



**UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"**

**A EXAUSTIVIDADE E A ESPECIFICIDADE COMO VALORES ÉTICOS NO
PROCESSO DE INDEXAÇÃO: aspectos conceituais e deontológicos**



Fundação de Amparo à Pesquisa do Estado de São Paulo

**Marília-SP
2016**

ISADORA VICTORINO EVANGELISTA

**A EXAUSTIVIDADE E A ESPECIFICIDADE COMO VALORES ÉTICOS NO
PROCESSO DE INDEXAÇÃO: ASPECTOS CONCEITUAIS E
DEONTOLÓGICOS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista - Unesp, como requisito para a obtenção do título de Mestre em Ciência da Informação.

Área de concentração: Informação, Tecnologia e Conhecimento

Linha de Pesquisa: Produção e Organização da Informação

Orientador: Prof. Dr. José Augusto Chaves Guimarães

Co-orientadora: Prof. Dra. Maria Graça de Melo Simões

Evangelista, Isadora Victorino.
E92e A exaustividade e a especificidade como valores éticos no processo de indexação: aspectos conceituais e deontológicos / Isadora Victorino Evangelista. – Marília, 2016.

143 f. ; 30 cm.

Orientador: José Augusto Chaves Guimarães.

Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2016.

Bibliografia: f. 86-92

1. Organização da informação. 2. Recuperação da informação. 3. Indexação. 4. Ética. I. Título.

CDD 029.7

ISADORA VICTORINO EVANGELISTA

**A EXAUSTIVIDADE E A ESPECIFICIDADE COMO VALORES ÉTICOS NO
PROCESSO DE INDEXAÇÃO: aspectos conceituais e deontológicos**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista - Unesp, como requisito para a obtenção do título de Mestre em Ciência da Informação.

BANCA EXAMINADORA

Prof. Dr. José Augusto Chaves Guimarães (Orientador)
Universidade Estadual Paulista – UNESP

Prof. Dr. Fábio Assis Pinho
Universidade Federal de Pernambuco – UFPE

Dra. Suellen de Oliveira Milani
Universidade Estadual Paulista – UNESP

Dedico este trabalho a Deus, porque se cheguei até aqui, é porque Ele tem me sustentado. “Porque dEle e por Ele, e para Ele, são todas as coisas; glória, pois, a Ele eternamente. Amém.” (Rm 11:36)

Dedico também, a minha avó Maria Cícera (in memoriam), que cuidou de mim quando pequena e continua a fazer, de onde está.

AGRADECIMENTOS

Durante a elaboração desta dissertação, algumas lágrimas foram derramadas. Poucas de nervosismo ou ansiedade, mas muitas de alegria. Certas pessoas contribuíram e muito, para que a segunda opção ocorresse em um número maior de vezes e é por isso que, sem nenhuma pieguice, eu as agradeço aqui.

Aos meus pais. Sei que a jornada pra me trazer até aqui não foi fácil, mas acredito que a cumpriram de forma magnífica. Tudo que sou, devo a vocês.

A toda minha família, por cumprirem tão bem o significado dessa palavra. Obrigada pelo apoio de sempre e por todas as frases de incentivo e segurança que recebi de vocês.

Ao Gian Luca, por ter me apoiado quando inicei meu Mestrado e por continuar me apoiando agora, quando estou concluindo. Por ter ficado feliz pela minha felicidade e incentivado todos os momentos da minha carreira, mesmo quando a distância doía.

A todos meus amigos de Assis, por entenderem que meus afastamentos se faziam necessários, mas que eu poderia contar com vocês em qualquer situação.

Ao meu orientador, Prof. Dr. José Augusto Chaves Guimarães, por ter acreditado em mim há cinco anos e me acompanhado desde então. Por ter me ajudado em tantas conquistas acadêmicas e por sempre, ser meu exemplo.

A minha co-orientadora, Dra. Maria Graça Melo Simões, por ter me recebido não apenas em seu país, mas em sua família. Obrigada por todas as contribuições que com toda certeza, enriqueceram não apenas este trabalho, mas no meu desenvolvimento enquanto pesquisadora.

A minha banca, Prof. Dr. Fábio de Assis Pinho e Dra. Suellen de Oliveira Milani. Minha trajetória de pesquisa foi iniciada por vocês que, juntamente com o professor José Augusto, sempre prezaram pelo desenvolvimento da ética no cenário da Biblioteconomia. Obrigada por terem iluminado meu caminho, para que eu chegasse até aqui.

Ao Grupo de Pesquisa Formação e Atuação Profissional em Organização da Informação, por a todo o momento provarem que o trabalho em equipe é sempre mais prazeroso.

Aos meus amigos Laura, Andrieli e Gilberto, por compartilharmos a mesma jornada e sempre nos ajudarmos durante ela. Obrigada pelo apoio de sempre, pelo teto quando necessário e por nunca me abandonarem quando precisei.

A Fundação de Amparo à Pesquisa do Estado de São Paulo pelo financiamento total de minha pesquisa não só no Brasil, mas também em Portugal. Obrigada pela confiança concedida sempre!

A todos os professores com quem pude compartilhar momentos durante esta trajetória. Obrigada por compartilharem o conhecimento de vocês de forma tão humilde e por fazerem a Pós-Graduação um espaço tão agradável para se aprender.

A Universidade Estadual Paulista e todos seus funcionários, que tiveram grande importância para que eu concluísse este trabalho com êxito.

A Universidade de Coimbra, por ter me aceito e concedido todas as ferramentas necessárias para que eu desenvolvesse minha pesquisa da melhor forma possível.

A Portugal, por ter superado todas as minhas expectativas e demonstrado que é possível se sentir em casa em mais de um país.

RESUMO

Considerando a natureza mediadora dos processos de organização da informação, destaca-se a necessidade de abordagem da dimensão ética desses processos, especialmente no que se refere às consequências para a recuperação da informação. Assim, coloca-se o problema relativo aos conceitos de exaustividade e especificidade na indexação, considerando-se estes não só como simples medidas de recuperação da informação mas também como valores éticos que condicionam a prática do indexador. Com base nos artigos recuperados na revisão bibliográfica e partindo da análise de conteúdo proposta por Bardin (2008), foram considerados os conceitos de exaustividade e especificidade, enquanto valores éticos na área da organização da informação. Os resultados foram analisados de modo a evidenciar as dimensões éticas que se associam à questão da exaustividade e da especificidade na prática da indexação, possibilitando um diagnóstico mais efetivo sobre essa questão. Conclui-se que essas características podem ser entendidas como além de medidas, pois afetam de forma direta a recuperação informacional, estabelecendo a quantidade e a qualidade dos termos representativos, ao mesmo tempo que podem também ser encarados enquanto valores éticos da atividade de indexação.

PALAVRAS-CHAVE: Organização da informação. Ética informacional. Indexação. Representação da Informação. Especificidade. Exaustividade. Valores éticos.

ABSTRACT

Considering the intermediary nature of the Information Organization processes, it is important to point out the ethical components in this context, specially the ones related to Information Retrieval. In this way, this study has the aim to explore two concepts – exhaustivity and specificity – in indexing which can be moving from simple Information Retrieval measures to indexer's ethical values. This study has the goal to analyse the concepts exhaustivity and specificity as ethical values in Information Organization applying the content analysis method proposed by Bardin (2008). The results discussion was developed in a way which could present a scheme stressing the ethical dimensions associated with exhaustivity and specificity in indexing, concluding that these characteristics can be seen beyond mere measures, because they affect directly information retrieval, establishing quantity and quality of representative terms, regarded this way, as ethical values for indexing.

KEYWORDS: Information Organization. Informational Ethics. Indexing. Information Representation. Specificity. Exhaustivity. Ethical values.

LISTA DE QUADROS

Quadro 1 - Ficha de descrição do artigo	58
Quadro 2 - Artigos não recuperados em texto integral	60
Quadro 3 - Artigos excluídos do corpus.....	62
Quadro 4 - Corpus do trabalho.....	64
Quadro 5 - Análise de conteúdo do termo exaustividade.....	69
Quadro 6 - Análise de conteúdo do termo especificidade.....	72
Quadro 7 - Sistematização de valores éticos a partir da análise de conteúdo do corpus.....	74

LISTA DE FIGURAS

Figura 1 - Porcentagem do corpus.....	56
Figura 2 - Artigos publicados por década	66
Figura 3 - Idioma dos artigos	67
Figura 4 - Artigos por periódico.....	68

LISTA DE ABREVIATURAS E SIGLAS

CI – Ciência da Informação

EDS – Ebsco Discovery Search

FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo

IFLA – International Federation of Library Associations and Institutions

LISA – Library and Information Science Abstracts

OC – Organização do Conhecimento

ORC – Organização e Representação do Conhecimento

RCAAP – Repositório Científico de Acesso Aberto de Portugal

SUMÁRIO

1 INTRODUÇÃO.....	14
2 ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO	18
3 ASPECTOS ÉTICOS NA ORGANIZAÇÃO DO CONHECIMENTO	24
4 A INDEXAÇÃO NA ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO: aspectos conceituais e procedimentais	35
5 METODOLOGIA	50
<i>5.1 CRITÉRIOS USADOS NA COLETA DE DADOS</i>	<i>51</i>
<i>5.2 FONTES DE OBTENÇÃO DE DADOS.....</i>	<i>52</i>
<i>5.3 AMOSTRA.....</i>	<i>55</i>
6 RESULTADOS E DISCUSSÃO.....	60
7 CONCLUSÕES.....	78
REFERÊNCIAS.....	85
APÊNDICE 01 - FICHAS DE ANÁLISE DE CONTEÚDO DOS TERMOS	92

1 INTRODUÇÃO

No contexto da Ciência da Informação, o conhecimento insere-se em uma dimensão processual e, uma vez produzido, registrado e para que tenha uma função social, é organizado a partir da utilização de instrumentos de representação documental, para que se torne disponível e acessível e, ainda, para que a sociedade possa dele se apropriar e produzir um novo conhecimento. Tal aspecto constitui aquilo que Guimarães (2008a) denomina movimento helicoidal do conhecimento na Ciência da Informação.

Nessa atividade intermediária de organizar e representar o conhecimento, tem-se, por um lado, uma abordagem descritiva, que se vale de representações de aspectos físicos, como é o caso da catalogação; e, por outro, uma abordagem temática, em que se analisam e se representam os assuntos do documento (aqui entendido como fruto de um conhecimento registrado e socializado). Nesse último contexto, tem-se a construção de classificação e de índices, que atuam como pontes (elementos remissivos) entre o documento e o usuário.

