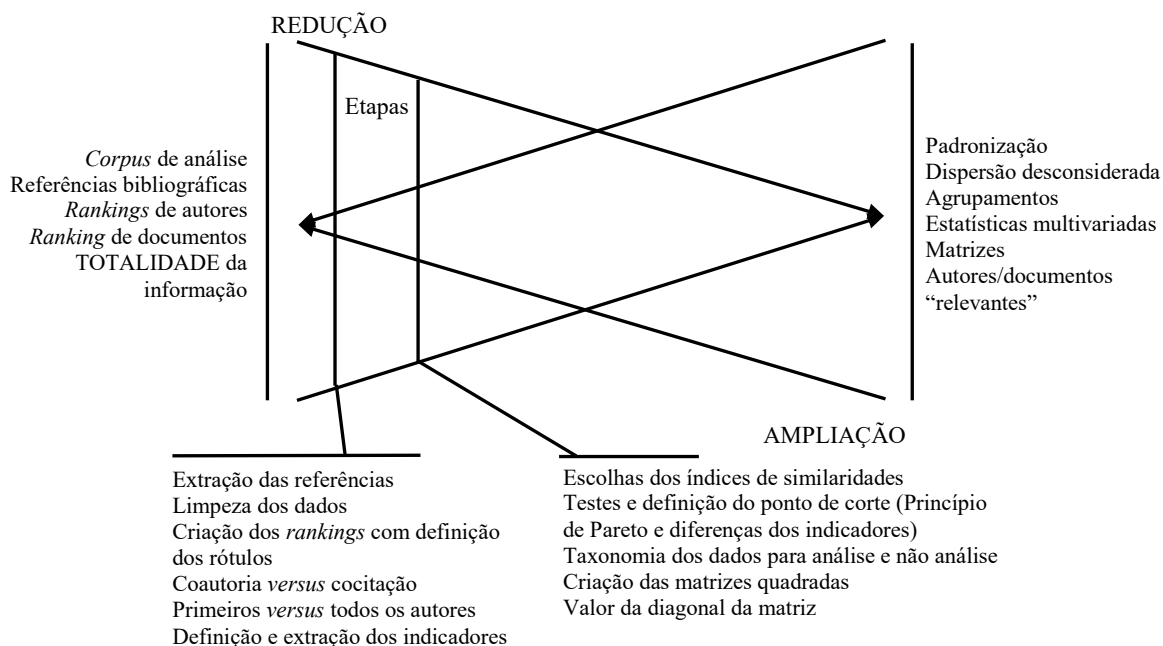
- i) Todos os autores das referências: 180 posições, com autores citados em pelo menos 12 artigos do *corpus*, com a elaboração de duas matrizes distintas a partir de dados de coautoria, mas apenas a matriz com dados sem os valores de coautoria foi utilizada na pesquisa. O Apêndice 8 apresenta um recorte da matriz com as 30 primeiras posições;

- ii) Primeiros autores das referências: 72 posições, com autores citados em pelo menos 11 artigos. O Apêndice 9 apresenta um recorte da matriz com os 30 autores citados em mais artigos citantes; e
- iii) Documentos: 109 posições, com documentos citados em pelo menos seis artigos. O Apêndice 10 apresenta um recorte da matriz com as 30 primeiras posições desses dados.

Há nas particularidades das unidades de contagem para estudos de cocitação a produção de informações por operações de seleção, extração e redução (LATOURE, 2006) que se manifesta: i) no trabalho de coleta e limpeza das referências para a criação de um *ranking* que proporciona a criação das matrizes de cocitação, ii) na elaboração de uma forma para definir um ponto de corte para a criação das matrizes, iii) na comparação dos dados de três indicadores relacionados aos autores (citação total, documentos citantes e cocitação do autor com ele mesmo, iv) na criação de dois *rankings* de autores que consideram as autorias das referências (todas *versus* primeiras), e v) na proposta de uma taxonomia para classificar a relação entre os dados de análise e dados de dispersão.

Todas essas etapas provocam a diminuição no volume dos dados, mas também possibilitam a criação de sentido, proporcionando a ligação entre a redução e a ampliação de informações, como sugerem os triângulos isósceles de Latour (2006). Esses triângulos isósceles invertidos são adaptados na Figura 18 e ilustram o início do caminho para a redução da informação em estudos de cocitação.

Figura 18. Triângulos isósceles da redução da informação em estudos de cocitação (duas primeiras fases).



Fonte: adaptado e ampliado de Latour (2006).

Os estudos de cocitação se mostram complexos e cheios de opções, desde a extração dos dados até a utilização dos indicadores advindos do *ranking* geral, principalmente para ACA. Objetivar e descrever, mesmo que exaustivamente, algumas escolhas metodológicas enriquecem a literatura sobre o tema e isso precisa ser planejado mesmo antes da extração das informações, já que a forma de extração é crucial para alcançar certo nível de detalhamento e domínio dos dados.

Nos dois próximos tópicos serão aprofundados elementos da primeira etapa da redução (Figura 18), na tentativa de desvendar o que se deixa de medir para contextualizar o que é analisado.

4.2 Sobre os dados não analisados na ACA e na DCA: mensuração, comparação e discussão sobre a perspectiva informacional

Os dados não analisados em um estudo de cocitação são elementos chave na problematização dessa pesquisa. Essas informações contextualizam a importância e a extensão dos dados analisados e fundamentam a particularidade de cada artigo citante que faz parte do *corpus*. Ou seja, as operações de seleção, extração e redução da produção de informações, que fala Latour (2006), permitem resolver a contradição entre a totalidade de dados e as informações analisadas, mas no sentido de também atribuir um nível de importância e significado ao que foi considerado dispersão, sejam os dados intermediários ou os dados absolutos.

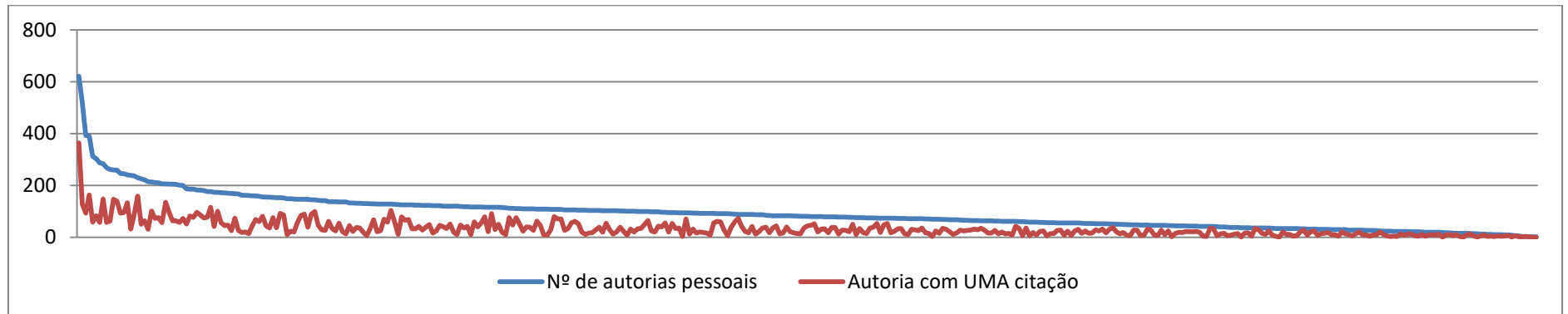
Essa discussão se inicia a partir de dois indicadores presentes na caracterização das referências: autores e documentos citados uma única vez no conjunto de dados. Essas informações isoladas não representam muito, mas ligadas aos citantes pela relação com os totais de autorias e documentos citados por cada artigo podem colaborar no reconhecimento do peso de cada documento junto ao *corpus* analisado.

As figuras 19 e 20 apresentam as distribuições dos autores e dos documentos citados uma única vez no *corpus* principal da análise, respectivamente. As linhas em azul das figuras demonstram o número máximo de autores e documentos (valores absolutos e não distintos) e a linha em vermelho destaca os valores únicos por artigo.

Essas figuras resumem os dados gerais apresentados na Tabela 5, que mostram que o total de autores e de documentos citados em apenas um dos artigos do *corpus* é de 13.934 e de 13.124¹⁵, com médias de 33,10 e 31,17 por artigo citante (N=421), respectivamente. Em relação aos valores de autores e documentos distintos, esses dados são de ordem percentual de 71,13% para autores, do total de 19.590 e 88,07% para os documentos, do total de 14.901.

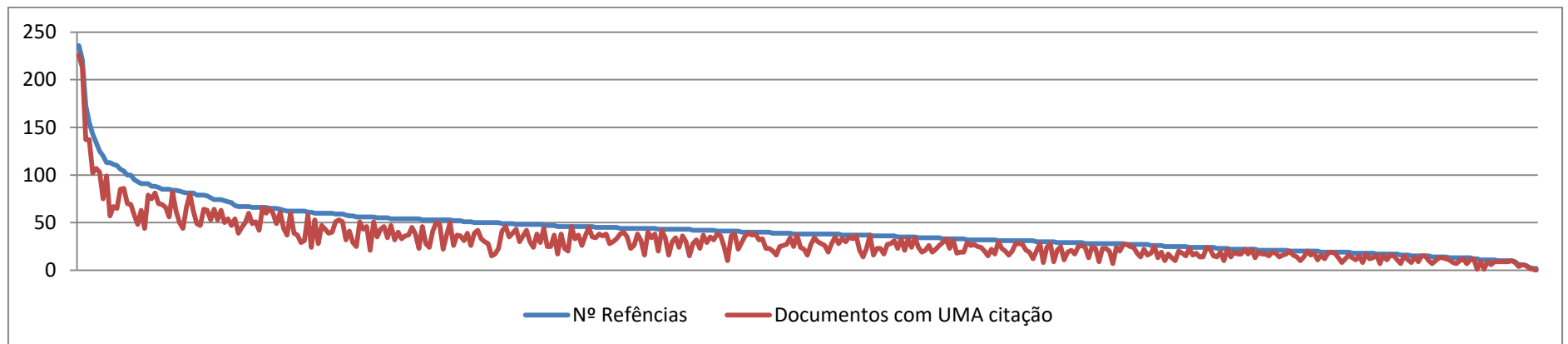
¹⁵ Há uma diferença na contagem dos documentos distintos citado em apenas um artigo citante e isso pode ter ocorrido por dois motivos: limpeza particionada dos documentos com autoria pessoal e limpeza dos documentos com outro tipo de autoria. Essa diferença é igual a 62 documentos e corresponde a 0,47% do outro total, 13186, que equivale a um erro com média de 0,15 por artigo citante e foi desconsiderado.

Figura 19. Distribuição da diferença entre o número total de autorias pessoais e autores citados apenas uma vez no *corpus*.



Fonte: dados da pesquisa.

Figura 20. Distribuição da diferença entre o número total de referências e documentos citados uma única vez no *corpus*.



Fonte: dados de pesquisa.

Esses dados iniciais já indicam as diferenças e o alcance entre a ACA e a DCA, uma vez que a dispersão absoluta é mais acentuada na contagem dos documentos do que na contagem dos autores. Isso se intensifica ainda mais quando é quantificada a diferença por artigo citante entre todas as citações de autores e de documentos com os citados apenas uma única vez diante dos dados do *corpus*.

No caso dos autores, 106 artigos citantes possuem 50% ou mais de citados que não se repetem em nenhum outro artigo citante, sendo que em três artigos esse percentual chega a 100%, o que os tornam únicos em termos de rastros referenciais relacionados aos autores. Vale a ressalva que esses valores são contados levando-se em consideração todos os autores de todas as referências. Considerando apenas os primeiros autores das referências, 6.913 autores são citados uma única vez no *corpus*, o que equivale a 74,03% dos 9.338 autores distintos, em uma proporção próxima aos dados incluindo autores de todas as posições das referências.

No caso dos documentos, 379 artigos citantes, dos 421 que formam o *corpus*, citam 50% ou mais trabalhos de maneira exclusiva, sendo que em 83 artigos o percentual é de 90% ou mais e em 18 artigos todos os trabalhos citados são únicos no contexto do *corpus*. Alguns desses artigos foram publicados em periódicos como o *Electronic Library* (quatro), *Library Management* (dois), *Knowledge Organization* (um) e até um no *Journal of the Association for Information Science and Technology*.

Essa enorme dispersão torna a produção científica da área de Organização do Conhecimento e da Recuperação da Informação muito rica pela diversidade de fontes e autores citados, mas é um *mar* de entropia para estudos de cocitação, pois se mostra evidente que a influência da elite de dados analisados não é grande em relação ao *corpus* e isso se perde ainda mais para estudos de cocitação de documentos. Trata-se de gerar dados que caracterizam mais o citante do que a área, pelo menos no contexto do *corpus* estudado, pois ainda que se proponha a analisar uma disciplina, como no caso desse estudo, análises de cocitação são sempre condicionadas às limitações das fontes de dados, como em qualquer EC. De qualquer forma, esses dados se evidenciam como uma variável interessante na compreensão da dinâmica da produção do conhecimento nas áreas citadas, como indica o enfoque normativo dos EC.

A apresentação desses dados marca uma análise sobre o que ficou sendo chamado no contexto dessa pesquisa de dispersão absoluta, ou seja, dados sem potencial

para uso em estudos de cocitação, dados únicos dos citantes, sem alcance mínimo no *corpus* de análise.

A partir de agora os dados são apresentados e discutidos tendo em vista o ponto de corte estabelecido no tópico anterior dessa seção. Além de apresentar os dados para responder ao segundo objetivo específico, será considerada também parte do quarto objetivo específico, que trata da comparação entre os dados da ACA e da DCA.

Os primeiros dados são de comparação genérica. No contexto desse estudo, a ACA que considera todos os autores das referências resulta na análise da relação de 180 autores, que correspondem a 0,92% do total de autores (19.590). Por esse motivo, é aconselhável ponderar os autores pela sua presença no *corpus*, seja pela contagem da citação total ou pela contagem do número de documentos que os citam. Considerando os primeiros autores das referências essa proporção de 0,77%, ou seja, a relação dos 72 sujeitos que fazem parte da análise diante do total de 9.338 autores distintos. Para os documentos a proporção é de 0,73% (109 documentos com relações analisadas diante de 14.901 documentos distintos).

Mais significativo do que essa relação genérica é a identificação dos rastros dos elementos (autores/documentos) diante dos artigos citantes do *corpus*. A Tabela 15 apresenta os valores de citação e de cocitação dos autores e dos documentos que foram utilizados na ACA e na DCA para evidenciar que parte desses documentos que não possuem relevância nas análises.

Tabela 15. Valores de citação e de cocitação dos autores e dos documentos utilizados na ACA e na DCA diante dos artigos citantes (N=421).

	Citação			Cocitação		
	Citam	Não citam	% *	Há cocitação	Não há cocitação	%*
Todos os autores (n=180)	369	52	87,65	334	87	79,33
Primeiros autores (n=072)	331	90	78,62	260	161	61,76
Documentos (n=109)	285	136	67,70	201	220	47,74

* percentual de alcance no *corpus* para um total de 421 citantes.

Fonte: dados da pesquisa.

As informações da tabela 15 evidenciam que as relações de cocitação analisadas não chegam a ocorrer em 80% dos artigos citantes, no melhor cenário, e essa proporção

diminui conforme as escolhas metodológicas de um estudo dessa natureza.

A ordem parece clara: o uso de todos os autores das referências torna um estudo de cocitação mais abrangente, o uso dos primeiros autores tem um alcance intermediário e o uso dos documentos possui o menor alcance (nos dados dessa pesquisa é exatamente de 47,74% dos documentos citantes). Oliveira e Grácio (2011) apresentam uma situação onde 15 documentos citantes não citam nenhum dos autores que formam a rede de cocitação apresentada no relato.

Essas diferenças não são meramente quantitativas em termos de dialética do conhecimento (KONDER, 1998), ou seja, há um impacto qualitativo nesses alcances, mas não se trata simplesmente de determinar qual a melhor forma de realizar um estudo - e o ponto de vista dessa contribuição é que isso deve ser abandonado em termos dos estudos de cocitação - é preciso um olhar focado nas especificidades e limitações das formas de operacionalizar um estudo. Os dados para a DCA, por exemplo, apresentam quatro artigos de autoria não pessoal (terceira parte do Apêndice 7), além das relações de cocitação de 11 documentos do autor "Hjørland, B" e cinco do autor "Olson, HA", informações que se perdem em uma ACA. Em termos do tamanho dos *rankings* de autores, por exemplo, se todo o trabalho de extração e limpeza feito nessa pesquisa se concentrasse apenas nas primeiras posições das referências era possível dobrar o alcance do estudo para a ACA considerando o primeiro autor e também para a DCA, que teriam matrizes mais robustas levando em consideração o ponto de corte estabelecido.

Todavia, parece necessário caracterizar mais algumas informações sobre os documentos citantes. Não cocitam nenhum dos autores e documentos das matrizes 85 artigos (a Tabela 16 apresenta a distribuição por periódico e anos de publicação), ou seja, 20,19% do total. Não cocitam pelo menos dois elementos das matrizes 66 artigos (15,68%) e não cocitam um dos elementos das matrizes de cocitação 81 artigos, ou seja, 19,24% do total de artigos citantes. Assim, 232 artigos (55,11%) colaboram com a dispersão do estudo em termos de pares de cocitação, sendo que 85 artigos de forma integral, alguns artigos citam autores e documentos isoladamente, ou seja, não há cocitação entre os autores/documentos analisados.

Os artigos citantes que cocitam autores ou documentos que formam as matrizes de cocitação são 336, ou seja, 79,81% do total. Os artigos que citam apenas um dos elementos somam 66 (15,68%), dois elementos são 81 artigos (19,24%) e cocitam

autores e documentos dos três arranjos são 189 artigos, ou seja, 44,89% dos 421 citantes cocitam elementos das três matrizes.

Tabela 16. Distribuição dos periódicos, por ano, dos artigos citantes que não cocitam nenhum dos autores e documentos das análises.

Periódicos	2015	2016	Total	%
Knowledge Organization	03	05	08	9,41
Journal Of The American Medical Informatics Association	02	04	06	7,06
Electronic Library	03	02	05	5,88
Qualitative & Quantitative Methods In Libraries	04	01	05	5,88
<i>Revista Ibero-Americana De Ciencia Da Informacao</i>	02	02	04	4,71
DESIDOC Journal Of Library & Information Technology	02	01	03	3,53
Journal Of Information Science	03	-	03	3,53
Library Management	03	-	03	3,53
<i>Perspectivas Em Ciencia Da Informacao</i>	02	01	03	3,53
Scientific And Technical Information Processing	01	02	03	3,53
<u>Scire-Representacion Y Organizacion Del Conocimiento</u>	02	01	03	3,53
Journal Of The Association For Information Science And Technology	01	01	02	2,35
Journal Of The Medical Library Association	01	01	02	2,35
<u>Profesional De La Informacion</u>	02	-	02	2,35
Program-Electronic Library And Information Systems	01	01	02	2,35
<u>Revista Espanola De Documentacion Cientifica</u>	02	-	02	2,35
Um artigo ¹⁶	18	11	29	34,12
Total	52	33	85	100

Fonte: dados de pesquisa.

Quanto aos dados da Tabela 16, o periódico com mais dados não inclusos nas matrizes de cocitação é também o periódico com mais artigos no *corpus*. Mas os dados intermediários de dispersão, principalmente os próximos aos dados da análise, devem conter informações desses artigos, uma vez que se trata de uma revista temática. Outros destaques ficam por conta de duas revistas de informação em saúde (destaque em **negrito**), duas revistas brasileiras (destaque em *itálico*) e três revistas espanholas (destaque em sublinhado). A heterogeneidade da WoS, com a inclusão de revistas regionais e técnicas, além da indexação inadequada, acarreta em uma dispersão mais acentuada que qualquer análise bibliométrica irá identificar, sem contar que os dados

¹⁶ African Journal Of Library Archives And Information Science; Anales De Documentacion; Biblios-Revista De Bibliotecologia Y Ciencias De La Informacion; College & Research Libraries; Collnet Journal Of Scientometrics And Information Management; Cuadernos De Documentacion Multimedia; Data Base For Advances In Information Systems; Education For Information; Health Information And Libraries Journal; Ibersid-Revista De Sistemas De Informacion Y Documentacion; Informacao & Sociedade-Estudos; Information Development; Information Processing & Management; Information Research-An International Electronic Journal; Information Systems Research; Information Technology & Management; International Journal Of Geographical Information Science; Jlis.It; Journal Of Documentation; Journal Of Information & Knowledge Management; Journal Of Knowledge Management; Journal Of Library Administration; Journal Of Web Librarianship; Library Hi Tech; Library Trends; Records Management Journal; Telematics And Informatics; Turkish Librarianship; World Patent Information.

são considerados a partir da cocitação e não da citação. Essas variáveis marcam o enfoque normativo dos EC (SILVEIRA; CAREGNATO, 2017b): demarcações geográficas e dinâmicas da produção do conhecimento.

O artigo “*The Web and the Pyramid: Hope Olson’s Vision of Connectedness in a World of Hierarchies*”, da autora Kwaśnik, um dos oito artigos citantes da revista *Knowledge Organization* que não cocitam autores ou documento da análise, cita um importante autor da área da Organização do Conhecimento, tratando sobre o tema relacionando com a Web. Contudo, ele segue uma narrativa referencial que não proporciona dados de cocitação relevantes para o contexto do *corpus* analisado nessa pesquisa e isso evidencia elementos do enfoque construtivista dos EC presentes em análises de cocitação, indicando variáveis relacionadas à narrativa e às marcas teóricas (SILVEIRA; CAREGNATO, 2017b), pois a citação tem relação direta com narrativa dos documentos citantes.

Outros aspectos que merecem importância em termos de dados não analisados em um estudo de cocitação são os dados intermediários de dispersão, principalmente os dados próximos às informações analisadas, que são retirados por uma abordagem quantitativa bem objetiva e, infelizmente, necessária aos estudos, mas com impacto certamente qualitativo.

Um pequeno experimento com os autores com mais citações totais e com mais documentos citantes, considerando o *ranking* com todas as posições das referências, evidenciaram alguns dados interessantes sobre os valores de pares de cocitação e oferece uma crítica à forma estabelecida por essa pesquisa para a definição do ponto de corte. Os autores “Hjørland, B” e “Croft, WB” possuem 191 e 136 citações totais, são citados em 68 e 60 artigos do *corpus*, respectivamente, e são cocitados apenas sete vezes.

A Tabela 17 apresenta a comparação das relações de cocitação desses autores com autores que fazem parte da matriz de análise (179 autores) e com os autores que fazem parte da dispersão intermediária dos dados (271 autores), seja a metade próxima à análise, isto é, sujeitos citados entre sete e 11 artigos. Os destaques em vermelho e azul demonstram que o autor “Croft, WB” é mais influente no *corpus*, não tendo relação de cocitação com apenas 6,70% dos autores da matriz de cocitação, enquanto “Hjørland, B” não apresenta relação com 20,67% dos outros 179 autores.

Tabela 17. Comparação dos valores de cocitação dos autores *Hjørland, B* e *Croft, WB* com dados da matriz e com dados de dispersão intermediária (todos os autores).

Valor do par de cocitação	Ocorrência a partir de dados da Matriz*		Ocorrência a partir dos dados intermediários de dispersão**	
	Hjørland, B	Croft, WB	Hjørland, B	Croft, WB
Zero	37	12	99	47
Um	39	15	81	68
Dois	31	17	44	45
Três	18	22	22	39
Quatro	12	14	11	27
Cinco	13	13	07	18
Seis	04	14	03	13
Sete	02	08	02	07
Oito	04	08	01	04
Nove	05	10	01	01
10	04	13	00	02
11	03	05	00	00
12	01	03	NA***	NA
13 ou mais	06	24	NA	NA

* relação com 179 autores da matriz quadrada

** relação com 271 autores da faixa de dispersão intermediária da análise, citados entre sete e 11 artigos

*** não se aplica

Fonte: dados da pesquisa.

Mas o que chama mais atenção são os dados com destaque em verde na Tabela 17. O autor “Croft, WB” possui relação de cocitação igual ou superior a aquela com “Hjørland, B” com 14 sujeitos que não fazem parte dos dados de análise de cocitação, enquanto “Hjørland, B” possui quatro casos com essa situação. Essa condição corrobora ainda mais as incertezas inerentes aos estudos e às dificuldades de estabelecer os autores/documentos que farão parte de uma análise.

Assim, parece necessário apresentar uma parte dos dados de dispersão intermediária que estão mais próximo aos dados analisados. No caso dos dados que considera todos os autores das referências, com ponto de corte em 12 documentos citantes, a inclusão de autores com 11 ou 10 documentos citantes é possível e é a tendência nos estudos dessa natureza buscar um número redondo. Uma iniciativa nesse sentido incluiria 60 autores na análise, sendo 33 citados em 11 artigos e 27 citados em 10, o que deixaria a análise com 240 autores. Pesquisadores da Organização do conhecimento e da Recuperação da Informação como “Fox, MJ”, “García Gutiérrez,

A”, “López-Huertas, MJ”, “Nicholas, D”, “Nielsen, J”, “Rowley, JE”, “van Eck, NJ”, “Budd, JM”, “Borgman, CL”, e “Star, SL”, para citar alguns, não ficariam de fora da análise.

No caso dos dados considerando os primeiros autores, que ficou com ponto de corte estabelecido em 11 documentos citantes, a inclusão de autores com 10 citantes (15 sujeitos) ou nove (26 sujeitos) também é possível e deixaria a matriz final de cocitação com 113 autores, tamanho similar a matriz da DCA e ainda bem menor do que a matriz criada com todos os autores. Pesquisadores como “Fox, MJ”, “Chen, C”, “Borgman, CL”, “Smucker, MD”, “Glänzel, W”, “Soergel, D”, “Börner, K”, “Amati, G”, “Savolainen, R”, “Bowker, GC” e “Miller, GA” fariam parte da matriz e enriqueceriam a análise. Essa é uma decisão de percepção do pesquisador diante da natureza e do domínio adequado dos dados, dos objetivos propostos, do tempo para operacionalizar e analisar as informações recuperadas.

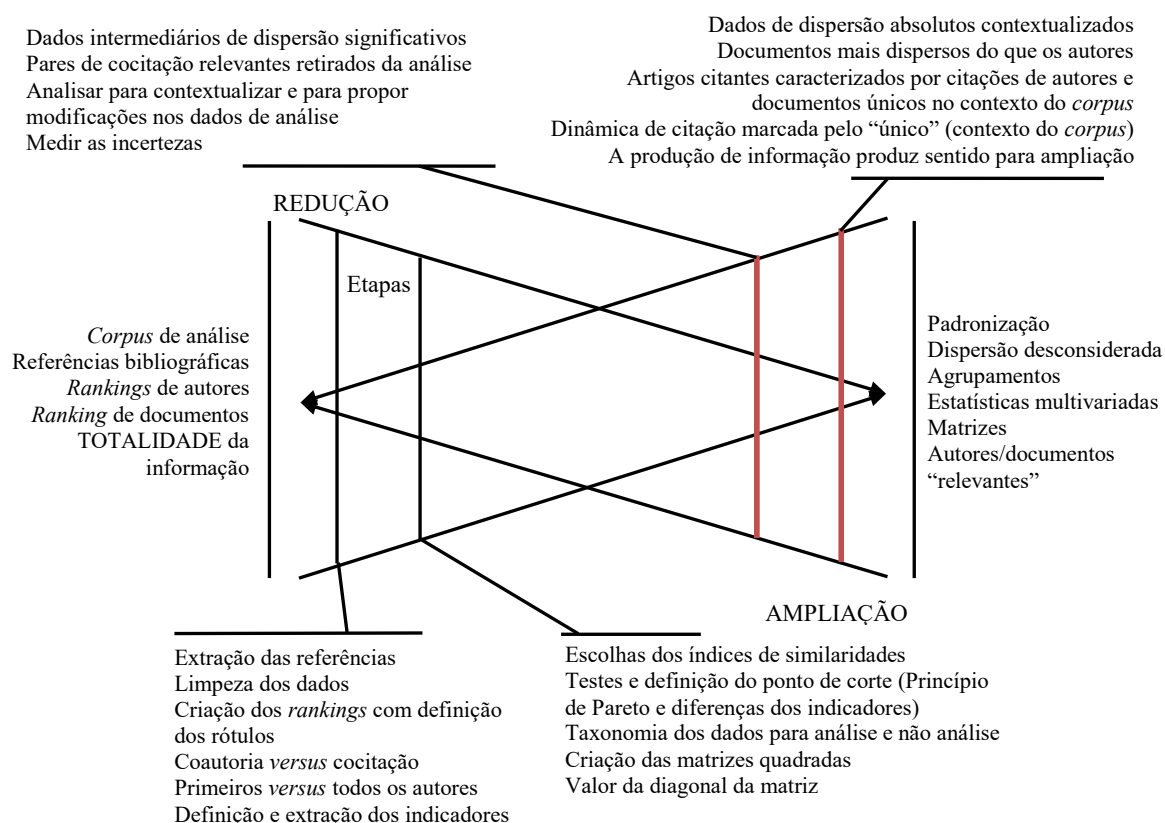
No caso dos documentos, que ficou com ponto de corte estabelecido em seis citações, a inclusão, levando em consideração cinco citações, acarretaria em mais 45 documentos na análise. McCain (1990) fez testes com cortes baixos, mas uma matriz com 154 posições e um corte de cinco citações pode ser bem irregular e prejudicar uma análise adequada. De qualquer forma documentos como “A statistical interpretation of term specificity and its application in retrieval”, de “Sparck Jones, K”, “Changes in relevance criteria and problem stages in task performance”, de “Vakkari, P”, “Co-citation analysis, bibliographic coupling, and direct citation”, de “Boyack, KW”, “Combining bibliometrics, information retrieval, and relevance theory, part 2”, de “White, HD”, “Discourse Analysis as a Research Method in Library and Information Science”, de “Frohmann, B”, “Domain analysis”, de “Hjørland, B”, “Human information behavior”, de “Wilson, TD”, “Knowledge Organization Systems (KOS)”, de “Zeng, ML” e “Standardization, Objectivity, and User Focus”, de “Olson, HA” entrariam na análise e enriqueceria a discussão, como também aumentaria a complexidade de utilizar a técnica e entender os agrupamentos.

Contextualizar os dados não analisados deveria ser parte de estudos de cocitação, seja para evidenciar a extensão do que será analisado e deixar os relatos mais descritivos e informativos, como a indicar parte dos dados intermediários de dispersão, principalmente os próximos aos dados analisados, ou indicar o número de documentos citantes que efetivamente cocitam autores ou documentos da análise, por exemplo; seja

para realizar ajustes. No caso específico dos dados dessa pesquisa, recomenda-se incluir autores citados em 10 documentos considerando os primeiros autores das referências.

Por fim, a Figura 21 amplia os triângulos isósceles apresentados na Figura 18, incluindo elementos que apontam para o vértice da ampliação.

Figura 21. Triângulos isósceles da redução da informação em estudos de cocitação (dados não analisados)



Fonte: Adaptado e ampliado de Latour (2006).

A contextualização dos dados não analisados, que ocorre pela produção de informação, oferece sentido à ampliação que fala Latour (1996) ao propor seus triângulos para a compreensão do fenômeno informacional.

No próximo tópico são apresentados dados que comparam o uso de todos os autores das referências com o uso dos primeiros.

4.3 A utilização dos primeiros e de todos os autores das referências em ACA: uma análise empírico-comparativa sobre as vantagens e as desvantagens _____

A relação entre o uso de todos os autores ou o uso dos primeiros autores das referências é tema recorrente na literatura sobre cocitação, como mostram alguns estudos (PERSSON, 2001; ROUSSEAU; ZUCCALA, 2004; ZHAO, 2006; SCHNEIDER; LARSEN; INGWERSEN, 2009).

Essa parte da tese objetiva comparar dados de cocitação pura de primeiro autor, com dados de cocitação pura de todos os autores, que considera todas as autorias das referências e descarta os dados de coautoria, conforme a classificação de Rousseau e Zuccala (2004). Vale salientar que a pesquisa gerou dados de cocitação de autor geral ou, como Zhao (2006) denomina, cocitação inclusiva de todos os autores, mas essa matriz não foi analisada.

Todavia, um dos pilares da problematização dessa pesquisa é a definição do ponto de corte das matrizes de cocitação levando em conta todos os dados do *corpus*, o que torna a comparação mais complicada e um pouco desproporcional, pois o uso de todos os autores implica em desenvolver uma ACA a partir de 19.590 autores distintos, enquanto o uso dos primeiros autores implica em uma ACA baseada em 9.338 autores distintos. Assim, proporcionalmente, com base na proposta do ponto de corte, as duas matrizes são parecidas diante do que se pretende analisar e em uma área com muita colaboração, o uso da cocitação exclusiva ou inclusiva de todos os autores é mais recomendável do que o uso do primeiro, se o objetivo é focar nas relações dos autores.

Entende-se, portanto, que são análises distintas de cocitação que apresentam suas vantagens e desvantagens, embora os dados evidenciem que os autores colaboram muito. A matriz de cocitação com todos os autores mostra que apenas 16 autores, de 180, não possuem relação de coautoria identificada no *corpus*: são eles: “Bates, MJ”, “Lancaster, FW”, “Beghtol, C”, “Dahlberg, I”, “Kuhlthau, CC”, “Mai, JE”, “Bar-Ilan, J”, “Buckland, MK”, “Dervin, B”, “Davis, FD”, “Frohmann, B”, “Porter, MF”, “Turney, PD”, “Wilson, P”, “Konstan, JA” e “Vickery, BC”. Os demais autores possuem na matriz 364 relações de coautoria, sendo que 189 são válidas a partir dos critérios apresentados nos procedimentos metodológicos, e isso traz ruído para a análise.

Os autores “Berners-Lee, T”, “Hendler, JA” e “Lassila, O”, por exemplo, têm valores altos de citações totais e documentos citantes, principalmente devido ao documento “The semantic web” (2001). O mesmo ocorre com “Baeza-Yates, R” e “Ribeiro-Neto, B”, autores do documento “Modern information retrieval” (1999), que sustenta parte significativa de suas participações nos dados. Esse é um problema que a cocitação pura de primeiro autor não tem, ainda que apenas seis autores, dos 72 da matriz de cocitação, não tenham citações oriundas de documentos publicados em coautoria, são eles: “Wilson, P”, “Svenonius, E”, “Frohmann, B”, “Dahlberg, I”, “Beghtol, C” e “Mai, JE.

Um caminho para comparar os dados é deixar as matrizes com o mesmo tamanho e contar as diferenças. Uma matriz de cocitação pura de primeiro autor, com os dados dessa pesquisa, por exemplo, precisaria ficar com 181 posições, com a inclusão de autores citados em pelo menos sete documentos, para uma comparação próxima em termos de tamanho com a matriz de todos os autores. Uma matriz de cocitação exclusiva de todos os autores, para ficar próxima à matriz de primeiro autor, ficaria com 74 posições, com a inclusão de autores citados em pelo menos 18 artigos. Somente essa diferença de corte com base nos valores de documentos citantes, sete e 18, para 12 e 11 documentos citantes de todos os autores e dos primeiros autores, respectivamente, já evidencia que são análises distintas.

Uma comparação possível é supor que todos os autores de uma matriz de cocitação pura de primeiro autor, respeitando o valor do ponto de corte de uma matriz de cocitação exclusiva de todos os autores, estejam na análise que inclui todos os autores, ou seja, no caso dos dados dessa pesquisa autores com pelo menos 12 documentos citantes que fazem parte da análise de cocitação utilizando os primeiros autores devem fazer parte da análise que inclui todas as autorias das referências; e essa condição pode ser percebida ainda na elaboração dos *rankings*. Ou seja, 66 autores da análise de cocitação pura de primeiro autor, que possuem 12 ou mais documentos citantes, fazem também parte da análise com todos os autores.

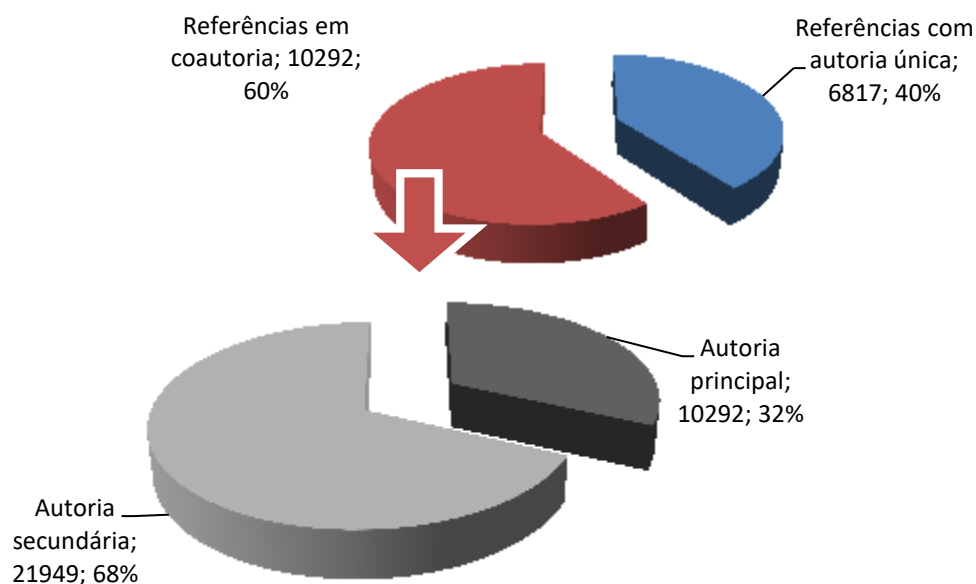
Os demais autores da análise de cocitação pura de primeiro autor, que possuem 11 documentos citantes, número abaixo do ponto de corte para a análise com todos os autores, que são “García Gutiérrez, A”, “Gnoli, C”, “Zhai, CX”, “Aitchison, J”, “Jones, Rosie” e “Li, X”, entram na análise exclusiva de todos os autores, exceto por “García Gutiérrez, A” e “Aitchison, J” . Ou seja, para os dados dessa pesquisa, 97,22% dos

autores da análise de cocitação pura de primeiro autor estão na análise que inclui todas as posições de autoria das referências.