Essa prática de construção de índices vem de longa data, pois, como destaca Witty (1973), os primeiros índices remontam à Mesopotâmia, na Antiguidade. Assim, com o aumento da produção informacional, a partir da invenção da imprensa no século XV e que se acelerou com o desenvolvimento das academias científicas (séc. XVIII), com a imprensa mecanizada (séc. XIX) e o avanço científico e tecnológico após a 2ª. Guerra Mundial, em particular com o desenvolvimento da Internet, a necessidade de acesso temático à informação tornou-se ainda mais evidente e necessária (GUIMARÃES, 2003). Os assuntos tornam-se cada vez mais específicos e inter-relacionados para responder às necessidades de um perfil de comunidades de usuários cada vez mais comprometidos com a multidisciplinaridade e com o *pormenor*, uma vez que esses usuários buscam maior diversidade de assuntos e possuem necessidades cada vez mais específicas. Isso leva a crer que, para uma satisfatória recuperação da informação, que possa atender às necessidades informacionais do usuário, é preciso uma indexação eficaz, que possibilite uma triangulação entre a tematicidade do documento, a representação feita pelo indexador e a necessidade informacional do usuário.

As atividades relacionadas à representação documental (em cujo âmbito se insere a indexação) não são neutras (GUIMARÃES; PINHO, 2007), de tal modo que o profissional da informação deve estar atento aos aspectos sociais, culturais, políticos, enfim, a todo um conjunto de valores e crenças que envolvem esses três polos do triângulo. Isso leva a

reconhecer uma dimensão ética inerente à organização e representação do conhecimento, em que os sistemas de *dever-ser* (COMPARATO, 2006) necessitam ser analisados constantemente.

O campo da Organização e Representação do Conhecimento encontra-se respaldado pelo conjunto de valores que se destacam em determinada sociedade, em um dado espaço e tempo, para reger a convivência dos que a integram, em busca do bem comum.

Valores como a privacidade, a equidade, o acesso à informação, o respeito ao usuário e a garantia cultural, entre outros, já vêm sendo apontados pela literatura da área há quase duas décadas e foram objeto de discussão e sistematização de Guimarães et al. (2008b). Estes apontaram que a precisão e a abrangência da representação temática poderiam representar valores éticos a serem perseguidos. Isso leva a se pensar no papel ético que podem desempenhar fatores como a exaustividade e a especificidade, indo além da tradicional concepção de medidas de recuperação da informação. Pode-se dizer que a busca pela exaustividade e pela especificidade na indexação vai ao encontro de um compromisso ético com o usuário (GUIMARÃES, 2000a), de tal modo que o catálogo de assuntos ou o índice evidenciem, da forma mais adequada possível, a diversidade temática do documento e, em cada tema, a profundidade com que é abordado.

Alguns estudos sobre a exaustividade e a especificidade no campo das políticas de indexação já foram propostos (CARNEIRO, 1985; FUJITA, 2003, 2009, GUIMARÃES, 2004; RUBI, 2004, 2008; BOCCATO, 2009). Estes autores introduziram essa discussão no terreno da ética na organização e representação do conhecimento, cujos aspectos se inserem no universo das decisões (ou dilemas) éticas do profissional indexador (DALHERG, 1992; GUIMARÃES, 2000a, 2006; GUIMARÃES, FERNÁNDEZ-MOLINA, 2003; BEGHTOL, 2002, 2005; OLSON, 2002; PINHO, 2006; MILANI, 2007; GUIMARÃES et al., 2008a, 2008c; LÓPEZ HUERTAS, 2008 e outros).

Considerando esses elementos, este estudo propõe-se a verificar como a exaustividade e a especificidade atuam no processo de indexação e quais são os fatores que lhes são determinantes no contexto da unidade informacional, enquanto valores éticos a serem buscados pelo indexador. Assim, colocam-se as seguintes questões: Em que situações? Em que contexto? A partir de que aspectos? Buscando responder a esses questionamentos, estabelece-se o seguinte problema de pesquisa: em que medida os conceitos de exaustividade e especificidade são encarados, teoricamente, como valores éticos, na literatura internacional da área de Ciência da Informação, e

como essa questão se coloca nos códigos de ética profissional do bibliotecário em distintos países?

Os conceitos de exaustividade e especificidade foram pesquisados na literatura internacional de Ciência da Informação, deles extraindo traços distintivos que os possam caracterizar enquanto valores éticos na área de organização da informação, em uma análise comparativa com sua dimensão deontológica a partir dos códigos de ética do profissional bibliotecário em distintos países. Sendo assim, pretende-se dar continuidade a uma trajetória de pesquisa iniciada em 2012, cujo objetivo foi atualizar e ampliar a categorização de valores e problemas éticos em organização da informação (GUIMARÃES, 2006a; PINHO, 2006; MILANI, 2006, 2007; GUIMARÃES et al., 2008).

Esta pesquisa se justifica pelo fato de: a) inserir-se no universo temático da linha de pesquisa Produção e Organização da Informação do Programa de Pós-Graduação em Ciência da Informação da UNESP, mais especificamente no âmbito das discussões sobre a “atuação profissional na área”; b) contribuir para literatura da área, que ainda é escassa, embora sua necessidade seja premente; c) inserir-se na temática de pesquisa sobre aspectos éticos em organização e representação do conhecimento iniciada e desenvolvida ao longo de duas décadas por Guimarães e seu grupo de pesquisa (FERNANDEZ-MOLINA; GUIMARÃES, 2002; GUIMARÃES; FERNÁNDEZ-MOLINA, 2003; GUIMARÃES et al., 2005, 2008, 2009; FERNÁNDEZ-MOLINA et al. (2005), GUIMARÃES, 2006; PINHO, 2006; GUIMARÃES; PINHO, 2007, 2008; GUIMARÃES; MILANI; PINHO, 2008; MILANI, 2010; PINHO, 2010); d) dar continuidade às pesquisas empreendidas por Guimarães et al. (2008) sobre os valores éticos inerentes ao profissional da informação nas atividades de organização e representação do conhecimento, aprofundando em um aspecto específico: a exaustividade e a especificidade como valores na indexação; e) contribuir para a área, de forma a apontar uma melhor maneira de se utilizar duas atividades importantes, para que haja interação entre a unidade informacional e o usuário.

Como objetivo geral, pretende-se analisar, na literatura internacional da Ciência da Informação, como vêm se construindo os conceitos de exaustividade e especificidade na indexação (mais especificamente, conceitos, características e objetivos), para neles identificar e sistematizar traços descritivos que contribuam para sua caracterização como valores éticos na representação da informação, comparando-os com os códigos de ética profissional bibliotecária.

Por sua vez, como objetivos específicos: a) identificar e sistematizar os elementos que caracterizam a construção conceitual de exaustividade e de especificidade na literatura internacional da Ciência da Informação; b) identificar como se colocam as questões ligadas à exaustividade e à especificidade na representação da informação nos códigos de ética do bibliotecário em âmbito internacional; c) analisar, comparativamente, como se colocam as questões da exaustividade e da especificidade na representação da informação enquanto valores éticos, em uma abordagem comparativa entre a literatura internacional da área e os códigos de ética profissional.

É importante ressaltar que, na presente pesquisa, foram utilizados os termos indexação e linguagens de indexação que caracterizam traços da abordagem anglo-americana, distinguindo-os dos termos da abordagem francesa: análise documentária e linguagens documentárias. Essa distinção se dá pela escolha de referencial teórico e pela adoção do termo na grade curricular, no momento de formação da autora.

A presente dissertação é composta dos seguintes capítulos: o primeiro capítulo trata da introdução; o segundo capítulo explora o campo da organização e representação do conhecimento; o terceiro aborda, de forma específica, os aspectos éticos que permeiam esse campo; o quarto traz uma construção teórica sobre o tema da indexação, indo ao encontro dos pressupostos teóricos da exaustividade e da especificidade; o quinto capítulo apresenta e descreve a metodologia da pesquisa; o sexto traz os resultados e a análise da pesquisa; no sétimo capítulo apresentam-se as conclusões.

2 ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

Este capítulo aborda os aspectos teóricos do campo da organização e representação do conhecimento, buscando definir este domínio a partir de seus autores. Antes de dar início às reflexões sobre o campo da organização e representação do conhecimento como um campo na Ciência da Informação (CI), vê-se necessário caracterizar a Ciência da Informação sob o seu aspecto conceitual.

Inicialmente, recorre-se a Harold Borko (1968), que, referindo-se às mudanças ocorridas na denominação da American Documentation Institute para American Society for Information Science – hoje American Society for Information Science and Technology – definiu Ciência da Informação como “[...] a disciplina que estuda as propriedades e o comportamento da informação, os fluxos informacionais e os meios de processar a informação para otimizar seu acesso e uso” (BORKO, 1968, p. 3, tradução nossa). Assim, partindo da perspectiva de que, para o referido autor, a informação constitui objeto da CI, busca-se conceituá-la.

Para Anderson (1983), a informação *per se* pode ser definida como um conjunto de estruturas de conhecimento que consistem em relações entre conceitos existentes na mente dos seres humanos, que aprenderam como representá-la por meio de documentos de diversos tipos, tarefa essencial para transmissão do fluxo informacional.

Essa concepção é substancialmente ampliada por Buckland (1991), que considera a informação como conhecimento e como processo, e também a caracteriza como coisa, que seriam os dados e documentos, considerados tipos informacionais, pois armazenam o conhecimento.

De acordo com Cleveland e Cleveland (2001), a preocupação com a informação se acentua a partir da década de 1950, em pleno pós-guerra, quando os avanços científicos e tecnológicos dão lugar à denominada “explosão informacional”, o que gerou a necessidade de criação de sistemas para controle e organização desses dados. Ainda de acordo com os autores, a informação se torna tão importante porque ela sempre esteve diretamente ligada ao conceito de controle, o que se intensificou no momento atual, considerado a Era da Informação.

O acesso à informação está fundamentalmente ligado a uma organização fidedigna, que corresponda aos objetivos de determinada instituição. Sendo assim, a falta dessa organização torna a informação caótica, não sendo possível atender às necessidades dos usuários que a buscam.