Assim, se as matrizes de cocitação utilizando os primeiros autores e todos os autores são construídas na proporcionalidade da extensão dessas abordagens, podemos afirmar que utilizar todos os autores nada mais é do que respeitar o tamanho do conjunto de dados que se apresenta e por isso é importante contextualizar a análise e objetivar o ponto de corte de inclusão dos autores a partir de todos os dados, mesmo os dados de dispersão absoluta.

De qualquer forma é importante apresentar alguns indicadores. A Figura 22 apresenta a proporção dos valores totais de referências dividido por autoria única e por coautoria. A parte inferior da figura divide as referências de coautoria por autoria principal e por autoria secundária. Esses dados partem das 39.058 autorias presentes no *ranking* geral de autores.

Figura 22. Proporção dos tipos de autorias do *ranking* de autores (N=39058).



Fonte: dados da pesquisa.

As proporções são parecidas com os dados apresentados por Carvalho e Caregnato (2016) e mostram que mais da metade das autorias (56,20%) são secundárias e isso exige exame e planejamento adequado para perceber a influência desses dados.

A Tabela 18 apresenta um recorte de 110 autores citados em 10 ou mais artigos considerando apenas as posições secundárias e os dados são robustos, principalmente pelos resultados do autor “Croft, WB”, que se mostrou influente na análise.

Tabela 18. Distribuição do *ranking* de autores a partir da posição secundária na referência (110 posições organizadas a partir de 10 documentos citantes).

AUTORES	Citação total	DC	AUTORES	Citação total	DC	AUTORES	Citação total	DC
Croft, WB	124	55	Yu, Y	16	15	Tang, J	15	11
Järvelin, K	59	33	Bawden, D	15	15	Chen, C	14	11
Dumais, ST	60	31	Nejdl, W	15	15	Mayr, P	14	11
Ribeiro-Neto, B	36	31	Jose, JM	36	14	Sugimoto, CR	14	11
Raghavan, P	30	30	Fox, EA	25	14	Motwani, R	13	11
Schutze, H	33	29	Soboroff, I	22	14	Zeng, ML	13	11
Spink, A	38	26	Zobel, J	22	14	Foo, S	12	11
Belkin, NJ	53	24	Giles, CL	20	14	Landauer, TK	12	11
Jordan, MI	31	24	Hu, X	18	14	Staab, S	11	11
Ng, AY	24	24	Albrechtsen, H	17	14	Wong, A	11	11
De Rijke, M	41	23	Kamps, J	17	14	Yang, Chung-Shu	11	11
Sanderson, M	27	21	Chan, LM	14	14	Tudhope, D	18	10
Walker, S	23	21	Jones, SA	14	14	Guimarães, JAC	17	10
Witten, IH	36	20	Lassila, O	14	14	McCallum, A	17	10
Buckley, C	24	20	Ruthven, I	26	13	Gonçalves, MA	16	10
Robertson, SE	24	20	Cool, C	19	13	Bruza, PD	13	10
Zhai, CX	37	19	Craswell, N	18	13	Chen, Z	13	10
Ounis, I	28	19	Granka, LA	17	13	Diaz, F	13	10
Ding, Y	28	18	Wolfram, D	17	13	He, B	13	10
Voorhees, EM	27	18	Romano, G	16	13	Li, R	13	10
Allan, J	29	17	Kekalainen, J	15	13	Card, SK	12	10
Hendler, JA	17	17	Song, D	14	13	Gabrilovich, E	12	10
Li, H	33	16	Winograd, T	14	13	Jones, Rosie	12	10
Nie, JY	29	16	Zhang, M	14	13	Kelly, D	12	10
Callan, JP	28	16	Konstan, JA	25	12	Steyvers, M	12	10
Macdonald, C	23	16	Glänzel, W	23	12	Tennis, JT	12	10
Moffat, A	21	16	Lalmas, M	21	12	Gatford, M	11	10
Metzler, D	20	16	Boyack, KW	19	12	Gilchrist, A	11	10
Teevan, J	20	16	Han, J	19	12	Nilan, MS	11	10
Lafferty, JD	25	15	Garcia-Molina, H	17	12	Pereira, FCN	11	10
White, RW	24	15	Chen, H	15	12	Brin, S	10	10
Baeza-Yates, R	22	15	Jones, GJF	15	12	Furnas, GW	10	10
Jansen, BJ	20	15	Zha, H	14	12	He, D	10	10
Marchionini, G	19	15	Clarke, CLA	13	12	Park, JR	10	10
Zhang, J	19	15	Saracevic, T	13	12	Strohman, T	10	10
MacFarlane, A	18	15	Vakkari, P	12	12	Van Rijsbergen, CJ	10	10
Li, X	17	15	Ingwersen, P	21	11			

Legenda: DC – documento citante.

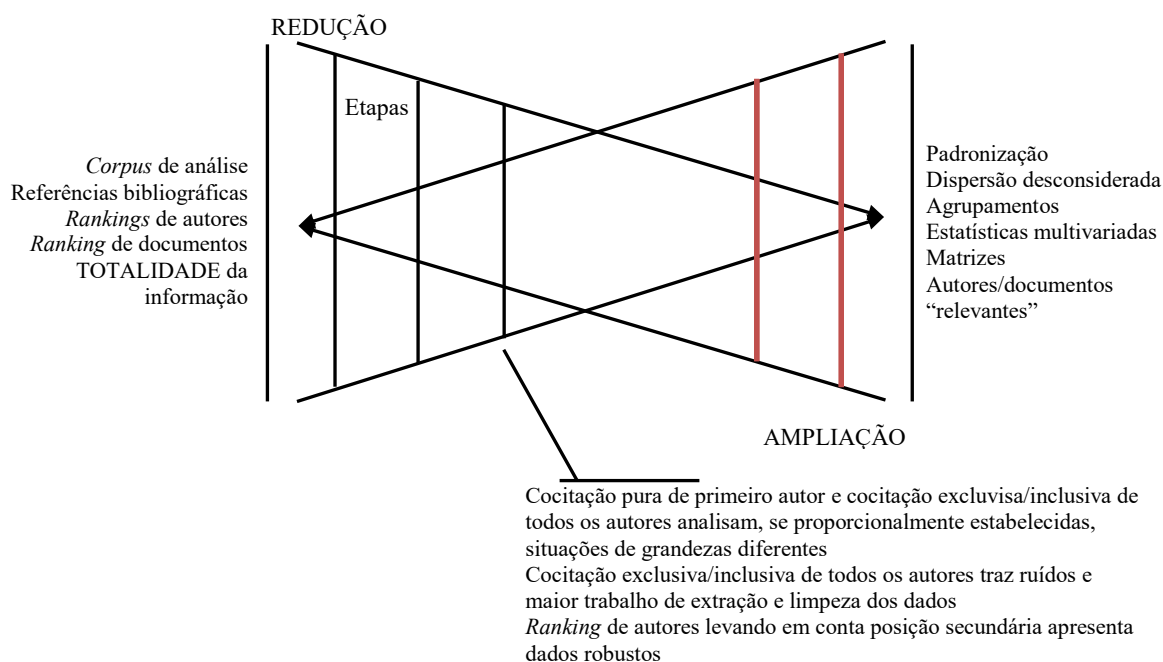
Fonte: dados de pesquisa.

A construção de um *ranking* considerando apenas as posições secundárias também foi um recurso utilizado por Carvalho e Caregnato (2016). Os dados se revelam

significativos e deixam mais claro a complexidade de lidar com referências, ainda que no enfoque normativo dos EC.

Por fim, mais uma etapa dos triângulos isósceles é realizada no sentido da redução, como apresentado na Figura 23.

Figura 23. Triângulos isósceles da redução da informação em estudos de cocitação (primeiro *versus* todos os autores)



Fonte: Adaptado e ampliado de Latour (2006).

O uso de todos os autores das referências é uma decisão significativa nos estudos de cocitação. Enquanto o uso do primeiro autor foca na relação bibliográfica e é adequado para demonstração de estruturais intelectuais (ROUSSEAU; ZUCCALA, 2004), o uso de todos os autores é mais abrangente, mas as relações de cocautoria produzem ruídos na análise, principalmente com autores com documentos muito relevantes para um domínio. Todavia, são abordagens para arranjos diferentes dos dados, sendo que o uso do primeiro autor torna a análise mais simples, mas um pouco menos robusta.

O próximo tópico aborda diferenças entre ACA e DCA.

4.4 DCA: o alcance, a relação com a ACA e as especificidades

O documento é a menor unidade de relacionamento para estudos de cocitação, mas é a unidade que efetivamente promove as relações. Alguns resultados já apresentados nessa pesquisa indicaram ser a DCA uma abordagem menos robusta em termos de dados do que a ACA, seja usando os primeiros ou todos os autores das referências, principalmente pela proporcionalidade buscada para definir os autores/documentos para formação das matrizes quadradas de cocitação nesse estudo, pois ainda que a matriz de cocitação de documentos possua 109 variáveis, diante das 72 posições da cocitação pura de primeiro autor, o ponto de corte é menor.

Mas há vantagens para se destacar sobre a DCA, suas especificidades podem assim ser enumeradas: a primeira é a inclusão de documentos sem autoria pessoal no *ranking*; a segunda é a identificação de relacionamentos de cocitação entre documentos do mesmo autor, como nos casos dos 11 documentos do autor “Hjørland, B” e dos cinco documentos autor “Olson, HA”, que fazem parte dos 109 documentos da DCA para essa pesquisa; a terceira é limpeza e a estruturação dos dados, que é mais simples, pois não há “coautoria”, não há mais de uma citação por documento citante; e por último, a DCA é mais clara para interpretar os dados, pois os títulos carregam elementos semânticos, já os autores, principalmente os mais interdisciplinares e com muitos documentos citados, trazem mais complexidade para a interpretação das relações nos agrupamentos.

A DCA busca, portanto, evidenciar a unidade bibliográfica enquanto insumo para visualizar domínios, com a questão autoral em segundo plano. Isso não significa que autores muitos citados presentes em estudos de cocitação não alcancem essa condição por poucos ou apenas um documento publicado.

Em termos de alcance, grande parte dos dados da DCA, sejam os dados de análise, os dados intermediários de dispersão e os dados absolutos de dispersão, já foram apresentados ao longo do texto, mas vale alguns destaques para em seguida propor as relações com a ACA.

O maior percentual de dispersão absoluta é dos documentos, 88,07% dos 14.901 documentos distintos são citados uma única vez no *corpus*. Falando especificamente dos 109 artigos que formam a matriz de cocitação de documentos, eles são citados em 285 (67,7%) artigos citantes e existe cocitação em 201 (47,74%), a partir de 929

citações totais. Quanto ao tipo de autoria, dos 1.715 documentos que formam o *ranking* do estudo, 680 (39,65%) são documentos de autoria única e 997 (60,35%) são documentos elaborados em coautoria. Os 109 documentos da matriz de cocitação são divididos em 64 documentos elaborados em autoria única e 41 em coautoria, uma relação inversa aos dados de todo o *ranking*.

Um aspecto significativo da DCA para os dados dessa pesquisa é o ponto de corte de construção da matriz. Com a inclusão de documentos a partir de seis citações, há a possibilidade da inclusão de alguns autores que não fazem parte de nenhuma das matrizes de ACA e isso aumenta a complexidade da análise. Esses dados compensam ser analisados apenas na interpretação e validação dos agrupamentos, ou seja, na fase final da DCA.

O filósofo “Wittgenstein, L”, por exemplo, que possui 15 citações no *corpus* principal, em sete artigos citantes, tem o documento “Philosophical Investigations” (em dois idiomas) citado seis vezes e está incluso na análise dos documentos, mesmo com o autor longe dos cortes da ACA para todos e primeiros autores. O autor “Smiraglia, RP”, como será mostrado no tópico 4.6 desse trabalho, a partir de dados do *corpus* secundário de análise, possui 44 documentos distintos citados, sendo que 34 possuem apenas uma citação e apenas um documento se destaca na análise que será apresentada quanto a caracterização de um agrupamento.

A DCA, portanto, apresenta um viés que pode ser escondido nas ACA: a identificação de documentos de destaque diante da técnica que privilegia a produção dos autores e pode refinar essas análises, como afirmam Chen, Ibekwe-Sanjuan e Hou (2010). Além disso, enquanto unidade temática imutável, diferente da unidade “autor” (entidade bibliográfica para estudos de citação ou cocitação) que se transforma pela variedade da produção bibliográfica do sujeito, os documentos definem com mais clareza as especialidades temáticas, como já salientado.

A Tabela 19 apresenta os autores que mais possuem documentos na matriz da DCA, considerando apenas a primeira posição das referências. Além do número de documentos, está indicado também o número de citações, que no contexto dos documentos indicam a distribuição pelos artigos citantes.

Tabela 19. Distribuição dos autores que possuem documentos incluídos na DCA, considerando a primeira posição de autoria (nº de documentos e citações totais).

AUTORES	Nº de documentos	Citações	AUTORES	Nº de documentos	Citações
1 Hjørland, B	11	97	19 Davis, FD	01	11
2 Olson, HA	05	43	20 Porter, MF	01	11
3 NAP**	04	33	21 Carpineto, C	01	10
4 Manning, CD	03	33	22 Case, DO	01	10
5 Salton, G	03	28	23 Dahlberg, I	01	10
6 Saracevic, T	02	24	24 Page, L	01	10
7 Ingwersen, P	02	23	25 Ponte, JM	01	10
8 Berners-Lee, T	02	21	26 Aitchison, J	01	09
9 Kuhlthau, CC	02	21	27 Bowker, GC	01	09
10 Marchionini, G	02	20	28 Broder, AZ	01	09
11 Jansen, BJ	02	18	29 Lavrenko, V	01	09
12 Lancaster, FW	02	18	30 Ranganathan, SR	01	09
13 Wilson, TD	02	17	31 Small, H	01	09
14 Robertson, SE	02	14	32 Borlund, P	01	08
15 Svenonius, E	02	14	33 Croft, WB	01	08
16 Buckley, C	02	13	34 Furner, J	01	08
17 Baeza-Yates, R	01	27	35 Golder, SA	01	08
18 Blei, DM	01	23	36 Järvelin, K	01	08
			Outros*	39	251

* 39 autores com um documento, que possui entre seis e sete citações
 ** Autoria não pessoal

Fonte: dados da pesquisa.

Os dados indicam que apenas 15 autores (excluindo os dados de autoria não pessoal) possuem mais de um documento na análise, com destaque para os autores “Hjørland, B” e “Olson, HA”. Os documentos de autoria não pessoal (bordas em amarelo) são duas normas ISO¹⁷, sendo uma recorrentemente citada em duas partes (ISO 25964-1 e ISO 25964-2), além do informe final do “*Functional Requirement for Bibliographic Records*”, publicado pela IFLA¹⁸.

O autor “Baeza-Yates, R” (bordas em vermelho), possui o documento mais citado da análise (*Modern information retrieval*, 1999) e parte considerável dos seus rastros na ACA ocorrem por conta desse documento, que foi publicado em coautoria com o autor “Ribeiro-Neto, B”. Os outros destaques ficam para a autora “Dahlberg, I”

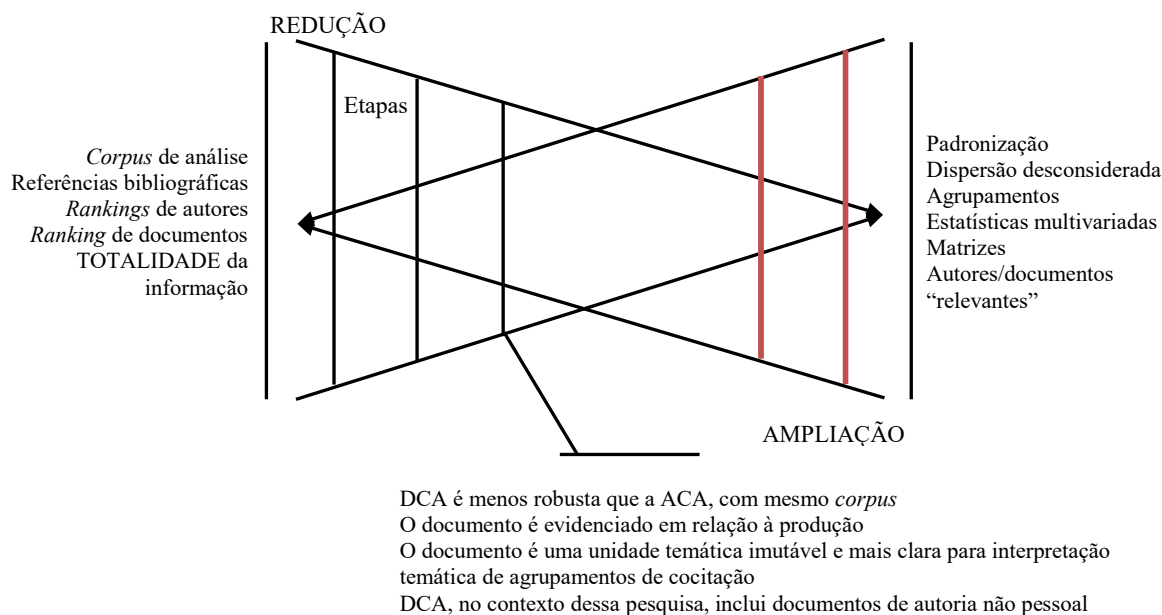
¹⁷ International Organization for Standardization.

¹⁸ International Federation of Library Associations and Institutions (IFLA).

(destaque em azul), pesquisadora influente e pioneira na área da Organização do conhecimento, mas que possui apenas um documento entre os 109 que fazem parte da DCA: *Knowledge organization: a new science?*; e para o autor “Croft, WB” (destaque em verde), que apesar de ser o segundo autor citado em mais documentos, considerando a análise com todos os autores das referências, possui apenas um documento utilizado na DCA: *Search engines*, 2010.

A Figura 24 ilustra a DCA como etapa da redução da produção de informações em estudos de cocitação, evidenciando algumas particularidades.

Figura 24. Triângulos isósceles da redução da informação em estudos de cocitação (análise de cocitação de documentos)



Fonte: Adaptado e ampliado de Latour (2006).

O documento é uma unidade de contagem para estudos de cocitação mais simples que o autor, principalmente quando do uso de todos os autores das referências, menos robusta em termos de grandes valores de pares de cocitação, mas pode colaborar com informações relevantes para ACA, como afirmam Chen, Ibekwe-Sanjuan e Hou (2010) e são eficazes para visualização de domínios específicos do conhecimento, como mostram Jankovic, Kaufmann e Kindler (2008), Ozçınar (2015), González-Alcaide *et al* (2016) e Trujillo e Long (2018).

4.5 Análise comparativa considerando valores da diagonal, matrizes normalizadas e técnicas de estatística multivariada

A complexidade dos estudos de cocitação se acentuam na aplicação das técnicas de estatísticas multivariadas, que vem na sequência das dificuldades de estabelecer os autores/documentos para a análise, definir os valores da diagonal das matrizes, normalizar os valores absolutos etc. Essa parte dos resultados busca apresentar dados gerais da complexidade dessa etapa, que como nas etapas anteriores não tem uma definição objetiva de aplicação.

A estrutura dos dados da aplicação da análise fatorial buscou identificar as diferenças quanto às especificidades das matrizes e os resultados são apresentados nas Tabelas 20 (matrizes de cocitação exclusiva de todos autores), 21 (matrizes de cocitação pura de primeiro autor) e 23 (matrizes de cocitação de documentos).

As matrizes de cocitação analisadas são as com valores absolutos (quatro de todos os autores das referências, quatro dos primeiros autores e duas de documentos), as normalizadas pelo r de Pearson (quatro de todos os autores, quatro com os primeiros autores e duas de documentos), as normalizadas pelo Cs (quatro de autores, sendo duas com todos os autores e duas com os primeiros e uma de documento) e as normalizadas pelo IJ (quatro de autores, sendo duas com todos os autores e duas com os primeiros e uma de documento). As matrizes absolutas de autores, seja para todas ou para as primeiras posições, tiveram diagonais definidas com zero, citação total, documentos citantes e cocitação do autor com ele mesmo. As matrizes de documentos foram criadas com zero na diagonal e valor de citação total, ainda que trabalhos como o de Ozçınar (2015) tenha optado por um cálculo considerando os três valores mais altos de cocitação do documento na matriz e que poderia ser uma opção utilizada que foi descartada pelo número de matrizes.

Os seguintes dados foram identificados na aplicação da análise fatorial da matrizes: valor mínimo e máximo de comunalidade, que determina a variância da variável original compartilhada com as demais (HAIR *et al*, 2009); percentual da variância total explicada e da principal componente gerada; número de variáveis (autores/documentos) que formam o principal agrupamento gerado e o percentual em

Tabela 20. Dados gerais da análise fatorial das matrizes de cocitação exclusiva de todos autores (dados absolutos e normalizados).

Matriz	Diagonal	Comunalidades		Variância Explicada (%)		Nº de variáveis da maior componente (% do total de autores = 180)		Número de componentes*	Observações	Número de interações**
		Mín.	Máx.	Componente Principal (%)	Total (%)					
Dados absolutos	Zero	0,606	0,921	30,46	79,42	76	42,22	18	Efetivamente 14 componentes e 05 autores*** se isolaram como componentes.	16
	Citação	0,552	0,931	18,90	74,13	47	26,11	33	Efetivamente 32 componentes e 13 autores se isolaram como componentes.	70
	Citante	0,683	0,950	24,15	84,28	55	30,56	29	Efetivamente 25 componentes e 08 autores se isolaram como componentes.	22
	Cocitação	0,632	0,966	31,08	85,22	73	40,56	20	Efetivamente 15 componentes e 03 autores se isolaram como componentes.	27
<i>r</i> de Pearson	Zero	0,824	0,996	59,15	97,15	119	66,11	07	Efetivamente 05 componentes e 01 autor se isolou como componente.	10
	Citação	0,666	0,987	48,58	92,97	110	61,11	13	Efetivamente 09 componentes e 04 autores se isolaram como componentes.	24
	Citante	0,770	0,996	52,45	95,59	116	64,44	11	Efetivamente 08 componentes e três autores se isolaram como componentes.	88
	Cocitação	0,829	0,997	57,91	96,91	119	66,11	07	Efetivamente 05 componentes e 01 autor se isolou como componente.	10
Cosseno de Salton	Citação	0,513	0,913	15,71	73,01	30	16,67	36	Efetivamente 34 componentes, 09 autores se isolaram e 01 autor não foi alocado em nenhuma componente.	52
	Citante	0,657	0,944	20,62	83,00	38	21,11	31	Efetivamente 29 componentes e 07 autores se isolaram como componentes.	54
Índice de Jaccard	Citação	0,270	0,752	7,63	55,89	18	10,00	45	09 autores se isolaram como componentes e 04 não aderiram a nenhuma componente.	77
	Citante	0,471	0,877	12,48	69,11	26	14,44	39	10 autores se isolaram como componentes.	42

* A partir da rotina do *software SPSS*

** Para gerar matriz rotacionada

*** Considerando a maior carga fatorial gerada para aderir a uma componente

Fonte: dados da pesquisa.

relação ao todo; número de componentes da matriz rotacionada; observação quanto ao número das componentes e aderência das variáveis na análise; e número de interações.

Para os dados das 12 matrizes que incluem todos os autores das referências, que contém 180 variáveis, 9.748 relações válidas de cocitação (um ou mais), 6.362 zeros (39,49% das 16.110 relações possíveis entre os autores), relação máxima de cocitação igual a 25, os resultados se mostraram bem distintos com implicações bem subjetivas de entendimento.

Bu et al (2017), por exemplo, quando aplicam o método MFTACA e comparam os resultados com uma ACA tradicional (Quadro 5) acham um ganho em termos de análise haver mais componentes e os resultados mostram na presente pesquisa, levando em conta a variância explicada e o número de agrupamentos, que os resultados mais consistentes são os com menos componentes, principalmente as quatro matrizes normalizadas pelo r de Pearson, que apresentam variância total explicada acima de 90%, mas também apresentam as principais componentes com mais 110 variáveis (autores), o que dificulta muito a análise. Na revisão de literatura há estudos que trabalham com um número menor do que esse no conjunto total de análise, o que ocorre inclusive com os dados dessa pesquisa utilizando a cocitação pura de primeiro autor. A matriz possui muitos zeros e Leydesdorff (2005) já alertou para a sensibilidade dessa medida de correlação quanto a essa condição dos dados, pois se trata de uma medida com aplicação global (OLIVEIRA; ALVES, 2017). O excesso de presença de “zeros” tem certa relação com a definição dos autores que farão parte da análise.

Como mencionado na Tabela 17, “Hjørland, B”, o autor citado em mais artigos na análise não tem nenhuma cocitação com 39 sujeitos da matriz de cocitação exclusiva de todos os autores, enquanto o segundo autor com mais documentos citantes, “Croft, WB”, tem relações de cocitação mais significativas com autores que nem fazem parte da matriz analisada. Esse é um problema recorrente dos estudos e fazer arranjos para se livrar dos “zeros”, que não seja aumentar o volume de dados, não é recomendado, pois esse valor faz parte da variabilidade do fenômeno que se estuda.

Os resultados, comparando o uso de citação total e documentos citantes na diagonal são esclarecedores. Com o uso do Cs e do IJ, por exemplo, o uso da citação total apresenta valores mais baixos de comunalidades das variáveis, valores mais baixos de variância, mais agrupamentos e principalmente não aderem alguns autores a nenhuma componente, nem de forma isolada, pois a carga fatorial é muito baixa. Esses

cálculos de similaridade devem ser operados com o valor da diagonal próximo aos valores de cocitação, pois interferem diretamente nos resultados. O uso do r de Pearson com citações totais na diagonal é uma saída quando não houver outra opção. Mesmo usando os dados absolutos, os resultados com citação total parecem piores, mas utilizável, diferente do que ocorre com Cs e IJ.

Quanto às matrizes com valores absolutos (recorte com 30 posições nos Apêndices 8, 9 e 10), os resultados são mais interessantes com zero e cocitação do autor com ele mesmo na diagonal, o que evidencia a importância de uma boa escolha para definir esse valor e que sejam valores próximos aos valores máximos de cocitação da matriz para cada autor. Os resultados isolaram menos autores como componentes, cinco e três para zero e cocitação do autor com ele mesmo, respectivamente, contra 13 e oito para o uso de citação total e número de documentos citantes, respectivamente. Zhou e Leydesdorff (2016) indicaram que os valores absolutos das matrizes não precisam ser normalizados e os resultados se mostraram consistentes, em menor grau com citação total na diagonal da matriz. Ainda assim, considera-se o uso do valor de documento citante mais apropriado do que a cocitação do autor com ele, pois esses dados dependem dos valores de citação de cada autor e não da distribuição do autor no *corpus* de análise, como evidenciaram os dados apresentados na Tabela 13.

Salienta-se que o rigor estatístico não deve ser o principal elemento para concluir algo sobre os agrupamentos e o domínio que se pretende visualizar, mas deve ser uma ferramenta para servir à subjetividade do pesquisador, que possui objetivos claros de pesquisa e máximo domínio possível dos dados brutos que sustentam todas as inferências estatísticas.

A matriz de cocitação pura de primeiro autor tem 1.398 relações válidas de cocitação, com 1.158 zeros (45,31% das 2.556 relações possíveis) e cocitação máxima igual a 21. Com 72 variáveis é mais simples avaliar as relações e apesar de proporcionalmente ter mais zeros do que a matriz com todos os autores, os dados normalizados pelo r de Pearson, com zero e cocitação do autor com ele mesmo na diagonal, se mostraram redondos, sem variável isolada e com componentes bem definidas. Esses dados podem ser visualizados no Apêndice 11.

A quarta componente dessa análise, com matriz normalizada pelo r de Person e zero na diagonal, é bem coerente e mostra a relação de autores ligados aos EC, inclusive cocitação. A Tabela 22 apresenta a componente gerada e mostra que a carga fatorial de

Tabela 21. Dados gerais da análise fatorial das matrizes de cocitação pura de primeiro autor (dados absolutos e normalizados)

Matriz	Diagonal	Comunalidades		Variância Explicada (%)		Nº de variáveis da maior componente (% do total de autores = 72)		Número de componentes*	Observações	Número de interações**
		Mínimo	Máximo	Componente Principal (%)	Total (%)					
Dados absolutos	Zero	0,550	0,865	26,70	74,34	22	30,56	09	08 componentes efetivas e 02 autores*** se isolaram como componentes.	07
	Citação	0,518	0,848	15,54	70,10	09	12,50	18	04 autores se isolaram como componentes.	29
	Citante	0,612	0,941	21,17	79,55	12	16,67	16	05 autores se isolaram como componentes.	21
	Cocitação	0,671	0,960	28,70	85,79	21	29,17	11	10 componentes efetivas e 03 autores se isolaram como componentes.	12
<i>r</i> de Pearson	Zero	0,648	0,994	55,38	95,49	30	41,67	04	Sem autores isolados e componentes redondas.	16
	Citação	0,590	0,977	39,99	86,84	31	43,06	08	03 autores se isolaram como componentes.	12
	Citante	0,612	0,992	46,13	91,72	32	44,44	07	06 componentes efetivas e 01 autor se isolou como componente.	12
	Cocitação	0,654	0,995	54,08	95,72	34	47,22	04	Sem autores isolados e componentes redondas.	15
Cosseno de Salton	Citação	0,440	0,822	14,75	67,99	08	11,11	18	04 autores se isolaram como componentes e 01 autor não aderiu a nenhuma componente	44
	Citante	0,587	0,938	20,63	79,08	11	15,28	17	03 autores se isolaram como componentes.	46
Índice de Jaccard	Citação	0,404	0,697	8,20	53,54	07	9,72	21	03 autores se isolaram como componentes e 04 autores não aderiram a nenhuma componente.	60
	Citante	0,499	0,820	13,17	64,82	07	9,72	20	04 autores se isolaram como componentes e 01 autor não aderiu a nenhuma componente.	30

* A partir da rotina do *software SPSS*

** Para gerar matriz rotacionada

*** Considerando a maior carga fatorial gerada para aderir a uma componente

Fonte: dados da pesquisa.

menor valor é de 0,766.

Tabela 22. Distribuição dos autores de uma componente (cocitação pura de primeiro autor; zero na diagonal, normalizada pelo r de Pearson)

Autores	Componentes e cargas fatoriais			
	C1	C2	C3	C4
Ding, Y	---	---	---	0,935
Garfield, E	---	-0,301	---	0,917
Small, H	---	-0,387	---	0,900
White, HD	0,313	---	---	0,892
Leydesdorff, L	---	-0,446	---	0,862
Yan, E	-0,428	---	---	0,858
Cronin, B	0,457	-0,396	---	0,766

Fonte: dados da pesquisa.

Quanto aos demais dados da Tabela 21, os resultados utilizando o IJ, as duas matrizes, e utilizando o Cs, com citação na diagonal, tiveram variáveis que não foram aderidas à nenhuma componente. Esses dados têm relação direta com os valores de comunalidades das variáveis, pois somente nesses casos o valor mínimo ficou abaixo de 0,500.

Os dados para a DCA, que possuem 109 variáveis, apresentam a maior proporcionalidade de zeros na matriz, com 78,90% das 5.886 relações de cocitação possíveis entre os documentos, ou seja, apenas 1.243 relações válidas, com cocitação máxima igual a nove. No caso dos documentos, a citação total tem a mesma função dos artigos citantes para os autores e evidencia a distribuição das unidades no *corpus*.

Os dados mostram situações próximas: dados absolutos com citação na diagonal e dados normalizados pelo Cs e IJ. Esses dados dividem as variáveis a partir de 26 componentes.

Os dados com citação na diagonal e normalizados pelo r de Pearson se apresentam como os mais consistentes, seja pelos valores das comunalidades das variáveis, seja por ter a componente principal um tamanho equilibrado, sendo que apenas um documento se isolou, mas com carga fatorial de 0,417.

A sexta componente gerada com esses dados apresenta cargas fatoriais altas, a partir de 0,606, e uma relação temática focada na “Relevância” da Recuperação da

Tabela 23. Dados gerais da análise fatorial das matrizes de cocitação de documentos (dados absolutos e normalizados).

Matriz	Diagonal	Comunalidades		Variância Explicada (%)		Nº de variáveis da maior componente (% do total de documentos = 109)	Número de componentes*	Observações	Número de interações**	
		Mínimo	Máximo	Componente Principal (%)	Total (%)					
Dados absolutos	Zero	0,453	0,854	15,41	73,16	18	16,51	20	18 componentes efetivas*** e 04 documentos se isolaram como componentes.	21
	Citação	0,520	0,955	11,76	80,33	09	8,26	26	01 documento se isolou como componente.	26
<i>r</i> de Pearson	Zero	0,447	0,990	41,16	92,45	37	33,94	10	09 componentes efetivas e 01 documento se isolou como componente.	14
	Citação ¹⁹	0,767	0,980	29,52	91,96	27	24,77	16	15 componentes efetivas e 01 documento se isolou como componente.	14
Cosseno de Salton	Citação	0,537	0,957	11,48	81,36	08	7,34	28	27 componentes efetivas e 03 documentos se isolaram como componentes	22
Índice de Jaccard	Citação	0,431	0,903	7,28	68,00	08	7,34	32	04 documentos se isolaram como componentes.	51

* A partir da rotina do *software SPSS*

** Para gerar matriz rotacionada

*** Considerando a maior carga fatorial gerada para aderir a uma componente

Fonte: dados da pesquisa.

¹⁹ Disponível no Apêndice 12.

Informação, com dois documentos sobre pesquisa na Web. Segue a lista dos documentos com as respectivas cargas fatoriais:

- Relevance Reconsidered [Saracevic, T, 1996] _____ 0,933;
- Sources of information on specific subjects [Bradford, SC, 1934] _____ 0,910;
- Relevance [Saracevic, T, 2007] _____ 0,907;
- Relevance [Mizzaro, S, 1997] _____ 0,868;
- The concept of relevance in IR [Borlund, P, 2003] _____ 0,835;
- How are we searching the World Wide Web? [Jansen, BJ, 2006] _____ 0,743;
- The Anatomy of a Large-scale Hypertextual Web Search Engine [Brin, S, 1998] _____ 0,606.

A distribuição completa dessas componentes encontra-se disponível no Apêndice 12 e evidencia coerência entre as variáveis, pelo menos do ponto de vista do rigor estatístico (cargas fatoriais), indicando também que, por mais que os zeros influenciem as análises e no caso da DCA foi o que predominou, as rotinas estatísticas consideram isso como parte da variabilidade natural dos dados.

Quanto à análise de cluster, a Tabela 24 apresenta alguns dados nos moldes da análise fatorial para os três conjuntos de dados de cocitação com as matrizes correspondentes. Os dados são mais simples e focados no dendrograma gerado pela rotina, que representa a hierarquia dos agrupamentos, como foco na parte divisiva da representação e não na aglomerativa. Assim, tendo por base que a rotina tinha faixa de soluções igual a dois para o número mínimo de cluster, no *software SSPS*, duas informações foram retiradas: o número inicial de divisão (a partir de dois) e o número total da segunda divisão da faixa anterior. Foi colocada também na Tabela 24 uma coluna chamada “nº de agrupamentos”, onde a partir do exame do dendrograma se estabeleceu uma divisão final para exame dos agrupamentos em futuras pesquisas e a condição principal dessa ação era não isolar nenhuma variável.

Existem diferenças nas análises, mas que não são acentuadas e claras como na análise de fatores. A divisão inicial em todos os arranjos dos dados de cocitação ficou em dois ou três. No caso dos dados de cocitação com todos os autores, os dados normalizados com Cs e IJ ficam com três divisões iniciais e para os dados de cocitação com primeiros autores apenas as matrizes com citação total na diagonal ficam com a divisão inicial igual a três, mostrando dados mais dispersos.

Tabela 24. Dados gerais da análise de *cluster* de todas as matrizes do estudo (dados absolutos e normalizados).

Matriz	Diagonal	Todos os autores			Primeiros autores			Documentos		
		Corte inicial	Primeira divisão de cada cluster**	Nº de agrupamentos***	Corte inicial	Primeira divisão de cada cluster**	Nº de agrupamentos	Corte inicial	Primeira divisão de cada cluster**	Nº de agrupamentos
Dados absolutos	Zero	02	04	12	02	04	08	02	07	12
	Citação	03	06	14	03	06	11	03	07	09
	Citante	03	06	16	02	04	09	NA*	NA	NA
	Cocitação	02	05	12	02	04	08	NA	NA	NA
<i>r</i> de Pearson	Zero	02	04	12	02	04	07	02	04	10
	Citação	02	04	16	02	04	09	02	04	12
	Citante	02	04	11	02	04	10	NA	NA	NA
	Cocitação	02	04	10****	02	04	08	NA	NA	NA
Cosseno de Salton	Citação	03	08	20	03	07	09	03	08	08
	Citante	03	06	18	02	04	10	NA	NA	NA
Índice de Jaccard	Citação	03	09	18	03	06	12	03	10	10
	Citante	03	08	15	02	04	08	NA	NA	NA

* NA: não se aplica

** Divisão de cada cluster a partir do corte inicial

*** Divisão subjetiva dos conglomerados

**** Apesar de grandes agrupamentos a divisão acarretaria no isolamento de autores, que ficaram isolados.

Fonte: dados da pesquisa.

Quanto às divisões dos *clusters* iniciais, os dados normalizados com Cs e IJ e os dados absolutos apresentam mais divisões, e com o *r* de Pearson a divisão é sempre por dois. Esses dados são pobres, mas evidenciam alguma diferença na aplicação da técnica, a partir dos arranjos estabelecidos, que podem ser melhor visualizados partindo da aglomeração das variáveis. O autor “Hjørland, B.” foi escolhido para a identificação de como ocorre a sua aglomeração com outros sujeitos da análise. O Quadro 13 apresenta a relação para os dados absolutos, a partir da cocitação de primeiro autor pura. Os Apêndices 13 e 14 apresentam o dendrograma completo das relações com citação total e zero na diagonal, respectivamente.

Quadro 13. Relação aglomerativa do autor *Hjørland, B.* (dados absolutos).

Citação total	Documento citante	Cocitação do autor com ele mesmo	Zero
	<u>Beghtol, C</u>	<u>Beghtol, C</u>	<u>Beghtol, C</u>
Broughton, V			
	Dahlberg, I		Dahlberg, I
			Frohmann, B
			Furner, J
			García Gutiérrez, A
<i>Gnoli, C</i>	<i>Gnoli, C</i>	<i>Gnoli, C</i>	<i>Gnoli, C</i>
	Guimarães, JAC		Guimarães, JAC
			Mai, JE
	<u>Olson, HA</u>	<u>Olson, HÁ</u>	<u>Olson, HÁ</u>
Ranganathan, SR			
	<u>Smiraglia, RP</u>	<u>Smiraglia, RP</u>	<u>Smiraglia, RP</u>
	<u>Tennis, JT</u>	<u>Tennis, JT</u>	<u>Tennis, JT</u>
Svenonius, E			
Vickery, BC			

Fonte: dados da pesquisa.

Observa-se que a matriz com zero na diagonal gera mais relações e que apenas o autor “Gnoli,C” se aglomera com “Hjørland, B” nas quatro condições. Os autores “Beghtol, C”, “Olson, HA”, “Smiraglia, RP” e “Tennis, JT” se aglomeram em três condições que demonstram que a matriz com valor de citação total atribuída na diagonal destoa das demais situações, como ocorreu nos resultados da análise fatorial.

A seguir, como última parte dos resultados, é apresentada uma proposta do agrupamento como unidade de análise em estudos de cocitação.

4.6 Uma proposta do agrupamento como unidade de análise em estudos de cocitação

Esse tópico responde ao último objetivo específico proposto na pesquisa e é baseado integralmente nos dados do *corpus* secundário da análise, que já foram utilizados no item desta seção que responde à proposta de definição de ponto de corte para a inclusão de autores ou documento em estudos de cocitação. A abordagem apresentada também é vista como uma maneira de objetivar a fase de interpretação e validação dos estudos de cocitação.

O agrupamento analisado é formado por 17 autores e foi desenvolvido a partir de uma ACA inclusiva de todos os autores a partir de uma matriz quadrada com 50 posições. A análise fatorial foi rodada com uma matriz normalizada pelo Cs, com zeros atribuídos à diagonal, utilizando a técnica de componentes principais, sem definir um número fixo de fatores, com rotação *varimax* e suprimindo coeficientes menores do que 0,5. A variância explicada da análise foi de 89,23% e a do agrupamento foi de 46,40%. As cargas fatorais ficaram entre 0,624 e 0,953.

Os seguintes autores formam o agrupamento: “López-Huertas, MJ”, “Albrechtsen, H”, “Thellefsen, TL”, “Thellefsen, MM”, “Frohmann, B”, “Guimarães, JAC”, “Buckland, MK”, “Hansson, J”, “Andersen, J”, “Friedman, A”, “Mai, JE”, “Dahlberg, I”, “Bates, MJ”, “Smiraglia, RP”, “Hodge, G”, “Hjørland, B”, e “Svenonius, E.

O primeiro aspecto necessário para o uso de um agrupamento enquanto unidade de análise em estudos de cocitação é determinar a sua dimensão em relação ao corpus de análise (N=151 artigos). Os autores que formam o agrupamento analisado são citados em 106 artigos (70,20%), mas somente em 69 (45,70%) há cocitação, ou seja, há pelo menos dois autores do agrupamento citados. Nesses 69 artigos citantes, em termos de cocitação, há citação de dois a 12 autores que formam o agrupamento e 82,61% desses artigos foram publicados no periódicos *Knowledge Organization* (42), *Journal of Documentation* (08) e *Journal of the Association for Information Science and Technology* (07), sendo que dois artigos citaram 12 autores e outros dois citaram nove autores do agrupamento. Os dados indicam ainda que quase 80% dos artigos citam pelo menos três autores e que 16 artigos (23,19%) citam seis ou mais. Utilizar os dados desses documentos e de suas fontes pode ajudar no desafio de rotular os agrupamentos,

como afirmam Chen, Ibekwe-Sanjuan e Hou (2010). O Quadro 14 apresenta alguns metadados dos 16 artigos que mais citam autores do agrupamento analisado.

Observa-se que seis autores - “Friedman, A”, “Guimarães, JAC”, “Hjørland, B”, “Smiraglia, RP”, “Thellefsen, TL”, e “Thellefsen, MM” - dos principais documentos citantes também fazem parte do agrupamento como citados. Os dados de autocitação não foram retirados da análise, pois esse não é um procedimento usual nos trabalhos aplicados de cocitação de autores e nem é recomendado por trabalhos metodológicos, mas evidencia-se como elemento para estudos futuros.

Quadro 14. Artigos que citaram seis ou mais autores do agrupamento analisado (dados do corpus secundário de análise).

Autores - Título - Data - Nº autores citados – Periódico
1. Thellefsen, M.; Thellefsen, T.; Sørensen, B. <i>A Pragmatic Semeiotic Perspective on the Concept of Information Need and Its Relevance for Knowledge Organization</i> – 2013 (12) – <i>Knowledge Organization</i> .
2. Friedman, A.; Thellefsen, M. <i>Concept theory and semiotics in knowledge organization</i> – 2011 (12) – <i>Journal of Documentation</i> .
3. Guimarães, JAC.; Tognoli, N. <i>Provenance as a Domain Analysis Approach in Archival Knowledge Organization</i> – 2015 (09) – <i>Knowledge Organization</i> .
4. Hjørland, B. <i>User-based and Cognitive Approaches to Knowledge Organization: A Theoretical Analysis of the Research Literature</i> – 2013 (09) – <i>Knowledge Organization</i> .
5. Smiraglia, R.; Heuvel, Cvd. <i>Classifications and concepts: towards an elementary theory of knowledge interaction</i> – 2013 (08) – <i>Journal of Documentation</i> .
6. Thellefsen, T.; Thellefsen, M.; Sørensen, B. <i>Emotion, Information, and Cognition, and Some Possible Consequences for Library and Information Science</i> – 2013 (08) – <i>Journal of the Association for Information Science and Technology</i> .
7. Guimarães, JAC.; Pinho, FA. <i>Male Homosexuality in Brazilian Indexing Languages: Some Ethical Questions</i> – 2012 (08) – <i>Knowledge Organization</i> .
8. Friedman, A; Smiraglia, R. <i>Nodes and arcs: concept map, semiotics, and knowledge organization</i> – 2013 (08) – <i>Journal of Documentation</i> .
9. Almeida, CC.; Fujita, MSL.; Reis, DM. <i>Peircean Semiotics and Subject Indexing: Contributions of Speculative Grammar and Pure Logic</i> – 2013 (08) – <i>Knowledge Organization</i> .
10. Castanha, RCG.; Grácio, MCC. <i>Bibliometrics Contribution to the Metatheoretical and Domain Analysis Studies</i> – 2014 (07) - <i>Knowledge Organization</i> .
11. Hjørland, B. <i>Is classification necessary after Google?</i> – 2012 (07) – <i>Journal of Documentation</i> .
12. Martínez-Ávila, M.; San Segundo, R.; Zurian, FA. <i>Challenges and opportunities for knowledge organization at the intersection with information technologies</i> – 2014 (06) – <i>Revista Española de Documentación Científica</i> .
13. Tennis, JT. <i>Foundational, First-Order, and Second-Order Classification Theory</i> – 2015 (06) – <i>Knowledge Organization</i> .
14. Martínez-Ávila, M. <i>Knowledge Organization in the Intersection with Information Technologies</i> – 2015 (06) – <i>Knowledge Organization</i> .
15. Almeida, CC. <i>The Methodological Influence of Peirce’s Pragmatism on Knowledge Organization</i> – 2012 (06) – <i>Knowledge Organization</i> .
16. Friedman, A. <i>The relationship between research method and visual display: a study of conference proceedings in the field of knowledge organization</i> – 2014 (06) - <i>Information Research</i> .

Fonte: dados da pesquisa.

Os artigos focam em aspectos conceituais da Organização do conhecimento e no relacionamento desse campo com outras áreas, além de algumas aplicações. Vale destacar os estudos sobre classificação e teoria do conceito. Um estudo de acoplamento bibliográfico poderia evidenciar com mais clareza a relação dos documentos.

Além desses dados, vale salientar que os 17 autores possuem 650 citações no corpus, sendo que 586 (90,15%) são nos 69 artigos do agrupamento. Assim, o agrupamento se mostra muito coeso na estrutura dos dados que o sustentam, como as fontes onde os citantes são publicados, que, apesar de a análise não ser no nível do conteúdo, demonstram, como sugerem Bu et al. (2017), uma ligação temática.

A Tabela 25 apresenta diversas informações que caracterizam a participação de cada autor no agrupamento e no *corpus* secundário de análise, que são:

- Valores absolutos de cocitação na proximidade por artigo (referências) na metade superior da matriz, considerando coautoria como cocitação;
- Valores absolutos de cocitação na proximidade por parágrafo (desconsiderando dados de coautoria) na parte inferior da matriz, onde são contabilizados os artigos, onde há a ocorrência do par de cocitação, e não os parágrafos;
- Número de citações que os autores receberam nos artigos que formam o agrupamento, número total de citações no *corpus* e a diferença desses valores;
- Número de citações oriundas de artigos citantes no agrupamento, número de citações oriundas de artigos no corpus e a diferença desses valores;
- Número de documentos distintos dos autores citados no agrupamento e no corpus de análise, destacando as datas de publicação mais antigas e as datas mais atuais.

Essas informações, ainda que descritivas, evidenciam a importância e a participação de cada autor no agrupamento. Essa abordagem pode colaborar com estudos nos quais os pesquisadores têm pouca familiaridade com o domínio, além de ajudar a tornar as comunicações dos resultados mais autônomas para a comunidade interessada.

Quanto às coautorias, as mais significativas da análise são entre os autores “Hjørland, B” e “Albrechtsen, H” e entre os autores “Thellefsen, MM” e “Thellefsen, TL”. Realizar estudos considerando todos os autores das referências enriquece a análise, pois salienta autores que colaboram muito (PERSSON, 2001; ROUSSEAU;

Tabela 25. Distribuição de dados relativos aos autores do agrupamento analisado (dados do *corpus* secundário de análise).

Autores*	Matrizes de cocitação (referência vs parágrafo)															Indicadores adicionais dos autores												
	Hjørland, B	Smiraglia, RP	Dahlberg, I	Mai, J-E	Hodge, G	Albrechtsen, H	Svenonius, E	López-Huertas, MJ	Guimarães, JAC	Frohmann, B	Buckland, MK	Thellefsen, MM	Bates, MJ	Andersen, J	Hansson, J	Thellefsen, TL	Friedman, A	Citações no corpus	Citações no agrupamento	Diferença	Citantes no corpus	Citantes no agrupamento	Diferença	Documentos citados no corpus (datas de publicação do mais antigo e do mais atual)	Documentos citados no agrupamento (datas de publicação do mais antigo e do mais atual)	Diferença		
Hjørland, B		19	23	20	14	19	14	13	10	10	09	10	07	09	07	08	08	228	202	26	70	59	11	59	1988-2015	56	1992-2015	03
Smiraglia, RP	14		09	08	06	06	07	06	03	03	03	03	02	02	02	02	05	68	62	06	29	24	05	48	1987-2015	44	=	04
Dahlberg, I	10	05		08	07	09	06	05	06	07	05	06	04	03	03	05	02	44	41	03	28	26	02	18	1974-2009	17	=	01
Mai, J-E	11	05	02		06	08	06	04	02	05	04	07	02	06	04	05	06	47	44	03	25	22	03	20	1997-2015	19	=	01
Hodge, G	05	01	02	01		04	04	03	-	03	04	04	04	04	04	03	03	25	16	09	23	16	07	04	1999-2013	02	2000-2013	02
Albrechtsen, H	12	03	01	04	-		02	04	05	06	05	05	04	02	03	05	02	23	23	-	19	19	-	05	1993-1998	05	=	-
Svenonius, E	06	03	01	02	-	-		02	-	03	03	02	-	05	04	01	04	25	25	-	18	18	-	08	1981-2004	08	=	-
López-Huertas, MJ	05	-	02	01	02	01	-		01	02	04	04	01	02	04	04	03	21	19	02	16	14	02	10	1996-2012	09	=	01
Guimarães, JAC	05	-	02	-	-	01	-	01		03	02	04	01	-	01	04	-	32	25	07	13	11	02	26	2000-2015	22	=	04
Frohmann, B	03	02	-	01	-	-	-	-	01		01	03	02	04	03	03	03	21	19	02	12	11	01	10	1983-2009	08	1983-2008	02
Buckland, MK	04	02	02	02	-	01	-	01	-	-		07	02	02	05	05	04	14	11	03	12	11	01	09	1969-2012	05	=	04
Thellefsen, MM	05	01	04	02	02	02	-	02	01	01	01		02	02	04	09	05	24	24	-	11	11	-	11	2003-2013	11	=	-
Bates, MJ	03	01	01	-	-	-	-	-	-	-	-	-		02	03	02	01	11	09	02	11	09	02	08	1989-2007	07	=	01
Andersen, J	02	-	-	04	01	01	-	-	-	01	-	01	-		04	02	01	12	12	-	10	10	-	07	2002-2008	07	=	-
Hansson, J	01	-	-	-	02	-	01	-	-	-	02	01	01	01		04	02	12	12	-	10	10	-	06	1997-2006	06	=	-
Thellefsen, TL	04	01	03	01	01	02	-	01	01	01	01	04	-	01	01		04	27	27	-	09	09	-	14	2002-2013	14	=	-
Friedman, A	01	02	02	01	02	-	-	01	-	-	-	-	-	-	01	01		15	15	-	10	10	-	05	2006-2013	05	=	-

OBSERVAÇÕES: * ordem definida a partir do número de artigos citantes | ** nº de artigos e não de parágrafos onde há ocorrência da cocitação nesse tipo de proximidade, além de coautorias desconsideradas |

Fonte: dados da pesquisa.

ZUCCALA, 2004; CARVALHO; CAREGNATO, 2016), mas acarreta um pouco de ruído dependendo do objetivo proposto, mesmo que essas relações sejam identificadas, pois a unidade documental é a mesma e exemplos com os dos autores “Baeza-Yates, R” e “Ribeiro-Neto, B” ilustram essa condição.

A ocorrência da cocitação no parágrafo mostra com mais clareza a relação entre os autores e se revela efetivamente temática, assim como na proximidade por sentença (LIU; CHEN, 2012). Os dados da Tabela 25 mostram que das 131 relações entre os autores (parte superior da matriz quadrada), dentre 136 possíveis (pois há autores que não são cocitados), 64,12% também ocorrem no parágrafo (parte inferior) da matriz quadrada e isso valida o agrupamento. Verificar a ocorrência dos pares de cocitação em outros tipos de proximidade se mostra uma alternativa interessante à ACA tradicional, principalmente na fase de validação e interpretação dos agrupamentos ou de um agrupamento de destaque, uma vez que operacionalizar estudos no texto completo se mostra complexo, principalmente para desambiguação dos nomes dos autores (DING et al, 2013) e pela forma como a citação ocorre no corpo dos textos.

Os dados da Tabela 25 evidenciam que autor “Hjørland, B” é o único do agrupamento cocitado, na proximidade por parágrafo, com todos os outros autores, e isso pode ser explicado pelos dados que complementam a tabela, como o número de citações (202), o número de artigos citantes (59) e, principalmente, o número de documentos citados (56), em uma produção bibliográfica que se estende de 1992 a 2015, abordando-se especificamente os dados do agrupamento. Já os autores “Svenonius, E”, “Guimarães, JAC” e “Bates, MJ” são os que menos se relacionam no nível do parágrafo, e isso pode ser explicado pela distribuição das citações nos 69 artigos citantes que formam o agrupamento. Ou seja, ampliar o nível de descrição ajuda a perceber esses detalhes que caracterizam o agrupamento.

Um destaque interessante é o autor “Hodge, G”, que possui apenas dois documentos citados no *corpus*, sendo que um deles está entre os mais citados. O documento foi publicado em 2000 e seu título é “*Systems of knowledge organization for digital libraries*” (ver Figura 25). Essa informação evidencia que a unidade bibliográfica é muito significativa para a área, mesmo para um autor pouco produtivo diante do contexto analisado.

Os dados evidenciam também que sete autores são citados apenas em artigos que formam o agrupamento (Albrechtsen, H; Svenonius, E; Thellefsen, MM; Andersen,

J;Hansson, J; Thellefsen, TL e Friedman, A) e que exceto pelos autores “Hjørland, B”, “Smiraglia, RP” e “Hodge, G”, a diferença se mostra pequena entre o número de citações, número de documentos citantes e documentos citados dos mesmos sete autores já indicados. Vale salientar que os artigos citantes do agrupamento também podem influenciar as relações dos outros agrupamentos gerados pela análise fatorial, principalmente os que citam apenas dois e três autores.

O autor “Hjørland, B”, por exemplo, apresentou uma carga fatorial alta em relação ao segundo agrupamento da análise. O número de documentos que o citam e de documentos de sua autoria citados na análise comprovam sua influência no contexto geral desse corpus, como na área da Organização do Conhecimento. Aplicar essa abordagem no segundo agrupamento poderia mostrar outros detalhes e oferecer uma base comparativa que enriqueceria a análise.

A diferença entre as citações e o número de documentos citantes mostra que os autores são citados mais de uma vez no mesmo artigo, como já salientado em pesquisa relacionadas (BU *et al*, 2017; CARVALHO *et al*, 2018) e em dados do *corpus* principal da análise, evidências que sustentam a proposta de ponto de corte da pesquisa. Essa informação é omitida por ACA tradicionais, pois a contagem dos pares é feita pela ocorrência, mesmo que um determinado autor possua 10 documentos citados em um mesmo artigo. Nos dados do agrupamento analisado, apenas os autores “Hodge, G” (16), “Buckland, MK” (11) e “Bates, MJ” (09) possuem citações iguais ao número de documentos citantes e apenas a última tem uma relação idêntica nos dados relativos ao *corpus*. Os Autores “Hjørland, B”, “Smiraglia, RP”, “Dahlberg, I” e “Mai, JE” são os que apresentam maior diferença nessa relação e estão também entre os cinco autores que mais possuem documentos distintos citados no agrupamento, respectivamente 56, 44, 17 e 19.

Outro dado que merece destaque se refere às datas de publicação dos documentos citados de cada autor, apresentadas na última parte da Tabela 25. O trabalho de Bu, Liu e Huang (2016) utilizou essa informação para qualificar os dados apresentados em uma ACA. No presente estudo, apenas as datas dos documentos mais antigos e mais atuais são apresentadas. Assim, há documentos que foram publicados a partir de 1969 e isso é, certamente, um elemento de destaque. Apenas o autor “Albrechtsen, H” não possui documento citado que tenha sido publicado no século 21. Há quatro autores (Hjørland, B, Smiraglia, RP, Mai, J-E e Guimarães, JAC) que

possuem documentos publicados em 2015, quatro (Hodge, G, Thellefsen, MM, Thellefsen, TL e Friedman, A) que possuem documentos publicados em 2013 e dois (López-Huertas, MJ e Buckland, MK) em 2012. Essas datas foram destacadas por fazerem parte do recorte temporal do *corpus* secundário da análise, e mostram a atualidade da pesquisa na área de Organização do Conhecimento, no contexto dos artigos utilizados por essa pesquisa.

Por fim, foram identificados os documentos citados dos 17 autores que formam o agrupamento analisado. Os autores possuem 252 documentos distintos citados no corpus, sendo que 230 (91,27%) são citados no agrupamento, totalizando 532 citações. A Tabela 26 apresenta a distribuição por frequência de citação dos documentos e a Figura 25 apresenta uma rede dos 32 artigos mais citados, que possuem pelo menos quatro citações.

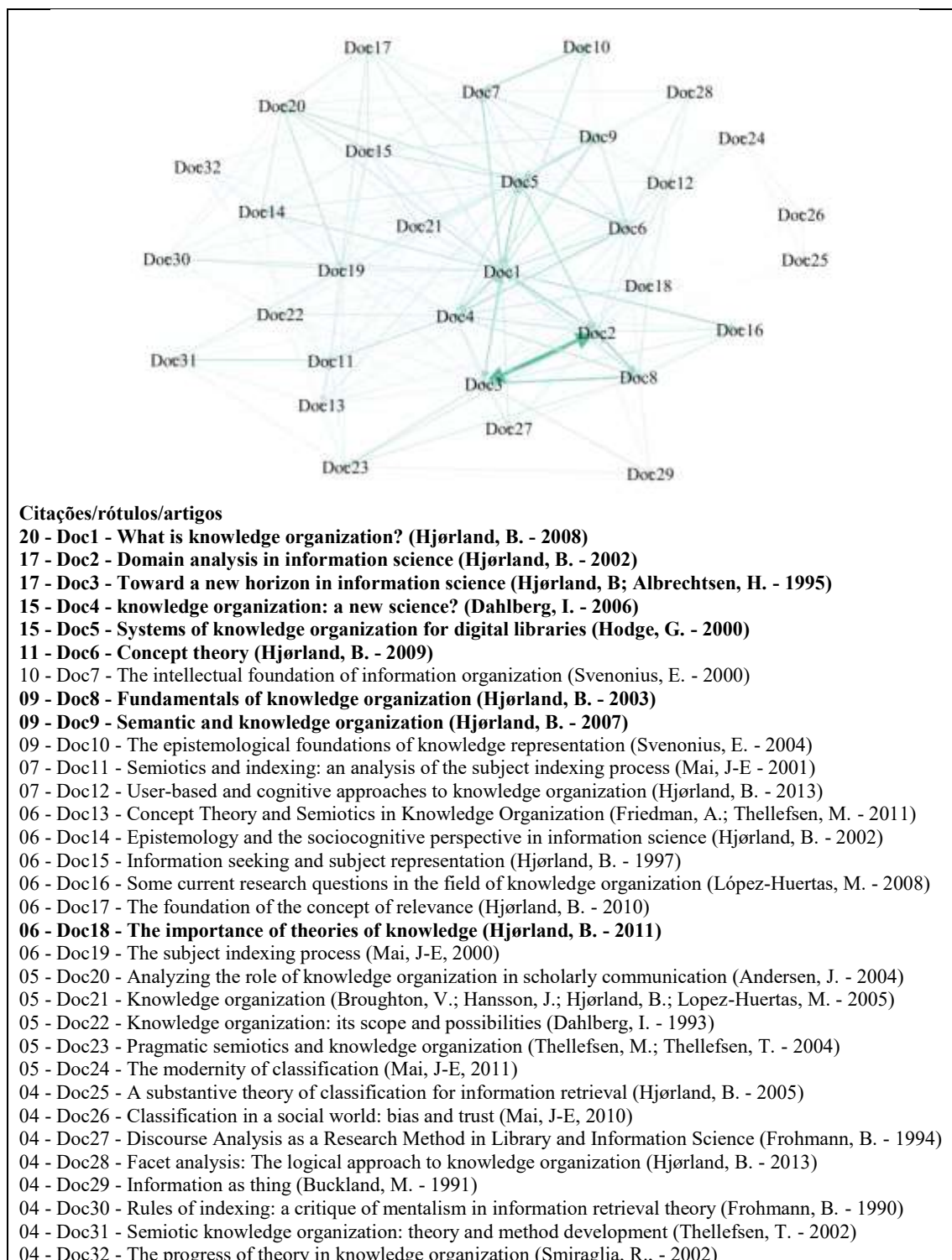
Tabela 26. Distribuição de frequências de citações dos documentos citados no agrupamento analisado (corpus secundário da análise).

Frequência de citações	Distribuição por documento			Distribuição por citação		
	Número de documentos	%	% acumulado	Citações	%	% acumulado
20 citações	01	00,43	00,43	20	03,76	03,76
17 citações	02	00,87	01,30	34	06,39	10,15
15 citações	02	00,87	02,17	30	05,64	15,79
11 citações	01	00,43	02,61	11	02,07	17,86
10 citações	01	00,43	03,04	10	01,88	19,74
09 citações	03	01,30	04,35	27	05,08	24,81
07 citações	02	00,87	05,22	14	02,63	27,44
06 citações	07	03,04	08,26	42	07,89	35,34
05 citações	05	02,17	10,43	25	04,70	40,04
04 citações	08	03,48	13,91	32	06,02	46,05
03 citações	20	08,70	22,61	60	11,28	57,33
02 citações	49	21,30	43,91	98	18,42	75,75
01 citação	129	56,09	100,00	129	24,25	100,00
Total	230	100,00		532	100,00	

Fonte: Dados da pesquisa.

Os dados da tabela mostram que há uma grande dispersão de documentos citados apenas uma vez nos 69 artigos que formam o agrupamento e que 32 artigos respondem por 46,05% das 532 citações. Essa informação demonstra a complexidade de lidar com dados de cocitação de autores, pois mesmo os mais citados, por produzirem muito, possuem documentos pouco citados e que ampliam consideravelmente o ruído informacional presente nos dados.

Figura 25. Rede e distribuição dos principais documentos citados na análise do agrupamento (04 ou mais citações – *corpus* secundário da análise).



Fonte: dados da pesquisa.

A Figura 25 apresenta um núcleo de oito documentos que se relacionam com mais força, com destaque para os documentos “Doc2” e “Doc3”, como se pode ser

observado na lista que complementa a figura. O autor “Hjørland, B” é responsável por seis desses documentos, sendo um deles em coautoria com “Albrechtsen, H”. Os outros são dos autores “Dahlberg, I” e “Hodge, G”, sendo que o último possui apenas dois documentos citados no agrupamento, como já salientado.

Desses oito documentos, três são sobre o que é a Organização do conhecimento e seus fundamentos: *What is knowledge organization?* (Hjørland, B. - 2008); *Knowledge organization: a new science?* (Dahlberg, I. - 2006); e *Fundamentals of knowledge organization* (Hjørland, B. - 2003). Outros documentos também abordam a área com esse enfoque sobre fundamentos e questões epistemológicas, além de relacionar a Organização do Conhecimento com a Ciência da Informação e outras disciplinas, como a Semiótica.

Outros temas decorrentes dos documentos mais citados são: classificação, indexação e teoria do conceito. Essa avaliação inicial é feita pelos títulos e, mesmo parecendo claro, não é simples rotular um agrupamento (CHEN; IBEKWE-SANJUAN; HOU, 2010) que se mostra muito complexo, como evidenciam os dados. Um caminho possível é verificar e entender como se dá o uso desses documentos nos artigos citantes, enquanto outro é verificar a citação ou cocitação na seção do documento citante por oferecer um rótulo ou um foco de uso.

Outra observação é que, dentre os documentos mais citados, há quatro que também são documentos citantes do *corpus* secundário, apresentados a seguir por ordem de citação: (i) *User-based and cognitive approaches to knowledge organization* (Hjørland, B. – 2013); (ii) *Concept Theory and Semiotics in Knowledge Organization* (Friedman, A.; Thellefsen, M. - 2011); (iii) *The importance of theories of knowledge* (Hjørland, B. - 2011); e (iv) *Facet analysis: The logical approach to knowledge organization* (Hjørland, B. - 2013).

Vale salientar que os autores “Guimarães, JAC” e “Bates, MJ” não possuem documentos no recorte de análise da Figura 1 e que os autores “Albrechtsen, H” e “Hansson, J” possuem apenas um documento citado na análise e publicado em coautoria, sendo que ambos não figuram na primeira posição das referências. Esses dados corroboram a importância de definir sobre o uso de todos os autores das referências e os ganhos ou perdas informacionais que acarreta.

Por fim, a unidade de contagem merece uma discussão com base nos dados. Contar cocitação por autores simplifica e agrega mais informações (230 documentos representados por 17 autores), pois há um destaque na produção, mas também pode ser um problema para rotular os agrupamentos posteriormente. O autor “Smiraglia, RP”, por exemplo, como já salientado, possui 34 documentos citados no agrupamento, com apenas uma citação dos 44 citados na análise, e apenas um documento está entre os mais citados listados na Figura 25 (*The progress of theory in knowledge organization*, 2002). Em um estudo de cocitação de documentos, esses dados poderiam ser desconsiderados, pois a dispersão é maior utilizando essa unidade de contagem. O ideal, como sugerem Chen, Ibekwe-Sanjuan e Hou (2010), é combinar as técnicas de cocitação de autores e cocitação de documentos. Além de “Smiraglia, RP”, os autores “Dahlberg, I” (11), “Guimarães, JAC” (21) e “Hjørland, B” (16) são outros destaques, com vários documentos citados uma única vez no agrupamento analisado.

Os resultados mostram que analisar agrupamentos pode ser uma abordagem válida em ACA, mesmo em um nível descritivo, e que parece fundamental dimensionar o agrupamento em relação aos dados gerais do corpus de análise.

Validar e interpretar agrupamentos em estudos de cocitação não é uma tarefa simples. Qualquer escolha na execução de um estudo dessa natureza pode implicar em elementos que deixam de ser analisados ou em muita informação para medir e condensar, e isso já se evidencia na escolha da unidade de contagem, principalmente as mais comuns: autores e documentos.

Na próxima seção são apresentadas as considerações finais da pesquisa e sugestões para futuros estudos.

5 CONSIDERAÇÕES FINAIS

O *corpus* de teste que ajudou na elaboração do projeto e que foi utilizado como *corpus* secundário de análise permitiu, além da publicação de alguns resultados, uma familiaridade com a ACA que vai além do que a literatura sobre cocitação pode oferecer. Isso ocorre porque os estudos focam nos agrupamentos ou em questões metodológicas, mas todo o trabalho de “mineração” dos pares é omitido, pois há uma excessiva geração de informação. No contexto dessa pesquisa, por exemplo, o tópico “4.5 Análise comparativa considerando valores da diagonal, matrizes normalizadas e técnicas de estatística multivariada” buscou resumir os resultados de 30 matrizes analisadas a partir de duas técnicas de estatísticas multivariadas e a parte relacionada à análise de *cluster* se mostra insuficiente quanto aos dados e discussão, sem contar que mais 12 matrizes inclusive de todos os autores, ou seja, que utiliza os dados de coautoria como cocitação, que estão prontas, foram retiradas da análise.

A ACA e a DCA não deveriam ser técnicas apenas para a identificação dos autores/documentos mais cocitados e o que isso pode representar em termos de rótulos que nomeiam uma disciplina ou uma subdisciplina do conhecimento. Deveriam ser também, em primeira instância, ainda que em caráter descritivo, abordagens para entender todo contexto representado pelos documentos citantes, mensurando o papel desses documentos na formação dos agrupamentos e medindo o que esses textos possuem de particular, os autores/documentos que não entram na compilação das matrizes. Entende-se que a pesquisa atingiu esse propósito, especialmente em definir os dados analisados, os dados intermediários de dispersão e os dados de dispersão absolutos.

A segunda instância de uma ACA são os agrupamentos amplamente estudados, que podem continuar sendo mapas intelectuais de autores cocitados e tratados isoladamente, mas que podem ser também instâncias de ligação com o que esses autores representam: seus documentos, sua produção intelectual. Nesse sentido, a noção de produção de informação que Latour (2006) propõe emerge como base para questionar o que uma ACA/DCA efetivamente mede e apresenta como resultado com seus processos de redução e seleção de informações.

Essa discussão está ligada ao enfoque normativo dos EC e ao conceito de informação, que parece ser um assunto inesgotável na CI, mas nem por isso é menor e desnecessário, ainda que o enfoque construtivista e a concepção sociocultural do EC (SILVEIRA, 2016) estejam em voga.

Assim, ampliar a descrição dos padrões estruturais (*corpus* de análise/citantes) que alicerçam as ACA, bem como apresentar mais elementos informativos que caracterizam os seus resultados (agrupamentos), sejam os documentos dos autores em um agrupamento específico ou uma DCA autônoma como recurso analítico, e mensurar e comparar dados desconsiderados em cada etapa de operacionalização da técnica, parece uma alternativa válida de pesquisa e discussão empírico-teórica no campo dos EC. Os resultados apresentados são robustos até a aplicação das técnicas de estatísticas multivariadas, mas mostrou-se problemático a partir disso.

As análises de cocitação se sustentam nas técnicas de mapeamento da ciência que não parecem ter nada de errado ou de definitivamente correto quanto à aplicação, mas se sustenta em decisões que parecem convenientes, como relacionar, por exemplo, 100 autores em um conjunto de milhares de dados e chamar isso de “estrutura intelectual de um domínio da ciência”, quando se pode estar captando apenas as relações de autores muito citados em uma parte específica dos documentos citantes. Os resultados deste trabalho mostraram essa condição.

Quanto à análise do agrupamento, último tópico dos resultados, os dados dos artigos citantes mostram a extensão dessas unidades e destacam os artigos que mais cocitam os autores, além de indicar as fontes desses documentos, que, no caso, são os periódicos e os autores. Os indicadores adicionais relacionados aos autores contextualizam a participação de cada um deles no agrupamento e no corpus de maneira geral, principalmente pela relação entre citação total, número de documentos citantes e documentos citados. A identificação das relações de cocitação na proximidade por parágrafo pode validar as cocitações recuperadas na lista de referências, pois, ainda que a análise não seja feita no nível do conteúdo, é bastante improvável que dois autores cocitados no mesmo parágrafo não tenham alguma relação temática.

Nos dados apresentados nessa parte da pesquisa, os indicadores adicionais serviram à descrição dos padrões estruturais de caracterização do agrupamento, esclarecendo a importância e a participação de cada autor. Sugere-se que essa abordagem seja utilizada na descrição mais detalhada dos resultados de ACA em

qualquer área e para todos os agrupamentos, pois amplia a autonomia informacional dos estudos.

Apresentar os documentos dos autores amplia o nível informacional e evidencia a natureza temática do agrupamento. Nos dados apresentados no último tópico dos resultados, a rede de cocitação dos documentos mostrou uma relação omitida pela ACA, qual seja, a ligação de documentos do mesmo autor, além de um documento muito citado de outro autor que parecia sem muita relevância no agrupamento. Combinar as técnicas de cocitação, ainda que como uma forma de interpretar um agrupamento, mostrou-se válido.

Para futuras pesquisas metodológicas no campo dos estudos de cocitação ficam as seguintes sugestões:

- Aprofundar os resultados quanto ao uso da análise de cluster e utilizar o escalonamento multidimensional com técnicas de estatísticas multivariadas para comparação com os resultados da análise fatorial;
- Utilizar outros valores na diagonal das matrizes, como a cocitação mais alta do autor/documento ou não atribuir valor, etc.;
- Comparar análises de cocitação desenvolvidas com matrizes de coocorrência e matrizes de ocorrência (binárias e com valores de citação);
- Desenvolver estudos de acoplamento bibliográfico de autores e documentos e comparar com análises de cocitação a partir do mesmo *corpus*; e
- Buscar aprofundar empírica e teoricamente os estudos de cocitação partir do enfoque construtivista dos EC.

Os dados utilizados nessa pesquisa já estão estruturados para realizar algumas das sugestões acima indicadas e o próximo passo é realizar um estudo de acoplamento bibliográfico, que é a técnica que evidencia as frentes de pesquisa de uma área.

Para futuras pesquisas aplicadas, principalmente no contexto brasileiro da CI, ficam as seguintes sugestões:

- ACA, DCA e acoplamento bibliográfico da área da CI no Brasil, a partir dos principais periódicos da área segundo a classificação Qualis-Periódicos²⁰; e
- Relacionar os resultados com os currículos dos cursos de graduação e pós-graduação da área da CI no Brasil, na intenção de buscar ainda mais sentido prático para os estudos de cocitação e acoplamento bibliográfico.

Enfim, espera-se que os resultados dessa pesquisa possam colaborar com o avanço dos estudos de cocitação no Brasil.

²⁰ [...] sistema usado para classificar a produção científica dos programas de pós-graduação no que se refere aos artigos publicados em periódicos científicos. Disponível em: <https://sucupira.capes.gov.br/sucupira/public/index.jsf>

REFERÊNCIAS

AHLGREN, P; JARNEVING, B.; ROUSSEAU, R. Requirements for a cocitation similarity measure, with special reference to Pearson's Correlation Coefficient. **Journal Of The American Society For Information Science And Technology**, v. 54, n. 6, p. 550–560, 2003. Disponível em: <https://doi.org/10.1002/asi.10242>. Acesso em: 14 fev. 2018.

BACHELARD, G. **Ensaio sobre o conhecimento aproximado**. Rio de Janeiro: Contraponto, 2004. 316p.

BECKER, J. L. **Estatística básica: transformando dados em informação**. Porto Alegre: Bookman, 2015. 488 p.