Cleveland e Cleveland (2001) propõem um ciclo do processo de criação e uso da informação. Para os autores, a informação é primeiramente criada e depois colocada em uso, formando um padrão distinto e repetitivo. Em um primeiro momento, há a criação de um novo conhecimento, registrado em um suporte, para posteriormente passar para uma avaliação editorial ou pelos pares e ser publicado. A posteriori, o documento passa por um controle bibliográfico, que envolve tarefas como indexação, condensação e classificação, para que possa ser disponibilizado em uma biblioteca. Em outro momento, o usuário manifesta uma necessidade informacional e sana sua dúvida a partir do documento retirado da biblioteca, que, após ser lido, inicia a criação de um novo conhecimento, estabelecendo, portanto, o ciclo.

Para Guimarães (2008a), esse movimento nada tem de cíclico, pois no ciclo se retorna ao ponto de origem. Para o autor, observa-se um movimento helicoidal em que o conhecimento, uma vez produzido e registrado, e para que tenha uma função social, é organizado a partir da utilização de instrumentos de representação documental, para que se torne disponível e acessível e, ainda, para que a sociedade possa dele se apropriar e, como consequência, produzir um novo conhecimento. Esse movimento ocorre, como destaca o autor, a partir da informação que, uma vez socializada, gerará nova informação, que subsidiará um novo conhecimento, em um movimento infinito.

Ainda de acordo com Guimarães (1999), três aspectos permeiam a realidade de uma unidade informacional: o fenômeno da globalização; o avanço de novas tecnologias, tornando as informações cada vez mais acessíveis, independentes de tempo ou espaço; e, ainda, um novo perfil de usuário, cada vez mais imediatista, exigente e interativo.

Esses aspectos podem ser observados quando se analisa a Internet, onde informações são disponibilizadas de forma massiva, estando acessíveis livremente ao usuário. Pode-se dizer, portanto, que a informação e o conhecimento ocupam cada vez mais espaço, e a biblioteca tem especial relevância (GUIMARÃES, 2000b).

Para que a informação seja organizada e, de fato, recuperada em unidades informacionais – e de forma essencial para a presente pesquisa, as bibliotecas –, é necessário que se recorra ao aporte teórico da Organização do Conhecimento, expressão utilizada pela primeira vez por Henry Evelyn Bliss, em dois de seus livros: *The Organization of Knowledge and the System of Sciences* (1929) e *Organization of Knowledge in Libraries and Subject Approach to Books* (1933). Posteriormente, o termo foi resgatado pelos alemães: Dagobert Soergel (1971) e

Ingetraut Dalhberg (1973), em suas respectivas teses de doutorado.

Hjørland (2008) conceitua a organização do conhecimento como sendo um campo de estudos sobre sistemas e processos que abrangem atividades como indexação, classificação e descrição documental, em ambientes como bibliotecas ou arquivos. No entanto não possui um corpo teórico próprio.

Anteriormente, Hjørland (1994) já havia se referido àquilo que denominou nove princípios que cercam a organização do conhecimento: percepções ingênuas em estruturas do conhecimento não são possíveis em ciências mais avançadas; categorizações e classificações devem unir assuntos parecidos e separar assuntos diferentes; para propósitos práticos, pode ser organizado de maneiras e níveis diferentes; as categorizações devem refletir seus objetivos; categorizações e classificações científicas sempre podem ser questionadas; o conceito de “polirepresentação” é importante; em certos graus, artes e ciências podem ser vistas de maneiras diferentes ao organizar o mesmo fenômeno; a natureza das disciplinas varia e a qualidade da produção em várias disciplinas possui sérios problemas.

Buscando definir o campo da Organização e Representação do conhecimento, Guimarães (2000b) afirma que ela se caracteriza por ser um conjunto de procedimentos que refletem sobre conhecimentos socializados – registrados – que variam de acordo com seu contexto de produção ou sua finalidade e que devem ser considerados no momento da sua representação.

Ainda buscando definições, Novellino (1996, p.38) diz que o “[...] desenvolvimento e avaliação de teorias para análise de determinadas áreas de assunto visando à elaboração de instrumentos e métodos para a representação das informações geradas nessas áreas de assunto.”

Barité (2001), em um texto seminal, sintetiza teoricamente o campo de organização do conhecimento, ao propor as seguintes premissas:

- o conhecimento é produzido e utilizado, sendo um dínamo social;
- o conhecimento decorre da informação e, quando socializado, pode se transformar em uma nova informação;
- a estrutura do conhecimento forma um sistema aberto;
- o conhecimento deve ser organizado para sua melhor utilização;
- a forma de organizar o conhecimento depende da sua produção e uso;

- a organização do conhecimento é feita artificialmente, passível de mudança dependendo do tempo, e de forma determinista, pois “faz lei” em um dado espaço e tempo;
- o conhecimento materializado possui usos distintos;
- o conhecimento é representado por meio de conceitos;
- a organização do conhecimento pode possuir fins científicos, funcionais ou de documentação, e
- as leis que organizam o conhecimento são uniformes, podendo ser utilizadas em qualquer disciplina.

Uma vez organizado e representado, o conhecimento produzido pela sociedade e por ela materializado se presta a ser recuperado, acessado e apropriado, como dínamo para a geração de um novo conhecimento. Assim, no momento da recuperação, como destaca Lancaster (1993), o que se busca é a satisfação de uma necessidade informacional, o que ultrapassa uma dimensão meramente quantitativa para assumir uma dimensão qualitativa por natureza. Assim, “relevância” e “pertinência” são conceitos inerentes a um sistema de organização e representação do conhecimento, já que uma recuperação que incorpore tais aspectos estará intrinsecamente ligada a documentos que contribuirão para sanar as dúvidas ou necessidades informacionais do usuário.

O autor demonstra esse momento da busca informacional em um centro de informação, a partir da necessidade informacional encontrada pelo usuário, que solicitará a ajuda de um especialista para que essa dúvida seja resolvida. Dessa forma, o usuário chega com um pedido para o bibliotecário, que irá confrontá-lo com uma estratégia de busca, valendo-se, por exemplo, de termos de indexação, para que, então, ocorra a recuperação de documentos *pertinentes* àquilo que o usuário busca. É possível perceber que muitos são os fatores que influenciam esse momento, mas, principalmente, a indexação, os elementos do vocabulário e as estratégias de busca (LANCASTER, 1993).

No âmbito da ORC, especial lugar ocupa a sua dimensão temática, que busca elementos do conteúdo documental para posterior recuperação, englobando processos como a classificação e a indexação.

Para Hjørland (1997), o assunto é o potencial informativo de um documento, e a análise

desses assuntos envolve a construção conceitual destes e seus potenciais informativos.

Segundo Guimarães (2009a), a representação de assuntos é um dos pontos fundamentais para o estudo da ORC, levando em conta que essa representação possui a capacidade de agregar valor ao documento e criar um substituto – termos representativos – do conhecimento gerado.

Essa questão de *substitutos do conhecimento (surrogates of knowledge)* também é abordada por Cubillo (1997), na medida em que um documento atua como representante de ideias, conhecimentos empíricos ou opiniões de um autor. Quando esse universo é representado por meio de termos, eles passam a ser substitutos do conhecimento.

Olson (2002) também trata essa questão de substitutos do documento, uma vez que, quando “se nomeiam” termos, se criam representantes do assunto tratado naquele documento. De acordo com a autora, quando se determinam esses substitutos, controla-se a forma com que se representam os assuntos do documento, portanto, controla-se o acesso.

De acordo com Novellino (1996), a principal característica da representação da informação e, mais especificamente, da representação temática da informação é a *substituição* de uma entidade linguística longa, como o texto de um documento, para uma forma reduzida, por meio dos termos, de forma a apresentar apenas a essência dos termos. Esse processo contém dois passos: a análise de assunto e a atribuição de conceitos.

Para Sousa (2013), a representação temática da informação é realizada por meio da análise de assunto, como forma de entender o que trata o documento e produzir a informação documentária que será disponibilizada nos catálogos da unidade informacional. O autor relata que esse processo é puramente intelectual, mas que os bibliotecários devem estar atentos às metodologias da operação. Dessa forma, é necessário que existam processos que amparem essa identificação.

Nesse contexto, tem-se, por um lado, a classificação que, para Cleveland e Cleveland (2001), é o arranjo das individualidades de um documento em grupos ou categorias, a partir de um critério pré-estabelecido, de tal modo que se possam separar as diferenças e agrupar as semelhanças, reunindo documentos que possuam o mesmo assunto, para facilitar no momento da busca informacional. Por outro lado, e inserindo-se no objeto desta pesquisa, tem-se a indexação, enquanto processo de tradução dos assuntos de um documento em termos documentais, controlados ou não por uma linguagem. Esse processo será tratado de maneira mais detalhada no quarto capítulo.

Todo esse universo vai muito além dos processos, instrumentos e produtos para atingir uma dimensão efetivamente profissional, em que o modo de agir, as crenças, as concepções, os valores têm função determinantes. Tem-se, aqui, uma dimensão ética inerente à organização do conhecimento.

3 ASPECTOS ÉTICOS NA ORGANIZAÇÃO DO CONHECIMENTO

A busca pelo bem-estar nas diferentes sociedades, ao longo da história, tem sido objeto da ética, que, pautando-se no estudo do comportamento humano a partir dos conceitos de bem e de mal, visa à coexistência em harmonia, independente de quais sejam as regras que qualificam essa boa convivência.

De acordo com Nalini (2004), ética pode ser conceituada como a ciência do comportamento moral, relacionada aos atos humanos na convivência em sociedade. Sua classificação como ciência decorre do fato de possuir leis e métodos próprios, assim como tem na moral e nas moralidades o seu objeto.

Inserindo-se no domínio comum dos costumes, a ética pode, conceitualmente, dividir-se em duas vertentes: a ética anterior, que designa o montante das normas, nas quais elas se enraizam; e a ética posterior, responsável pelo que está jusante delas, ou seja, a vista de aplicar as leis e normas em situações concretas (NALINI, 2004).

Para Maliandi (2004), deve-se ter em mente que a ética é a tematização do *ethos*, o fenômeno da moralidade, um substantivo que nomeia uma disciplina em particular. Essa tematização ocorre por meio de um caráter reflexivo. Com efeito, a ética é uma das formas em que o homem se auto-observa: suas atitudes, convicções, crenças e modos de agir, aqui não se considerando o homem apenas individualmente, mas também um grupo social. A Ética é, ainda, ontológica, ou seja, um pensar que tem sua morada no homem enquanto ser.

Na Filosofia, a ética se define por meio de uma premissa principal – se deve fazer – e duas perguntas consideradas primordiais: o que se deve fazer e por que se deve fazer? Dessa forma, os princípios que regem a vida humana devem ser fundamentados a partir de normas (MALIANDI, 2004).