BORNMANN, L.; DANIEL, H.-D. What do citation counts measure? A review of studies on citing behavior. **Journal of Documentation**, v. 64, n. 1, p. 45-80, 2008. Disponível em: <https://doi.org/10.1108/00220410810844150>. Acesso em: 31 jul. 2015.

BRAAM, R. R.; MOED, H. F.; van RAAN, A. F. J. Mapping of science by combined co-citation and word analysis. I. Structural aspects. **Journal Of The American Society For Information Science**. v. 42, n. 4, p. 233-251, 1991a. Disponível em: [https://doi.org/10.1002/\(SICI\)1097-4571\(199105\)42:4233::AID-ASI13.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-4571(199105)42:4233::AID-ASI13.0.CO;2-I). Acesso em: 23 ago. 2016.

BRAAM, R. R.; MOED, H. F.; van RAAN, A. F. J. Mapping of science by combined co-citation and word analysis. II. Dynamical aspects. **Journal Of The American Society For Information Science**. v. 42, n. 4, p. 252-266, 1991b. Disponível em: [https://doi.org/10.1002/\(SICI\)1097-4571\(199105\)42:4252::AID-ASI23.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-4571(199105)42:4252::AID-ASI23.0.CO;2-G). Acesso em: 23 ago. 2016.

BRAGA, J. L. Para começar um projeto de pesquisa. **Comunicação & Educação**, v.10, n. 3, p. 288-296, 2005. Disponível em: <http://dx.doi.org/10.11606/issn.2316-9125.v10i3p288-296>. Acesso em: 30 set. 2015.

BU, Y. *et al.* MFTACA: an author co-citation analysis method combined with metadata in full text. In: INTERNATIONAL CONFERENCE ON SCIENTOMETRICS AND INFORMETRICS – ISSI, 16., 2017, Wuhan, China. **Proceedings...** Wuhan: International Society of Scientometrics and Informetrics – ISSI, 2017.

BU, Y.; LIU, T-Y; HUANG, W-B. MACA: a modified author co-citation analysis method combined with general descriptive metadata of citations. **Scientometrics**, v. 108, n. 1, p. 143-166, 2016. Disponível em: <http://doi.org/10.1007/s11192-016-1959-5>. Acesso em: 11 maio 2017.

CALLON, M. et al. From translations to problematic networks: an introduction to co-word analysis. **Social Science Information**, v. 22, n. 2, p. 191-235, 1983. Disponível em: <http://doi.org/10.1177/053901883022002003>. Acesso em: 23 ago. 2016.

CASSILLAS, J.; ACEDO, F. Evolution of the Intellectual Structure of Family Business Literature: A Bibliometric Study of FBR. **Family Business Review**, v. 20, n. 2, jun. 2007. Disponível em: <https://doi.org/10.1111/j.1741-6248.2007.00092.x>. Acesso em: 25 abr. 2018.

CARVALHO, R. A. et al. A relação entre referências e menções: estudo exploratório em artigos na base de dados BRAPCI. **Prisma.Com**, n. 36, p. 99- 115, 2018. Disponível em: <http://ojs.letras.up.pt/index.php/prisma.com/article/view/3928/3677>. Acesso em: 24 abr. 2018.

CARVALHO, R. A.; CAREGNATO, S. E. Análise de cocitação de autores – ACA: estudo exploratório comparando proximidade nas referências, seção do artigo e parágrafo. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, Marília – SP. **Anais...** Marília: ANCIB, 2017. Disponível: brapci.inf.br/index.php/article/download/59098. Acesso em: 10 dez. 2017.

CARVALHO, R. A.; CAREGNATO, S. E. Primeiro vs todos os demais autores citados: estudo empírico em artigos na base BRAPCI In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 5., 2016, São Paulo. **Anais...** São Paulo: USP, 2016. Disponível em: www.ebbc.inf.br/ebbc5/index.php/main/download/77. Acesso em: 30 nov. 2016.

CHEN, C. Citespace II: detecting and visualizing emerging trends and transient patterns in scientific literature. **Journal of The American Society For Information Science And Technology**, v. 57, n. 3, p. 359–377, 2006. Disponível em: <https://doi.org/10.1002/asi.20317>. Acesso em: 18 abr. 2018.

CHEN, C. *et al.* Mapping scientometrics (1981-2001). **Proceedings of the Association for Information Science and Technology**, v.39, n. 1, p.25-34, 2002. Disponível em: <https://doi.org/10.1002/meet.1450390103>. Acesso em: 08 nov. 2017.

CHEN, C. Science mapping: a systematic review of the literature. **Journal of Data and Information Science**, v.2, n.2, 2017. Disponível em: <https://doi.org/10.1515/jdis-2017-0006>. Acesso em: 16 abr. 2017.

CHEN, C.; IBEKWE-SANJUAN, F.; HOU, J. The structure and dynamics of co-citation clusters: a multiple-perspective co-citation analysis. **Journal of the American Society for Information Science and Technology**, v. 61, n. 7, p. 1386–1409, 2010. Disponível em: <http://doi.org/10.1002/asi.21309>. Acesso em: 20 out. 2016.

CO-CITATION ANALYSIS. c2013. Disponível em: <http://diging.github.io/tethne/doc/0.6.1-beta/tutorial.cocitation.html>. Acesso em: 20 jan. 2019.

CRONIN, B. Agreement and divergence on referencing practice. **Journal of Information Science**, v. 3, n. 1, p.27-33, 1981. Disponível em: <https://doi.org/10.1177/016555158100300104>. Acesso em: 28 fev. 2018.

CRONIN, B. **The citation process**: the role and significance of citations in scientific communication. London: Taylor Graham, 1984.

CULNAN, , M. J.; O'REILLY III, C. A. CHATMAN, J. A. Intellectual structure of research in organizational behavior, 1972-1984: A co-citation analysis. **Journal of the American Society for Information Science**, v. 41, n. 6, p.453-458, 1990. Disponível em: [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6453::AID-ASI133.0.CO;2-E](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6453::AID-ASI133.0.CO;2-E). Acesso em: 14 fev. 2018.

CUNHA, M. B.; CAVALCANTI, C. R. de O. **Dicionário de Biblioteconomia e Arquivologia**. Briquet de Lemos: Brasília, DF, 2008. 451p.

CUSTÓDIO, P. Á. G. R. **Dissertações e Teses da Pós-Graduação em Educação da Unesp/Campus de Marília**: um estudo das citações e cocitações (2004 a 2009). 2012. 146 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2012. Disponível em: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/Custodio_P_A_G_da_R_mestrado_CI_2012.pdf. Acesso em: 18 ago. 2016.

DING, Y. *et al.* The distribution of references across texts: some implications for citation analysis. **Journal of Informetrics**, n. 7, p. 583-592, 2013. Disponível em: <https://doi.org/10.1016/j.joi.2013.03.003>. Acesso em: 20 jun. 2016.

EOM, S. B. **Author cocitation analysis**: quantitative methods for mapping the intellectual structure of an academic discipline. Hershey: Information Science Reference, 2009. 347 p.

EOM, S. B.; FARRIS, R. S. The contributions of organizational science to the development of decision support systems research subspecialties. **Journal Of The American Society For Information Science**. v. 47, n. 12, p. 941-952, 1996. Disponível: <https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-4571%28199612%2947%3A12%3C941%3A%3AAID-ASI7%3E3.0.CO%3B2-2>. Acesso em: 14 fev. 2018.

FERREIRA, F. A. F. Mapping the field of arts-based management: Bibliographic coupling and cocitation analyses. **Journal of Business Research**, v. 85, p. 348-357, 2018. Disponível em: <https://doi.org/10.1016/j.jbusres.2017.03.026>. Acesso em: 07 nov. 2017.

FIELD, A. **Descobrimos a estatística usando o SPSS**. 2.ed. Porto Alegre: Artmed, 2009. 687 p.

GADOTTI, M. A dialética: concepção e método. In: _____. **Concepção dialética da educação: um estudo introdutório**. 3.ed. São Paulo: Autores Associados; Cortez, 1984. p.15-38. (Educação Contemporânea)

GARCÍA-LILLO, F.; ÚBEDA-GARCÍA, M.; MARCO-LAJARA, B. The intellectual structure of research in hospitality management: a literature review using bibliometric methods of the journal *International Journal of Hospitality Management*. **International Journal of Hospitality Management**, n.52, p. 121–130, 2016. Disponível em: <http://dx.doi.org/10.1016/j.ijhm.2015.10.007>. Acesso em: 11 maio 2017.

GARFIELD, E. Can citation indexing be automated? **Essays of an Information Scientist**, v.1, p. 84-90, 1962. Disponível em: <http://garfield.library.upenn.edu/essays/V1p084y1962-73.pdf>. Acesso em 10 abr. 2018.

GARFIELD, E. **From bibliographic coupling to co-citation analysis via algorithmic historio-bibliography**. 2001. Disponível em: <http://garfield.library.upenn.edu/papers/drexelbelvergriffith92001.pdf>. Acesso em: 10 abr. 2018.

GLÄNZEL, W. **Bibliometrics as a research field: a course on theory and application of bibliometric indicators**. 2003. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.5311&rep=rep1&type=pdf>. Acesso em: 01 nov. 2017.

GLEICK, J. **A informação: uma história, uma teoria, uma enxurrada**. Companhia das Letras: São Paulo, 2013. 521 p.

GONZÁLEZ-ALCAIDE, G. *et al.* Co-citation analysis of articles published in substance Abuse journals: intellectual structure and research fields (2001–2012). **Journal Of Studies On Alcohol And Drugs**, v. 77, n. 5, p.710–722, 2016. Disponível em: <https://doi.org/10.15288/jsad.2016.77.710>. Acesso em: 19 set. 2016.

GRÁCIO, M. C. C. Acoplamento bibliográfico e análise de cocitação: revisão teórico-conceitual. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 21, n. 47, p. 82-99, set./dez., 2016. Disponível: <http://www.doi.org/10.5007/1518-2924.2016v21n47p82>. Acesso em: 12 set. 2016.

GRÁCIO, M. C. C.; OLIVEIRA, E. F. T. de. Análise de cocitação de autores: um estudo teórico-metodológico dos indicadores de proximidade, aplicados ao GT7 da ANCIB. **Liinc em Revista**, Rio de Janeiro, v. 9, n. 1, p. 196-213, 2013a. Disponível em: <https://doi.org/10.18617/liinc.v9i1.527>. Acesso em: 18 ago. 2016

GRÁCIO, M. C. C.; OLIVEIRA, E. F. T. de. Estudos de análise de cocitação de autores: uma abordagem teórico-metodológica para a compreensão de um domínio. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013b, Florianópolis – SC. **Anais...** Florianópolis: ANCIB, 2013b. Disponível em: enancib.ibict.br/index.php/enancib/xivenancib/paper/viewFile/4331/3454. Acesso em: 04 jun. 2016.

GRÁCIO, M. C. C.; OLIVEIRA, E. F. T. de. Indicadores de proximidades em análise de cocitação de autores: um estudo comparativo entre coeficiente de Correlação de Pearson e Cosseno de Salton. **Informação & Sociedade: Estudos**, 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: 04 jun. 2016.

GRAEML, A. R.; MACADA, M. A.; ROSSONI, L. Redes sociais e intelectuais em administração da informação: uma análise cientométrica do período 1997-2006.

Informação & Sociedade: Estudos, João Pessoa, v.20, n.1, p. 95-110, jan./abr. 2010.

Disponível: www.periodicos.ufpb.br/index.php/ies/article/download/2182/7223. Acesso em: 18 ago. 2016.

HAIR, J. F. *et al.* **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009. 688 p.

HE, Q. Knowledge discovery through co-word analysis. **Library Trends**, v. 48, n. 1, p.133-159, 1999. Disponível em:

https://www.ideals.illinois.edu/bitstream/handle/2142/8267/librarytrendsv48i1i_opt.pdf.

Acesso em: 18 abr. 2018.

HJØRLAND, B. Citation analysis: A social and dynamic approach to knowledge organization. **Information Processing and Management**, n. 49, p.1313–1325, 2013.

Disponível em: <http://dx.doi.org/10.1016/j.ipm.2013.07.001>. Acesso em: 11 jun. 2017.

HJØRLAND, B. Domain analysis in Information Science: eleven approaches – traditional as well as innovative. **Jornal of Documentation**, v. 58, n. 4, p.422-462, 2002.

Disponível em: <http://doi.org/10.1108/00220410210431136>. Acesso em 12 set. 2016.

HU, C.-P. *et al.* A journal co-citation analysis of library and information science in China. **Scientometrics**, v. 86, n. 3, 2011. Disponível em:

<https://doi.org/10.1007/s11192-010-0313-6>. Acesso em: 18 abr. 2018.

JANKOVIC, M. P.; KAUFMANN, M.; KINDLER, C. H. Active research fields in anesthesia: a document co-citation analysis of the anesthetic literature. **Anesthesia & Analgesia**, v. 106, n. 5, p.1524-1533, 2008. Disponível em:

<http://doi.org/10.1213/ane.0b013e31816d18a1>. Acesso em: 22 ago. 2016.

JEONG, Y.; SONG, M.; DING, Y. Content-based author co-citation analysis. **Journal of Informetrics**, n.8, p.197-211, 2014. Disponível em:

<https://doi.org/10.1016/j.joi.2013.12.001>. Acesso em: 20 set. 2014.

KESSLER, M. M. Bibliographic coupling between scientific papers. **American Documentation**, p.10-25, 1963a. Disponível em:

<https://doi.org/10.1002/asi.5090140103>. Acesso em: 15 ago. 2016.

- KESSLER, M. M. Bibliographic coupling extended in time: ten case histories. **Information Storage and Retrieval**, v.1, n.4, p.169-187, 1963b. Disponível em: [https://doi.org/10.1016/0020-0271\(63\)90016-0](https://doi.org/10.1016/0020-0271(63)90016-0). Acesso em: 22 ago. 2016.
- KHASSEH, A. A. *et al.* Intellectual structure of knowledge in iMetrics: a co-word analysis. **Information Processing and Management**, v. 53, n. 3, p.705–720, 2017. Disponível em: <http://dx.doi.org/10.1016/j.ipm.2017.02.001>. Acesso em: 04 abr. 2018.
- KHASSEH, A. A.; SOHEILI, F; CHELAK, A. M. An author co-citation analysis of 37 years of iMetrics. **The Electronic Library**, v. 36, n. 2, p.319-337, 2018. Disponível em: <https://doi.org/10.1108/EL-09-2016-0191>. Acesso em: 04 abr. 2018.
- KONDER, L. **O que é dialética?** 28.ed. São Paulo: Brasiliense, 1998. 87p. (Primeiros Passos; v.23)
- KREUZMAN, H. A co-citation analysis of representative authors in philosophy: Examining the relationship between epistemologists and philosophers of science. **Scientometrics**, v. 51, n. 3, p525-539, 2001. Disponível em: <https://doi.org/10.1023/A:101277830>. Acesso em: 23 ago. 2016.
- LATOURE, B. Redes que a razão desconhece: laboratórios, bibliotecas, coleções. In: BARATIN, M.; JACOB, C. (dir.). **O poder das bibliotecas: a memória dos livros no ocidente**. 2.ed. Rio de Janeiro: UFRJ, 2006. p.21-44.
- LEYDESDORFF, L. et al. Citations: Indicators of Quality? The Impact Fallacy. **Frontiers in Research Metrics and Analytics**, 2016. Disponível em: <https://arxiv.org/pdf/1603.08452>. Acesso em: 03 abr. 2016.
- LEYDESDORFF, L. On the normalization and visualization of Author Co-Citation Data: Salton's Cosine versus the Jaccard Index. **Journal Of The American Society For Information Science And Technology**, v. 59, n.1, p.77–85, 2008. Disponível em: <https://doi.org/10.1002/asi.20732>. Acesso em: 14 fev. 2018.
- LEYDESDORFF, L. Similarity Measures, Author Cocitation Analysis, and Information Theory. **Journal Of The American Society For Information Science And Technology**, v. 56, n.7, p. 769–772, 2005. Disponível em: <https://doi.org/10.1002/asi.20130>. Acesso em: 20 out. 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://doi.org/10.1002/asi.20335>. Acesso em: 14 fev. 2018.
- LI, X.; QIAO, H.; WANG, S. Exploring evolution and emerging trends in business model study: a co-citation analysis. **Scientometrics**, v. 111, n. 2, p. 869–887, 2017. Disponível em: <https://doi.org/10.1007/s11192-017-2266-5>. Acesso em: 07 nov. 2017.

LIMA, L. S. **Produção científica em Organização do conhecimento**: uma análise de domínio via cocitações de autores. Marília, 2015, 82 f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista. 2015. Disponível em: <https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/DISSERTACAOARQUIVOMESTRE-LIDYANESILVALIMA.pdf>. Acesso em: 18 ago. 2016.

LIU, M. Progress in documentation the complexities of citation practice: a review of citation studies. **Journal of Documentation**, v.49, n.4, p. 370-408, 1993. Disponível em: <https://doi.org/10.1108/eb026920>. Acesso em: 22 ago. 2016.

LIU, S.; CHEN, C. The proximity of co-citation. **Scientometrics**, n.91, p.495-511. 2012. Disponível em: <http://doi.org/10.1007/s11192-011-0575-7>. Acesso em: 04 jun. 2016.

LUCAS, E. O.; GARCIA-ZORITA, J. C. Produção científica sobre o capital social: estudo por acoplamento bibliográfico. **Em Questão**, v. 20, n.3, Edição Especial, 2014. Disponível em: <http://seer.ufrgs.br/index.php/EmQuestao/article/view/49122/32415>. Acesso em: 15 ago. 2018.

MACHADO, R. das N. **Estrutura intelectual da literatura científica do Brasil e outros países dos BRICS**: uma análise de cocitação de periódicos na área de célula-tronco. 2015. 364f. Tese (Doutorado em Ciência da Informação) - Escola de Comunicação, Universidade Federal do Rio de Janeiro. Instituto Brasileiro de Informação em Ciência e Tecnologia, Programa de Pós-graduação em Ciência da Informação, Rio de Janeiro, 2015. Disponível em: <http://ridi.ibict.br/handle/123456789/884>. Acesso em: 18 set. 2016.

MARX, W.; BORNMANN, L. Change of perspective: bibliometrics from the point of view of cited references—a literature overview on approaches to the evaluation of cited references in bibliometrics. **Scientometrics**, v.109, n. 2, p. 1397–1415, 2016. Disponível em: <https://doi.org/10.1007/s11192-016-2111-2>. Acesso em: 19 set. 2016.

MATTOS, A. M.; DIAS, E. W. Análise de cocitação de autores: questões metodológicas. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 10., 2010, Rio de Janeiro – RJ. **Anais...** Rio de Janeiro: ANCIB, 2010. Disponível em: <http://www.brapci.inf.br/index.php/article/view/0000010589/61b12f45382ea8d227377501e6c62a94>. Acesso em: 23 mar. 2016.

McCAIN, K. Mapping author intellectual space: a technical overview. **Journal of the American Society for Information Science**. v.41, n.6, p.433-443. 1990. Disponível em: [http://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6:433::AID-ASII13.0.CO:2-Q](http://doi.org/10.1002/(SICI)1097-4571(199009)41:6:433::AID-ASII13.0.CO:2-Q). Acesso em: 30 jun. 2016.

McCAIN, K. Mapping economics through the journal literature: an experiment in journal cocitation analysis. **Journal of The American Society For Information**

Science, v. 42, n.4, p.290-296, 1991. Disponível em:
[https://doi.org/10.1002/\(SICI\)1097-4571\(199105\)42:4290::AID-ASI53.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199105)42:4290::AID-ASI53.0.CO;2-9).
Acesso em: 18 abr. 2018.

MILMAN, B. L.; GAVRILOVA, Y. A. Analysis of citation and co-citation in Chemical Engineering, **Scientometrics**, v.27, n. 1, p.53-74, 1993. Disponível em:
<https://doi.org/10.1007/BF02017755>. Acesso em: 23 ago. 2016.

MUÑOZ-MUÑOZ, A. M.; MIRÓN-VALDIVIESO, M. D. Analysis of collaboration and co-citation networks between researchers studying violence involving women. **Information Research**, v.22, n. 2, 2017. Disponível em:
<http://www.webcitation.org/6r5U5gOsZ>. Acesso em: 07 nov. 2017.

NI, C.; SUGIMOTO, C. R.; JIANG, J. Venue-author-coupling: a measure for identifying disciplines through author communities. **Journal Of The American Society For Information Science And Technology**, v. 64, n. 2, p.265–279, 2013. Disponível em: <https://doi.org/10.1002/asi.22630>. Acesso em: 14 fev. 2018.

OLIVEIRA, E. F. T. de.; GRÁCIO, M. C. C. Visibilidade dos pesquisadores no GT7 da ANCIB: um estudo de cocitações. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 12., 2011, Brasília – DF. **Anais...** Brasília – DF: ANCIB, 2011. Disponível em:
<http://enancib.ibict.br/index.php/enancib/enancibXII/paper/view/605>. Acesso em: 28 ago. 2016.

OLIVEIRA, E. F. T. de.; GRÁCIO, M. C. C. Visibilidade dos pesquisadores no periódico **Scientometrics** a partir da perspectiva brasileira: um estudo de cocitação. **Em Questão**, Porto Alegre, v. 18, Edição Especial, p. 99 - 113,/dez. 2012. Disponível em:
<http://seer.ufrgs.br/EmQuestao/article/view/33044>. Acesso em: 18 ago. 2016.

OLIVEIRA, E. F. T.; ALVES, B. H. Cosseno de Salton, Índice de Jaccard e Correlação de Pearson: comparando índices normalizados e absolutos em análise de cocitação de autores. **Em Questão**, v.23, Número especial, p. 235-253. 2017. Disponível em:
<http://dx.doi.org/10.19132/1808-5245230.235-253>. Acesso em: 16 nov. 2017.

OSAREH, F.; McCAIN, K. W. The structure of iranian chemistry research, 1990–2006: an author cocitation analysis. **Journal Of The American Society For Information Science And Technology**, v. 59, n.13, p.2146–2155, 2008. Disponível em:
<https://doi.org/10.1002/asi.20923>. Acesso em: 14 fev. 2018.

ÖZÇINAR, H. Mapping teacher education domain: a document co-citation analysis from 1992 to 2012. **Teaching and Teacher Education**, v. 47, p. 42-61, 2015. Disponível em: <https://doi.org/10.1016/j.tate.2014.12.006>. Acesso em: 18 abr. 2018.

PERSSON, O. All author citations versus first author citations. **Scientometrics**, v.50, n.2, p. 339-344. 2001. Disponível em: <https://doi.org/10.1023/A:1010534009428>. Acesso em 30 abr. 2016.

PERSSON, O. The Intellectual base and research fronts of JASIS 1986-1990. **Journal of the American Society for Information Science**, v. 45, n. 1, p. 31-38, 1994. Disponível em: [https://doi.org/10.1002/\(SICI\)1097-4571\(199401\)45:131::AID-ASI43.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-4571(199401)45:131::AID-ASI43.0.CO;2-G). Acesso em: 14 fev. 2018.

PIOVEZAN, L. B.; FUJITA, M. S. L. Análise de cocitação de autores: uma aplicação em estudos de indexação. **Em Questão: Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS**, v. 21, n. 1, 2015. Disponível em: <http://dx.doi.org/10.19132/1808-5245211.110-129>. Acesso em: 18 ago. 2016.

POPPER, K. R. **Conhecimento objetivo**: uma abordagem evolucionária. Belo Horizonte: Itatiaia; São Paulo: USP, 1975. 394p. (Espírito no Nosso Tempo; v.13)

REITZ, J. M. **ODLIS - Online Dictionary for Library and Information Science**. c2014. Disponível em: http://www.abc-clio.com/ODLIS/odlis_c.aspx. Acesso em: 13 mar. 2015.

ROMANCINI, R. O que uma citação? Análise de citações na ciência. **Intexto**, v. 2, n. 23, p. 20-35, 2010. Disponível em: <http://seer.ufrgs.br/intexto/article/view/15885>. Acesso em: 13 mar. 2015.

ROUSSEAU, R.; ZUCCALA, A. A classification of author co-citations: definitions and search strategies. **Journal of the American Society for Information Science and Technology**, v.55, n.6, p.513–529, 2004. Disponível em: <http://doi.org/10.1002/asi.10401>. Acesso em: 04 jun. 2016.

SANTAREM, L. G. S. **Caracterização dos pesquisadores em Tratamento Temático da Informação**: um estudo da produção científica por meio da análise de domínio. 2010. 112 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Disponível em: <https://repositorio.unesp.br/handle/11449/93617>. Acesso em: 18 ago. 2016.

SCHNEIDER, J. W.; LARSEN, B.; INGWERSEN, P. A comparative study of first and all author co-citation counting, and two different matrix generation approaches applied for author co-citation analyses. **Scientometrics**, v. 80, n. 1, p. 105–132, 2009. Disponível em: <https://link.springer.com/article/10.1007/s11192-007-2019-y>. Acesso em: 04 jun. 2016.

SERRA, F. et al. A pesquisa em administração estratégica nos primeiros anos do século XXI: Um estudo bibliométrico de citação e cocitação no *Strategic Management Journal* entre 2001 e 2007. **Revista Eletrônica de Estratégia & Negócios**, v. 5, p. 257-274. 2012. Disponível em: http://globadvantage.ipleiria.pt/files/2012/08/Working_paper-103_globadvantage.pdf. Acesso em: 20 fev. 2019.

SILVA, J. A. da.; BIANCHI, M. de L. P. Cientometria: a métrica da ciência. **Paidéia**, v.11, n.20, p.5-10, 2001. Disponível em: <http://migre.me/ly7Go>. Acesso em: 06 jul. 2014.

SILVA, R. R. G. de. Informação, ciberespaço e consciência. **TransInformação**, Campinas, v.18, n.3, p.191-202, set./dez. 2006. Disponível em: <http://revista.ibict.br/pscib/index.php/pscib/article/view/552>. Acesso em: 10 abr. 2007.

SILVEIRA, M. A. A. **Produção e distinção no domínio da Organização e representação do conhecimento no Brasil**. Porto Alegre, 2016, 154 f. Tese (Doutorado em Comunicação e Informação) – Programa de Pós-Graduação em Comunicação e Informação, Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul. 2016. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/153353/001014715.pdf>. Acesso em: 30 ago. 2017.

SILVEIRA, M. A. A.; BAZI, R. E. R. As referências nos estudos de citação: algumas questões para discussão. **DataGramZero**, v.10, n. 4, 2009. Disponível em: http://www.brapci.inf.br/_repositorio/2010/01/pdf_b5ff85e7c9_0007879.pdf. Acesso em: 16 jan. 2018.

SILVEIRA, M. A. A.; CAREGNATO, S. E. Demarcações epistemológicas dos estudos de citação: o fenômeno da citação. **Informação & Sociedade**, v. 27, n.3, p.145-154, 2017a. Disponível em: <https://doi.org/10.22478/ufpb.1809-4783.2017v27n3.33847>. Acesso em: 20 mar. 2018.

SILVEIRA, M. A. A.; CAREGNATO, S. E. Demarcações epistemológicas dos estudos de citação: teorias das citações. **Em Questão**, Porto Alegre. v. 23, n. 3, p. 250-275, 2017b. Disponível em: <http://dx.doi.org/10.19132/1808-5245233.250-275>. Acesso em: 16 fev. 2018.

SMALL, H. G. A co-citation model of a scientific specialty: a longitudinal study of collagen research. **Social Studies of Science**, v. 7, n. 2, p. 139-166, 1977. Disponível em: <https://doi.org/10.1177/030631277700700202>. Acesso em: 23 ago. 2016.

SMALL, H. G. Co-citation in the scientific literature: a new measure of the relationship between two documents. **Journal of the American Society for Information Science**, v. 24, n. 4, p. 265-269, 1973. Disponível em: <http://doi.org/10.1002/asi.4630240406>. Acesso em: 15 ago. 2016.

SMALL, H. G.; GREENLEE, E. A co-citation study of AIDS research. **Communication Research**, v. 16, n. 5, p. 642-666, 1989. Disponível em: <https://doi.org/10.1177/009365089016005006>. Acesso em: 07 nov. 2017.

SPINAK, E. **Diccionario enciclopédico de bibliometría, cientometría e informetría**. Caracas: UNESCO, 1996. 245 p.

TAHAMTAN, I.; BORNMANN, L. Core elements in the process of citing publications: conceptual overview of the literature. **Journal of Informetrics**, v.12, n. 1, p. 203-216, 2018. Disponível em: <https://doi.org/10.1016/j.joi.2018.01.002>. Acesso em: 16 jan. 2018.

TAN, P. N.; KUMAR, V.; SRIVASTAVA, J. Selecting the right objective measure for association analysis. **Information Systems**, v. 29, p. 293-313, 2004. Disponível em: [http://www.doi.org/10.1016/S0306-4379\(03\)00072-3](http://www.doi.org/10.1016/S0306-4379(03)00072-3). Acesso em: 17 maio 2017.

TRUJILLO, C. M.; LONG, T. M. Document co-citation analysis to enhance transdisciplinary research. **Science Advances**, v. 4, n.1, 2018. Disponível em: <http://doi.org/10.1126/sciadv.1701130>. Acesso em: 03 mar. 2018.

TSAY, M.; XU, H.; WU, C. Journal co-citation analysis of semiconductor literature. **Scientometrics**, v. 57, n. 7, 2003. Disponível em: <https://doi.org/10.1023/A:1023667318934>. Acesso em: 18 abr. 2018.

URBIZAGÁSTEGUI ALVARADO, R. A bibliometria: história, legitimação e estrutura. In: TOUTAIN, L. M. B. B. (Org.). **Para entender a Ciência da Informação**. Salvador: EDUFBA, 2007. p.185-217. (Saladeaula; v.6).

VANZ, S. A. S. **As redes de colaboração científica no Brasil: (2004-2006)**. Porto Alegre, 2009. 204 f. Tese (Doutorado em Comunicação e Informação) – Programa de Pós-Graduação em Comunicação e Informação, Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul. 2009. Disponível em: www.lume.ufrgs.br/bitstream/handle/10183/17169/000711634.pdf. Acesso em: 30 jun. 2015.

VANZ, S. A. S.; CAREGNATO, S. E. Estudos de Citação: uma ferramenta para entender a comunicação científica. **Em Questão**, v. 9, n. 2, p. 247-259, 2003. Disponível em: <http://seer.ufrgs.br/index.php/EmQuestao/article/view/75/35>. Acesso em: 16 jan. 2018.

WHITE, H. D. Author Cocitation Analysis and Pearson's *r*. **Journal Of The American Society For Information Science And Technology**, v. 54, n. 13, p.1250–1259, 2003. Disponível em: <https://doi.org/10.1002/asi.10325>. Acesso em: 14 fev. 2018.

WHITE, H. D.; GRIFFITH, B. C. Author cocitation: a literature measure of intellectual structure. **Journal Of The American Society For Information Science**, v. 32, n. 3, p. 163-171, 1981. Disponível em: <https://doi.org/10.1002/asi.4630320302>. Acesso em: 18 abr. 2018.

WHITE, H. D.; McCAIN, K. W. Visualizing a discipline: an author co-citation analysis of Information Science, 1972–1995. **Journal Of The American Society For Information Science**, v. 49, n. 4, p.327–355, 1998. Disponível em: [https://doi.org/10.1002/\(SICI\)1097-4571\(19980401\)49:4327::AID-ASI43.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-4571(19980401)49:4327::AID-ASI43.0.CO;2-4). Acesso em: 08 ago. 2016.

ZHAO, D. Towards all-author co-citation analysis. **Information Processing and Management**, v. 42, n. 6, p. 1578-1591, 2006. Disponível em: <https://doi.org/10.1016/j.ipm.2006.03.022>. Acesso em: 12 ago. 2017.

ZHAO, D.; STROTMANN, A. Evolution of research activities and intellectual influences in information science 1996–2005: introducing author bibliographic-coupling analysis. **Journal Of The American Society For Information Science And Technology**, v. 59, n. 13, p. 2070–2086, 2008a. Disponível em: <https://doi.org/10.1002/asi.20910>. Acesso em: 14 fev. 2018.

ZHAO, D.; STROTMANN, A. Information science during the first decade of the web: an enriched author cocitation analysis. **Journal Of The American Society For Information Science And Technology**, v. 59, n. 6, p.916–937, 2008b. Disponível em: <https://doi.org/10.1002/asi.20799>. Acesso em: 14 fev. 2018.

ZHAO, D.; STROTMANN, A. The knowledge base and research front of Information Science 2006–2010: an author cocitation and bibliographic coupling analysis. **Journal Of The Association For Information Science And Technology**, v. 65, n. 5, p.995–1006, 2014. Disponível em: <https://doi.org/10.1002/asi.23027>. Acesso em: 14 fev. 2018.

ZHAO, H.; ZHANG, F.; KWON, J. Corporate social responsibility research in international business journals: an author co-citation analysis. **International Business Review**, v. 27, n. 2, p. 389-400. Disponível em: <https://doi.org/10.1016/j.ibusrev.2017.09.006>. Acesso em: 13 abr. 2018.

ZHOU, Q.; LEYDESDORFF, L. The normalization of occurrence and co-occurrence matrices in bibliometrics using Cosine similarities and Ochiai coefficients. **Journal Of The Association For Information Science And Technology**, v67, n. 11, p. 2805-2814, 2016. Disponível em: <https://doi.org/10.1002/asi.23603>. Acesso em: 14 fev. 2018.

Apêndice 1. Expressões extraídas dos títulos e das palavras-chaves dos 151 artigos que formam o *corpus* secundário de dados (ocorrência igual ou maior que dois) _____

Knowledge organization* ²¹	68	Simple Knowledge Organization System	02
Classif*	26	Archival knowledge organization	02
Knowledge organization system*	20	Communication Sciences	02
Information Science*	14	Decision-Making	02
Ontolog*	13	Dewey Decimal Classification	02
Informat*	13	Digital environments	02
Knowledge	11	Ethical Issues	02
Concept*	09	Facet analysis	02
Document*	08	Folksonomy	02
Relation*	08	Health	02
Thesaur*	07	Information technologies	02
Epistem*	07	Knowledge Map*	02
Taxonom*	06	Learning	02
Semantic web	06	Library resources	02
Ethics	05	Morality	02
Domain Analysis	05	Organization of information	02
Knowledge Representation	05	Paul Otlet	02
Information Retrieval	05	Social Tagging	02
Art*	04	Subject headings	02
Work*	04	Topic Map*	02
Semiotics	04	Vocabularies	02
Controlled vocabular*	04	AGROVOC	02
Logic	04	Articles	02
Semantic*	04	Concept theory	02
Indexing	04	Context	02
Knowledge management*	04	Desclassification	02
Metadata*	04	Design*	02
Tag*	04	Ethical Values	02
Categor*	03	Facet*	02
Terminolog*	03	Humanities	02
Bibliometric*	03	Keyphrase*	02
Catalog*	03	K-State*	02
Library and Information Science	03	Language*	02
Data	03	Linked data	02
Digital librar*	03	Metaphysic*	02
Digital network	03	Music	02
Information management	03	Peirce	02
Information Organization	03	Ranganathan	02
Linked open data	03	Semantic Relation*	02
Metatheor*	03	Social Security*	02
Theor*	03	Subject Indexing	02
Visual distinctive language*	03	Technolog*	02
Artificial intelligence	02	Terms	02
Classification schemes	02	Topic model*	02
Entit*	02	Value*	02
Representation	02	Web Of Science (WoS)	02
Research	02		
Schemes	02		
Search	02		

²¹ * O uso do asterisco indica a presença de plurais ou outras variações da palavra/termo.

Apêndice 2. *Unitermos* extraídos dos títulos e das palavras-chaves dos 151 artigos que formam o *corpus* secundário de dados (ocorrência igual ou maior que dois) _____

* O uso do asterisco indica a presença de plurais ou outras variações da palavra/termo.