Marchioni (2008) define a ética enquanto sendo “a arte do bom” e justifica essa escolha porque os seres humanos em busca da felicidade almejam coisas boas; e a ética verifica o que é bom e reflete sobre o fato.

Segundo Blackburn (1997), em seu Dicionário Oxford de Filosofia, a ética consiste no “[...] estudo dos conceitos envolvidos no raciocínio prático: o bem, a ação correta, o dever, a obrigação, a virtude, a liberdade, a racionalidade e a escolha.” (BLACKBURN, 1997, p. 129). Para o autor, em situações como aborto, eutanásia, tratamento dado aos animais, dentre outras, a questão ética se torna cotidianamente visível na sociedade, atingindo problemas de ordem legal,

política, social e do meio ambiente.

Para Japiassú (2008), a ética pode ser entendida como

Parte da filosofia prática que tem por objetivo elaborar uma reflexão sobre os problemas fundamentais da moral [...], mas fundada em um estudo metafísico do conjunto das regras de conduta consideradas como universalmente válidas [...] a ética está preocupada em detectar os princípios de uma vida conforme a sabedoria filosófica, em elaborar uma reflexão sobre as razões de se desejar a justiça e a harmonia e sobre os meios de alcançá-la. (JAPIASSÚ, 2008, p. 97)

Ao longo do tempo, a ética foi estudada por diferentes filósofos de diferentes escolas de pensamento. Pinho (2006) sistematiza essa trajetória, destacando os trabalhos de: a) Thomas Hobbes (1588-1679) e sua forma de entender a conduta do ser humano como uma forma de preservar a si; b) Baruch Espinosa (1632-1677) e a necessidade em desejar o bem a si como forma de conservar seu bem-estar; c) John Locke (1632-1704) e sua defesa à experiência enquanto método de busca da conservação do ser; d) Gottfried Wilhelm Leibniz (1646-1716), que produziu uma doutrina onde existem verdades inatas; e) David Hume (1711-1776), que criou a filosofia fenomenista e refletiu sobre os *móveis* da conduta humana; f) Immanuel Kant (1724-1804), autor da obra *Crítica da Razão Pura*, demonstrando que o cérebro – razão – prevalece, e que essa é a origem das ações éticas, regidas por *imperativos categóricos*; g) Jeremy Bentham (1748-1832), a moral enquanto uma das fontes que trazem prazer e dor; h) Henri Bergson (1859-1941), que determina a ética enquanto uma finalidade a ser perseguida, e definiu moral aberta e moral fechada; i) Max Scheler (1874-1928) e Edward Von Hartmann (1842-1906), que discorreram sobre as noções de bem enquanto valores éticos.

Guimarães et al. (2009b) destacam que a ética, como campo de estudos, surgiu na Grécia, na Antiguidade Clássica, a partir de Aristóteles, que a definiu enquanto disciplina filosófica; posteriormente, foi incorporada à ética cristã durante toda Idade Média; e o distanciamento só ocorreu com o Iluminismo, chegando no século XX, com Immanuel Kant e a criação do imperativo categórico, enquanto disciplina baseada no livre-arbítrio e em valores constituídos a partir das conveniências do tempo e sociedade.

Como objeto de reflexão da ética tem-se a moral. Para Maliandi (2004), a reflexão sobre o que é bom/mau, sobre o que se deve ou não fazer, é complementada a partir da moral. Ela pressupõe princípios e os aplica em determinadas situações, e é aqui que a pergunta determinada pela ética sobre o que se deve fazer é respondida. A moral ainda julga a partir do

caráter de atos particulares, ou seja, os valores.

Ainda de acordo com Maliandi (2004), a moral é, então, a parte concreta da ética, que remete a fundamentos universais, como os apresentados por Kant, que são criados a partir dos costumes de determinada comunidade, expressão de uma determinada exigência para a ação. A moral possui profunda relação com os aspectos deontológicos da ética.

Dessa forma, a ética é a reflexão teórica sobre a boa convivência humana, que possui sua parte prática nas questões morais, a partir de valores morais determinados de acordo com a cultura e costumes de determinadas sociedades.

Guimarães et al. (2009b) destacam que a diferença entre ética e moral ocorre na medida em que a moral se encontra relacionada com as ações que são desenvolvidas em um determinado universo, em uma delimitação espaço-temporal, enquanto a ética busca refletir sobre a moral e, portanto, acompanha a moral ao longo dos anos e das diferentes sociedades e culturas.

Segundo o Dicionário de Ética e Filosofia Moral (2007), a moral caracteriza-se pelas suas funções de designar as normas – os princípios do que é permitido e do que é proibido - e a obrigação subjetiva que um sujeito tem em relação às ditas normas.

Já no Dicionário Oxford de Filosofia (BLACKBURN, 1997), a moral se restringe a estudos, como os kantianos, em que os sistemas se baseiam em noções de dever, obrigação e princípios de conduta, enquanto a ética se refere a estudos mais aristotélicos, baseados em noções de virtudes.

Como destaca Japiassú (2008), a moral relaciona-se à “[...] teoria dos valores que regem a ação ou conduta humana, tendo um caráter normativo ou prescritivo [...] que diz respeito aos costumes, valores e normas de conduta específicos de uma sociedade ou cultura” (JAPIASSÚ, 2008, p. 193).

Segundo Nalini (2004), a moral é o objeto da ética e se caracteriza por ser um dos aspectos do comportamento humano, diretamente ligada aos costumes, que assumem uma dimensão normativa (do latim *mores*) em sua prática.

Se moral é o conjunto de normas e valores de uma sociedade ou cultura, vê-se necessário conceituar o que a literatura traz sobre valores.

Japiassú (2008) conceitua valor como possuindo um caráter positivo, ligado àquilo que é bom, útil, positivo e, ainda, do ponto de vista da axiologia, a algo que deva ser realizado. Eles acabam sendo os fundamentos da moral.

Para Blackburn (1997), quando se considera um valor, leva-se em conta sua capacidade nas tomadas de decisão ou, ainda, o quanto se está inclinado a utilizá-lo como orientação a todos os indivíduos.

Relacionado aos valores, Milani (2010) destaca que, no âmbito da ética, os valores passam a ser considerados enquanto normas ou princípios que são designados a partir de uma sociedade composta por seres humanos que agem conscientemente e voluntariamente. Embora os valores dependam culturalmente de cada sociedade, alguns deles são estendidos globalmente, e o julgamento moral (ou a reflexão ética) será admitido a partir dos cidadãos, que buscam um convívio em harmonia.

Relativamente ao estudo dos valores morais em universos específicos, como os entornos profissionais, tem-se o campo da deontologia, que se caracteriza por ser um conjunto de preceitos morais que devem ser seguidos por um grupo de profissionais e serão definidos a partir de um código ou, ainda, denominados enquanto ética profissional.

Para Souza (2009, p. 135), o termo deontologia é um resignificado da forma como é pensada a ideia de dever ou obrigação, tendo adquirido “[...] o sentido de representar os deveres que são assumidos pelas pessoas ao exercerem papéis ou atribuições profissionais como decorrência de sua atuação na sociedade”. O termo pode ainda ser entendido enquanto ciência do dever e como os profissionais em exercício levam em consideração a maneira de se portar diante dos usuários.

De acordo com Guimarães et al. (2009b), a ética profissional busca se restringir aos valores, que têm por objetivo um “agir profissional correto e adequado”, mesmo que isso interfira nos valores pessoais; e, ainda, demonstrar as “responsabilidades profissionais”, para que ocorra uma maior credibilidade dos serviços prestados pelo profissional à sociedade.

Segundo Gomes (2009),

O comportamento ético vindica a opção pela prática do bem comum, da paz e da felicidade social. Em termos deontológicos, qualquer profissão deve ser exercida com a consciência da escolha de ações pautadas em objetivos irmanados à opção por um estar ético no mundo, que impõe o respeito à cidadania e a resistência às ações de violência, sejam elas simbólicas ou materiais, como também de abuso de poder (GOMES, 2009, p. 149).

A deontologia, etimologicamente considerada como “ciência do dever”, tem seu uso mais atualmente ligado ao exercício de determinada profissão, enquanto “[...] código moral das regras e procedimentos próprios a determinada categoria profissional” (JAPIASSÚ, 2008, p. 68).

Ao analisar o Código de Ética brasileiro da profissão do bibliotecário, Gomes (2009) identificou valores relacionados ao bom exercício da profissão: as práticas que dignifiquem, moral, ética e profissionalmente, o trabalho do bibliotecário; o respeito às leis e normas vigentes e aos colegas de trabalho, sendo leal e solidário e contribuindo para sua qualificação profissional; a contribuição ao desenvolvimento do país; o excelente atendimento ao público, sem diferenciações; assegurar que seu trabalho seja desenvolvido de forma excelente e que seja remunerado de forma justa; orientar no momento da execução de pesquisas, como também a normalização de trabalhos intelectuais.

A autora cita como deveres: a não convivência com erros, alertando de forma respeitosa quem os cometem; evitar os comentários maldosos, acusando apenas quando houver provas, para que seja elaborada uma denúncia junto ao Conselho; respeitar as ideias e as soluções trazidas pelos demais colegas de trabalho; e não substituir colegas que tenham desistido da vaga por dignidade profissional, caso essas condições se mantenham (GOMES, 2009).

No âmbito internacional, Fernandez-Molina e Guimarães (2002) analisaram 16 códigos de ética de profissionais da informação dos 5 continentes, identificando os seguintes valores intrínsecos à profissão: a prioridade dos interesses do usuário; a objetividade no fornecimento dos serviços (sem *bias*); a atualidade e a precisão no fornecimento da informação; a ausência de censura na seleção de materiais da informação (ou a informação de sua existência, quando imprescindível); a separação das crenças pessoais e dos serviços profissionais; e a manutenção da competência profissional.

Guimarães (2000a), ao considerar as demandas de um mundo cada vez mais globalizado, alerta para a necessidade de uma nova postura do bibliotecário, assumindo cinco instâncias de compromissos éticos: com o usuário, com a organização, com a informação, com a profissão e consigo mesmo.