Knowledge - 134	Bibliograph* - 06	Hierarch* - 03	Gaming - 02
Organiz* - 110	Heading* - 06	Icon* - 03	Geographic - 02
Information - 75	Linked - 06	Interdisciplinar* - 03	Global - 02
Classif* - 50	Open - 06	Journal* - 03	Government - 02
Science* - 35	Resourc* - 06	Keyphrase* - 03	Health - 02
System - 32	Semiotics - 06	Learning - 03	Higher - 02
Librar* - 18	Topic* - 06	Media - 03	Hive* - 02
Ontolog* - 17	Approach* - 05	Metatheor* - 03	Homosexuality - 02
Semant* - 17	Art* - 05	Multilingual - 03	Interaction - 02
Concept* - 15	Author* - 05	National - 03	Issues - 02
Languag* - 15	Controlled - 05	Peirce* - 03	Kaba - 02
Domain* - 14	Human* - 05	Philosophy - 03	Kos - 02
Representat* - 14	Logic* - 05	Scholar* - 03	K-stat - 02
Theor* - 14	Music* - 05	Search - 03	Levels - 02
Subject* - 13	Schem* - 05	Structur* - 03	Lexical - 02
Ethic* - 12	Scientific - 05	Text - 03	Literature - 02
Relation* - 12	Technolog* - 05	Univers* - 03	Markup - 02
Research* - 12	Work* - 05	Activit* - 02	Mass - 02
Vocabular* - 12	Articles - 04	Agrovoc - 02	Mesh - 02
Web - 12	Brazil* - 04	Arrangement - 02	Morality - 02
Analysis - 11	Communication - 04	Bakhtin* - 02	Multidisciplinary - 02
Data - 11	Cultur* - 04	Basic-level - 02	Objects - 02
Indexing - 11	Group* - 04	British - 02	Otlet - 02
Social - 11	Integrat* - 04	Cluster* - 02	Paul - 02
Archiv* - 10	Intelligence - 04	Co-author* - 02	Perspective - 02
Map* - 10	Metadata - 04	Cognit* - 02	Practic* - 02
Digital - 09	Model* - 04	Collection - 02	Proceedings - 02
Document* - 09	Standards - 04	Congress - 02	Product* - 02
Epistem* - 09	Value* - 04	Creation - 02	Profession* - 02
Facet* - 09	Academ* - 03	Decision-making - 02	Public - 02
Tag* - 09	Analytic - 03	Description - 02	Ranganathan - 02
Term* - 09	Artificial - 03	Disambiguation - 02	Security - 02
Thesaur* - 09	Automat* - 03	Display - 02	Site - 02
Management - 08	Bibliometric* - 03	Education - 02	Space* - 02
Retrieval - 08	Cartographic - 03	Engineering - 02	Specificity - 02
Taxonom* - 08	Content - 03	Entit* - 02	Tool* - 02
Visual* - 08	Context* - 03	Evaluation - 02	Traditional - 02
Catalog* - 07	Decimal - 03	Extraction - 02	Typ* - 02
Categor* - 07	Design* - 03	Field - 02	Use* - 02
Chin* - 07	Distinctive - 03	Folksonomy - 02	Voice* - 02
Method* - 07	Environments - 03	Formal - 02	World - 02
Network* - 07	Graph* - 03		

APÊNDICE 3. Distribuição dos periódicos com até quatro artigos publicados que formam o *corpus* principal de análise _____

Periódicos	2015	2016	Total
INFORMACAO & SOCIEDADE-ESTUDOS	02	02	04
INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE	02	02	04
LIBRARY TRENDS	04	--	04
PERSPECTIVAS EM CIENCIA DA INFORMACAO	03	01	04
PROFESIONAL DE LA INFORMACION	02	02	04
PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS	01	03	04
REVISTA IBERO-AMERICANA DE CIENCIA DA INFORMACAO	02	02	04
SCIENTIFIC AND TECHNICAL INFORMATION PROCESSING	02	02	04
BIBLIOS-REVISTA DE BIBLIOTECOLOGIA Y CIENCIAS DE LA INFORMACION	02	01	03
EDUCATION FOR INFORMATION	01	02	03
INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT	02	01	03
LIBRARY MANAGEMENT	03	--	03
REVISTA ESPANOLA DE DOCUMENTACION CIENTIFICA	02	01	03
REVISTA GENERAL DE INFORMACION Y DOCUMENTACION	01	02	03
IBERSID-REVISTA DE SISTEMAS DE INFORMACION Y DOCUMENTACION	01	01	02
INFORMATION DEVELOPMENT	02	--	02
JLIS.IT	01	01	02
JOURNAL OF ACADEMIC LIBRARIANSHIP	01	01	02
JOURNAL OF LIBRARIANSHIP AND INFORMATION SCIENCE	--	02	02
JOURNAL OF THE MEDICAL LIBRARY ASSOCIATION	01	01	02
MALAYSIAN JOURNAL OF LIBRARY & INFORMATION SCIENCE	01	01	02
TURKISH LIBRARIANSHIP	--	02	02
WORLD PATENT INFORMATION	01	01	02
AFRICAN JOURNAL OF LIBRARY ARCHIVES AND INFORMATION SCIENCE	01	--	01
ANALES DE DOCUMENTACION	--	01	01
ANNALS OF LIBRARY AND INFORMATION STUDIES	--	01	01
AUSTRALIAN ACADEMIC & RESEARCH LIBRARIES	--	01	01
BIBLIOTHEK FORSCHUNG UND PRAXIS	01	--	01
CANADIAN JOURNAL OF INFORMATION AND LIBRARY SCIENCE-REVUE CANADIENNE DES SCIENCES DE L INFORMATION ET DE BIBLIOTHECONOMIE	--	01	01
COLLEGE & RESEARCH LIBRARIES	01	--	01
COLLNET JOURNAL OF SCIENTOMETRICS AND INFORMATION MANAGEMENT	--	01	01
CUADERNOS DE DOCUMENTACION MULTIMEDIA	--	01	01
DATA BASE FOR ADVANCES IN INFORMATION SYSTEMS	01	--	01
EVIDENCE BASED LIBRARY AND INFORMATION PRACTICE	01	--	01
HEALTH INFORMATION AND LIBRARIES JOURNAL	--	01	01
INFORMATION & CULTURE	01	--	01
INFORMATION RESOURCES MANAGEMENT JOURNAL	01	--	01
INFORMATION SYSTEMS RESEARCH	01	--	01
INFORMATION TECHNOLOGY & MANAGEMENT	01	--	01
JOURNAL OF INFORMATION & KNOWLEDGE MANAGEMENT	--	01	01
JOURNAL OF LIBRARY ADMINISTRATION	01	--	01
JOURNAL OF WEB LIBRARIANSHIP	--	01	01
LIBRARY & INFORMATION SCIENCE RESEARCH	01	--	01
LIBRARY REVIEW	--	01	01
RECORDS MANAGEMENT JOURNAL	--	01	01
SOCIAL SCIENCE COMPUTER REVIEW	--	01	01
SOUTH AFRICAN JOURNAL OF INFORMATION MANAGEMENT	01	--	01
TELEMATICS AND INFORMATICS	01	--	01
VINE-JOURNAL OF INFORMATION AND KNOWLEDGE MANAGEMENT	--	01	01
Total	50	44	94

APÊNDICE 4. Distribuição dos periódicos com até dois artigos publicados que formam o *corpus* secundário de análise

Periódicos	Ano de publicação					Total
	2011	2012	2013	2014	2015	
EL PROFESIONAL DE LA INFORMACIÓN	--	02	--	--	--	02
INFORMAÇÃO & SOCIEDADE: ESTUDOS	--	02	--	--	--	02
INFORMATION RESEARCH	--	01	--	01	--	02
INFORMATION PROCESSING AND MANAGEMENT	--	--	02	--	--	02
QUALITATIVE AND QUANTITATIVE METHODS IN LIBRARIES	--	--	--	--	02	02
THE ELECTRONIC LIBRARY	--	01	--	01	--	02
ASLIB PROCEEDINGS	--	01	--	--	--	01
EVIDENCE BASED LIBRARY AND INFORMATION PRACTICE	--	--	--	--	01	01
INVESTIGACIÓN BIBLIOTECOLÓGICA	--	--	--	--	01	01
INFORMATION TECHNOLOGY AND LIBRARIES	--	--	01	--	--	01
JOURNAL OF LIBRARIANSHIP AND INFORMATION SCIENCE	01	--	--	--	--	01
LIBRARY QUARTERLY	--	--	01	--	--	01
ONLINE INFORMATION REVIEW	--	--	--	--	01	01
PROGRAM: ELECTRONIC LIBRARY AND INFORMATION SYSTEMS	--	01	--	--	--	01
REVISTA ESPAÑOLA DE DOCUMENTACIÓN CIENTÍFICA	--	--	--	01	--	01
Total	01	08	04	03	05	21

APÊNDICE 5. Artigos que formam o *corpus* principal de análise

Publicados em 2015

1. Choi, Y. A Complete Assessment of Tagging Quality: A Consolidated Methodology. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23198
2. Sugiyama, K; Kan, MY. A comprehensive evaluation of scholarly paper recommendation using potential citation papers. INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES. DOI: 10.1007/s00799-014-0122-2
3. Tax, N; Bockting, S; Hiemstra, D. A cross-benchmark comparison of 87 learning to rank methods. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2015.07.002
4. Brasileiro, FS; Loureiro, JMM; Freire, GHD. A historical-epistemological reflection of the social perspective in the field of Information Science. INVESTIGACION BIBLIOTECOLOGICA. DOI: 10.1016/j.ibbai.2016.02.018
5. Xavier, BM; da Silva, AD; Gomes, GRR. A hybrid architecture for document indexing of the Municipal Official Gazette in Cachoeiro de Itapemirim. TRANSINFORMACAO. DOI: 10.1590/0103-37862015000100008
6. Marijan, R; Leskovar, R. A library's information retrieval system (In) effectiveness: case study. LIBRARY HI TECH. DOI: 10.1108/LHT-07-2015-0071
7. Lifante, MPA; Navarro, CC; Gonzalez, FJG. A new proposal to improve the description of astronomical resources The case of historical star catalogues. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-10-2013-0129
8. Ko, Y. A new term-weighting scheme for text classification using the odds of positive and negative class probabilities. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23338
9. Szostak, R. A Pluralistic Approach to the Philosophy of Classification. LIBRARY TRENDS. DOI: 10.1353/lib.2015.0007
10. Olijnyk, NV. A quantitative examination of the intellectual profile and evolution of information security from 1965 to 2015. SCIENTOMETRICS. DOI: 10.1007/s11192-015-1708-1
11. Vallez, M; Pedraza-Jimenez, R; Codina, L; Blanco, S; Rovira, C. A semi-automatic indexing system based on embedded information in HTML documents. LIBRARY HI TECH. DOI: 10.1108/LHT-12-2014-0114
12. Ajao, O; Hong, J; Liu, WR. A survey of location inference techniques on Twitter. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551515602847
13. Inthiran, A; Alhashmi, SM; Ahmed, PK. A user study on the information search behaviour of medical students. MALAYSIAN JOURNAL OF LIBRARY & INFORMATION SCIENCE.
14. Saldanha, GS. ABOUT THE BIBLIOLOGY BETWEEN PEIGNOT, OTLET AND ESTIVALS: three approaches of a long "metaepistemological" discourse of the knowledge organization. INFORMACAO & SOCIEDADE-ESTUDOS.
15. Bernaoui, R; Hassoun, M. Algerian university libraries and the digital age: new communication behaviors. LIBRARY MANAGEMENT. DOI: 10.1108/LM-09-2014-0114
16. Lu, K; Mao, J. An Automatic Approach to Weighted Subject Indexing-An Empirical Study in the Biomedical Domain. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23290
17. O'Brien, H; Cairns, P. An empirical evaluation of the User Engagement Scale (UES) in online news environments. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2015.03.003
18. Kelly, M. An Evidence Based Methodology to Facilitate Public Library Non-fiction Collection Development. EVIDENCE BASED LIBRARY AND INFORMATION PRACTICE. DOI: 10.18438/B8PW2P
19. Onyancha, OB. An informetrics view of the relationship between internet ethics, computer ethics and cyberethics. LIBRARY HI TECH. DOI: 10.1108/LHT-04-2015-0033
20. Liao, JW; Li, L; Liu, XY. An integrated, ontology-based agricultural information system. INFORMATION DEVELOPMENT. DOI: 10.1177/0266666913510716
21. de Gemmis, M; Lops, P; Semeraro, G; Musto, C. An investigation on the serendipity problem in recommender systems. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2015.06.008
22. Yoo, D; Choi, K; Lee, H; Suh, Y. An Ontology-based Co-creation Enhancing System for Idea Recommendation in an Online Community. DATA BASE FOR ADVANCES IN INFORMATION SYSTEMS. DOI: 10.1145/2804075.2804077
23. Tamine, L; Chouquet, C; Palmer, T. Analysis of biomedical and health queries: Lessons learned from TREC and CLEF evaluation benchmarks. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23351
24. Martinez-Avila, D; Ferreira, M; Magro, JL. Application of Critical Race Theory to knowledge organization and representation. SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO.
25. Vuotto, A; Bogetti, C; Fernandez, G. Application of TF-IDF factor in the semantic analysis of a documentary collection. BIBLIOS-REVISTA DE BIBLIOTECOLOGIA Y CIENCIAS DE LA INFORMACION. DOI: 10.5195/biblios.2015.227
26. Lai, CH. Applying Knowledge Flow Mining to Group Recommendation Methods for Task-Based Groups. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23185
27. Hjørland, B. Are relations in thesauri "context-free, definitional, and true in all possible worlds?". JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23253
28. Quan, TT; Luong, XH; Nguyen, TC; Cheung, HS. Argumentation-based schema matching for multiple digital libraries. ONLINE INFORMATION REVIEW. DOI: 10.1108/OIR-02-2014-0023
29. Cavicchi, S. Authority control. Operational aspects in a university environment and new experiences. JLIIS.IT. DOI: 10.4403/jlis.it-10340
30. Cohen, AM; Smalheiser, NR; McDonagh, MS; Yu, C; Adams, CE; Davis, JM; Yu, PS. Automated confidence ranked classification of randomized controlled trial articles: an aid to evidence-based medicine. JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION. DOI: 10.1093/jamia/ocu025
31. Petraki, E; Kapetis, C; Yannakoudakis, EJ. Automated thesaurus population and management. QUALITATIVE & QUANTITATIVE METHODS IN LIBRARIES.
32. Joorabchi, A; English, M; Mahdi, AE. Automatic mapping of user tags to Wikipedia concepts: The case of a Q&A website - StackOverflow. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551515586669
33. Alejo-Machado, OJ; Fernandez-Luna, JM; Huete, JF. Bibliometric study of the scientific research on "Learning to Rank" between 2000 and 2013. SCIENTOMETRICS. DOI: 10.1007/s11192-014-1467-4
34. Glanzel, W. Bibliometrics-aided retrieval: where information retrieval meets scientometrics. SCIENTOMETRICS. DOI: 10.1007/s11192-014-1480-7

35. Adler, MA. Broker of Information, the "Nation's Most Important Commodity": The Library of Congress in the Neoliberal Era. *INFORMATION & CULTURE*. DOI: 10.7560/IC50102
36. Liang, SS; de Rijke, M. Burst-aware data fusion for microblog search. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2014.10.008
37. Zarebska, H. Case law as a source of information. *QUALITATIVE & QUANTITATIVE METHODS IN LIBRARIES*.
38. Meireles, MRG; Cendon, BV. Citation-Based Document Categorization: An Approach Using Artificial Neural Networks. *QUALITATIVE & QUANTITATIVE METHODS IN LIBRARIES*.
39. Hjørland, B. Classical Databases and Knowledge Organization: A Case for Boolean Retrieval and Human Decision-Making During Searches. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23250
40. Fang, H. Classifying Research Articles in Multidisciplinary Sciences Journals into Subject Categories. *KNOWLEDGE ORGANIZATION*.
41. Abbasi, MK; Frommholz, I. Cluster-based polyrepresentation as science modelling approach for information retrieval. *SCIENTOMETRICS*. DOI: 10.1007/s11192-014-1478-1
42. White, HD. Co-cited author retrieval and relevance theory: examples from the humanities. *SCIENTOMETRICS*. DOI: 10.1007/s11192-014-1483-4
43. Kim, S; Ko, Y; Oard, DW. Combining Lexical and Statistical Translation Evidence for Cross-Language Information Retrieval. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23153
44. Moholkar, M; Roy, S; Buddle, K. Comparative study of Indian patent databases. *WORLD PATENT INFORMATION*. DOI: 10.1016/j.wpi.2015.03.001
45. Cassettari, RRB; Pinto, AL; Rodrigues, RS; Leticia-Silvana-dos Santos. Comparison of Zipf's law in textual content and oral discourse. *PROFESIONAL DE LA INFORMACION*. DOI: 10.3145/epi.2015.mar.09
46. Liu, DR; Chen, YH; Shen, MX; Lu, PJ. Complementary QA Network Analysis for QA Retrieval in Social Question-Answering Websites. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23155
47. Kulinich, AA. Conceptual Frameworks of Ontologies for Ill-Structured Problem Domains. *SCIENTIFIC AND TECHNICAL INFORMATION PROCESSING*. DOI: 10.3103/S0147688215060027
48. Neto, ALD; Cordeiro, RID. CONTRIBUTIONS TO ANALYSIS, DESCRIPTION AND ARCHIVAL INFORMATION REPRESENTATION OF NATIONAL AGENCY NEWSREELS. *INFORMACAO & SOCIEDADE-ESTUDOS*.
49. Lawrie, D; Mayfield, J; McNamee, P; Oard, DW. Cross-Language Person-Entity Linking from 20 Languages. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23254
50. Shirts, BH; Salama, JS; Aronson, SJ; Chung, WK; Gray, SW; Hindorff, LA; Jarvik, GP; Plon, SE; Stoffel, EM; Tarczy-Hornoch, PZ; Van Allen, EM; Weck, KE; Chute, CG; Freimuth, RR; Grundmeier, RW; Hartzler, AL; Li, RL; Peissig, PL; Peterson, JF; Rasmussen, LV; Starren, JB; Williams, MS; Overby, CL. CSER and eMERGE: current and potential state of the display of genetic information in the electronic health record. *JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION*. DOI: 10.1093/jamia/ocv065
51. Lee, WC. Culture and Classification: An Introduction to Thinking about Ethical Issues of Adopting Global Classification Standards to Local Environments. *KNOWLEDGE ORGANIZATION*.
52. Seshadri, K; Shalinie, SM; Kollengode, C. Design and evaluation of a parallel algorithm for inferring topic hierarchies. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2015.06.006
53. Sinha, B; Chandra, S; Garg, M. Development of ontology from Indian agricultural e-governance data using IndoWordNet: a semantic web approach. *JOURNAL OF KNOWLEDGE MANAGEMENT*. DOI: 10.1108/JKM-10-2014-0441
54. Abrahamson, JA; Rubin, VL. Differences over discourse structure differences: a reply to Urquhart and Urquhart. *JOURNAL OF DOCUMENTATION*. DOI: 10.1108/JD-02-2014-0037
55. Bode, EC; de Sousa, RTB. Digital preservation, information retrieval and Language. *REVISTA IBERO-AMERICANA DE CIENCIA DA INFORMACAO*.
56. Zong, N; Lee, S; Kim, HG. Discovering expansion entities for keyword-based entity search in linked data. *JOURNAL OF INFORMATION SCIENCE*. DOI: 10.1177/0165551514562704
57. Avila-Garcia, L; Ortiz-Repiso, V; Rodriguez-Mateos, D. Discovery tools: one-stop shopping?. *REVISTA ESPANOLA DE DOCUMENTACION CIENTIFICA*. DOI: 10.3989/redc.2015.1.1178
58. Cordula, FR; de Araujo, WJ. Discussions about acceptance of the free software for management and creation of referencial database for papers. *BIBLIOS-REVISTA DE BIBLIOTECOLOGIA Y CIENCIAS DE LA INFORMACION*. DOI: 10.5195/biblios.2015.257
59. Zhang, L; Olson, HA. Distilling Abstractions: Genre Redefining Essence versus Context. *LIBRARY TRENDS*. DOI: 10.1353/lib.2015.0015
60. Ravana, SD; Taheri, MS; Rajagopal, P. Document-based approach to improve the accuracy of pairwise comparison in evaluating information retrieval systems. *ASLIB JOURNAL OF INFORMATION MANAGEMENT*. DOI: 10.1108/AJIM-12-2014-0171
61. Smiraglia, RP. Domain Analysis of Domain Analysis for Knowledge Organization: Observations on an Emergent Methodological Cluster. *KNOWLEDGE ORGANIZATION*.
62. Rodriguez, AFZ; Figuerola, CG; Berrocal, JLA. Editing content in a collaborative environment: the case of the Spanish Wikipedia. *SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO*.
63. Macias-Galindo, D; Cavedon, L; Thangarajah, J; Wong, W. Effects of domain on measures of semantic relatedness. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23303
64. Alqaraleh, S; Ramadan, O; Salamah, M. Efficient watcher based web crawler design. *ASLIB JOURNAL OF INFORMATION MANAGEMENT*. DOI: 10.1108/AJIM-02-2015-0019
65. Baca, M; Gill, M. Encoding Multilingual Knowledge Systems in the Digital Age: the Getty Vocabularies. *KNOWLEDGE ORGANIZATION*.
66. Secundo, G; Magnier-Watanabe, R; Heisig, P. Engineering knowledge and information needs in Italy and Japan: bridging the gap between theory and practice. *JOURNAL OF KNOWLEDGE MANAGEMENT*. DOI: 10.1108/JKM-01-2015-0029
67. Suvorov, RE; Sochenkov, IV. Establishing the Similarity of Scientific and Technical Documents Based on Thematic Significance. *SCIENTIFIC AND TECHNICAL INFORMATION PROCESSING*. DOI: 10.3103/S0147688215050081
68. McQueen, K. Ethical Issues of Knowledge Organization in Designing a Metadata Schema for the Leo Kottke Archives. *KNOWLEDGE ORGANIZATION*.
69. da Silva, AP; Guimaraes, JAC; Tognoli, NB. Ethical Values in Archival Arrangement and Description: An Analysis of Professional Codes of Ethics. *KNOWLEDGE ORGANIZATION*.

70. Hoerber, O; Khazaei, T. Evaluating citation visualization and exploration methods for supporting academic search tasks. *ONLINE INFORMATION REVIEW*. DOI: 10.1108/OIR-10-2014-0259
71. Kenter, T; Balog, K; de Rijke, M. Evaluating document filtering systems over time. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2015.03.005
72. Mohamed, KA; Hassan, A. Evaluating federated search tools: usability and retrievability framework. *ELECTRONIC LIBRARY*. DOI: 10.1108/EL-12-20130211
73. Martin-Moncunill, D; Sicilia-Urban, MA; Garcia-Barriocanal, E; Sanchez-Alonso, S. Evaluating the degree of domain specificity of terms in large terminologies The case of AGROVOC. *ONLINE INFORMATION REVIEW*. DOI: 10.1108/OIR-02-2015-0052
74. Martin-Moncunill, D; Garcia-Barriocanal, E; Sicilia, MA; Sanchez-Alonso, S. Evaluating the Practical Applicability of Thesaurus-Based Keyphrase Extraction in the Agricultural Domain: Insights from the VOA3R Project. *KNOWLEDGE ORGANIZATION*.
75. Lewandowski, D. Evaluating the Retrieval Effectiveness of Web Search Engines Using a Representative Query Sample. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23304
76. Wu, YJ; Lehman, A; Dunaway, DJ. Evaluations of a Large Topic Map as a Knowledge Organization Tool for Supporting Self-Regulated Learning. *KNOWLEDGE ORGANIZATION*.
77. McTavish, J. Everyday life classification practices and technologies Applying domain-analysis to lay understandings of food, health, and eating. *JOURNAL OF DOCUMENTATION*. DOI: 10.1108/JD-08-2014-0105
78. Chang, YW; Huang, MH; Lin, CW. Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. *SCIENTOMETRICS*. DOI: 10.1007/s11192-015-1762-8
79. Yip, JYT; Lee, RWB; Tsui, E. Examining knowledge audit for structured and unstructured business processes: a comparative study in two Hong Kong companies. *JOURNAL OF KNOWLEDGE MANAGEMENT*. DOI: 10.1108/JKM-10-2014-0420
80. Ozdemiray, AM; Altıngöve, IS. Explicit Search Result Diversification Using Score and Rank Aggregation Methods. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23259
81. Stocker, A; Richter, A; Kaiser, C; Softic, S. Exploring barriers of enterprise search implementation: a qualitative user study. *ASLIB JOURNAL OF INFORMATION MANAGEMENT*. DOI: 10.1108/AJIM-03-2015-0035
82. Madalli, DP; Balaji, BP; Sarangi, AK. Faceted Ontological Representation for a Music Domain. *KNOWLEDGE ORGANIZATION*.
83. Szulc, J. Formal ontologies and their role in the integration of library resources. *QUALITATIVE & QUANTITATIVE METHODS IN LIBRARIES*.
84. Tennis, JT. Foundational, First-Order, and Second-Order Classification Theory. *KNOWLEDGE ORGANIZATION*.
85. Lambe, P. From Cataloguers to Designers: Paul Otlet, Social Impact and a More Proactive Role for Knowledge Organisation Professionals. *KNOWLEDGE ORGANIZATION*.
86. Zare, MN; Azzar, SA; Mardani, AH; Arein, MA. From change to organizational transformation A survey of Tehran Medical Science University's libraries. *LIBRARY MANAGEMENT*. DOI: 10.1108/LM-06-2014-0074
87. De Sabbata, S; Mizzaro, S; Reichenbacher, T. Geographic dimensions of relevance. *JOURNAL OF DOCUMENTATION*. DOI: 10.1108/JD-12-2013-0167
88. Bishop, BW; Moulaison, HL; Burwell, CL. Geographic Knowledge Organization: Critical Cartographic Cataloging and Place-Names in the Geoweb. *KNOWLEDGE ORGANIZATION*.
89. Gu, MS; Hwang, J. Geosemantic information retrieval and its performance evaluation. *JOURNAL OF INFORMATION SCIENCE*. DOI: 10.1177/0165551515586717
90. Ruocco, M; Ramampiaro, H. Geo-temporal distribution of tag terms for event-related image retrieval. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2014.09.001
91. Marteleto, RM; Carvalho, LD. Health as a Knowledge Domain and Social Field: Dialogues with Birger Hjorland and Pierre Bourdieu. *KNOWLEDGE ORGANIZATION*.
92. Yaraghi, N; Du, AY; Sharman, R; Gopal, RD; Ramesh, R. Health Information Exchange as a Multisided Platform: Adoption, Usage, and Practice Involvement in Service Co-Production. *INFORMATION SYSTEMS RESEARCH*. DOI: 10.1287/isre.2014.0547
93. Hoenkamp, E; Bruza, P. How Everyday Language Can and Will Boost Effective Information Retrieval. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23279
94. Brychcin, T; Konopik, M. HPS: High precision stemmer. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2014.08.006
95. Mathiesen, K. Human rights as a topic and guide for LIS research and practice. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23293
96. Yan, EJ; Zhu, Y. Identifying entities from scientific publications: A comparison of vocabulary- and model-based methods. *JOURNAL OF INFORMETRICS*. DOI: 10.1016/j.joi.2015.04.003
97. Hendahewa, C; Shah, C. Implicit search feature based approach to assist users in exploratory search tasks. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2015.06.004
98. Colace, F; De Santo, M; Greco, L; Napolitano, P. Improving relevance feedback-based query expansion by the use of a weighted word pairs approach. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23331
99. Ajuwon, GA; Popoola, SO. Influence of motivational factors on utilisation of Internet health information resources by resident doctors in Nigeria. *ELECTRONIC LIBRARY*. DOI: 10.1108/EL-12-2012-0159
100. Gonzalez-Rodriguez, D; Kostakis, V. Information literacy and peer-to-peer infrastructures: An autopoietic perspective. *TELEMATICS AND INFORMATICS*. DOI: 10.1016/j.tele.2015.01.001
101. Gil, MDC. Information management in public administrations in the context of information retrieval. *IBERSID-REVISTA DE SISTEMAS DE INFORMACION Y DOCUMENTACION*.
102. Raghavan, KS; Apoorva, KH; Jivrajani, A. Information Retrieval as a Domain: Visualizations Based on Two Data Sets. *KNOWLEDGE ORGANIZATION*.
103. Esteveao, JSB; Strauhs, FD. Information retrieval in institutional repositories: Proposal of an ontology for Historical Organizational Memory. *QUALITATIVE & QUANTITATIVE METHODS IN LIBRARIES*.
104. Furner, J. Information Science Is Neither. *LIBRARY TRENDS*. DOI: 10.1353/lib.2015.0009
105. Du, SH; Feng, CC; Guo, L. Integrative representation and inference of qualitative locations about points, lines, and polygons. *INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE*. DOI: 10.1080/13658816.2015.1004333

106. Gregori, L; Losito, L; Siritto, P. Knowledge mapping and visualization as a common ground between librarianship and scholarly communication: Qualitative and quantitative methods for improving semantic categorization and retrieval. *QUALITATIVE & QUANTITATIVE METHODS IN LIBRARIES*.
107. Martinez-Avila, D. Knowledge Organization in the Intersection with Information Technologies. *KNOWLEDGE ORGANIZATION*.
108. Wu, QQ; Zhang, HB; Lan, J. K-State automaton burst detection model based on KOS: Emerging trends in cancer field. *JOURNAL OF INFORMATION SCIENCE*. DOI: 10.1177/0165551514551500
109. Kardara, M; Papadakis, G; Papaikononou, A; Tserpes, K; Varvarigou, T. Large-scale evaluation framework for local influence theories in Twitter. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2014.06.002
110. Ghosh, K; Parui, SK; Majumder, P. Learning combination weights in data fusion using Genetic Algorithms. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2014.12.002
111. Behnert, C; Borst, T. LibRank: New Approaches to Relevance Ranking in Library Information Systems - A Research Project Funded by the German Research Foundation (DFG). *BIBLIOTHEK FORSCHUNG UND PRAXIS*. DOI: 10.1515/bfp-2015-0052
112. Ambrozic, M; Zumer, M. Libraries and the Library System of Slovenia. *LIBRARY TRENDS*. DOI: 10.1353/lib.2015.0021
113. Urbano, C; Zhang, Y; Downey, K; Klingler, T. Library Catalog Log Analysis in E-book Patron-Driven Acquisitions (PDA): A Case Study. *COLLEGE & RESEARCH LIBRARIES*. DOI: 10.5860/crl.76.4.412
114. Contreras, RI. Linguistic storm: an essential information retrieval tool to update researchers. *QUALITATIVE & QUANTITATIVE METHODS IN LIBRARIES*.
115. Huurdeman, HC; Kamps, J; Samar, T; de Vries, AP; Ben-David, A; Rogers, RA. Lost but not forgotten: finding pages on the unarchived web. *INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES*. DOI: 10.1007/s00799-015-0153-3
116. Jia, JZ; Zhao, J. Mapping Analysis of Pre-coordinated Classes in DDC and CLC. *KNOWLEDGE ORGANIZATION*.
117. Gonzalez-Teruel, A; Gonzalez-Alcaide, G; Barrios, M; Abad-Garcia, MF. Mapping recent information behavior research: an analysis of co-authorship and co-citation networks. *SCIENTOMETRICS*. DOI: 10.1007/s11192-015-1548-z
118. Zitt, M. Meso-level retrieval: IR-bibliometrics interplay and hybrid citation-words methods in scientific fields delineation. *SCIENTOMETRICS*. DOI: 10.1007/s11192-014-1482-5
119. Ramesh, P; Vivekavardhan, J; Bharathi, K. Metadata Diversity, Interoperability and Resource Discovery Issues and Challenges. *DESIDOC JOURNAL OF LIBRARY & INFORMATION TECHNOLOGY*. DOI: 10.14429/djlit.35.3.8074
120. Tognoli, NB. Metatheory and Diplomats: building a new model of analysis for the identification of archival records. *SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO*.
121. Pena, GA; Pena, CN. Method for Selecting Specialized Terms from a General Language Corpus. *KNOWLEDGE ORGANIZATION*.
122. Zhitomirsky-Geffet, M; Daya, Y. Mining query subtopics from social tags. *INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL*.
123. Boberic-Krsticev, D; Tesendic, D. Mixed approach in creating a university union catalogue. *ELECTRONIC LIBRARY*. DOI: 10.1108/EL-02-2014-0026
124. Karlsson, A; Hammarfelt, B; Steinhauer, HJ; Falkman, G; Olson, N; Nelhans, G; Nolin, J. Modeling uncertainty in bibliometrics and information retrieval: an information fusion approach. *SCIENTOMETRICS*. DOI: 10.1007/s11192-014-1481-6
125. Akhigbe, BI; Afolabi, BS; Adagunodo, ER. Modelling User-Centered Attributes: The Web Search Engine as a Case. *KNOWLEDGE ORGANIZATION*.
126. Azmi, AM; Aljafari, EA. Modern information retrieval in Arabic - catering to standard and colloquial Arabic users. *JOURNAL OF INFORMATION SCIENCE*. DOI: 10.1177/0165551515585720
127. Gross, T. Naming and Reframing: A Taxonomy of Attacks on Knowledge Organization. *KNOWLEDGE ORGANIZATION*.
128. Sonkar, SK; Bhatnagar, V; Challa, RK. Need of Intelligent Search in Dynamic Social Network. *INFORMATION RESOURCES MANAGEMENT JOURNAL*. DOI: 10.4018/IRMJ.2015040104
129. Sturges, P. Neuroscience advances and their implications for Information Science. *PROFESIONAL DE LA INFORMACION*. DOI: 10.3145/epi.2015.mar.10
130. Shoemaker, E. No One Can Whistle a Symphony: Seeking a Catalogers' Code of Ethics. *KNOWLEDGE ORGANIZATION*.
131. Smalheiser, NR; Shao, WX; Yu, PS. Nuggets: findings shared in multiple clinical case reports. *JOURNAL OF THE MEDICAL LIBRARY ASSOCIATION*. DOI: 10.3163/1536-5050.103.4.002
132. Schulz, M; Winter, P; Choi, SKT. On the relevance of reports-Integrating an automated archiving component into a business intelligence system. *INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT*. DOI: 10.1016/j.ijinfomgt.2015.07.005
133. Oh, KE; Joo, S; Jeong, EJ. Online Consumer Health Information Organization: Users' Perspectives on Faceted Navigation. *KNOWLEDGE ORGANIZATION*.
134. Ozmen, MU. Online information retrieval behaviour and economics of attention. *ONLINE INFORMATION REVIEW*. DOI: 10.1108/OIR-05-2015-0139
135. Gonzalez, MEL. Organization knowledge on the digital network. *INVESTIGACION BIBLIOTECOLOGICA*. DOI: 10.1016/j.ibbai.2016.02.037
136. Liu, JJ; Belkin, NJ. Personalizing Information Retrieval for Multi-session Tasks: Examining the Roles of Task Stage, Task Type, and Topic Knowledge on the Interpretation of Dwell Time as an Indicator of Document Usefulness. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23160
137. de Sales, R; Murguia, EI. Political determinations in the scientific production of Information Science in Brazil: impact of the Classification of Knowledge areas of the CNPq. *SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO*.
138. Bawden, D; Robinson, L; Siddiqui, T. Potentialities or Possibilities: Towards Quantum Information Science?. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23192
139. Zhang, XM; Liu, JJ; Cole, M; Belkin, N. Predicting Users' Domain Knowledge in Information Retrieval Using Multiple Regression Analysis of Search Behaviors. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23218
140. Tan, EM; Goh, DHL. Presenting social media information on mobile devices using multiple contexts. *ASLIB JOURNAL OF INFORMATION MANAGEMENT*. DOI: 10.1108/AJIM-09-2014-0124
141. Vulic, I; De Smet, W; Tang, J; Moens, MF. Probabilistic topic modeling in multilingual settings: An overview of its methodology and applications. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2014.08.003
142. Stokes, P; Urquhart, C. Profiling information behaviour of nursing students: part 2: derivation of profiles. *JOURNAL OF DOCUMENTATION*. DOI: 10.1108/JD-07-2013-0091

143. Guimaraes, JAC; Tognoli, NB. Provenance as a Domain Analysis Approach in Archival Knowledge Organization. KNOWLEDGE ORGANIZATION.
144. Ravana, SD; Rajagopal, P; Balakrishnan, V. Ranking retrieval systems using pseudo relevance judgments. ASLIB JOURNAL OF INFORMATION MANAGEMENT. DOI: 10.1108/AJIM-03-2015-0046
145. Behnert, C; Lewandowski, D. Ranking Search Results in Library Information Systems - Considering Ranking Approaches Adapted From Web Search Engines. JOURNAL OF ACADEMIC LIBRARIANSHIP. DOI: 10.1016/j.acalib.2015.07.010
146. Gilchrist, A. Reflections on Knowledge, Communication and Knowledge Organization. KNOWLEDGE ORGANIZATION.
147. Sharma, RS. Reflections on the role of information retrieval for development: learning from the missed opportunities of the past. INFORMATION DEVELOPMENT. DOI: 10.1177/0266666915571518
148. Felquer, LV; Huertas, MJL; Gimenez, L. Relevance of the retrieval and use of information from the user standpoint of the Agronomy and Veterinary Library. Northeastern National University (Corrientes, Argentina). REVISTA GENERAL DE INFORMACION Y DOCUMENTACION. DOI: 10.5209/rev_RGID.2015.v25.n2.51233
149. Assumpcao, FS; Santos, PLVAD. Representation in the bibliographic domain: a look at MARC 21 Formats. PERSPECTIVAS EM CIENCIA DA INFORMACAO. DOI: 10.1590/1981-5344/2054
150. DaL Evedove, PR; Fujita, MSL. Research on indexing policy in Brazil: progress and challenges. SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO.
151. Dasgupta, S; Pal, P; Mazumdar, C; Bagchi, A. Resolving authorization conflicts by ontology views for controlled access to a digital library. JOURNAL OF KNOWLEDGE MANAGEMENT. DOI: 10.1108/JKM-10-2014-0435
152. Cleverley, PH; Burnett, S. Retrieving haystacks: a data driven information needs model for faceted search. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551514554522
153. Mutschke, P; Mayr, P. Science models for search: a study on combining scholarly information retrieval and scientometrics. SCIENTOMETRICS. DOI: 10.1007/s11192-014-1485-2
154. Lucas, EDO; de Lara, MLG. Scientific production networks in Knowledge Organization: possible author and citation visualization and analysis. SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO.
155. Hussain, A. Search Engines as an Effective Tool for Library Professionals. DESIDOC JOURNAL OF LIBRARY & INFORMATION TECHNOLOGY. DOI: 10.14429/djlit.35.6.8804
156. Berget, G; Sandnes, FE. Searching databases without query-building aids: implications for dyslexic users. INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL.
157. Martinez-Costa, C; Cornet, R; Karlsson, D; Schulz, S; Kalra, D. Semantic enrichment of clinical models towards semantic interoperability. The heart failure summary use case. JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION. DOI: 10.1093/jamia/ocu013
158. Pattuelli, MC; Miller, M. Semantic network edges: a human-machine approach to represent typed relations in social networks. JOURNAL OF KNOWLEDGE MANAGEMENT. DOI: 10.1108/JKM-11-2014-0453
159. Wells, MJ. Social semiotics as theory and practice in library and information science. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-01-2014-0018
160. Jayasinghe, GK; Webber, W; Sanderson, M; Dharmasena, LS; Culpepper, JS. Statistical comparisons of non-deterministic IR systems using two dimensional variance. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2015.06.005
161. Bruyaka, O; James, T; Cook, DF; Barkhi, R. Strategic complementarities in M&As: evidence from the US information retrieval services industry. INFORMATION TECHNOLOGY & MANAGEMENT. DOI: 10.1007/s10799-014-0194-0
162. Luo, ZC; Yu, Y; Osborne, M; Wang, T. Structuring Tweets for improving Twitter search. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23332
163. Rodrigues, MR; Cervantes, BMN. Subject analysis and concept maps: similarities in processes. PERSPECTIVAS EM CIENCIA DA INFORMACAO. DOI: 10.1590/1981-5344/2423
164. Kar, M; Nunes, S; Ribeiro, C. Summarization of changes in dynamic text collections using Latent Dirichlet Allocation model. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2015.06.002
165. Karanikolas, NN. Supervised learning for building stemmers. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551515572528
166. Ram, S. Tag Cloud Application and Information Retrieval System: Visualisation to Create Information Literacy. DESIDOC JOURNAL OF LIBRARY & INFORMATION TECHNOLOGY. DOI: 10.14429/djlit.35.1.8036
167. Shaila, SG; Vadeivel, A. TAG term weight-based N gram Thesaurus generation for query expansion in information retrieval application. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551515581567
168. Joho, H; Jatowt, A; Blanco, R. Temporal information searching behaviour and strategies. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2015.03.006
169. Cleverley, PH; Burnett, S. The Best of Both Worlds: Highlighting the Synergies of Combining Manual and Automatic Knowledge Organization Methods to Improve Information Search and Discovery. KNOWLEDGE ORGANIZATION.
170. Guimaraes, JAC; Matos, DFDO; dos Santos, AY; de Sales, R. The conceptual dimension of knowledge organization in the scientific universe of ISKO: a domain-analysis of ISKO-Brazil, ISKO-Spain, NASKO and ISKO France conferences. SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO.
171. Absar, R; Guastavino, C. The Design and Formative Evaluation of Nonspeech Auditory Feedback for an Information System. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23282
172. Cole, C; Beheshti, J; Abuhimed, D; Lamoureux, I. The end game in Kuhlthau's ISP Model: Knowledge construction for grade 8 students researching an inquiry-based history project. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23300
173. Arboit, AE; Guimaraes, JAC. The Ethics of Knowledge Organization and Representation from a Bakhtinian Perspective. KNOWLEDGE ORGANIZATION.
174. Nhacuongue, JA; Ferneda, E. The field of information science: contributions, challenges and perspectives. PERSPECTIVAS EM CIENCIA DA INFORMACAO. DOI: 10.1590/1981-5344/1932
175. Oravec, JA. The Moral Imagination in an Era of "Gaming Academia": Implications of Emerging Reputational Issues in Scholarly Activities for Knowledge Organization Practices. KNOWLEDGE ORGANIZATION.
176. Ribeiro, CJS; Pereira, DV. The publication of open government data: A proposal for the revision of class in Social Security Regarding the Controlled Vocabulary of the Electronic Government. TRANSINFORMACAO. DOI: 10.1590/0103-37862015000100007
177. Alarcon, MRO. The RDA (Resource Description and Access) and the adaption to change in bibliographic systems in Spain. REVISTA IBERO-AMERICANA DE CIENCIA DA INFORMACAO.

178. Martinez-Avila, D; Guimaraes, JAC; Pinho, FA; Fox, MJ. The Representation of Ethics and Knowledge Organization in the WoS and LISTA Databases. KNOWLEDGE ORGANIZATION.
179. Xie, X; Sonnenwald, DH; Fulton, C. The role of memory in document re-finding. LIBRARY HI TECH. DOI: 10.1108/LHT-06-2014-0050
180. Murray, TE. The Specialist. JOURNAL OF LIBRARY ADMINISTRATION. DOI: 10.1080/01930826.2015.1038933
181. Wolfram, D. The symbiotic relationship between information retrieval and informetrics. SCIENTOMETRICS. DOI: 10.1007/s11192-014-1479-0
182. Adams, B; Janowicz, K. Thematic signatures for cleansing and enriching place-related linked data. INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE. DOI: 10.1080/13658816.2014.989855
183. Hjørland, B. Theories are Knowledge Organizing Systems (KOS). KNOWLEDGE ORGANIZATION.
184. Albrechtsen, H. This is Not Domain Analysis. KNOWLEDGE ORGANIZATION.
185. Lu, K; Shen, B; Ju, D. To seize the emerging historical opportunity of the networked knowledge. LIBRARY MANAGEMENT. DOI: 10.1108/LM-08-2014-0086
186. Amjad, T; Ding, Y; Daud, A; Xu, J; Malic, V. Topic-based heterogeneous rank. SCIENTOMETRICS. DOI: 10.1007/s11192-015-1601-y
187. Weissenberger, L. Toward a universal, meta-theoretical framework for music information classification and retrieval. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-08-2013-0106
188. Solomou, G; Koutsomitropoulos, D. Towards an evaluation of semantic searching in digital repositories: a DSpace case-study. PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS. DOI: 10.1108/PROG-07-2013-0037
189. Shalabi, R; Elfatraty, A. Towards improving XML search by using structure clustering technique. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551514560523
190. Zhang, JS; Sun, YC; Jara, AJ. Towards semantically linked multilingual corpus. INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT. DOI: 10.1016/j.ijinfomgt.2015.01.004
191. Weissenberger, LK. Traditional Musics and Ethical Considerations of Knowledge and Documentation Processes. KNOWLEDGE ORGANIZATION.
192. Onyancha, OB; Ngoepe, M; Maluleka, J. Trends, Patterns, Challenges and Types of Archival Research in Sub-Saharan Africa. AFRICAN JOURNAL OF LIBRARY ARCHIVES AND INFORMATION SCIENCE.
193. Alonso-Lifante, MP; Chain-Navarro, C. Typology of historical astronomical resources and map libraries and archives in Spain: state of the art. REVISTA ESPANOLA DE DOCUMENTACION CIENTIFICA. DOI: 10.3989/redc.2015.3.1207
194. Civilcharran, S; Hughes, M; Maharaj, MS. Uncovering Web search tactics in South African higher education. SOUTH AFRICAN JOURNAL OF INFORMATION MANAGEMENT. DOI: 10.4102/sajim.v17i1.644
195. Arif, ASM; Du, JT; Lee, I. Understanding tourists' collaborative information retrieval behavior to inform design. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23319
196. Soergel, D. Unleashing the Power of Data Through Organization: Structure and Connections for Meaning, Learning and Discovery. KNOWLEDGE ORGANIZATION.
197. Vallez, M; Pedraza-Jimenez, R; Codina, L; Blanco, S; Rovira, C. Updating controlled vocabularies by analysing query. ONLINE INFORMATION REVIEW. DOI: 10.1108/OIR-06-2015-0180
198. Qasim, J; Khan, AM. Use of e-journals by the scientists of CSIR-Institute of Genomics and Integrative Biology (IGIB), Delhi, India A study. ELECTRONIC LIBRARY. DOI: 10.1108/EL-07-2014-0107
199. Gurikar, R; Gurikar, R. Use of Library Resources and Services by Karanatak University Research Scholars: A Case Study. DESIDOC JOURNAL OF LIBRARY & INFORMATION TECHNOLOGY. DOI: 10.14429/djlit.35.5.8275
200. Liu, JJ. User assessment of search task difficulty: Relationships between reasons and ratings. LIBRARY & INFORMATION SCIENCE RESEARCH. DOI: 10.1016/j.lisr.2015.05.003
201. Lee, JH; Clarke, RI; Kim, YS. Video game information needs and game organization: differences by sex and age. INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL.
202. Yang, SL; Wang, FF. Visualizing information science: Author direct citation analysis in China and around the world. JOURNAL OF INFORMETRICS. DOI: 10.1016/j.joi.2015.01.001
203. Alqudsi, TM; Alkhaledi, RM. Website usability: the case of Kuwaiti middle school students. ELECTRONIC LIBRARY. DOI: 10.1108/EL-12-2013-0216
204. Bowen, T; Evans, MM. What does knowledge look like? Drawing as a means of knowledge representation and knowledge construction. EDUCATION FOR INFORMATION. DOI: 10.3233/EFI-150947
205. Beak, J. Where is Childrens' Voice in KO?. KNOWLEDGE ORGANIZATION.
206. Xu, HL; Yuan, H; Ma, BJ; Qian, Y. Where to go and what to play: Towards summarizing popular information from massive tourism blogs. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551515603323
207. Chifu, AG; Hristea, F; Mothe, J; Popescu, M. Word sense discrimination in information retrieval: A spectral clustering-based approach. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2014.10.007

Publicados em 2016

1. Amador, SR; Perez, MD; Perez, MJLH; Penas, DA. A case study of knowledge organization patterns within Curricular Information Systems. INVESTIGACION BIBLIOTECOLOGICA. DOI: 10.1016/j.ibbai.2016.02.005
2. Jett, J; Sacchi, S; Lee, JH; Clarke, RI. A conceptual model for video games and interactive media. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23409
3. Dang, EKF; Luk, RWP; Allan, J. A context-dependent relevance model. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23419
4. Golub, K; Soergel, D; Buchanan, G; Tudhope, D; Lykke, M; Hiom, D. A framework for evaluating automatic indexing or classification in the context of retrieval. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23600
5. Vlachidis, A; Tudhope, D. A knowledge-based approach to Information Extraction for semantic interoperability in the archaeology domain. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23485
6. Ye, Z; Huang, JX. A learning to rank approach for quality-aware pseudo-relevance feedback. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23430

7. Selvaretnam, B; Belkhatir, M. A linguistically driven framework for query expansion via grammatical constituent highlighting and role-based concept weighting. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2015.04.002
8. Marcondes, CH; da Costa, LC. A Model to Represent and Process Scientific Knowledge in Biomedical Articles with Semantic Web Technologies. *KNOWLEDGE ORGANIZATION*.
9. Zhao, BQ; Xu, SH; Lin, SJ; Luo, XN; Duan, L. A new visual navigation system for exploring biomedical Open Educational Resource (OER) videos. *JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION*. DOI: 10.1093/jamia/ocv123
10. Park, Y; Hwang, H; Lee, SG. A novel algorithm for scalable k-nearest neighbour graph construction. *JOURNAL OF INFORMATION SCIENCE*. DOI: 10.1177/0165551515594728
11. Cafe, LMA; de Barros, CM. A PANORAMA OF NATIONAL AND INTERNATIONAL RESEARCH ON MUSIC INFORMATION. *Informacao & Sociedade-Estudos*.
12. Denning, J; Pera, MS; Ng, YK. A readability level prediction tool for K-12 books. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23417
13. Menard, E; Khashman, N; Kochkina, S; Torres-Moreno, JM; Velazquez-Morales, P; Zhou, F; Jourlin, P; Rawat, P; Peinl, P; Pontes, EL; Brunetti, I. A Second Life for TIARA: From Bilingual to Multilingual!. *KNOWLEDGE ORGANIZATION*.
14. Paredes-Valverde, MA; Valencia-Garcia, R; Rodriguez-Garcia, MA; Colomo-Palacios, R; Alor-Hernandez, G. A semantic-based approach for querying linked data using natural language. *JOURNAL OF INFORMATION SCIENCE*. DOI: 10.1177/0165551515616311
15. Shaw, R; Rabinowitz, A; Golden, P; Kansa, E. A sharing-oriented (d)esign strategy for networked knowledge organization systems. *INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES*. DOI: 10.1007/s00799-015-0164-0
16. Borlund, P. A study of the use of simulated work task situations in interactive information retrieval evaluations A meta-evaluation. *JOURNAL OF DOCUMENTATION*. DOI: 10.1108/JD-06-2015-0068
17. Zhou, D; Lawless, S; Wu, X; Zhao, WY; Liu, JX. A study of user profile representation for personalized cross-language information retrieval. *ASLIB JOURNAL OF INFORMATION MANAGEMENT*. DOI: 10.1108/AJIM-06-2015-0091
18. Chen, SJ; Zeng, ML; Chen, HH. Alignment of conceptual structures in controlled vocabularies in the domain of Chinese art: a discussion of issues and patterns. *INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES*. DOI: 10.1007/s00799-015-0163-1
19. Packalen, S; Henttonen, P. Ambiguous Labels: Facet Analysis of Class Names in Finnish Public-Sector Functional Classification Systems. *KNOWLEDGE ORGANIZATION*.
20. Kilinc, D. An accurate toponym-matching measure based on approximate string matching. *JOURNAL OF INFORMATION SCIENCE*. DOI: 10.1177/0165551515590097
21. Netto, CM; Lima, GABD; Pierozzi, I. An Application of Facet Analysis Theory and Concept Maps for Faceted Search in a Domain Ontology: Preliminary Studies. *KNOWLEDGE ORGANIZATION*.
22. Gaona-Garcia, PA; Stoitsis, G; Sanchez-Alonso, S; Biniari, K. An Exploratory Study of User Perception in Visual Search Interfaces Based on SKOS. *KNOWLEDGE ORGANIZATION*.
23. Godert, W. An ontology-based model for indexing and retrieval. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23420
24. Felici, JM; Jimenez, AG; Humanes, ML. Analysis and reformulation of the knowledge organization of Communication field: application for UNESCO code. *REVISTA GENERAL DE INFORMACION Y DOCUMENTACION*. DOI: 10.5209/rev_RGID.2016.v26.n1.53049
25. de Almeida, CC; Farias, MCQD. Analysis of semiotic theories in Brazilian information science: theoretical referents. *SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO*.
26. Kobrinskii, BA. Approaches to the Construction of Cognitive Linguistic-Image Models of Knowledge Representation for Medical Intelligent Systems. *SCIENTIFIC AND TECHNICAL INFORMATION PROCESSING*. DOI: 10.3103/S0147688216050026
27. Reichenbacher, T; De Sabbata, S; Purves, RS; Fabrikant, SI. Assessing Geographic Relevance for Mobile Search: A Computational Model and Its Validation via Crowdsourcing. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23625
28. Flores, FN; Moreira, VP. Assessing the impact of Stemming Accuracy on Information Retrieval - A multilingual perspective. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2016.03.004
29. Xu, B; Lin, HF; Lin, Y. Assessment of Learning to Rank Methods for Query Expansion. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23476
30. Kanan, T; Fox, EA. Automated Arabic Text Classification With P-Stemmer, Machine Learning, and a Tailored News Article Taxonomy. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23609
31. Walters, WH. Beyond use statistics: Recall, precision, and relevance in the assessment and management of academic libraries. *JOURNAL OF LIBRARIANSHIP AND INFORMATION SCIENCE*. DOI: 10.1177/0961000615572174
32. Mostafa, SP; Segundo, JES; Sabbag, DMA. BIBLIOGRAPHIC DESCRIPTION IN SEMANTIC WEB ERA: for a new document notion. *Informacao & Sociedade-Estudos*.
33. Arcila-Calderon, C; Barbosa-Caro, E; Cabezuelo-Lorenzo, F. Big data techniques: Large-scale text analysis for scientific and journalistic research. *PROFESIONAL DE LA INFORMACION*. DOI: 10.3145/epi.2016.jul.12
34. Ridenour, L. Boundary Objects: Measuring Gaps and Overlap Between Research Areas. *KNOWLEDGE ORGANIZATION*.
35. Hughes, E; Marshall, A. Can Twitter improve your health? An analysis of alcohol consumption guidelines on Twitter. *HEALTH INFORMATION AND LIBRARIES JOURNAL*. DOI: 10.1111/hir.12133
36. Choi, Y; Syn, SY. Characteristics of tagging behavior in digitized humanities online collections. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23472
37. Meireles, MRG; Ferraro, G; Geva, S. Classification and information management for patent collections: a literature review and some research questions. *INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL*.
38. Bates, J; Fodeh, SJ; Brandt, CA; Womack, JA. Classification of radiology reports for falls in an HIV study cohort. *JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION*. DOI: 10.1093/jamia/ocv155
39. Gnoli, C. Classifying Phenomena Part 1: Dimensions. *KNOWLEDGE ORGANIZATION*.
40. Sousa, CD; Soares, AL; Pereira, CS. Collaborative conceptualisation processes in the development of lightweight ontologies. *VINE-JOURNAL OF INFORMATION AND KNOWLEDGE MANAGEMENT*. DOI: 10.1108/VJKMS-03-2015-0022
41. Madalli, D; Sulochana, A; Singh, AK. COMAT: core ontology of matter. *PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS*. DOI: 10.1108/PROG-01-2015-0002

42. Lund, H; Ornager, S. Company taxonomy development: the case of an international emergency response organization. *ASLIB JOURNAL OF INFORMATION MANAGEMENT*. DOI: 10.1108/AJIM-09-2015-0140
43. Lu, C; Zhang, CZ; He, DQ. Comparative analysis of book tags: a cross-lingual perspective. *ELECTRONIC LIBRARY*. DOI: 10.1108/EL-03-2015-0042
44. Meireles, MRG; Cendon, BV; de Almeida, PEM. Comparison of document categorization process using key words and citations in a restricted knowledge domain. *TRANSINFORMACAO*. DOI: 10.1590/231808892016002800007
45. Savolainen, R. Conceptual growth in integrated models for information behaviour. *JOURNAL OF DOCUMENTATION*. DOI: 10.1108/JD-09-2015-0114
46. Chou, S; Dai, ZT. Construction and application of specialty-term information for document re-ranking. *ONLINE INFORMATION REVIEW*. DOI: 10.1108/OIR-04-2015-0128
47. Ko, YM; Song, MS; Lee, SJ. Construction of the structural definition-based terminology ontology system and semantic search evaluation. *LIBRARY HI TECH*. DOI: 10.1108/LHT-08-2016-0090
48. Doslu, M; Bingol, HO. Context sensitive article ranking with citation context analysis. *SCIENTOMETRICS*. DOI: 10.1007/s11192-016-1982-6
49. Perez, MD; Pena, DA; Font, RJR; Carrillo-Calvet, HA. Curriculum information management system and institutional knowledge. *REVISTA GENERAL DE INFORMACION Y DOCUMENTACION*. DOI: 10.5209/rev_RGID.2016.v26.n1.53053
50. da Silva, MB; Neves, DAD. Database prototyping: the use of theory of faceted classification in data modeling. *REVISTA IBERO-AMERICANA DE CIENCIA DA INFORMACAO*.
51. Arslan, A. DeASCIification approach to handle diacritics in Turkish information retrieval. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2015.08.004
52. Mu, TT; Goulermas, JY; Korkontzelos, I; Ananiadou, S. Descriptive document clustering via discriminant learning in a co-embedded space of multilevel similarities. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23374
53. Blazquez-Ochando, M. Design of an aggregator for managing informative big data. *PROFESIONAL DE LA INFORMACION*. DOI: 10.3145/epi.2016.jul.17
54. Enrique, LEP; Suarez, LLCS. Design of information architecture of the web site of the Faculty of Industrial Engineering and Tourism of the Universidad Central "Marta Abreu" de Las Villas (Cuba). *CUADERNOS DE DOCUMENTACION MULTIMEDIA*. DOI: 10.5209/CDMU.53363
55. AlNoamany, Y; Weigle, MC; Nelson, ML. Detecting off-topic pages within TimeMaps in Web archives. *INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES*. DOI: 10.1007/s00799-016-0183-5
56. Diriba, C; Meshesha, M; Tesfaye, D. Developing a Knowledge-Based System for Diagnosis and Treatment of Malaria. *JOURNAL OF INFORMATION & KNOWLEDGE MANAGEMENT*. DOI: 10.1142/S0219649216500362
57. Moraes, JBE; Lima, LM; Caprioli, MS. Discourse analysis and Information Science: theoretical contributions to information organization and representation. *SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO*.
58. Lian, TH; Yu, CH; Wang, W; Yuan, QJ; Hou, ZP. Doctoral Dissertations on Tourism in China: A Co-Word Analysis. *KNOWLEDGE ORGANIZATION*.
59. Hjørland, B. Does the Traditional Thesaurus Have a Place in Modern Information Retrieval?. *KNOWLEDGE ORGANIZATION*.
60. Neto, MRA; Lima, MHTD. Domain of knowledge organization represented in the BRAPCI base: a statistical analysis. *INVESTIGACION BIBLIOTECOLOGICA*. DOI: 10.1016/j.ibbai.2016.10.005
61. Ahmad, P; Brogan, M. E-book user behaviour in academic libraries: The role of user agents in perception and satisfaction. *MALAYSIAN JOURNAL OF LIBRARY & INFORMATION SCIENCE*. DOI: 10.22452/mjlis.vol21no3.6
62. Abainia, K; Ouamour, S; Sayoud, H. Effective language identification of forum texts based on statistical approaches. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2015.12.003
63. Rivera, IM; Romero, L. EFFECTIVENESS OF CONCEPT MAPPING IN DEVELOPING INFORMATION SEARCH AND EVALUATION SKILLS. *ANALES DE DOCUMENTACION*. DOI: 10.6018/analesdoc.19.1.231871
64. Fattahi, R; Parirokh, M; Dayyani, MH; Khosravi, A; Zareivenovel, M. Effectiveness of Google keyword suggestion on users' relevance judgment A mixed method approach to query expansion. *ELECTRONIC LIBRARY*. DOI: 10.1108/EL-03-2015-0035
65. Halpern, Y; Horng, S; Choi, Y; Sontag, D. Electronic medical record phenotyping using the anchor and learn framework. *JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION*. DOI: 10.1093/jamia/ocw011
66. Yalcinkaya, Y. Electronic Word: Information Communicative Approach to Technologization of the Word. *TURKISH LIBRARIANSHIP*.
67. Smiraglia, RP. Empirical Methods for Knowledge Evolution across Knowledge Organization Systems. *KNOWLEDGE ORGANIZATION*.
68. Chuang, HM; Chang, CH; Kao, TY; Cheng, CT; Huang, YY; Cheong, KP. Enabling maps/location searches on mobile devices: constructing a POI database via focused crawling and information extraction. *INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE*. DOI: 10.1080/13658816.2015.1133820
69. Lhadj, LS; Boughanem, M; Amrouche, K. Enhancing information retrieval through concept-based language modeling and semantic smoothing. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23553
70. Garcia-Marco, FJ. Enhancing the Visibility and Relevance of Thesauri in the Web: Searching for a Hub in the Linked Data Environment. *KNOWLEDGE ORGANIZATION*.
71. Niazi, MG; Kamran, MKA. Evaluating Iranian state university websites using WebQEM. *ELECTRONIC LIBRARY*. DOI: 10.1108/EL-07-2014-0112
72. Goker, A; Butterworth, R; MacFarlane, A; Ahmed, TS; Stumpf, S. Expeditions through image jungles The commercial use of image libraries in an online environment. *JOURNAL OF DOCUMENTATION*. DOI: 10.1108/JD-01-2014-0019
73. Rahimi, R; Shakery, A; King, I. Extracting translations from comparable corpora for Cross-Language Information Retrieval using the language modeling framework. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2015.08.001
74. Das, S; Roy, S. Faceted Ontological Model for Brain Tumour Study. *KNOWLEDGE ORGANIZATION*.
75. Daoud, CM; de Moura, ES; Carvalho, A; da Silva, AS; Fernandes, D; Rossi, C. Fast top-k preserving query processing using two-tier indexes. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2016.03.005

76. Chatterjee, U; Kumar, V; Madalli, DP. Formalizing Food Ingredients for Data Analysis and Knowledge Organization. COLLNET JOURNAL OF SCIENTOMETRICS AND INFORMATION MANAGEMENT. DOI: 10.1080/09737766.2016.1213970
77. Burscher, B; Vliegthart, R; de Vreese, CH. Frames Beyond Words: Applying Cluster and Sentiment Analysis to News Coverage of the Nuclear Power Issue. SOCIAL SCIENCE COMPUTER REVIEW. DOI: 10.1177/0894439315596385
78. Chao, Z. From query analysis to user information needs: a study of campus map searches. LIBRARY HI TECH. DOI: 10.1108/LHT-12-2014-0110
79. de Souza, EG; Bezerra, DA. Functional Requirements for Bibliographic Records in the context of the Semantic Web: The contribution of Paul Otlet. TRANSINFORMACAO. DOI: 10.1590/2318-08892016000200002
80. Colvin, E; Kraft, DH. Fuzzy Retrieval for Software Reuse. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23584
81. Svenonius, E. Gems from our Digitization Project. KNOWLEDGE ORGANIZATION.
82. Momtazi, S; Lindenberg, F. Generating query suggestions by exploiting latent semantics in query logs. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551515594723
83. Jensen, S; Liu, XZ; Yu, YY; Milojevic, S. Generation of topic evolution trees from heterogeneous bibliographic networks. JOURNAL OF INFORMETRICS. DOI: 10.1016/j.joi.2016.04.002
84. Campos, R; Dias, G; Jorge, A; Nunes, C. GTE-Rank: A time-aware search engine to answer time-sensitive queries. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2015.07.006
85. Wiley, LK; Tarczy-Hornoch, P; Denny, JC; Freimuth, RR; Overby, CL; Shah, N; Martin, RD; Sarkar, IN. Harnessing next-generation informatics for personalizing medicine: a report from AMIA's 2014 Health Policy Invitational Meeting. JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION. DOI: 10.1093/jamia/ocv111
86. Chowdhury, GG. How to Improve the Sustainability of Digital Libraries and Information Services?. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23599
87. Desrochers, N; Laplante, A; Martin, K; Quan-Haase, A; Spiteri, L. Illusions of a "Bond": tagging cultural products across online platforms. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-09-2015-0110
88. Capkin, C. Impact of Metadata on Full-text Information Retrieval Performance: An Experimental Research on a Small Scale Turkish Corpus. TURKISH LIBRARIANSHIP.
89. Ghosh, K; Chakraborty, A; Parui, SK; Majumder, P. Improving Information Retrieval Performance on OCR'd Text in the Absence of Clean Text Ground Truth. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2016.03.006
90. Binding, C; Tudhope, D. Improving interoperability using vocabulary linked data. INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES. DOI: 10.1007/s00799-015-0166-y
91. Amini, I; Martinez, D; Li, XD; Sanderson, M. Improving patient record search: A meta-data based approach. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2015.07.005
92. Renteria-Agualimpia, W; Lopez-Pellicer, FJ; Lacasta, J; Zarazaga-Soria, FJ; Muro-Medrano, PR. Improving the geospatial consistency of digital libraries metadata. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551515597364
93. Nemati, S; Naghsh-Nilchi, AR. Incorporating social media comments in affective video retrieval. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551515593689
94. Chebil, W; Soualmia, LF; Omri, MN; Darmoni, SJ. Indexing biomedical documents with a possibilistic network. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23435
95. Wang, YS; Lee, JS; Choi, IC. Indexing by Latent Dirichlet Allocation and an Ensemble Model. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23444
96. Lopes Fujita, MS; dos Santos, LBP. Indexing policy in University libraries: Diagnostic and analytical study with participatory research. TRANSINFORMACAO. DOI: 10.1590/231808892016002800005
97. Albertson, D; Johnston, MP. Individual differences of science teachers and interactive video retrieval. ELECTRONIC LIBRARY. DOI: 10.1108/EL-11-2014-0198
98. Ibarra, R; Avila, E. Inducing Academic Databases Direct User over Popular Search Engines. QUALITATIVE & QUANTITATIVE METHODS IN LIBRARIES.
99. Bellot, P; Moriceau, V; Mothe, J; SanJuan, E; Tannier, X. INEX Tweet Contextualization task: Evaluation, results and lesson learned. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2016.03.002
100. Sales, OMM; Pinto, VB; de Sousa, MRF. Information architecture: study and analysis of data Public Medical PDF (Portugues base (PubMed)). BIBLIOS-REVISTA DE BIBLIOTECOLOGIA Y CIENCIAS DE LA INFORMACION.
101. Huang, RH; Li, BY; Zhou, LH. Information literacy instruction in Chinese universities: MOOCs versus the traditional approach. LIBRARY HI TECH. DOI: 10.1108/LHT-02-2016-0013
102. Jarvelin, A; Keskestalo, H; Sormunen, E; Saastamoinen, M; Kettunen, K. Information retrieval from historical newspaper collections in highly inflectional languages: A query expansion approach. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23379
103. Savolainen, R. Information seeking and searching strategies as plans and patterns of action A conceptual analysis. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-03-2016-0033
104. de Andrade, J; de Lara, MLG. Interoperability and Mapping Between Knowledge Organization Systems: Metathesaurus-Unified Medical Language System of the National Library of Medicine. KNOWLEDGE ORGANIZATION.
105. Raieli, R. Introducing Multimedia Information Retrieval to libraries. JLIST. DOI: 10.4403/jlist.11530
106. Athukorala, K; Glowacka, D; Jacucci, G; Oulasvirta, A; Vreeken, J. Is Exploratory Search Different? A Comparison of Information Search Behavior for Exploratory and Lookup Tasks. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23617
107. Leydesdorff, L; Heimeriks, G; Rotolo, D. Journal portfolio analysis for countries, cities, and organizations: Maps and comparisons. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23551
108. Force, DC; Zhang, J. Knowledge discovery from within An examination of records management and electronic records management syllabi. RECORDS MANAGEMENT JOURNAL. DOI: 10.1108/RMJ-11-2015-0034
109. Scharnhorst, A; Smiraglia, RP; Gueret, C; Salah, AAA. Knowledge Maps of the UDC: Uses and Use Cases. KNOWLEDGE ORGANIZATION.
110. MacFarlane, A. Knowledge Organisation and its Role in Multimedia Information Retrieval. KNOWLEDGE ORGANIZATION.
111. Jimenez, AG. Knowledge organization for journalism documentation: state of the art and prospective. SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO.

112. Pando, DA; de Almeida, CC. Knowledge Organization in the Context of Postmodernity from the Theory of Classification Perspective. KNOWLEDGE ORGANIZATION.
113. Alajmi, B; Rehman, SU. Knowledge organization trends in library and information education: Assessment and analysis. EDUCATION FOR INFORMATION. DOI: 10.3233/EFI-160084
114. Santos, PM; Rover, AJ. Knowledge representation through ontologies: an application in the electronic democracy field. PERSPECTIVAS EM CIENCIA DA INFORMACAO. DOI: 10.1590/1981-5344/2523
115. Sabbar, C; Xie, I. Language in the information-seeking context A study of US scholars using non-English sources. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-07-2014-0094
116. Blank, I; Rokach, L; Shani, G. Leveraging metadata to recommend keywords for academic papers. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23571
117. de Sales, R. Library classification and archive classification: differences and similarities in knowledge organization. SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO.
118. Suki, NM; Suki, NM. Library patrons' emotions after information retrieval: effects of perceived self-efficacy. PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS. DOI: 10.1108/PROG-07-2014-0045
119. Hidalgo-Delgado, Y; Senso, JA; Leiva-Mederos, A; Hipola, P. Management of archival materials with Linked Data and federated queries. REVISTA ESPANOLA DE DOCUMENTACION CIENTIFICA. DOI: 10.3989/redc.2016.3.1299
120. Suominen, A; Toivanen, H. Map of Science with Topic Modeling: Comparison of Unsupervised Learning and Human-Assigned Subject Classification. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23596
121. Mai, JE. Marginalization and Exclusion: Unraveling Systemic Bias in Classification. KNOWLEDGE ORGANIZATION.
122. Martinez-Avila, D; Semidao, R; Ferreira, M. Methodological Aspects of Critical Theories in Knowledge Organization. KNOWLEDGE ORGANIZATION.
123. Tennis, JT. Methodological Challenges in Scheme Versioning and Subject Ontogeny Research. KNOWLEDGE ORGANIZATION.
124. Reyes, AAM; Pena, CN. Methods and trends of biomedical and genomic information retrieval based on semantic relations of thesauri and MeSH. INVESTIGACION BIBLIOTECOLOGICA. DOI: 10.1016/j.ibbai.2016.02.006
125. Martinez-Avila, D; Beak, J. Methods, Theoretical Frameworks and Hope for Knowledge Organization. KNOWLEDGE ORGANIZATION.
126. Soulier, L; Tamine, L; Shah, C. MineRank: Leveraging users' latent roles for unsupervised collaborative information retrieval. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2016.05.002
127. Balakrishnan, V; Mehmood, Y; Nagappan, Y. Moving beyond text highlights: inferring users' interests to improve the relevance of retrieval. INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL.
128. Nzomo, P; Ajiferuke, I; Vaughan, L; McKenzie, P. Multilingual Information Retrieval & Use: Perceptions and Practices Amongst Bi/Multilingual Academic Users. JOURNAL OF ACADEMIC LIBRARIANSHIP. DOI: 10.1016/j.acalib.2016.06.012
129. Sharma, A. Nanotechnology Ontology: Semantic Access to Information in the Nano World. DESIDOC JOURNAL OF LIBRARY & INFORMATION TECHNOLOGY. DOI: 10.14429/djlit.36.1.9198
130. Aleahmad, A; Karisani, P; Rahgozar, M; Oroumchian, F. OLFinder: Finding opinion leaders in online social networks. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551515605217
131. Iandoli, L; Quinto, I; De Liddo, A; Shum, SB. On online collaboration and construction of shared knowledge: Assessing mediation capability in computer supported argument visualization tools. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23481
132. Alexiev, V; Isaac, A; Lindenthal, J. On the composition of ISO 25964 hierarchical relations (BTG, BTP, BTI). INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES. DOI: 10.1007/s00799-015-0162-2
133. Atanassova, I; Bertin, M; Lariviere, V. On the composition of scientific abstracts. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-09-2015-0111
134. Clemmensen, ML; Borlund, P. Order effect in interactive information retrieval evaluation: an empirical study. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-04-2015-0051
135. Clarke, SGD. Origins and Trajectory of the Long Thesaurus Debate. KNOWLEDGE ORGANIZATION.
136. Silveira, NC; Saldanha, GS. Own name in Knowledge Organization Epistemology: A Philosophical-Theoretical Debate. KNOWLEDGE ORGANIZATION.
137. Albertson, D; Ju, B. Perceived self-efficacy and interactive video retrieval. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-09-2015-0109
138. Bhattacharya, S; Yang, C; Srinivasan, P; Boynton, B. Perceptions of presidential candidates' personalities in twitter. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23377
139. El Sherif, R; Pluye, P; Gore, G; Granikov, V; Hong, QN. Performance of a mixed filter to identify relevant studies for mixed studies reviews. JOURNAL OF THE MEDICAL LIBRARY ASSOCIATION. DOI: 10.3163/1536-5050.104.1.007
140. Chakrabarty, A; Roy, S. Personalised Information Recommender Using Framework for Ontology Alignment Among Digital Libraries. DESIDOC JOURNAL OF LIBRARY & INFORMATION TECHNOLOGY. DOI: 10.14429/djlit.36.4.9600
141. Ridi, R. Phenomena or Noumena? Objective and Subjective Aspects in Knowledge Organization. KNOWLEDGE ORGANIZATION.
142. Fox, MJ. Priorities of Arrangement or a "Hierarchy of Oppressions?": Perspectives on Intersectionality in Knowledge Organization. KNOWLEDGE ORGANIZATION.
143. Zhang, PY; Soergel, D. Process patterns and conceptual changes in knowledge representations during information seeking and sensemaking: A qualitative user study. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551515615834
144. Peres, M; Miranda, A; Simeao, E. Promotion of information competencies: training for scientific research in higher education and communities. REVISTA IBERO-AMERICANA DE CIENCIA DA INFORMACAO.
145. Saastamoinen, M; Jarvelin, K. Queries in authentic work tasks: the effects of task type and complexity. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-09-2015-0119
146. Bodoff, D; Raban, D. Question types and intermediary elicitation. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23388
147. Shah, C; Hendaheva, C; Gonzalez-Ibanez, R. Rain or Shine? Forecasting Search Process Performance in Exploratory Search Tasks. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23484
148. Mendoza, M; Marin, M; Gil-Costa, V; Ferrarotti, F. Reducing hardware hit by queries in web search engines. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2016.04.008