Trazendo a questão dos compromissos éticos para o universo específico da organização e

representação da informação, Guimarães et al. (2008b) realizaram uma pesquisa, no período de 1994 a 2005, a partir da literatura internacional da área, em periódicos como: *Journal of the American Society for Information Science and Technology*, *Journal of Documentation*, *Cataloging and Classification Quarterly*, *The Indexer*, *Ethics and Information Technology* e *Knowledge Organization*. Identificaram valores que permeiam a atividade profissional do bibliotecário, dividindo-os em três grupos: valores maiores (ou supervalores), que permeiam toda a atividade informacional; aqueles antes havidos como requisitos profissionais, na medida em que integram a essência do fazer profissional na área; e aqueles antes havidos como meras medidas da ORC, mas que hoje se integram ao universo axiológico da área. Ainda, identificaram problemas – ou não valores – presentes na profissão, divididos em dois grupos: problemas que permeiam o mundo atual e aqueles que afetam diretamente as atividades da ORC, quando do exercício profissional.

Utilizando a mesma metodologia, mas ampliando o universo de pesquisa para os anais da International Society for Knowledge Organization e para o capítulo norte-americano - North American Symposium on Knowledge Organization -, Guimarães, Milani e Evangelista (2015) ampliaram esse universo até o ano de 2012. Dessa forma, com a união das duas pesquisas, chega-se ao seguinte grupo específico de valores:

- Respeito à diversidade do usuário - garantia cultural, multiculturalidade (adequação da indexação às linguagens dos usuários, atenção às minorias étnicas, preservação da identidade das comunidades discursivas), multilinguismo, preservação da privacidade do usuário, postura mediadora na indexação/equidade.
- Respeito à diversidade da linguagem de indexação - hospitalidade cultural, respeito à natureza não estática das estruturas de organização do conhecimento, respeito ao propósito e às bases filosóficas específicas de cada linguagem de indexação (área temática, comunidade usuária, valores institucionais).
- Respeito à diversidade do documento - busca pela ausência de censura; percepção de traços políticos inerentes à coleção; respeito à diversidade de conteúdo e de abordagens do documento; respeito ao direito autoral; competência profissional (eficiência, cooperação, confiabilidade dos processos desenvolvidos e dos produtos gerados); *precisão* (entendida aqui como parte da especificidade), *exaustividade*, facilidade de uso dos produtos gerados e consistência.

No que diz respeito aos problemas, foram identificados os seguintes:

- Quanto à linguagem - categorias dicotômicas, categorias remanescentes, gerando empecilhos ao fluxo informacional, falta de garantia cultural, moralidade classicista determinando as estruturas principais das classificações, terminologia preconceituosa, traduções inadequadas, vocabulários controlados, desatualizados, com lacunas, discriminatórios e anglo cêntricos.
- Quanto à postura profissional - *biases* na representação de assunto, censura, desconhecimento das necessidades informacionais do usuário, despreocupação com questões de gênero, direcionamento informacional (com o objetivo de produzir silêncio, dificultando a recuperação), distorção da informação, falta de espírito profissional coletivo, marginalização na representação, negligência, parcialidade não declarada, utilização de hierarquias reforçando preconceitos, crença em linguagens estáticas, crença na necessidade de linguagens universais, crença na neutralidade dos processos, produtos e instrumentos de representação, crença na objetividade como sinal de imparcialidade na representação de assunto, crença na tradição cartesiana como única lógica de classificação.
- crença no bem - fazer do indexador como algo intrínseco ao profissional; manutenção da visão ocidental em contextos orientais; presunções teóricas (tomar postulados como verdade sem reflexão) quanto à unidade ou sistema informacional (falta de organização informacional em sites de busca, gerando inacessibilidade); sistemas de browsing genéricos; difamação; veiculação inadvertida de plágios; violação da privacidade (incluindo vigilância, monitoramento informacional e censura); violência (quando a unidade de informação viola os direitos do usuário, a ponto de causar-lhe dano).

Essa necessidade de um olhar ético sobre as atividades de organização e representação do conhecimento já havia sido alertada por Dahlberg (1992), em um editorial do periódico *International Classification*, denominado “Ethics and knowledge organization: in memory of Dr. S. R. Ranganathan in his centenary year”. Nele, a autora refere-se às questões éticas que se encontram presentes no dia a dia, atingindo inevitavelmente as atividades de organização do conhecimento, em que se deve dar ouvido à “voz na consciência”, para evitar situações

manifestamente antiéticas: (1) quando os profissionais se utilizam de índices elaborados pelo próprio autor sem nenhuma reflexão a respeito; (2) quando o indexador não tem zelo suficiente no momento de suas representações, induzindo a erros; (3) ou quando se faz um investimento considerável em um sistema ou instrumento de indexação sem que esse dê os resultados esperados ou prometidos, levando à suspeição de fraude.

Hudon (1997) traz à tona a questão do multiculturalismo ao ponderar que um tesouro multilíngue deve ser aquele em que todas as culturas a que se destina se sintam nele representadas, de tal modo que a transferência de um termo para uma diferente língua tenha equivalência naquela cultura, pois a tradução como ato mecânico de pura correspondência pode levar a equívocos por conta do diferente significado cultural que um termo ou conceito pode ter em diferentes culturas.

Partindo de uma crítica a um pensamento grego dicotômico, mutuamente excludente, linear e teleológico, inspirado por Parmênides, Platão e Aristóteles, ainda presente nas classificações dos dias atuais, Olson (1999) alerta para um fato: embora o processo de classificar seja natural, as classes não o são. Assim, Olson (2009) observa influências políticas e econômicas no estabelecimento de classificações científicas, como é o caso da utilização da escala Fahrenheit, nos Estados Unidos, enquanto o Celsius é a medida de temperatura utilizada no resto do mundo; ou, ainda, uma escala que meça apenas a velocidade do vento ser considerada o suficiente para medir tragédias causadas por furacões, como acontece com a escala Saffir- Simpson.

Em seu livro *The Power to Name*, Olson (2002) deixa claro como a representação do conhecimento é imprescindível para a recuperação da informação e como os profissionais da informação estão imbuídos de poder no momento em que “nomeia” a informação. Esse poder de nomear reflete aspectos, como os valores sociais. Segundo a mesma autora, o poder de nomear e o poder de deixar “*unnamed*” refere-se à falta de neutralidade na representação, visto que o catálogo é uma ferramenta subjetiva, construída por humanos passíveis de erros. A autora afirma que, apesar de os estudos em representação já terem mais de 70 anos, os problemas continuam os mesmos: termos muito gerais ou a falta deles e confusões no momento de estabelecer os números de classificação. Para Olson, as classificações têm pretensões de ser universais mas são culturalmente específicas e refletem as culturas dominantes do universo em que estão inseridas. Tais aspectos levam à necessidade de uma abordagem ética na

organização e representação do conhecimento, como se observa nas temáticas ligadas ao feminismo, em que Olson (2007) pontua que a mulher não é representada da mesma forma que o homem nas classificações e, portanto, acaba sendo marginalizada sob a ótica da lógica. Esse raciocínio hierárquico é fortemente visto em ferramentas de organização do conhecimento, como a Classificação Decimal de Dewey e o *Anglo American Cataloging Rules*.

Para a autora, existem três correntes de pensamento feminista na representação da informação: um grupo que quer incluir a mulher na lógica tradicional; outro que rejeita completamente a lógica tradicional; e outro, utilizado por Olson, em que se abandona o pensamento de que existe uma única forma correta de representação e atribuem-se novos modelos. Dessa forma, propõe a autora uma forma de representação conectada, focada em relações, como já pode ser visto nas relações associativas dos tesouros e cabeçalhos de assunto; na Classificação Decimal Universal, na qual os pontos e notações estabelecem as facetas daquele assunto; no *Functional Requirement for Bibliographic Records* e em seu relacionamento entre entidades; e nas folksonomias, que são formas de representação por meio de tags informadas pelo próprio usuário daquela informação (OLSON, 2007).

Beghtol (2002, 2005) traz a noção de “*cultural warrant*” ou garantia cultural. Para a autora, deve-se lidar diariamente com o problema de prover acesso à informação em instância tanto local quanto global, independente do indivíduo, cultura, domínio, a qualquer momento e por qualquer propósito. Essas preocupações têm sido mais discutidas atualmente, devido ao amplo desenvolvimento de novas tecnologias.

Para a autora, conceitualmente falando, cultura são as crenças coletivas e atividades de determinado grupo, e, geralmente, seus aspectos centram-se em seus valores, histórias, linguagem, memória coletiva, atividades sociais, preferências, práticas, etc. Embora quase sempre uma cultura diga respeito à nacionalidade do indivíduo, diferentes culturas podem coexistir dentro de um mesmo país, principalmente quando dizem respeito à religião, grupo econômico e grupos de discussão. Isso demonstra que um mesmo indivíduo pode fazer parte de diferentes culturas ao mesmo tempo e que as necessidades informacionais de cada cultura são complexas e não claramente entendidas.

A garantia cultural é aquela que permite que sistemas e ferramentas de representação do conhecimento sejam amplamente úteis para seus usuários, por se basearem nos valores, pressupostos e predisposições da cultura da comunidade usuária, pois esses sistemas e

ferramentas são, de fato, artefatos culturais (BEGHTOL, 2002).

Beghtol (2005) alerta para a questão das *bias* enquanto ponto de vista, ressaltando que em um bom sistema de representação do conhecimento se deve ter em mente todos os tipos de tendências possíveis, para que sejam evitadas distorções ou privilégios. E para que isso seja possível, é necessário que o sistema seja concebido a partir de uma “*cultural inclusiveness*”, buscando adequar-se culturalmente ao maior número de possíveis grupos de usuários.

Antonio Garcia-Gutiérrez (2002), ao postular a necessidade de uma epistemografia interativa, por meio de uma ética transcultural de mediação, refere-se a culturas de fronteira, em que os sistemas de organização e representação do conhecimento possam ser mais abertos e flexíveis, e não haja limites para o estabelecimento de novas identidades. Para ele, a indexação deve ser entendida enquanto um processo hermenêutico, ou seja, interpretativo. Outro termo cunhado pelo autor é o que ele denomina “*exomemória*”, refletindo sobre as memórias externas construídas pelo profissional no momento da mediação (e aqui se entendendo a organização e a representação como atividades mediadoras por natureza).

Essa influência da cultura nos processos de organização e representação do conhecimento é também tratada por López-Huertas (2008a), ao analisar os estudos de gêneros no Uruguai e Espanha, concluindo que os aspectos socioculturais afetam a representação de áreas específicas, sendo necessário que os sistemas sejam construídos com base em garantia literária, com a criação e incorporação de novas categorias gerais e específicas (em que termos como *gênero* podem criar novas terminologias em diferentes disciplinas), e contemplando as diferenças culturais no momento da elaboração de um sistema de organização do conhecimento.