149. Hider, P; Dalgarno, B; Bennett, S; Liu, YH; Gerts, C; Daws, C; Spiller, B; Mitchell, P; Parkes, R; Macaulay, R. Reindexing a Research Repository from the Ground up: Adding and Evaluating Quality Metadata. AUSTRALIAN ACADEMIC & RESEARCH LIBRARIES. DOI: 10.1080/00048623.2016.1204589
150. Xia, LX; Wang, ZY; Chen, C; Zhai, SS. Research on feature-based opinion mining using topic maps. ELECTRONIC LIBRARY. DOI: 10.1108/EL-11-2014-0197
151. Sheble, L. Research Synthesis Methods and Library and Information Science: Shared Problems, Limited Diffusion. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23499
152. Hudon, M. Research Trajectories in Knowledge Organization OUTRO TÍTULO Indexing Languages and Subject Representation: A Fascinating and Inexhaustible Field of Research. KNOWLEDGE ORGANIZATION.
153. Oh, S; Yang, SW; Pomerantz, JP; Wildemuth, BM; Fox, EA. Results of a digital library curriculum field test. INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES. DOI: 10.1007/s00799-015-0151-5
154. Hjørland, b. Knowledge Organization (KO). KNOWLEDGE ORGANIZATION.
155. Fricke, M. Reviews of Concepts in Knowledge Organization OUTRO TÍTULO Logical Division. KNOWLEDGE ORGANIZATION.
156. Lin, WC; Ke, SW; Tsai, CF. SAFQuery: a simple and flexible advanced Web search interface. ELECTRONIC LIBRARY. DOI: 10.1108/EL-12-2014-0207
157. Theng, YL; Lee, EA; Chu, SKW; Lee, CWY; Chiu, MML; Chan, RCH. Scaffolding in information search: Effects on less experienced searchers. JOURNAL OF LIBRARIANSHIP AND INFORMATION SCIENCE. DOI: 10.1177/0961000615595455
158. Ali, S; Gul, S. Search engine effectiveness using query classification: a study. ONLINE INFORMATION REVIEW. DOI: 10.1108/OIR-07-2015-0243
159. Hsin, CT; Cheng, YH; Tsai, CC. Searching and sourcing online academic literature Comparisons of doctoral students and junior faculty in education. ONLINE INFORMATION REVIEW. DOI: 10.1108/OIR-11-2015-0354
160. Vakkari, P. Searching as learning: A systematization based on literature. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551515615833
161. Zhu, YJ; Yan, EJ. Searching bibliographic data using graphs: A visual graph query interface. JOURNAL OF INFORMETRICS. DOI: 10.1016/j.joi.2016.09.005
162. Atwan, J; Mohd, M; Rashaideh, H; Kanaan, G. Semantically enhanced pseudo relevance feedback for Arabic information retrieval. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551515594722
163. Menard, E; Dorey, J. SINCERITY: A Search Engine for Image Retrieval. CANADIAN JOURNAL OF INFORMATION AND LIBRARY SCIENCE-REVUE CANADIENNE DES SCIENCES DE L INFORMATION ET DE BIBLIOTHECONOMIE.
164. Zhang, J; Zhai, SS; Liu, HX; Stevenson, JA. Social network analysis on a topic-based navigation guidance system in a public health portal. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23468
165. Chung, WY; Zeng, D. Social-Media-Based Public Policy Informatics: Sentiment and Network Analyses of US Immigration and Border Security. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23449
166. Lauruhn, M; Groth, P. Sources of Change for Modern Knowledge Organization Systems. KNOWLEDGE ORGANIZATION.
167. Wiemann, S; Bernard, L. Spatial data fusion in Spatial Data Infrastructures using Linked Data. INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE. DOI: 10.1080/13658816.2015.1084420
168. Suri, VR; Ekbia, HR. Spatial mediations in historical understanding: GIS and epistemic practices of history. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23562
169. Tudhope, D; Binding, C. Still Quite Popular After all Those Years The Continued Relevance of the Information Retrieval Thesaurus. KNOWLEDGE ORGANIZATION.
170. Aljanaki, A; Wiering, F; Veltkamp, RC. Studying emotion induced by music through a crowdsourcing game. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2015.03.004
171. Vilares, J; Alonso, MA; Doval, Y; Vilares, M. Studying the effect and treatment of misspelled queries in Cross-Language Information Retrieval. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2015.12.010
172. Tuamsuk, K; Kaewboonma, N; Chansanam, W; Leopenwong, S. Taxonomy of Folktales from the Greater Mekong Sub-region. KNOWLEDGE ORGANIZATION.
173. Zhitomirsky-Geffet, M; Bar-Ilan, J; Levene, M. Testing the stability of "wisdom of crowds" judgments of search results over time and their similarity with the search engine rankings. ASLIB JOURNAL OF INFORMATION MANAGEMENT. DOI: 10.1108/AJIM-10-2015-0165
174. Nahotko, M. Text genres in information organization. INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL.
175. Chen, J; Wang, D; Lu, Q; Xu, ZY. THC-DAT helps in reading a multi-topic document Results from a user-centered evaluation of a within-document analysis tool. LIBRARY HI TECH. DOI: 10.1108/LHT-07-2016-0081
176. Adler, M. The Case for Taxonomic Reparations. KNOWLEDGE ORGANIZATION.
177. Borovik, MA; Shemberko, LV. The Challenges of Information Retrieval in Social Sciences and Humanities and Ways to Overcome Information Barriers. SCIENTIFIC AND TECHNICAL INFORMATION PROCESSING. DOI: 10.3103/S0147688216020039
178. Vaidya, P; Harinarayana, NS. The Comparative and Analytical Study of LibraryThing Tags with Library of Congress Subject Headings. KNOWLEDGE ORGANIZATION.
179. Madani, F; Weber, C. The evolution of patent mining: Applying bibliometrics analysis and keyword network analysis. WORLD PATENT INFORMATION. DOI: 10.1016/j.wpi.2016.05.008
180. Suaiden, E. The future of information systems in an open and transparent government environment. IBERSID-REVISTA DE SISTEMAS DE INFORMACION Y DOCUMENTACION.
181. Fujita, MSL; dos Santos, LBP. The logical and hierarchical structure of indexing languages used by university libraries. SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO.
182. Pennington, DR. The most passionate cover I've seen: emotional information in fan-created U2 music videos. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-07-2015-0086
183. Leydesdorff, L; Bornmann, L. The operationalization of "fields" as WoS subject categories (WCs) in evaluative bibliometrics: The cases of "library and information science" and "science & technology studies". JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23408

184. Chase, D; Trapasso, E; Tolliver, R. The Perfect Storm: Examining User Experience and Conducting a Usability Test to Investigate a Disruptive Academic Library Web Site Redevelopment. *JOURNAL OF WEB LIBRARIANSHIP*. DOI: 10.1080/19322909.2015.1124740
185. Wolfram, D. The Power to Influence: An Informetric Analysis of the Works of Hope Olson. *KNOWLEDGE ORGANIZATION*.
186. Guedes, RD; Moura, MA. The Principle of Semantic Warrant and the Study of Language: Conceptual Reflections. *KNOWLEDGE ORGANIZATION*.
187. Vernon, NA; Evans, MM; Frissen, I. The relationship between dimensions of personality and Library Anxiety in graduate students. *EDUCATION FOR INFORMATION*. DOI: 10.3233/EFI-160083
188. Mostafa, SP; Sabbag, D. The relationship between power and knowledge in knowledge representation and organization. *SCIRE-REPRESENTACION Y ORGANIZACION DEL CONOCIMIENTO*.
189. Vaidya, P; Harinarayana, NS. The role of social tags in web resource discovery: an evaluation of user-generated keywords. *ANNALS OF LIBRARY AND INFORMATION STUDIES*.
190. Kempf, AO; Neubert, J. The Role of Thesauri in an Open Web: A Case Study of the STW Thesaurus for Economics. *KNOWLEDGE ORGANIZATION*.
191. Wojdyski, BW; Kalyanaraman, S. The three dimensions of website navigability: Explication and effects. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23395
192. Ferro, N; Silvello, G; Keskustalo, H; Pirkola, A; Jarvelin, K. The twist measure for IR evaluation: Taking user's effort into account. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23416
193. Kwasnik, BH. The Web and the Pyramid: Hope Olson's Vision of Connectedness in a World of Hierarchies. *KNOWLEDGE ORGANIZATION*.
194. Guimaraes, JAC; Pinho, FA; Milani, SO. Theoretical Dialogs About Ethical Issues in Knowledge Organization: Garcia Gutierrez, Hudon, Beghtol, and Olson. *KNOWLEDGE ORGANIZATION*.
195. Simoes, MD; de Freitas, MCV; Rodriguez-Bravo, B. Theory of Classification and Classification in Libraries and Archives: Convergences and Divergences. *KNOWLEDGE ORGANIZATION*.
196. Higgins, M. Totally Invisible: Asian American Representation in the Dewey Decimal Classification, 1876-1996. *KNOWLEDGE ORGANIZATION*.
197. Hu, X; Lee, JH. Towards global music digital libraries A cross-cultural comparison on the mood of Chinese music. *JOURNAL OF DOCUMENTATION*. DOI: 10.1108/JD-01-2016-0005
198. Bimba, AT; Idris, N; Al-Hunaiyyan, A; Mahmud, RB; Abdelaziz, A; Khan, S; Chang, V. Towards knowledge modeling and manipulation technologies: A survey. *INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT*. DOI: 10.1016/j.ijinfomgt.2016.05.022
199. Lopatovska, I; Sessions, D. Understanding academic reading in the context of information-seeking. *LIBRARY REVIEW*. DOI: 10.1108/LR-03-2016-0026
200. Zhu, YJ; Yan, EJ; Song, M. Understanding the evolving academic landscape of library and information science through faculty hiring data. *SCIENTOMETRICS*. DOI: 10.1007/s11192-016-2033-z
201. Magdy, W; Elsayed, T. Unsupervised adaptive microblog filtering for broad dynamic topics. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2015.11.004
202. ChanLin, LJ; Hung, WH. Usability and evaluation of a library mobile web site. *ELECTRONIC LIBRARY*. DOI: 10.1108/EL-07-2015-0119
203. Ma, CC; Cao, SJ; Gu, TH. Usability evaluation with tasks characterized by the information search process The China National Knowledge Infrastructure. *ELECTRONIC LIBRARY*. DOI: 10.1108/EL-05-2015-0074
204. Lee, JH; Price, R. User experience with commercial music services: An empirical exploration. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23433
205. Choi, E; Shah, C. User motivations for asking questions in online Q&A services. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23490
206. Lee, JH; Cho, H; Kim, YS. Users' Music Information Needs and Behaviors: Design Implications for Music Information Retrieval Systems. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23471
207. Abdulhayoglu, MA; Thijs, B; Jeuris, W. Using character n-grams to match a list of publications to references in bibliographic databases. *SCIENTOMETRICS*. DOI: 10.1007/s11192-016-2066-3
208. Melucci, M. Utilising a statistical inequality for efficiently finding term sets. *INFORMATION PROCESSING & MANAGEMENT*. DOI: 10.1016/j.ipm.2016.04.011
209. Gomez-Nunez, AJ; Vargas-Quesada, B; Chinchilla-Rodriguez, Z; Batagelj, V; Moya-Anegon, F. Visualization and analysis of SCImago Journal & Country Rank structure via journal clustering. *ASLIB JOURNAL OF INFORMATION MANAGEMENT*. DOI: 10.1108/AJIM-12-2015-0205
210. Yang, SL; Han, RZ; Wolfram, D; Zhao, YH. Visualizing the intellectual structure of information science (2006-2015): Introducing author keyword coupling analysis. *JOURNAL OF INFORMETRICS*. DOI: 10.1016/j.joi.2015.12.003
211. Jachimczyk, A; Chrapek, M; Chrapek, Z. Web directories: selected features and their impact on directory quality. *PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS*. DOI: 10.1108/PROG-08-2014-0060
212. Moulahi, B; Tamine, L; Ben Yahia, S. When time meets information retrieval: Past proposals, current plans and future trends. *JOURNAL OF INFORMATION SCIENCE*. DOI: 10.1177/0165551515607277
213. Gonzalez-Agirre, A; Rigau, G; Agirre, E; Aletras, N; Stevenson, M. Why are these similar? Investigating item similarity types in a large digital library. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*. DOI: 10.1002/asi.23482
214. Suki, NM. Willingness of patrons to use library public computing facilities: insights from Malaysia. *ELECTRONIC LIBRARY*. DOI: 10.1108/EL-01-2015-0008

APÊNDICE 6. Artigos que formam o *corpus* secundário de análise

Publicados em 2011

1. Huvila, I. The complete information literacy? Unforgetting creation and organization of information. JOURNAL OF LIBRARIANSHIP AND INFORMATION SCIENCE. DOI: 10.1177/0961000611418812
2. Szostak, R. Complex Concepts Into Basic Concepts. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.21635
3. Rodrigues, ALC. Conceptual selection in the organization of knowledge domains in human and social sciences: the case of the culture. PERSPECTIVAS EM CIENCIA DA INFORMACAO. DOI: 10.1590/S1413-99362011000200009
4. Vital, LP; Cafe, LMA. Ontology and taxonomy: differences. PERSPECTIVAS EM CIENCIA DA INFORMACAO. DOI: 10.1590/S1413-99362011000200008
5. Soilen, KS; Alipour, PB. Semantics, Classifications and Evidence in a Model for Global Catastrophic Risks. KNOWLEDGE ORGANIZATION.
6. Friedman, A; Thellefsen, M. Concept theory and semiotics in knowledge organization. JOURNAL OF DOCUMENTATION. DOI: 10.1108/00220411111145034
7. Estrada, LMM. Topic Maps from a Knowledge Organization Perspective. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2011-1-43
8. Benson, AC. OntoPhoto and the Role of Ontology in Organizing Knowledge. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2011-2-79
9. Pattuelli, MC. Mapping People-centered Properties for Linked Open Data. KNOWLEDGE ORGANIZATION.
10. Bufrem, LS; Breda, SM. The presence of logic in the domain of knowledge organization: interdisciplinary aspects of college curricula. PERSPECTIVAS EM CIENCIA DA INFORMACAO. DOI: 10.1590/S1413-99362011000100011
11. Hjørland, B. The Importance of Theories of Knowledge: Indexing and Information Retrieval as an Example. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.21451
12. Feinberg, M. How information systems communicate as documents: the concept of authorial voice. JOURNAL OF DOCUMENTATION. DOI: 10.1108/00220411111183573
13. McTavish, JR; Neal, DR; Wathen, CN. Is What You See What You Get? Medical Subject Headings and their Organizing Work in the Violence Against Women Research Literature. KNOWLEDGE ORGANIZATION.
14. Bounhas, I; Elayeb, B; Evrard, F; Slimani, Y. Organizing Contextual Knowledge for Arabic Text Disambiguation and Terminology Extraction. KNOWLEDGE ORGANIZATION.
15. Kaipainen, M; Hautamaki, A. Epistemic Pluralism and Multi-Perspective Knowledge Organization: Explorative Conceptualization of Topical Content Domains. KNOWLEDGE ORGANIZATION.
16. Tognoli, NB; Guimaraes, JAC. The Archival Knowledge Organization: some renewal perspectives from the Canadian scientific approaches. PERSPECTIVAS EM CIENCIA DA INFORMACAO. DOI: 10.1590/S1413-99362011000100003
17. Gutierrez, AG. Desclassification in knowledge organization: a post-epistemological essay. TRANSINFORMACAO. DOI: 10.1590/S0103-37862011000100001

Publicados em 2012

1. Widen, G; Hansen, P. Managing collaborative information sharing: bridging research on information culture and collaborative information behaviour. INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL.
2. Julien, CA; Tirilly, P; Leide, JE; Guastavino, C. Constructing a true LCSH tree of a science and engineering collection. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.22749
3. Pontes, FV; Lima, GABD. Knowledge organization in digital environments: faceted classification theory applied. PERSPECTIVAS EM CIENCIA DA INFORMACAO. DOI: 10.1590/S1413-99362012000400003
4. Buckland, MK. Interrogating Spatial Analogies Relating to Knowledge Organization: Paul Otlet and Others. LIBRARY TRENDS. DOI: 10.1353/lib.2012.0039
5. Hapke, T. Wilhelm Ostwald's Combinatorics as a Link between In-formation and Form. LIBRARY TRENDS. DOI: 10.1353/lib.2012.0041
6. de Barros, CM; Cafe, LMA. Semiotics studies in the Information Science: interdisciplinarity reports. PERSPECTIVAS EM CIENCIA DA INFORMACAO. DOI: 10.1590/S1413-99362012000300003
7. Sanchez-Cuadrado, S; Colmenero-Ruiz, MJ; Moreira, JA. Thesauri: Standards and guidelines. PROFESIONAL DE LA INFORMACION.
8. Mendez, E; Greenberg, J. LINKED DATA FOR OPEN VOCABULARIES AND HIVE'S GLOBAL FRAMEWORK. PROFESIONAL DE LA INFORMACION.
9. La Barre, KA; Tilley, CL. The Elusive Tale: Leveraging the Study of Information Seeking and Knowledge Organization to Improve Access to and Discovery of Folktales. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.21710
10. Noruzi, A. FRBR and Tillet's Taxonomy of Bibliographic Relationships. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2012-6-409
11. Pinho, FA; Guimaraes, JAC. Male Homosexuality in Brazilian Indexing Languages: Some Ethical Questions. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2012-5-363
12. Fox, MJ; Reece, A. Which Ethics? Whose Morality?: An Analysis of Ethical Standards for Information Organization. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2012-5-377
13. Tennis, JT. A Convenient Verisimilitude or Oppressive Internalization?: Characterizing the Ethical Augments Surrounding Hierarchical Structures in Knowledge Organization Systems. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2012-5-394

14. da Silva, AKA; Barbosa, RR; Duarte, EN. SOCIAL NETWORK OF CO-AUTHORING IN INFORMATION SCIENCE: study on the thematic area "Knowledge Organization And Representation". INFORMACAO & SOCIEDADE-ESTUDOS.
15. de Souza, MB; Cafe, L. ONTOLOGY: approaches in the theses and dissertations from the Brazilian public universities.. INFORMACAO & SOCIEDADE-ESTUDOS.
16. Zschocke, T. Resolving controlled vocabulary in DITA markup: a case example in agroforestry. PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS. DOI: 10.1108/00330331211244869
17. Couzinet, V. Knowledge Organization in Information and Communication Sciences, a French Exception?. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2012-4-259
18. Gnoli, C. Metadata About What? Distinguishing Between Ontic, Epistemic, and Documental Dimensions in Knowledge Organization. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2012-4-268
19. Doria, OD. The Role of Activities Awareness in Faceted Classification Development. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2012-4-283
20. Ragaller, I; Rafferty, P. Biases in the classification of Welsh art material Dispersion, dilettantism and depreciation. ASLIB PROCEEDINGS. DOI: 10.1108/00012531211244563
21. Marijuan, PC; del Moral, R; Navarro, J. Scientomics: An Emergent Perspective in Knowledge Organization. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2012-3-153
22. Souza, RR; Tudhope, D; Almeida, MB. Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2012-3-179
23. Kim, JA. Understanding Knowledge Representation in the Knowledge Management Environment: Evaluation of Ontology Visualization Methods. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2012-3-193
24. de Almeida, CC. The Methodological Influence of Peirce's Pragmatism on Knowledge Organization. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2012-3-204
25. Freitas, JL; Gabriel, RF; Buftem, LS. Theoretical Approximations Between Brazilian and Spanish Authors' Production in the Field of Knowledge Organization in the Production of Journals on Information Science in Brazil. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2012-3-216
26. Hjørland, B. Is classification necessary after Google?. JOURNAL OF DOCUMENTATION. DOI: 10.1108/00220411211225557
27. Lee, HL. Epistemic foundation of bibliographic classification in early China A Ru classicist perspective. JOURNAL OF DOCUMENTATION. DOI: 10.1108/00220411211225593
28. Chen, SJ; Chen, HH. Mapping multilingual lexical semantics for knowledge organization systems. ELECTRONIC LIBRARY. DOI: 10.1108/02640471211221386
29. Fujita, MSL; Lacruz, MDA; Diaz, RG. The current state of indexing in library tasks. PERSPECTIVAS EM CIENCIA DA INFORMACAO. DOI: 10.1590/S1413-99362012000100006

Publicados em 2013

1. Pisanski, J; Pisanski, T; Zumer, M. A new approach to complex web site organization. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551513482270
2. Hjørland, B. Citation analysis: A social and dynamic approach to knowledge organization. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2013.07.001
3. Yoo, D; Choi, K; Suh, Y; Kim, G. Building and evaluating a collaboratively built structured folksonomy. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551513480309
4. Hajibayova, L. Basic-level categories: A review. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551513481443
5. Gilchrist, A; Zeng, ML; Clarke, SD; Isaac, A; Lambe, P; Vernau, J. Logic and the Organization of Information - an appreciation of the book of this title by Martin Fricke. A set of short essays. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551513480310
6. Brunt, RM. Information Management of British Military Intelligence: The Work of the Documentalists, 1909-1945. LIBRARY TRENDS. DOI: 10.1353/lib.2013.0048
7. Laqua, D. Intellectual Exchange and the New Information Order of the Interwar Years: The British Society for International Bibliography, 1927-1937. LIBRARY TRENDS. DOI: 10.1353/lib.2013.0047
8. Sha, XQ; Chang, KTT; Zhang, C; Zhang, CH. Knowledge popularity in a heterogeneous network: Exploiting the contextual effects of document popularity in knowledge management systems. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.22879
9. Thellefsen, T; Thellefsen, M; Sorensen, B. Emotion, information, and cognition, and some possible consequences for library and information science. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.22858
10. Rodriguez-Barcenas, G; Lopez-Huertas, MJ. Saaty's analytic hierarchies method for knowledge organization in decision making. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.22823
11. Nguyen, SH; Chowdhury, G. Interpreting the knowledge map of digital library research (1990-2010). JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.22830
12. de Souza, CG; Ferreira, MLA. Researchers profile, co-authorship pattern and knowledge organization in information science in Brazil. SCIENTOMETRICS. DOI: 10.1007/s11192-012-0882-7
13. Drabinski, E. Queering the Catalog: Queer Theory and the Politics of Correction. LIBRARY QUARTERLY. DOI: 10.1086/669547
14. Hjørland, B. Facet analysis: The logical approach to knowledge organization. INFORMATION PROCESSING & MANAGEMENT. DOI: 10.1016/j.ipm.2012.10.001
15. Zong, QJ; Shen, HZ; Yuan, QJ; Hu, XW; Hou, ZP; Deng, SG. Doctoral dissertations of Library and Information Science in China: A co-word analysis. SCIENTOMETRICS. DOI: 10.1007/s11192-012-0799-1
16. Donahue, T. Animated Subject Maps for Book Collections. INFORMATION TECHNOLOGY AND LIBRARIES. DOI: 10.6017/ital.v32i2.2892

17. Hansson, J. The Materiality of Knowledge Organization: Epistemology, Metaphors and Society. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-6-384
18. Marras, C. Structuring Multidisciplinary Knowledge: Aquatic and Terrestrial Metaphors. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-6-392
19. Lopez-Huertas, M. Reflexions on Multidimensional Knowledge: Its Influence on the Foundation of Knowledge Organization. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-6-400
20. Pattuelli, C; Rubinow, S. The knowledge organization of DBpedia: a case study. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-07-2012-0084
21. Herre, H. Formal Ontology and the Foundation of Knowledge Organization. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-5-332
22. Kleineberg, M. The Blind Men and the Elephant: Towards an Organization of Epistemic Contexts. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-5-340
23. Thellefsen, M; Thellefsen, T; Sorensen, B. A Pragmatic Semeiotic Perspective on the Concept of Information Need and Its Relevance for Knowledge Organization. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-4-213
24. de Almeida, CC; Fujita, MSL; dos Reis, DM. Peircean Semiotics and Subject Indexing: Contributions of Speculative Grammar and Pure Logic. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-4-225
25. Mai, JE. Ethics, Values and Morality in Contemporary Library Classifications. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-4-242
26. Marchese, C; Smiraglia, RP. Boundary Objects: CWA, an HR Firm, and Emergent Vocabulary. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-4-254
27. Fox, MJ; Reece, A. The Impossible Decision: Social Tagging and Derrida's Deconstructed Hospitality. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-4-260
28. Adler, M; Tennis, JT. Toward a Taxonomy of Harm in Knowledge Organization Systems. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-4-266
29. Smiraglia, RP; van den Heuvel, C. Classifications and concepts: towards an elementary theory of knowledge interaction. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-07-2012-0092
30. Channon, MG. The Unification of Concept Representations: An Impetus for Scientific Epistemology. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-2-83
31. Marcondes, CH. Knowledge Organization and Representation in Digital Environments: Relations Between Ontology and Knowledge Organization. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-2-115
32. Sienkiewicz, U; Kijenska-Dabrowska, I. Knowledge Creation and Commercialization Activities in Polish Public HEUs in the Area of Technical and Engineering Sciences. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-2-136
33. Friedman, A; Smiraglia, RP. Nodes and arcs: concept map, semiotics, and knowledge organization. JOURNAL OF DOCUMENTATION. DOI: 10.1108/00220411311295315
34. Hjørland, B. User-based and Cognitive Approaches to Knowledge Organization: A Theoretical Analysis of the Research Literature. KNOWLEDGE ORGANIZATION.
35. Corrochano, MD; Lopez-Huertas, MJ; Jimenez-Contreras, E. Spanish Research in Knowledge Organization (2002-2010). KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-1-28
36. Tennis, JT. Ethos and Ideology of Knowledge Organization: Toward Precepts for an Engaged Knowledge Organization. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-1-42
37. Campos, MLD; Campos, MLM; Davila, AMR; Gomes, HE; Campos, LM; Oliveira, LDE. Information Sciences Methodological Aspects Applied to Ontology Reuse Tools: A Study Based on Genomic Annotations in the Domain of Trypanosomatides. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2013-1-50

Publicados em 2014

1. Friedman, A. The relationship between research method and visual display: a study of conference proceedings in the field of knowledge organization. INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL.
2. Martinez-Gonzalez, MM; Alvite-Diez, ML. On the evaluation of thesaurus tools compatible with the Semantic Web. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551514545603
3. Ma, XY; Cahier, JP. Graphically structured icons for knowledge tagging. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551514542388
4. Freitas, JL; do Nascimento, BS; Bufrem, LS. Knowledge organization in the dynamics of research in the scientific literature of articles at Brapci. TRANSINFORMACAO. DOI: 10.1590/0103-37862014000300006
5. Hu, ZY; Fang, S; Liang, T. Empirical study of constructing a knowledge organization system of patent documents using topic modeling. SCIENTOMETRICS. DOI: 10.1007/s11192-014-1328-1
6. Golub, K; Tudhope, D; Zeng, ML; Zumer, M. Terminology registries for knowledge organization systems: Functionality, use, and attributes. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23090
7. Budd, J. Organizing Acts and Objects: Metaphysical Foundations. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-6-419
8. Martinez-Avila, D; San Segundo, R; Zurian, FA. Challenges and opportunities for knowledge organization at the intersection with information technologies. REVISTA ESPANOLA DE DOCUMENTACION CIENTIFICA. DOI: 10.3989/redc.2014.3.1112
9. Hjørland, B. Is Facet Analysis Based on Rationalism? A Discussion of Satija (1992), Tennis (2008), Herre (2013), Mazzocchi (2013b), and Dousa & Ibekwe-SanJuan (2014). KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-5-369
10. Gutierrez, AG. Declassifying Knowledge Organization. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-5-393
11. Golub, K; Lykke, M; Tudhope, D. Enhancing social tagging with automated keywords from the Dewey Decimal Classification. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-05-2013-0056
12. Yu, F; Qiu, JP; Lou, W. Library resources semantization based on resource ontology. ELECTRONIC LIBRARY. DOI: 10.1108/EL-05-2012-0056
13. Szostak, R. Classifying the Humanities. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-4-263

14. Kumar, BLV; Nikam, K. Development of an Information Support System for Yogic Science using Knowledge Organization Systems. KNOWLEDGE ORGANIZATION.
15. Saldanha, GS. The Philosophy of Language and Knowledge Organization in the 1930s: Pragmatics of Wittgenstein and Ranganathan. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-4-296
16. Moura, MA. Emerging Discursive Formations, Folksonomy and Social Semantic Information Spaces (SSIS): The Contributions of the Theory of Integrative Levels in the Studies carried out by the Classification Research Group (CRG). KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-4-304
17. Gutierrez, AG; Martinez-Avila, D. Critical Organization of Knowledge in Mass Media Information Systems. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-3-205
18. Pieterse, V; Kourie, DG. Lists, Taxonomies, Lattices, Thesauri and Ontologies: Paving a Pathway Through a Terminological Jungle. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-3-217
19. Losee, R. Combining High Metainformation with High Information Content: The Information-Metainformation Utility Hypothesis. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-2-123
20. Zhao, YL; Nunes, JMB; Deng, ZH. Construction and Evolution of a Chinese Information Science and Information Service (CIS&IS) Onto-Thesaurus. KNOWLEDGE ORGANIZATION.
21. Meireles, MRG; Cendon, BV; de Almeida, PEM. Bibliometric Knowledge Organization: A Domain Analytic Method Using Artificial Neural Networks. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-2-145
22. Szostak, R. Classifying for Social Diversity. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-2-160
23. Castanha, RCG; Gracio, MCC. Bibliometrics Contribution to the Metatheoretical and Domain Analysis Studies. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-2-171
24. Brascher, M. Semantic Relations in Knowledge Organization Systems. KNOWLEDGE ORGANIZATION.
25. De Santis, R; de Souza, RF. Classifying Popular Songs: Possibilities and Challenges. KNOWLEDGE ORGANIZATION.
26. Ma, XY; Cahier, JP. An Exploratory Study on Semantic Arrangement of VDL-Based Iconic Knowledge Tags. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-1-14
27. Moulaison, HL; Dykas, F; Budd, JM. Foucault, the Author, and Intellectual Debt: Capturing the Author-Function Through Attributes, Relationships, and Events in Knowledge Organization Systems. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-1-30
28. Giunchiglia, F; Dutta, B; Maltese, V. From Knowledge Organization to Knowledge Representation. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-1-44
29. Kempf, AO; Ritze, D; Eckert, K; Zapilko, B. New Ways of Mapping Knowledge Organization Systems: Using a Semi-Automatic Matching Procedure for Building up Vocabulary Crosswalks. KNOWLEDGE ORGANIZATION. DOI: 10.5771/0943-7444-2014-1-66

Publicados em 2015

1. Zhang, L; Olson, HA. Distilling Abstractions: Genre Redefining Essence versus Context. LIBRARY TRENDS. DOI: 10.1353/lib.2015.0015
2. Szostak, R. A Pluralistic Approach to the Philosophy of Classification. LIBRARY TRENDS. DOI: 10.1353/lib.2015.0007
3. Gonzalez, MEL. Organization knowledge on the digital network. INVESTIGACION BIBLIOTECOLOGICA. DOI: 10.1016/j.ibbai.2016.02.037
4. Szulc, J. Formal ontologies and their role in the integration of library resources. QUALITATIVE & QUANTITATIVE METHODS IN LIBRARIES.
5. Gregori, L; Losito, L; Siritto, P. Knowledge mapping and visualization as a common ground between librarianship and scholarly communication: Qualitative and quantitative methods for improving semantic categorization and retrieval. QUALITATIVE & QUANTITATIVE METHODS IN LIBRARIES.
6. Hjørland, B. Classical Databases and Knowledge Organization: A Case for Boolean Retrieval and Human Decision-Making During Searches. JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23250
7. Hjørland, B. Are relations in thesauri "context-free, definitional, and true in all possible worlds?". JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY. DOI: 10.1002/asi.23253
8. Wu, QQ; Zhang, HB; Lan, J. K-State automaton burst detection model based on KOS: Emerging trends in cancer field. JOURNAL OF INFORMATION SCIENCE. DOI: 10.1177/0165551514551500
9. Kelly, M. An Evidence Based Methodology to Facilitate Public Library Non-fiction Collection Development. EVIDENCE BASED LIBRARY AND INFORMATION PRACTICE. DOI: 10.18438/B8PW2P
10. McTavish, J. Everyday life classification practices and technologies Applying domain-analysis to lay understandings of food, health, and eating. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-08-2014-0105
11. Jia, JZ; Zhao, J. Mapping Analysis of Pre-coordinated Classes in DDC and CLC. KNOWLEDGE ORGANIZATION.
12. Wu, YJ; Lehman, A; Dunaway, DJ. Evaluations of a Large Topic Map as a Knowledge Organization Tool for Supporting Self-Regulated Learning. KNOWLEDGE ORGANIZATION.
13. Soergel, D. Unleashing the Power of Data Through Organization: Structure and Connections for Meaning, Learning and Discovery. KNOWLEDGE ORGANIZATION.
14. Cleverley, PH; Burnett, S. The Best of Both Worlds: Highlighting the Synergies of Combining Manual and Automatic Knowledge Organization Methods to Improve Information Search and Discovery. KNOWLEDGE ORGANIZATION.
15. Lambe, P. From Cataloguers to Designers: Paul Otlet, Social Impact and a More Proactive Role for Knowledge Organisation Professionals. KNOWLEDGE ORGANIZATION.
16. Gilchrist, A. Reflections on Knowledge, Communication and Knowledge Organization. KNOWLEDGE ORGANIZATION.
17. Gross, T. Naming and Reframing: A Taxonomy of Attacks on Knowledge Organization. KNOWLEDGE ORGANIZATION.
18. Martinez-Avila, D; Guimaraes, JAC; Pinho, FA; Fox, MJ. The Representation of Ethics and Knowledge Organization in the WoS and LISTA Databases. KNOWLEDGE ORGANIZATION.
19. Beak, J. Where is Childrens' Voice in KO?. KNOWLEDGE ORGANIZATION.
20. Weissenberger, LK. Traditional Musics and Ethical Considerations of Knowledge and Documentation Processes. KNOWLEDGE ORGANIZATION.

21. Lee, WC. Culture and Classification: An Introduction to Thinking about Ethical Issues of Adopting Global Classification Standards to Local Environments. KNOWLEDGE ORGANIZATION.
22. Oravec, JA. The Moral Imagination in an Era of "Gaming Academia": Implications of Emerging Reputational Issues in Scholarly Activities for Knowledge Organization Practices. KNOWLEDGE ORGANIZATION.
23. Arboit, AE; Guimaraes, JAC. The Ethics of Knowledge Organization and Representation from a Bakhtinian Perspective. KNOWLEDGE ORGANIZATION.
24. McQueen, K. Ethical Issues of Knowledge Organization in Designing a Metadata Schema for the Leo Kottke Archives. KNOWLEDGE ORGANIZATION.
25. da Silva, AP; Guimaraes, JAC; Tognoli, NB. Ethical Values in Archival Arrangement and Description: An Analysis of Professional Codes of Ethics. KNOWLEDGE ORGANIZATION.
26. Fang, H. Classifying Research Articles in Multidisciplinary Sciences Journals into Subject Categories. KNOWLEDGE ORGANIZATION.
27. Bishop, BW; Moulaison, HL; Burwell, CL. Geographic Knowledge Organization: Critical Cartographic Cataloging and Place-Names in the Geoweb. KNOWLEDGE ORGANIZATION.
28. Baca, M; Gill, M. Encoding Multilingual Knowledge Systems in the Digital Age: the Getty Vocabularies. KNOWLEDGE ORGANIZATION.
29. Tennis, JT. Foundational, First-Order, and Second-Order Classification Theory. KNOWLEDGE ORGANIZATION.
30. Martin-Moncunill, D; Garcia-Barriocanal, E; Sicilia, MA; Sanchez-Alonso, S. Evaluating the Practical Applicability of Thesaurus-Based Keyphrase Extraction in the Agricultural Domain: Insights from the VOA3R Project. KNOWLEDGE ORGANIZATION.
31. Hjørland, B. Theories are Knowledge Organizing Systems (KOS). KNOWLEDGE ORGANIZATION.
32. Martin-Moncunill, D; Sicilia-Urban, MA; Garcia-Barriocanal, E; Sanchez-Alonso, S. Evaluating the degree of domain specificity of terms in large terminologies The case of AGROVOC. ONLINE INFORMATION REVIEW. DOI: 10.1108/OIR-02-2015-0052
33. Madalli, DP; Balaji, BP; Sarangi, AK. Faceted Ontological Representation for a Music Domain. KNOWLEDGE ORGANIZATION.
34. Ribeiro, CJS; Pereira, DV. The publication of open government data: A proposal for the revision of class in Social Security Regarding the Controlled Vocabulary of the Electronic Government. TRANSINFORMACAO. DOI: 10.1590/0103-37862015000100007
35. Martinez-Avila, D. Knowledge Organization in the Intersection with Information Technologies. KNOWLEDGE ORGANIZATION.
36. Guimaraes, JAC; Tognoli, NB. Provenance as a Domain Analysis Approach in Archival Knowledge Organization. KNOWLEDGE ORGANIZATION.
37. Marteleto, RM; Carvalho, LD. Health as a Knowledge Domain and Social Field: Dialogues with Birger Hjørland and Pierre Bourdieu. KNOWLEDGE ORGANIZATION.
38. Smiraglia, RP. Domain Analysis of Domain Analysis for Knowledge Organization: Observations on an Emergent Methodological Cluster. KNOWLEDGE ORGANIZATION.
39. Wells, MJ. Social semiotics as theory and practice in library and information science. JOURNAL OF DOCUMENTATION. DOI: 10.1108/JD-01-2014-0018

APÊNDICE 7. Autores e documentos das matrizes de cocitação

Parte 1 – *Ranking* de 180 autores que considera todas as posições das referências organizados a partir dos valores de documentos citantes.

	Autores	Citações totais no corpus	Documentos citantes	Autoria única	Coautoria	Autoria principal	Autoria secundária
1	Hjørland, B	191	68	155	36	31	05
2	Croft, WB	136	60	03	133	09	124
3	Järvelin, K	77	45	03	74	15	59
4	Baeza-Yates, R	60	45	00	60	38	22
5	Salton, G	62	44	12	50	40	10
6	Robertson, SE	69	42	11	58	34	24
7	Spink, A	74	41	05	69	31	38
8	Saracevic, T	68	40	48	20	07	13
9	Belkin, NJ	95	38	20	75	22	53
10	Jansen, BJ	66	37	08	58	38	20
11	Manning, CD	49	36	00	49	35	14
12	Voorhees, EM	59	34	21	38	11	27
13	Olson, HA	106	33	75	31	16	15
14	Dumais, ST	65	33	01	64	04	60
15	Smiraglia, RP	73	32	54	19	08	11
16	Ingwersen, P	53	32	13	40	19	21
17	Buckley, C	49	32	02	47	23	24
18	Marchionini, G	49	32	26	23	04	19
19	Zhang, J	46	32	06	40	21	19
20	Ribeiro-Neto, B	37	32	00	37	01	36
21	Bates, MJ	47	30	46	01	01	00
22	Schutze, H	35	30	01	34	01	33
23	Raghavan, P	30	30	00	30	00	30
24	Tennis, JT	54	29	38	16	04	12
25	Blei, DM	51	28	06	45	33	12
26	Metzler, D	40	28	00	40	20	20
27	Sanderson, M	37	28	04	33	06	27
28	Vakkari, P	47	27	23	24	12	12
29	White, RW	70	26	00	70	46	24
30	Ng, AY	26	26	02	24	00	24
31	Lancaster, FW	36	25	26	10	06	04
32	Van Rijsbergen, CJ	29	25	18	11	01	10
33	Ding, Y	46	24	04	42	14	28
34	Berners-Lee, T	39	24	15	24	16	08
35	Ounis, I	39	24	00	39	11	28
36	Furner, J	32	24	22	10	03	07
37	Zeng, ML	32	24	08	24	11	13
38	Jordan, MI	31	24	00	31	00	31
39	De Rijke, M	41	23	00	41	00	41
40	Allan, J	39	23	04	35	06	29
41	Li, X	28	23	00	28	11	17
42	Zhai, CX	56	22	03	53	16	37
43	Witten, IH	43	23	00	43	07	36
44	Kelly, D	40	22	13	27	15	12

	Autores	Citações totais no corpus	Documentos citantes	Autoria única	Coautoria	Autoria principal	Autoria secundária
45	Beghtol, C	32	22	31	01	00	01
46	Teevan, J	31	22	00	31	11	20
47	Walker, S	24	22	00	24	01	23
48	White, HD	50	21	29	21	16	05
49	Kuhlthau, CC	36	21	33	03	03	00
50	Joachims, T	34	21	13	21	09	12
51	Dahlberg, I	30	21	30	00	00	00
52	Mai, JE	32	20	32	00	00	00
53	Chen, C	28	20	06	22	08	14
54	Bawden, D	26	20	06	20	05	15
55	Li, Y	25	20	01	24	17	07
56	Harman, DK	23	20	13	10	02	08
57	Guimarães, JAC	40	19	12	28	11	17
58	Callan, JP	38	19	03	35	07	28
59	Wilson, TD	38	19	32	06	03	03
60	Borlund, P	35	20	19	16	12	04
61	Ruthven, I	34	19	03	31	05	26
62	Hu, X	32	19	01	31	13	18
63	Garfield, E	30	19	18	12	08	04
64	Sparck Jones, K	26	19	09	17	09	08
65	Ellis, D	25	19	15	10	06	04
66	Hendler, JA	19	19	02	17	00	17
67	Li, H	36	18	00	36	03	33
68	Fox, EA	35	18	01	34	09	25
69	Lafferty, JD	34	18	00	34	09	25
70	Nie, JY	33	18	04	29	00	29
71	Zobel, J	29	18	05	24	02	22
72	Moffat, A	27	18	00	27	06	21
73	Jones, Rosie	23	18	01	22	10	12
74	MacFarlane, A	22	18	02	20	02	18
75	Soboroff, I	30	17	01	29	07	22
76	Macdonald, C	29	17	00	29	06	23
77	Wildemuth, BM	28	17	08	20	12	08
78	Bar-Ilan, J	24	17	07	17	14	03
79	Chen, H	21	17	00	21	06	15
80	Svenonius, E	21	17	21	00	00	00
81	Page, L	19	17	00	19	10	09
82	Jones, SA	18	17	00	18	04	14
83	Park, JR	17	17	03	14	04	10
84	Leydesdorff, L	51	16	12	39	23	16
85	Glänzel, W	44	16	03	41	18	23
86	Small, H	33	16	20	13	06	07
87	Wolfram, D	28	16	05	23	06	17
88	Xu, J	27	16	00	27	20	07
89	Han, J	25	16	00	25	06	19
90	McCallum, A	25	16	04	21	04	17
91	Singhal, A	19	16	06	13	04	09
92	Buckland, MK	18	16	14	04	04	00
93	Chan, LM	16	16	02	14	00	14
94	Yan, E	29	15	02	27	16	11
95	Chowdhury, GG	21	15	12	09	01	08
96	Dervin, B	21	15	16	05	05	00
97	Toms, EG	20	15	06	14	03	11

	Autores	Citações totais no corpus	Documentos citantes	Autoria única	Coautoria	Autoria principal	Autoria secundária
98	Fidel, R	19	15	10	09	07	02
99	Kamps, J	19	15	02	17	00	17
100	Van Raan, AFJ	18	15	05	13	01	12
101	Brin, S	17	15	00	17	07	10
102	Zhang, M	17	15	00	17	03	14
103	Broder, AZ	16	15	09	07	04	03
104	Cronin, B	16	15	07	09	05	04
105	Yu, Y	16	15	00	16	00	16
106	Nejdl, W	15	15	00	15	00	15
107	Strohman, T	15	15	00	15	05	10
108	Jose, JM	37	14	00	37	01	36
109	Boyack, KW	34	14	00	34	15	19
110	Mayr, P	28	14	07	21	07	14
111	Tudhope, D	28	14	00	28	10	18
112	Cool, C	25	14	02	23	04	19
113	Albrechtsen, H	24	14	07	17	00	17
114	Ranganathan, SR	22	14	21	01	01	00
115	Clarke, CLA	21	14	00	21	08	13
116	Giles, CL	21	14	00	21	01	20
117	He, B	21	14	00	21	08	13
118	Tang, J	21	14	00	21	06	15
119	Kekalainen, J	20	14	01	19	04	15
120	Granka, LA	19	14	01	18	01	17
121	Sugimoto, CR	19	14	00	19	05	14
122	Agichtein, E	18	14	00	18	08	10
123	Hearst, MA	18	14	09	09	03	06
124	Savoy, J	18	14	10	08	01	07
125	Gao, J	17	14	00	17	06	11
126	Jones, GJF	17	14	00	17	02	15
127	Landauer, TK	17	14	00	17	05	12
128	Azzopardi, L	16	14	04	12	05	07
129	Song, D	16	14	00	16	02	14
130	Amati, G	15	14	03	12	07	05
131	Lassila, O	15	14	01	14	00	14
132	Lavrenko, V	15	14	01	14	09	05
133	Winograd, T	15	14	01	14	00	14
134	Martínez-Ávila, D	29	13	02	27	16	11
135	Soergel, D	28	13	11	17	05	12
136	Frohmann, B	23	13	23	00	00	00
137	Davis, FD	22	13	13	09	06	03
138	Lalmas, M	22	13	00	22	01	21
139	Oard, DW	22	13	03	19	05	14
140	Bizer, C	19	13	00	19	10	09
141	Craswell, N	19	13	00	19	01	18
142	Byström, K	18	13	04	14	10	04
143	Gilchrist, A	18	13	05	13	02	11
144	Bruza, PD	17	13	01	16	03	13
145	Mitra, M	17	13	00	17	04	13
146	Carpineto, C	16	13	00	16	16	00
147	Chen, Z	16	13	00	16	03	13
148	Broughton, V	15	13	14	01	01	00
149	Case, DO	15	13	14	01	01	00
150	Savolainen, R	15	13	09	06	01	05

	Autores	Citações totais no corpus	Documentos citantes	Autoria única	Coautoria	Autoria principal	Autoria secundária
151	Turney, PD	15	13	08	07	04	03
152	Wang, P	15	13	03	12	08	04
153	Wilson, P	15	13	14	01	00	01
154	Jacob, EK	14	13	08	06	02	04
155	Porter, MF	14	13	12	02	01	01
156	Furnas, GW	13	13	00	13	03	10
157	He, D	13	13	00	13	03	10
158	Klavans, R	29	12	00	29	11	18
159	Liu, Jingjing	27	12	00	27	18	09
160	Gnoli, C	26	12	16	10	06	04
161	Konstan, JA	26	12	00	26	01	25
162	Börner, K	25	12	06	19	08	11
163	Diaz, F	22	12	04	18	05	13
164	Hersh, WR	22	12	02	20	11	09
165	Larsen, B	21	12	04	17	05	12
166	Joho, H	20	12	00	20	14	06
167	Moya-Anegón, F	20	12	00	20	06	14
168	Carterette, BA	19	12	03	16	08	08
169	Garcia-Molina, H	17	12	00	17	00	17
170	Hembrooke, HA	17	12	00	17	04	13
171	Jones, CB	16	12	00	16	04	12
172	Smucker, MD	16	12	00	16	13	03
173	Webber, W	16	12	01	15	07	08
174	Gabrilovich, E	15	12	00	15	03	12
175	Lykke, M	15	12	02	13	04	09
176	Romano, G	16	13	00	16	00	16
177	Scholer, F	15	12	00	15	05	10
178	Zha, H	14	12	00	14	00	14
179	Radlinski, F	13	12	00	13	04	09
180	Vickery, BC	13	12	12	01	01	00

Parte 2 – *Ranking* de 72 autores que considera a primeira posição das referências organizados a partir dos valores de documentos citantes.

	Autores	Citações totais no corpus	Documentos citantes	Autoria principal	Autoria única
1	Hjørland, B	186	66	31	155
2	Salton, G	52	40	40	12
3	Robertson, SE	45	34	34	11
4	Olson, HA	91	33	16	75
5	Baeza-Yates, R	38	33	38	00
6	Saracevic, T	55	31	07	48
7	Manning, CD	35	31	35	00
8	Smiraglia, RP	62	30	08	54
9	Bates, MJ	47	30	01	46
10	Jansen, BJ	46	29	38	08
11	Belkin, NJ	42	27	22	20
12	Blei, DM	39	27	33	06
13	Spink, A	36	27	31	05
14	Ingwersen, P	32	25	19	13

	Autores	Citações totais no corpus	Documentos citantes	Autoria principal	Autoria única
15	Marchionini, G	30	25	04	26
16	Tennis, JT	42	24	04	38
17	Voorhees, EM	32	23	11	21
18	Lancaster, FW	32	22	06	26
19	Berners-Lee, T	31	22	16	15
20	White, RW	46	21	46	00
21	Kuhlthau, CC	36	21	03	33
22	Beghtol, C	31	21	00	31
23	Dahlberg, I	30	21	00	30
24	White, HD	45	20	16	29
25	Mai, JE	32	20	00	32
26	Vakkari, P	35	19	12	23
27	Zhang, J	27	19	21	06
28	Buckley, C	25	19	23	02
29	Wilson, TD	35	18	03	32
30	Borlund, P	31	19	12	19
31	Furner, J	25	18	03	22
32	Metzler, D	20	18	20	00
33	Zeng, ML	19	18	11	08
34	Garfield, E	26	17	08	18
35	Svenonius, E	21	17	00	21
36	Van Rijsbergen, CJ	19	17	01	18
37	Kelly, D	28	16	15	13
38	Joachims, T	22	16	09	13
39	Bar-Ilan, J	21	16	14	07
40	Ellis, D	21	16	06	15
41	Buckland, MK	18	16	04	14
42	Järvelin, K	18	16	15	03
43	Small, H	26	15	06	20
44	Dervin, B	21	15	05	16
45	Wildemuth, BM	20	15	12	08
46	Xu, J	20	15	20	00
47	Ranganathan, SR	22	14	01	21
48	Sparck Jones, K	18	14	09	09
49	Fidel, R	17	14	07	10
50	Leydesdorff, L	35	13	23	12
51	Frohmann, B	23	13	00	23
52	Davis, FD	19	13	06	13
53	Ding, Y	18	13	14	04
54	Li, Y	18	13	17	01
55	Yan, E	18	13	16	02
56	Carpineto, C	16	13	16	00
57	Broughton, V	15	13	01	14
58	Case, DO	15	13	01	14
59	Broder, AZ	13	13	04	09
60	Guimarães, JAC	23	12	11	12
61	Harman, DK	15	12	02	13
62	Wilson, P	14	12	00	14
63	Porter, MF	13	12	01	12
64	Vickery, BC	13	12	01	12
65	Croft, WB	12	12	09	03
66	Cronin, B	12	12	05	07
67	García Gutiérrez, A	24	11	04	20

	Autores	Citações totais no corpus	Documentos citantes	Autoria principal	Autoria única
68	Gnoli, C	22	11	06	16
69	Zhai, CX	19	11	16	03
70	Aitchison, J	13	11	12	01
71	Jones, Rosie	11	11	10	01
72	Li, X	11	11	11	00

Parte 3 – Ranking de 109 documentos organizados a partir do número de citações totais (CT). Legendas restantes: AU – autoria única; CoA – coautoria e ANP – autoria não pessoal

	Autores	Ano	Títulos	CT	AU	CoA	ANP
1	Baeza-Yates, R	1999	Modern information retrieval	27	-- -	27	---
2	Blei, DM	2003	Latent Dirichlet allocation	23	-- -	23	---
3	Manning, CD	2008	Introduction to information retrieval	18	-- -	18	---
4	Saracevic, T	2007	Relevance - Saracevic	17	17	---	---
5	Ingwersen, P	2005	The turn	17	-- -	17	---
6	Hjørland, B	2002	Domain analysis in information science	15	15	---	---
7	Hjørland, B	1995	Toward a New Horizon in Information Science	15	01	14	---
8	Van Rijsbergen, CJ	1979	Information Retrieval	14	14	---	---
9	Berners-Lee, T	2001	The semantic web	14	-- -	14	---
10	Olson, HA	2002	The Power to Name - 2002	13	13	---	---
11	Hjørland, B	2008	What is Knowledge Organization (KO)?	13	13	---	---
12	Lancaster, FW	1993	Indexing and abstracting	12	12	---	---
13	Salton, G	1975	A vector space model for automatic indexing	11	-- -	11	---
14	Porter, MF	1980	An algorithm for suffix stripping	11	11	---	---
15	Marchionini, G	2006	Exploratory search	11	11	---	---
16	Kuhlthau, CC	1991	Inside the search process	11	11	---	---
17	Davis, FD	1989	Perceived usefulness, perceived ease of use and user acceptance of information technology	11	11	---	---
18	Bates, MJ	1989	The design of browsing and berrypicking techniques for the online search interface	11	11	---	---
19	Ponte, JM	1998	A language modeling approach to information retrieval	10	-- -	10	---
20	Carpineto, C	2012	A survey of automatic query expansion in information retrieval	10	-- -	10	---
21	Jansen, BJ	2006	How are we searching the World Wide Web?	10	-- -	10	---
22	NAP	2013	ISO 25964-2	10	-- -	---	10
23	Dahlberg, I	2006	Knowledge organization: a new science?	10	10	---	---
24	Case, DO	2012	Looking for information	10	10	---	---
25	Kuhlthau, CC	2004	Seeking meaning	10	10	---	---
26	Page, L	1999	The PageRank citation ranking	10	-- -	10	---
27	Broder, AZ	2002	A taxonomy of Web search	09	09	---	---
28	Manning, CD	2009	An introduction to information retrieval	09	-- -	09	---

	Autores	Ano	Títulos	CT	AU	CoA	ANP
29	Small, H	1973	Co-citation in the scientific literature	09	09	---	---
30	Olson, HA	2000	Difference, Culture and Change	09	09	---	---
31	Marchionini, G	1995	Information seeking in electronic environments	09	09	---	---
32	Wilson, TD	1999	Models in information behaviour research	09	09	---	---
33	Ranganathan, SR	1937	Prolegomena to Library Classification	09	08	01	---
34	Lavrenko, V	2001	Relevance based language models	09	-- -	09	---
35	Bowker, GC	2000	Sorting Things Out	09	-- -	09	---
36	Salton, G	1987	Term weighting approaches in automatic text retrieval	09	-- -	09	---
37	Hjørland, B	2011	The importance of theories of knowledge	09	09	---	---
38	Aitchison, J	2000	Thesaurus construction and use	09	-- -	09	---
39	Järvelin, K	2002	Cumulated gain-based evaluation of IR techniques	08	-- -	08	---
40	Furner, J	2007	Dewey Deracialized	08	08	---	---
41	NAP	2004	Funcional Requirement for Bibliographic Records	08	-- -	---	08
42	Salton, G	1983	Introduction to Modern Information Retrieval – Salton	08	-- -	08	---
43	NAP	2011	ISO 25964-1	08	-- -	---	08
44	Robertson, SE	1996	Okapi at TREC-3	08	-- -	08	---
45	Wilson, TD	1981	On user studies and information needs	08	08	---	---
46	Jansen, BJ	2000	Real life, real users, and real needs	08	-- -	08	---
47	Croft, WB	2010	Search engines	08	-- -	08	---
48	Borlund, P	2003	The concept of relevance in IR	08	08	---	---
49	Golder, SA	2006	Usage patterns of collaborative tagging systems	08	-- -	08	---
50	Vakkari, P	2001	A theory of the task-based information retrieval process	07	07	---	---
51	Belkin, NJ	1980	Anomalous states of knowledge as a basis for information retrieval	07	07	---	---
52	NAP	2005	ANSI/NISO Z39.19 – 2005	07	-- -	---	07
53	Strauss, AL	1990	Basics of qualitative research	07	-- -	07	---
54	Hjørland, B	2009	Concept Theory - Hjørland	07	07	---	---
55	Smiraglia, RP	2012	Epistemology of Domain Analysis	07	07	---	---
56	Buckley, C	2000	Evaluating evaluation measure stability	07	-- -	07	---
57	White, RW	2009	Exploratory search - White	07	-- -	07	---
58	Hjørland, B	2003	Fundamentals of Knowledge Organization	07	07	---	---
59	Olson, HA	2007	How We Construct Subjects	07	07	---	---
60	Berners-Lee, T	2007	Linked data - Lee	07	07	---	---
61	Olson, HA	1998	Mapping beyond Dewey's boundaries	07	07	---	---
62	Berman, S	1993	Prejudices and antipathies	07	07	---	---
63	Hulme, EW	1911	Principles of Book Classification	07	07	---	---
64	Saracevic, T	1996	Relevance Reconsidered	07	07	---	---
65	Hjørland, B	2007	Semantics and Knowledge Organization	07	07	---	---
66	Mai, JE	2001	Semiotics and indexing	07	07	---	---
67	Hodge, G	2000	Systems of knowledge organization for digital libraries	07	07	---	---
68	Byström, K	1995	Task complexity affects information seeking and use	07	-- -	07	---

	Autores	Ano	Títulos	CT	AU	CoA	ANP
69	Wildemuth, BM	2004	The effects of domain knowledge on search tactic formulation	07	07	---	---
70	Svenonius, E	2004	The epistemological foundations of knowledge representation	07	07	---	---
71	Svenonius, E	2000	The Intellectual Foundation of Information Organization	07	07	---	---
72	Shannon, CE	1948	The mathematical theory of communication	07	01	06	---
73	Anderson, JD	2001	The nature of indexing	07	--	07	---
74	Olson, HA	2001	The Power to Name - 2001	07	07	---	---
75	Foskett, AC	1982	The subject approach to information	07	07	---	---
76	Zeng, ML	2004	Trends and issues in establishing interoperability among knowledge organization systems	07	--	07	---
77	Wilson, P	1978	Two kinds of power	07	07	---	---
78	White, HD	1998	Visualizing a discipline	07	--	07	---
79	Ellis, D	1989	A behavioural approach to information retrieval and system design	06	06	---	---
80	Zhai, CX	2004	A study of smoothing methods for language models applied to information retrieval	06	--	06	---
81	Joachims, T	2005	Accurately interpreting clickthrough data as implicit feedback	06	--	06	---
82	Buckley, C	1994	Automatic query expansion using smart	06	01	05	---
83	Kessler, MM	1963	Bibliographic coupling between scientific papers	06	06	---	---
84	Levenshtein, VI	1966	Binary codes capable of correcting deletions, insertions and reversals	06	06	---	---
85	Hjørland, B	2013	Citation analysis - Hjørland	06	06	---	---
86	Jacob, EK	2004	Classification and categorization	06	06	---	---
87	Ingwersen, P	1996	Cognitive perspectives of information retrieval interaction	06	06	---	---
88	Hjørland, B	2013	Facet analysis	06	06	---	---
89	Blondel, VD	2008	Fast unfolding of communities in large networks	06	--	06	---
90	Manning, CD	1999	Fondations of statistical natural language processing	06	--	06	---
91	Deerwester, SC	1990	Indexing by Latent Semantic Analysis	06	--	06	---
92	Buckland, MK	1991	Information as thing	06	06	---	---
93	Wei, X	2006	LDA-based document models for ad-hoc retrieval	06	--	06	---
94	Bizer, C	2009	Linked data – the story so far	06	--	06	---
95	Garshol, LM	2004	Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all	06	06	---	---
96	Wittgenstein, L	1979	Philosophical Investigations	06	06	---	---
97	Briet, S	1951	Qu'est-ce que la documentation?	06	06	---	---
98	Mizzaro, S	1997	Relevance - Mizzaro	06	06	---	---
99	Rocchio, JJ	1971	Relevance Feedback in Information Retrieval	06	06	---	---
100	Robertson, SE	1988	Relevance weighting of search terms	06	--	06	---
101	Bradford, SC	1934	Sources of information on specific subjects	06	06	---	---
102	Brin, S	1998	The Anatomy of a Large-scale Hypertextual Web Search Engine	06	--	06	---
103	Broughton, V	2006	The Need for a Faceted Classification as the Basis of All Methods of Information Retrieval	06	06	---	---
104	Hjørland, B	2013	Theories of Knowledge Organization	06	06	---	---

	Autores	Ano	Títulos	CT	AU	CoA	ANP
105	Otlet, P	1934	Traité de documentacion	06	06	---	---
106	Tennis, JT	2003	Two Axes of Domain for Domain Analysis	06	06	---	---
107	Lu, C	2010	User tags versus expertassigned subject terms	06	-- -	06	---
108	Hjørland, B	2013	User-based and cognitive approaches to knowledge organization	06	06	---	---
109	Lancaster, FW	1995	Vocabulary Control for Information Retrieval	06	06	---	---

APÊNDICE 11. Componentes da análise fatorial de duas matrizes de cocitação de primeiro autor (diagonal zero e cocitação do autor com ele mesmo, normalizadas pelo r de Pearson).

Diagonal: zero Normalizada pelo r de Pearson					Diagonal: cocitação do autor com ele mesmo Normalizada pelo r de Pearson				
AUTORES	Compones e cargas				AUTORES	Compones e cargas			
	C1	C2	C3	C4		C1	C2	C3	C4
Lancaster, FW	0,967				Lancaster, FW	0,963			
Wilson, P	0,922				Wilson, P	0,944			
Furner, J	0,922				Furner, J	0,940			
Buckland, MK	0,892				Buckland, MK	0,922			
Vickery, BC	0,887				Mai, JE	0,902			
Ranganathan, SR	0,882				Vickery, BC	0,901			
Zeng, ML	0,880				Zeng, ML	0,893			
Mai, JE	0,877				Olson, HA	0,890			
Dahlberg, I	0,868				Ranganathan, SR	0,890			
Broughton, V	0,868				Svenonius, E	0,887			
Smiraglia, RP	0,867				Hjørland, B	0,886			
Svenonius, E	0,867				Broughton, V	0,884			
Olson, HA	0,866				Dahlberg, I	0,880			
Frohmann, B	0,864				Beghtol, C	0,875			
Beghtol, C	0,860				Tennis, JT	0,874			
García Gutiérrez, A	0,856				Smiraglia, RP	0,871			
Tennis, JT	0,856				Gnoli, C	0,868			
Gnoli, C	0,851				Guimarães, JAC	0,859			
Guimarães, JAC	0,850				García Gutiérrez, A	0,855			
Aitchison, J	0,830				Frohmann, B	0,848			
Hjørland, B	0,815				Jones, Rosie	-0,848			
Jones, Rosie	-0,801				Joachims, T	-0,841			
Joachims, T	-0,776				Aitchison, J	0,804			
Berners-Lee, T	0,760				Voorhees, EM	-0,791			
Li, Y	-0,746				Metzler, D	-0,785			
Li, X	-0,728				Li, Y	-0,781			
Voorhees, EM	-0,724				Li, X	-0,761			
Metzler, D	-0,718				Buckley, C	-0,754			
Buckley, C	-0,699				Berners-Lee, T	0,749			
Xu, J	-0,678				Xu, J	-0,741			
Ellis, D	0,918				Harman, DK	-0,740			
Davis, FD	0,909				Zhai, CX	-0,707			
Wilson, TD	0,903				Carpineto, C	-0,689			
Dervin, B	0,901				Porter, MF	-0,633			
Bates, MJ	0,901				Ellis, D	0,929			
Marchionini, G	0,900				Davis, FD	0,919			
Case, DO	0,898				Saracevic, T	0,918			
Kuhlthau, CC	0,897				Bates, MJ	0,915			
Vakkari, P	0,894				Case, DO	0,914			
Saracevic, T	0,891				Dervin, B	0,911			
Belkin, NJ	0,886				Wilson, TD	0,910			
Ingwersen, P	0,874				Marchionini, G	0,907			
Wildemuth, BM	0,867				Kuhlthau, CC	0,906			
Spink, A	0,864				Vakkari, P	0,901			
Borlund, P	0,860				Belkin, NJ	0,884			

Jansen, BJ	0,857	Ingwersen, P	0,877
Fidel, R	0,825	Spink, A	0,877
Kelly, D	0,814	Wildemuth, BM	0,862
Bar-Ilan, J	0,794	Borlund, P	0,856
White, RW	0,792	Fidel, R	0,851
Broder, AZ	0,759	Jansen, BJ	0,843
Järvelin, K	0,697	Kelly, D	0,806
Zhang, J	0,673	Bar-Ilan, J	0,800
<hr/>		White, RW	0,771
Sparck Jones, K	0,923	Broder, AZ	0,749
Van Rijsbergen, CJ	0,883	Järvelin, K	0,666
Croft, WB	0,853	Zhang, J	0,628
Manning, CD	0,836	<hr/>	
Salton, G	0,794	Sparck Jones, K	0,924
Baeza-Yates, R	0,784	Van Rijsbergen, CJ	0,905
Robertson, SE	0,780	Manning, CD	0,815
Porter, MF	0,714	Croft, WB	0,803
Carpineto, C	0,692	Salton, G	0,752
Harman, DK	0,685	Baeza-Yates, R	0,746
Zhai, CX	0,653	Robertson, SE	0,715
Blei, DM	0,642	Blei, DM	0,532
<hr/>		<hr/>	
Ding, Y	0,935	Ding, Y	0,955
Garfield, E	0,917	White, HD	0,951
Small, H	0,900	Garfield, E	0,939
White, HD	0,892	Small, H	0,924
Leydesdorff, L	0,862	Yan, E	0,890
Yan, E	0,858	Leydesdorff, L	0,888
Cronin, B	0,766	Cronin, B	0,785

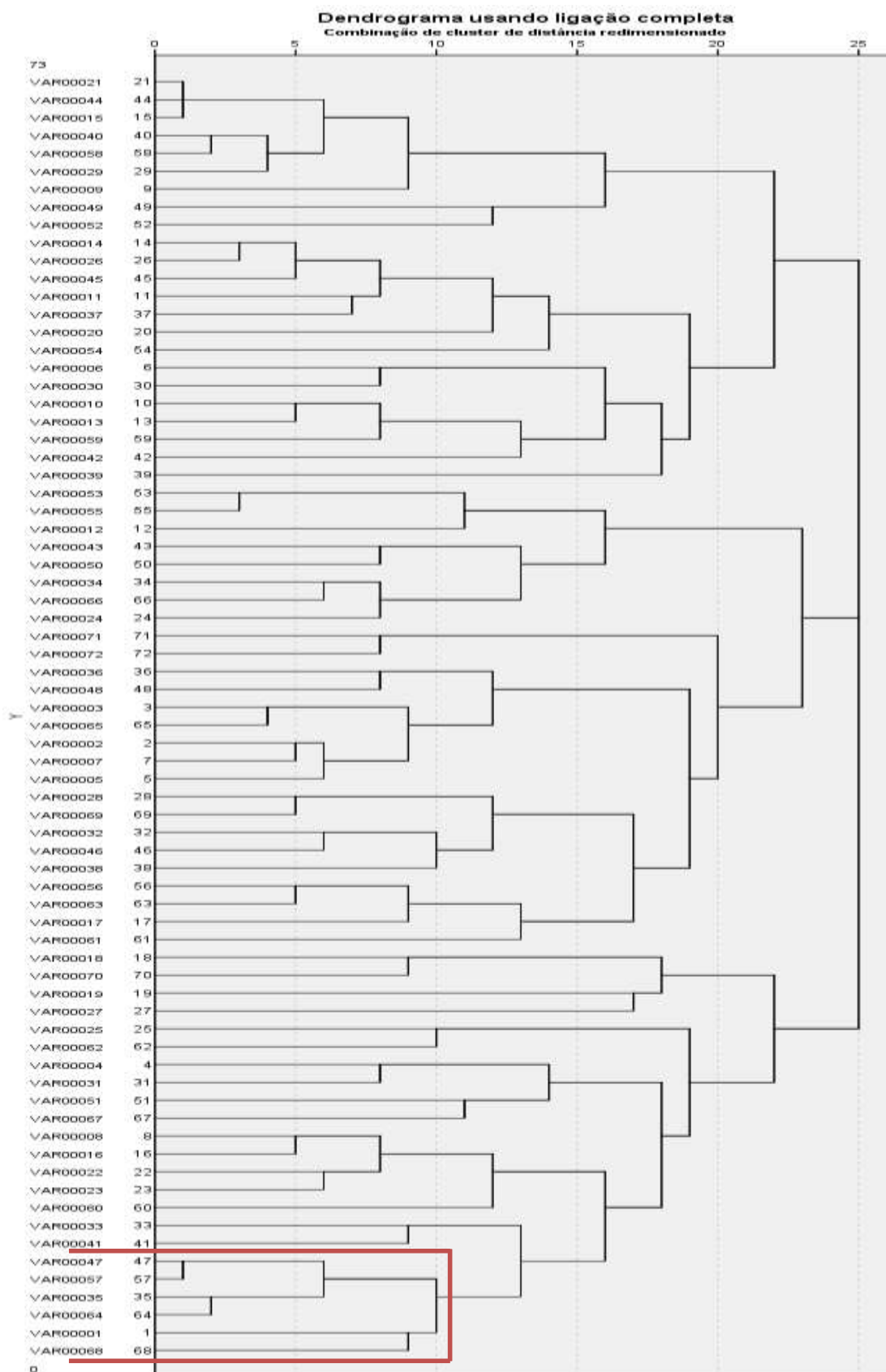
APÊNDICE 12. Componentes da análise fatorial de uma matriz de cocitação de documentos (diagonal com citação total, normalizada pelo r de Pearson).

Documentos	Componentes	
VAR00032 Models in information behaviour research	-0,93	
VAR00025 Seeking meaning	-0,93	
VAR00016 Inside the search process	-0,93	
VAR00031 Information seeking in electronic environments	-0,92	
VAR00018 The design of browsing and berrypicking techniques for the online search interface	-0,91	
VAR00051 Anomalous states of knowledge as a basis for information retrieval	-0,91	
VAR00079 A behavioural approach to information retrieval and system design	-0,91	
VAR00068 Task complexity affects information seeking and use	-0,91	
VAR00050 A theory of the task-based information retrieval process	-0,90	
VAR00024 Looking for information	-0,90	
VAR00045 On user studies and information needs	-0,87	
VAR00005 The turn	-0,85	
VAR00015 Exploratory search	-0,80	
VAR00057 Exploratory search - White	-0,77	C1
VAR00069 The effects of domain knowledge on search tactic formulation	-0,76	
VAR00053 Basics of qualitative research	-0,74	
VAR00087 Cognitive perspectives of information retrieval interaction	-0,73	
VAR00027 A taxonomy of Web search	-0,72	
VAR00017 Perceived usefulness, perceived ease of use and user acceptance of information technology	-0,62	
VAR00059 How We Construct Subjects	0,50	
VAR00030 Difference, Culture and Change	0,50	
VAR00074 The Power to Name - 2001	0,49	
VAR00010 The Power to Name - 2002	0,49	
VAR00035 Sorting Things Out	0,49	
VAR00061 Mapping beyond Dewey's boundaries	0,48	
VAR00040 Dewey Deracialized	0,48	
VAR00062 Prejudices and antipathies	0,46	
VAR00080 A study of smoothing methods for language models applied to information retrieval	0,92	
VAR00034 Relevance based language models	0,91	
VAR00082 Automatic query expansion using smart	0,91	
VAR00019 A language modeling approach to information retrieval	0,86	
VAR00099 Relevance Feedback in Information Retrieval	0,85	
VAR00014 An algorithm for suffix stripping	0,85	C2
VAR00020 A survey of automatic query expansion in information retrieval	0,72	
VAR00093 LDA-based document models for ad-hoc retrieval	0,68	
VAR00002 Latent Dirichlet allocation	0,59	
VAR00090 Foundations of statistical natural language processing	0,53	
VAR00081 Accurately interpreting clickthrough data as implicit feedback	0,47	
VAR00036 Term weighting approaches in automatic text retrieval	0,88	
VAR00008 Information Retrieval	0,84	
VAR00047 Search engines	0,80	
VAR00091 Indexing by Latent Semantic Analysis	0,79	C3
VAR00001 Modern information retrieval	0,75	
VAR00039 Cumulated gain-based evaluation of IR techniques	0,71	
VAR00044 Okapi at TREC-3	0,70	
VAR00013 A vector space model for automatic indexing	0,67	

VAR00042 Introduction to Modern Information Retrieval - Salton	0,63	
VAR00100 Relevance weighting of search terms	0,58	
VAR00028 An introduction to information retrieval	0,57	
VAR00073 The nature of indexing	0,56	
VAR00026 The PageRank citation ranking	0,54	
VAR00003 Introduction to information retrieval	0,49	
<hr/>		
VAR00095 Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all	0,92	
VAR00065 Semantics and Knowledge Organization	0,89	
VAR00085 Citation analysis - Hjørland	0,89	
VAR00088 Facet analysis	0,83	
VAR00037 The importance of theories of knowledge	0,82	C4
VAR00108 User-based and cognitive approaches to knowledge organization	0,74	
VAR00104 Theories of Knowledge Organization	0,61	
VAR00096 Philosophical Investigations	0,57	
VAR00011 What is Knowledge Organization (KO)?	0,46	
<hr/>		
VAR00022 ISO 25964-2	-0,93	
VAR00043 ISO 25964-1	-0,90	
VAR00094 Linked data – the story so far	-0,83	
VAR00038 Thesaurus construction and use	-0,82	
VAR00076 Trends and issues in establishing interoperability among knowledge organization systems	-0,82	C5
VAR00109 Vocabulary Control for Information Retrieval	-0,81	
VAR00009 The semantic web	-0,81	
VAR00052 ANSI/NISO Z39.19 – 2005	-0,75	
<hr/>		
VAR00064 Relevance Reconsidered	0,93	
VAR00101 Sources of information on specific subjects	0,91	
VAR00004 Relevance - Saracevic	0,91	
VAR00098 Relevance - Mizzaro	0,87	C6
VAR00048 The concept of relevance in IR	0,84	
VAR00021 How are we searching the World Wide Web?	0,74	
VAR00102 The Anatomy of a Large-scale Hypertextual Web Search Engine	0,61	
<hr/>		
VAR00007 Toward a New Horizon in Information Science	0,86	
VAR00055 Epistemology of Domain Analysis	0,84	
VAR00058 Fundamentals of Knowledge Organization	0,82	
VAR00006 Domain analysis in information science	0,79	
VAR00023 Knowledge organization: a new science?	0,75	C7
VAR00106 Two Axes of Domain for Domain Analysis	0,73	
VAR00054 Concept Theory - Hjørland	0,55	
VAR00060 Linked data - Lee	0,54	
VAR00041 Functional Requirement for Bibliographic Records	0,49	
<hr/>		
VAR00070 The epistemological foundations of knowledge representation	0,79	
VAR00071 The Intellectual Foundation of Information Organization	0,77	
VAR00033 Prolegomena to Library Classification	0,76	
VAR00012 Indexing and abstracting	0,64	C8
VAR00097 Qu'est-ce que la documentation?	0,59	
VAR00063 Principles of Book Classification	0,58	
VAR00075 The subject approach to information	0,57	
VAR00077 Two kinds of power	0,56	
<hr/>		
VAR00029 Co-citation in the scientific literature	-0,95	
VAR00083 Bibliographic coupling between scientific papers	-0,93	C9
VAR00078 Visualizing a discipline	-0,84	
VAR00089 Fast unfolding of communities in large networks	-0,66	
<hr/>		
VAR00084 Binary codes capable of correcting deletions, insertions and reversals	0,74	C10

VAR00046 Real life, real users, and real needs	0,72	
VAR00105 Traité de documentacion	0,86	
VAR00072 The mathematical theory of communication	0,78	C11
VAR00092 Information as thing	0,53	
VAR00107 User tags versus expertassigned subject terms	-0,95	C12
VAR00049 Usage patterns of collaborative tagging systems	-0,92	
VAR00086 Classification and categorization	0,73	C13
VAR00067 Systems of knowledge organization for digital libraries	0,60	
VAR00066 Semiotics and indexing	0,78	C14
VAR00056 Evaluating evaluation measure stability	0,46	
VAR00103 The Need for a Faceted Classification as the Basis of All Methods of Information Retrieval	0,47	C15

APÊNDICE 13. Dendrograma de análise de cluster (primeiros autores, dados absolutos e diagonal com citação total).



APÊNDICE 14. Dendrograma de análise de cluster (primeiros autores, dados absolutos e diagonal com citação total).