Para a autora, a pesquisa acerca das questões culturais torna-se premente no campo de organização e representação do conhecimento, de tal modo que aspectos como o multilinguismo, a diversidade sociocultural, a busca pela universalidade sem prejuízo da diversidade, a integração de diferenças socioculturais e a representação de minorias, entre outros, tornam-se cada dia mais prementes, carecendo de uma ampla abordagem ética a respeito (LÓPEZ-HUERTAS, 2008b). Deve-se atentar ao que está por trás dos sistemas utilizados globalmente, e se eles refletem pontos de vista e crenças de economia e culturas dominantes, sendo imprescindível zelar pelas necessidades informacionais de culturas marginalizadas e sua efetiva representação em sistemas globais, pois esses usuários possuem os mesmos direitos de acesso à informação.

Pinho (2010) demonstrou sua preocupação em relação aos aspectos éticos da Organização do Conhecimento, evidenciando a falta de precisão das linguagens de indexação brasileiras, no que tange às temáticas relativas à homossexualidade masculina. Para o autor, considerando que a representação do conhecimento deve ser realizada da maneira mais precisa possível, como evitar que um grupo que tenta se solidificar e é culturalmente marginalizado sofra eufemismos e metáforas nas linguagens de representação? O autor comparou termos encontrados em periódicos científicos renomados, relacionados a homossexuais masculinos, com termos definidos pelos principais vocabulários controlados brasileiros, concluindo que essas linguagens não possuem um nível de precisão adequado no que tange ao universo da homossexualidade masculina.

Em sentido semelhante, Milani (2010) analisa a representação da questão feminina em linguagens documentais brasileiras. Considerando um grupo de termos relativos ao âmbito das questões femininas, fez-se uma comparação com as principais linguagens documentais brasileiras, verificando uma série de *biases* - tendenciosidades - nos termos e em suas hierarquias, revelando uma visão retrógrada da mulher na sociedade.

Tendo em vista que as reflexões éticas têm conquistado seu espaço no âmbito da organização do conhecimento, evidenciando, cada vez mais, sua relevância, demonstrar-se-á, a seguir, como se constrói o campo da indexação, enquanto prática da representação temática do conhecimento, para que, posteriormente, se analisem os conceitos de exaustividade e especificidade enquanto valores éticos.

4 A INDEXAÇÃO NA ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO: aspectos conceituais e procedimentais

O presente capítulo refere-se aos aspectos teóricos do processo da indexação, ao mesmo tempo em que se procede ao paralelo destes com os aspectos semióticos considerados neste campo, indo ao encontro dos princípios teóricos da exaustividade e da especificidade.

Em termos históricos, Cleveland e Cleveland (2001) afirmam que a civilização se inicia quando é capaz de reconhecer sistemas por meio de símbolos, prova disso são os desenhos pictóricos nas paredes das cavernas e a evolução deles para mensagens cada vez mais complexas. Considerando o processo da indexação enquanto tradução de assuntos para termos – símbolos – controlados ou não, percebe-se que esse tipo de comunicação está presente desde o início das civilizações, sendo considerado um marco no desenvolvimento do homem moderno.

O interesse da indexação remonta à Mesopotâmia, onde os documentos eram guardados em envelopes com fins de preservação. Para que estes não fossem abertos sem necessidade, buscou-se colocar selos dizendo sobre o quê os envelopes diziam (WITTY, 1973). Posteriormente, países como a China e a Índia começaram a manter registros do que acontecia na sociedade, de forma a ordenar a memória, como, por exemplo, nascimentos, casamentos e mortes, se expandindo, mais tarde, para ofícios religiosos, documentos governamentais e contratos entre pessoas, demonstrando que esses registros deveriam ser organizados para tornar seu acesso facilitado (CLEVELAND; CLEVELAND, 2001).

A norma ABNT/12676 (1992), no seu parágrafo 3.4, descreve indexação como o “[...] ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação.” De acordo com a mesma Norma, entende-se por indexação um processo que consiste na análise conceitual, identificação de assuntos de um documento e a tradução desses assuntos em termos representativos, que podem ou não ser controlados, dependendo da linguagem de indexação que se usa na sua representação. Essa operação tem por fim “[...] disponibilizar, com objetividade, o conteúdo dos documentos, o pensamento dos autores, em uma palavra, abrir caminho para que se venha a obter uma informação de qualidade” (MENDES; SIMÕES, 2002, p. 11).

Enquanto processo, a indexação tem sua importância evidenciada quando, a partir dela, a utilidade de um documento é reconhecida, e tal situação apenas ocorre quando um documento

sanar as necessidades de informação de quem o busca (BONURA, 1994). Para Lancaster “[...] os processos de indexação identificam o assunto de que trata o documento” (1993, p. 1), ideia também partilhada por Borko e Bernier (1978), quando afirmam que os assuntos são as principais ideias do documento, aquelas tratadas como centrais pelo autor.

O processo da indexação envolve fases, como análise de assunto, representação desses assuntos por meio de termos, traduzidos posteriormente por uma linguagem controlada, aproximando-se do processo de classificação na medida em que os dois processos tenham como objetivo a recuperação informacional de maneira efetiva e partem da análise temática do documento (CLEVELAND; CLEVELAND, 2001).

Esses instrumentos usados para a representação e recuperação da informação são denominados linguagens de indexação. Segundo Langridge (1969), são os meios através dos quais o indexador traduz a linguagem natural do autor, encontrada no documento, em uma linguagem artificial, distinta da linguagem de busca do usuário no momento da recuperação. Esse processo é denominado, por Borko e Bernier (1978), “guide words” e consiste na tradução dos assuntos do documento, que se encontra em linguagem natural para uma linguagem controlada, através de uma linguagem de indexação, dando origem a pontos de acesso que remetem para o respectivo documento (LANCASTER, 1993).

Essa tradução é realizada com base em importantes razões, das quais se salientam: estabelecer relações entre documentos representados e recuperados por meio dos mesmos termos ou, ainda, para se evitar dispersão de termos ou termos correspondentes sendo indexados de forma diferente (BORKO; BERNIER, 1978, ANDERSON, 1983).

Um vocabulário controlado é basicamente uma lista autorizada de termos que compõem em sua forma uma relação semântica, destinada principalmente a controlar sinônimos, diferenciar palavras que se escrevem da mesma forma e reunir termos que possuem significados interligados entre si. Essa lista autorizada deve obedecer a algumas regras, que devem estar expressas no momento da busca, de forma a tornar a indexação consistente; é uma linguagem artificial baseada na linguagem natural, que possui, dadas as suas características, grande complexidade, principalmente no que se refere à ambiguidade. As principais características de uma linguagem controlada são: representar conceitualmente um assunto e guiar o usuário no índice; ser o mais perto possível da linguagem natural; possuir frases pré-coordenadas; obter controle de sinônimos; definir termos ambíguos quando necessário; e ter notas remissivas (LANCASTER, 1993;

CLEVELAND; CLEVELAND, 2001).

Para Langridge (1969), uma simples lista ordenada alfabeticamente não é o suficiente para um sistema de indexação, é necessário que haja relações entre a disposição desses termos, fato que faz a diferença entre um dicionário e um tesouro, por exemplo.

Segundo Lancaster (1993), a indexação é um processo subjetivo, pois o mesmo documento poderá ser indexado de modo diferente, seja na situação em que ele é indexado pelo mesmo indexador em momentos diferentes, ou quando é indexado por diferentes indexadores, pois podem divergir na seleção dos conceitos no momento da análise do documento. Com o propósito de contornar essa subjetividade, é importante fazer algumas perguntas ao documento, no momento da análise, tais como: Qual o tema do documento? Por que foi incorporado ao nosso acervo? Quais de seus aspectos serão de interesse para os usuários? Considerando que existe um conjunto de termos representativos de um documento, as necessidades dos usuários e os objetivos da instituição para a qual se indexa, é justificável que um documento possa ser representado de diferentes formas, dependendo destes fatores.

Anderson (1983) aponta que diferentes termos irão provocar reações diferentes em diferentes pessoas, por questões culturais ou linguísticas. Quando um item é indexado, espera-se que os termos eleitos pelo indexador evoquem conceitos idênticos àqueles que os usuários conhecem.

Entre outros produtos que resultam da prática da indexação, destacam-se os índices, dada a sua importância na recuperação da informação.

Borko e Bernier (1978) destacam a importância da indexação a partir da imprescindibilidade dos seus produtos, em particular o índice. Neste contexto, os autores mencionados afirmam:

Sem a indexação, nós não poderíamos utilizar o telefone facilmente, pedir comida em restaurantes, organizar viagens, encontrar materiais em bibliotecas, agendar compromissos, localizar correspondências, encontrar ruas, ou fazer outras coisas que agora nós fazemos normalmente. Interações sociais, como as que conceituamos atualmente, seriam impossíveis sem indexação. Índices são guias convenientes, eles não apenas nos guiam aos campos de nosso interesse, eles também nos propiciam uma visão geral – um campo que nenhuma outra ferramenta bibliográfica pode proporcionar” (BORKO; BERNIER, 1978, p. 3, tradução nossa).

Na Idade Média, com o surgimento das universidades, a utilização de índices começou a se tornar mais evidente. Exemplos de sua utilização, ainda que incipiente, são as listas alfabéticas

elaboradas por advogados, para ressaltar as partes relevantes de estatutos, e a utilização pelos clérigos das partes em branco da bíblia, para referenciar as partes significativas das escrituras (CLEVELAND; CLEVELAND, 2001).

Outro fato que demonstra a importância dos índices, no século XVII, é a introdução de índices nos livros escolares. Neste mesmo século, as palavras *index* (índice) e *table* (tabela) começam a ser utilizadas de formas distintas, situação que não ocorria até então na língua inglesa. Enquanto a palavra *table* é utilizada para apresentar a estrutura dos conteúdos de um livro, através do registro dos títulos de seus capítulos, já a palavra *index* é empregada para as entradas alfabéticas ou de classificação, usualmente encontradas no final dos livros (BORKO; BERNIER, 1978). Também nesse século, com a emergência dos periódicos, a importância do uso dos índices é evidenciada (CLEVELAND; CLEVELAND, 2001).

No século XIX, William Frederick Pole (1882) criou o Poole's Indexes, cujas entradas eram geradas a partir das palavras-chave e títulos dos artigos indexados. Em 1876, Charles Ammi Cutter criou o Rules for a Printed Dictionary Catalog, obra na qual se apresentam os princípios que deveriam ser considerados na criação dos pontos de acesso, entre os quais se destacam os destinados às entradas para os catálogos alfabéticos de assuntos e para os catálogos sistemáticos. Em 1880, John Shaw Billings, da Biblioteca Médica do Exército dos Estados Unidos, criou o primeiro índice para a literatura médica. É neste âmbito que as primeiras noções de índice e indexação são apresentadas da forma como são entendidas hoje no universo da Ciência da Informação (BORKO; BERNIER, 1978).

O índice tem por principal objetivo “[...] caracterizar e prover acesso à informação, representada em itens indexáveis e/ou informação sobre esses itens” (ANDERSON, 1983, p. 3, tradução nossa).

De acordo com Lancaster (1993), “[...] o propósito principal da elaboração dos índices é construir representações de documentos publicados de modo a serem incluídas em algum tipo de base de dados”. Hjørland (2001) diz que muitas são as possibilidades de se encontrar o assunto de um livro, como, por exemplo: as palavras expressas pelo autor no interior do documento; a forma com que o usuário vê os assuntos contidos no documento; a versão do editor, quando se trata de uma publicação seriada; e a visão do indexador, que será demonstrada por meio dos termos escolhidos para representar aquele documento, no caso, um índice.

De modo geral, os índices se referem a um domínio de determinado conhecimento,

diferentemente dos catálogos – utilizados para guiar o usuário aos títulos e autores dos trabalhos – e das classificações – que relacionam um trabalho com outros em uma coleção. Por meio de um sistema notacional, os índices guiam o usuário à tematicidade do documento, por meio de termos representativos (BORKO; BERNIER, 1978). Eles têm a função de direcionar o leitor para os documentos que contêm os assuntos específicos que ele busca. Além disso, irá direcionar o usuário a outros assuntos associados com os da sua busca, demonstrando as relações entre diferentes tópicos e auxiliando-o quando ele tiver dúvidas a respeito do assunto procurado, através de notas remissivas, por exemplo. É considerado um bom índice aquele que cria substitutos, de forma que represente de maneira adequada o documento com um todo. Para tanto, é necessário que no momento da indexação o profissional seja o substituto do leitor, conseguindo identificar quais os principais pontos que eventualmente lhe poderão interessar (BONURA, 1994).

Para que o processo de “substituição” resulte de forma efetiva, é necessário abandonar o pensamento da universalidade, pois, como apontado por Olson (2001), o leitor não é uniforme, existem diferentes tipos de leitores, assim como diferentes necessidades informacionais, mesmo que os leitores busquem pelo mesmo assunto. De acordo com essa autora, é necessário abandonar a ideia do absoluto e de verdades singulares, dando voz às minorias, o que não é possível quando se utilizam discursos dominantes excludentes na representação dos conceitos. Uma forma de se evitar esse tipo de dominação consiste em encontrar espaços de negociação entre os diferentes discursos, ou seja, seria ideal utilizar uma terminologia feminista para representar o universo das mulheres. Mas, para adequar-se à realidade atual, é preciso encontrar tais espaços de negociação em linguagens mais tradicionalmente utilizadas, como a CDD ou a CDU.

A autora demonstra que, apesar de a universalidade ser um princípio seguido por alguns teóricos da indexação (entre os quais Cleverdon, 1960), não é possível atingir esse tipo de ideal, uma vez que a universalidade não deve ser imposta a um público, composto por uma diversidade cultural e linguística. Uma atitude como essa poderia causar uma inconsistência entre a linguagem utilizada pelo usuário no momento da busca e a linguagem utilizada pelo indexador no momento da representação.

O índice possui dois propósitos básicos: minimizar o tempo gasto na busca da informação e maximizar o sucesso da busca do usuário, recuperando documentos pertinentes que coincidem com a necessidade informacional do usuário. O índice é uma ferramenta efetiva quando nele são

registrados os assuntos de forma específica e concisa, de modo a responder aos diferentes perfis de usuários, que terão necessidades informacionais diferentes e que buscarão, de outra maneira, a informação (CLEVELAND; CLEVELAND, 2001; FETTERS, 2008).

A indexação envolve todas as complexidades inerentes à criação de um índice, além da manutenção de um vocabulário controlado, e esse processo será satisfatório caso as necessidades dos usuários sejam atendidas (ROTHMAN, 1983).

A partir da década de 1950, surge a indexação automática como um campo emergente, justificada pela necessidade do acesso à informação de forma mais rápida, pela explosão informacional ocorrida nesta época e pelo surgimento de uma nova área de estudos denominada linguística computacional. Esta nova matéria teve como precursores Noam Chomsky e Zellig Harris, e se preocupava com a análise computacional em estruturas e semânticas linguísticas, dentre outros fatores (BORKO; BERNIER, 1978).

Segundo Borges (2010), o pioneiro desta nova forma de construção de índices foi H. P. Luhn, que, em 1959, desenvolveu o método Keyword in Context – KWIC. De acordo com esse método, as palavras-chave que apareciam no título do documento se tornavam um termo representativo. Além desse método, surge também o modelo Selective Listing in Combination – SLIC, no qual os termos são organizados alfabeticamente a partir de palavras não redundantes; o Preserved Context Indexing System – PRECIS, em que os termos são alterados de forma sistemática, para que todos ocupem posições de entrada; e o Nasted Phrase Indexing System – NEPHIS, caracterizado por ser um índice no qual os termos são articulados com seus vizinhos por meio de pontuações ou palavras funcionais.

Leiva (2008) afirma que a expressão “indexação automática” pode referir-se a três conceitos: programas que analisam o processo de indexação realizado de forma intelectual, podendo proporcionar notas explicativas e termos relacionados, para auxiliar o trabalho do indexador; programas de indexação que analisam e extraem os termos, mas demandam que um profissional os valide; e a indexação automática propriamente dita, em que programas de computador analisam, extraem e formulam os termos representativos, sem que passem por nenhum profissional a posteriori.

Segundo Borko e Bernier (1978), o principal fundamento em que a indexação automática se abriga é aquele em que as ideias são transmitidas por meio de palavras e seu arranjo. Partindo desse fundamento, entende-se por indexação automática o processo em que o assunto de um

documento é extraído pela análise mecânica, com base nas palavras do próprio texto. No entanto, ainda de acordo com os mesmos autores, a indexação automática requer alguma intervenção humana.

Essa vertente da indexação também pode ser denominada “natural language indexing”, já que as palavras selecionadas para indexação são aquelas que aparecem no texto, não sendo adicionadas palavras ou introduzidas notas remissivas com o propósito de se elaborar qualquer vocabulário controlado (ROTHMAN, 1983).

Segundo Borges (2010), a indexação assistida por computador é uma forma de conciliar a informática com a área da documentação. O autor afirma que a indexação automática seguirá, basicamente, quatro tarefas: contar todas as palavras dispostas no texto; compará-las com uma lista de palavras não autorizadas; eliminar palavras consideradas não significativas, como artigos ou proposições; e ordenar essas palavras por ordem de frequência, sem identificar, por exemplo, os conceitos dessas palavras, o que pode ser denominado indexação por extração automática. Esses fatores também podem ser verificados na premissa proposta por Anderson (1983): “(...) máquinas indexam símbolos, enquanto humanos indexam conceitos” (ANDERSON, 1983, p. 5, tradução nossa).

No caso da indexação humana, os termos definidos trazem alguma ligação com seus respectivos conceitos, o que não ocorre com a indexação realizada por máquinas, pois elas não possuem essa percepção.

A indexação realizada por meio da máquina não tem em seu cerne a preocupação com conceitos, ela considerará apenas os símbolos existentes no documento e os manipulará de acordo com as instruções informadas previamente; fará o cruzamento das palavras presentes no texto com uma lista de palavras construída a priori. Neste processo, todas aquelas que não coincidirem serão automaticamente eliminadas (ANDERSON, 1983).

Outra prática que norteia a indexação automática é aquela que se prende com a frequência das palavras. Quanto mais vezes uma palavra aparecer num texto, maior será a probabilidade de que essa palavra seja representativa do assunto tratado naquele documento. O programa pode contar quantas palavras o documento possui e calcular o número médio de vezes que um termo aparece e, assim, ser considerado para servir na indexação; o programa pode ainda contar quantas vezes determinada palavra aparece em todo o documento e fazer uma lista de palavras representativas daquele documento, uma lista de termos. Existe ainda a possibilidade do

truncamento, que consiste na retirada de prefixos e sufixos das palavras do documento, contabilizando a frequência de palavras a partir de sua raiz morfológica (BORKO; BERNIER, 1978).

A prática de apenas elencar os termos mais utilizados em um documento pode resultar num grande número de termos e muitos deles não serem necessariamente considerados guias para os assuntos do recurso. Isso ocorre porque o ato de indexar vai muito além que apenas selecionar palavras recorrentes (BORKO; BERNIER, 1978).

Para Rothman (1983), embora a indexação automática tenha auxiliado de alguma forma o processo de indexação, principalmente no que diz respeito à agilidade do processo, ela ainda possui alguns sérios problemas para resolver, como: abreviações ou pontuações nas palavras (hífen, por exemplo); sinônimos com conceitos diferentes, como, por exemplo, cidades diferentes com nomes iguais; a utilização de termos desnecessários ou a não utilização de termos que seriam importantes em uma pesquisa; nomes próprios ou títulos, que podem ter sido alterados após a criação do documento. Outra questão é aquela relacionada ao julgamento, pois apenas os humanos podem reconhecer, relativamente, a importância de determinados termos, como, por exemplo, no que se refere às questões contextuais.

Embora Keyser (2012) afirme que a indexação automática seja um meio efetivo de economizar tempo e recursos, ela não apresenta uma visão geral coerente com os termos indexados, não pode resolver problemas, como é o caso dos sinônimos ou erros ortográficos, além de não considerar o contexto do documento e do usuário, tal como afirma Rothman (1983), posição com a qual esta dissertação concorda.

Outra potencialidade da indexação automática é a análise sintática e semântica. Nessas análises, identificam-se por hipóteses, no grupo de palavras mais utilizadas, conteúdos significantes. Na análise sintática, verifica-se a classe gramatical da palavra e sua relação numa frase, enquanto na análise semântica verificam-se as relações entre termos ou palavras simples. Perante tal situação, fica evidente, portanto, a necessidade de uma comunicação entre linguísticos e cientistas da informação (BORKO; BERNIER, 1978).

Bonura (1994) refere que, embora a indexação automática seja uma forma de agilizar o processo de indexação, economizando tempo, é importante que o indexador tenha pleno controle do software utilizado, sendo-lhe dada a possibilidade de fazer alterações nos formatos e nos métodos utilizados, editando quando for necessário e verificando sempre a exatidão do índice.

Pelo exposto, infere-se que a indexação automática pode até substituir o indexador, mas isso deve ser evitado.

O ato de indexar envolve diversas tomadas de decisões que exigem do indexador o conhecimento de conceitos, capacidade de percepção e competências para gerenciar problemas, com vistas à solução. As principais tarefas do indexador centram-se na seleção de assuntos específicos, na transformação desses assuntos em termos de representação, na utilização de linguagens de indexação, na elaboração de notas remissivas quando necessário, dentre outros aspectos que circundam a representação temática (BONURA, 1994).

Mendes e Simões (2002), baseadas em um documento proposto pela IFLA - Principles Underlying Subject Languages - propõem uma série de princípios para se realizar uma indexação de qualidade, dentre as quais se destacam: a análise do conteúdo (representar de forma fidedigna o que o autor expõe); os termos eleitos para representar os conceitos, de acordo com a comunidade usuária; e a busca contínua de critérios, como a coerência e a uniformidade. As autoras reforçam que a escolha dos termos, nesta operação, é uma decisão terminológica que se deve buscar com precisão.

De acordo com Bonura (1994), para uma boa indexação, o indexador deve atender a cinco princípios: precisão, que se encontra diretamente ligada à especificidade; profundidade, relacionada com a exaustividade; concisão ou clareza dos termos escolhidos; elaboração de notas remissivas; e aspectos lógicos dos cabeçalhos.

Borko e Bernier (1978) trazem algumas implicações éticas intrínsecas relacionadas com o profissional indexador: minimizar os erros o máximo possível (para isso, precisa estar consciente dos erros mais comuns, para poder evitá-los); e reduzir ao máximo o tempo de busca do usuário, assim como sua falta de conhecimento relativamente a um assunto.

Os autores ressaltam que os indexadores são humanos e, portanto, passíveis de não entenderem ou entenderem de forma errônea os materiais que lhes são dados para indexar, de terem lapsos de concentração, situação que pode concorrer para uma indexação de má qualidade, provocada por erros conscientes ou inconsistentes. Podem, ainda, contribuir para uma indexação demasiado consistente e, deste modo, resultar em uma indexação inflexível, com adoção de termos que não são os preferidos do usuário (BORKO; BERNIER, 1978).

Cleveland e Cleveland (2001) afirmam que o trabalho do indexador deve ser pautado por normas profissionais e éticas, pois os especialistas podem ser importantes censores da

informação, mesmo que não intencionalmente, quando não reconhecem as *biases*, tendo como tradução mais próxima uma tendenciosidade no momento da representação, o que pode evidenciar preconceitos e visões de mundo consideradas incorretas por parte do indexador.

Uma das pioneiras relativamente às *biases* no momento da representação do conhecimento é Hazel Bell. Em artigo publicado em 1991, ela refere como esse fator pode ocorrer na indexação e afirma que o trabalho do indexador é visto como puramente objetivo, o que, na verdade, não acontece desde os primórdios. Para tanto, a autora apresenta cinco possíveis origens de *biases*: quando opiniões consideradas demasiadamente fortes são apresentadas pelo autor e, posteriormente, reforçadas pelo indexador; os preconceitos do próprio indexador introduzidos durante a indexação; quando as imposições e os preconceitos são introduzidos pela própria linguagem de indexação; quando os termos, por uma questão de barreira linguística, acabam por possuir lacunas sociais, políticas ou que excluem determinados grupos; e, por último, quando uma parte importante do texto é suprimida no momento da indexação.

Outra autora que também se utiliza do termo *biases* é Fetters (2008), quando esclarece que, no momento de elaboração do índice, é necessário que todos os preconceitos sejam deixados de lado, mantendo-se neutro, evitando possíveis vieses, tradução não tão abrangente do termo *biases*.

Milani (2014) demonstra que, embora alguns dos sentidos para a palavra *biases* sejam de vertente negativa – ligada à discriminação das minorias, visando a uma representação objetiva –, ela também pode se apresentar enquanto uma forma de se garantir a especificidade da representação, sendo conceitualmente entendida de forma positiva enquanto viés ou direcionamento.

Esse processo de indexação possui características importantes para a sua realização, a exaustividade e a especificidade, tradicionalmente entendidas enquanto medidas da indexação. No presente estudo, busca-se entendê-las além de meras medidas, mas enquanto características que afetam a prática da indexação e, em particular, a recuperação informacional, assumindo neste contexto o seu valor ético.

No que tange à exaustividade, Simões (2011) a conceitua: “[...] o número de noções que são extraídas e que caracterizam um documento” (SIMÕES, 2011, p. 42).

Borko e Bernier (1978) evidenciam a necessidade da exaustividade quando ressaltam que a seleção de termos por parte do indexador deve ocorrer de forma geral. Essa escolha deve partir

de uma seleção do indexador, pois um único trabalho pode discutir assuntos específicos, mas em diferentes níveis de generalização; por isso, todos esses assuntos devem ser selecionados pelo indexador, de maneira a contribuir para a boa qualidade da indexação.

Anderson (1983) refere que uma busca informacional só será efetiva se a indexação tiver sido exaustiva. Assim, se um sistema utiliza o termo “computador” para todos os documentos relacionados com esse conceito, mesmo quando se trate de assuntos particulares, no momento da pesquisa o usuário irá recuperar documentos sobre “computador” (assunto geral) e também documentos que tratem assuntos mais específicos relacionados com ele.

Já de acordo com Langridge (1969), uma indexação sumarizada e uma indexação exaustiva são dois extremos, e o indexador deverá optar por um deles. Uma indexação sumarizada irá meramente explicitar qual é o principal assunto de um documento, reduzindo o potencial informativo de um documento. A indexação exaustiva reflete todos os assuntos pertinentes que pode conter um documento, resultando, deste modo, em diversas entradas de assunto.

Segundo Lancaster (1993), a exaustividade pode ser considerada uma decisão política da indexação, não assumindo, no entanto, limite para quantidade de termos que devem ser empregados. A exaustividade corresponde “[...] ao número de termos atribuídos em média [...] a indexação exaustiva implica o emprego de termos em número suficiente para abranger o temático do documento de modo bastante completo” (LANCASTER, 1993, p. 23). Ainda de acordo com o autor, quanto maior for o número de termos utilizados na hora da representação de um documento, maior será a probabilidade de acesso no momento da recuperação. Tal situação ocorre devido ao fato de se criarem mais pontos de acesso, aumentando, por este meio, a capacidade de recuperação de um item. Uma indexação exaustiva irá demonstrar todos os assuntos específicos de um documento, bem como os assuntos considerados mais gerais (LANCASTER, 1993).

A mesma visão compartilham Cleveland e Cleveland (2001), ao referirem que quanto mais exaustiva for a indexação de um item, maior será a probabilidade de recuperação. Os autores também concordam que a exaustividade depende da política de indexação. Além das necessidades do usuário, o tempo que o indexador pode desprender para realizar uma indexação e o financiamento são aspectos importantes a ponderar. Segundo os mesmos autores, uma vez que uma indexação exaustiva recupere uma grande variedade de documentos, é necessário que o

indexador esteja atento à pertinência dos mesmos, identificando quando esse fator poderá afetar negativamente a recuperação.

A capacidade que o sistema possui para identificar os assuntos relevantes é denominada capacidade de revocação. Esta refere que o principal objetivo do sistema é recuperar tudo o que se encontra relacionado com as questões informacionais do usuário. Essa medida é realizada a partir da divisão do número total dos documentos recuperados pelo número de documentos relevantes, ou seja, os documentos passíveis de atender às necessidades do usuário, encontrados na unidade informacional (CLEVELAND; CLEVELAND, 2001).

Para Bonura (1994), a exaustividade diz respeito ao número de assuntos que podem ser extraídos de um livro e incluídos em um índice. Para tanto, a completude deve ser indispensável, incluindo o máximo possível dos tópicos que foram extraídos de um documento.

Foskett (1973) apresenta a exaustividade como uma decisão administrativa que diz respeito à extensão de todo o conteúdo temático de determinado documento, para que, posteriormente, seja especificado. Dessa forma, é possível ver o conteúdo global do documento, como sendo seu tema principal, e os subtemas que podem ser encontrados nele, constituindo-se estes, de acordo com o contexto, em secundários.

Saindo de uma dimensão mais quantitativa rumo a uma abordagem mais qualitativa, tem-se a especificidade. Segundo Borko e Bernier (1978), a especificidade diz respeito ao nível de pormenor com que os assuntos de um documento são tratados, traduzindo, deste modo, a especificidade que o autor lhe dotou. Apresenta o seguinte exemplo: se um autor trata sobre os perigos de dirigir no inverno, a forma mais específica de se indexar esse assunto seria: riscos de colisões entre automóveis dirigindo em ruas congeladas, e não “colisões” ou “inverno”.

Para Anderson (1983), a especificidade concerne à relação gênero/espécie entre os termos que serão dados pelo indexador, como, por exemplo, um termo como “computador IBM 370” será indexado especificamente, enquanto termos como “computador” ou “máquinas” seriam termos mais gerais. Para o autor, itens indexados de forma mais específica permitem uma recuperação com maior precisão, sendo relacionados ao que o usuário solicita.

De acordo com Simões (2011), o conceito de especificidade está ligado ao “[...] grau de exatidão com o qual são extraídos os conceitos dos documentos, assim como o grau de exatidão que é usado na representação desses mesmos conceitos” (SIMÕES, 2011, p. 42).

Segundo Langridge (1969), a especificidade promove consistência à indexação. Por

exemplo: se um termo como “arma” está indexado a um livro de administração de negócios, este não estará indexado de forma precisa, porque não expressa o verdadeiro sentido do termo; no entanto, se o termo for “mercado de armas”, o usuário que busca informação sobre organizações de armamentos conseguirá recuperar os documentos de que precisa.

Para Lancaster (1993), quanto mais especializados forem os usuários de um centro de informação, maiores são as probabilidades de a indexação ser realizada de forma específica, indo ao encontro dos interesses do grupo. Para tanto, o autor diz que um documento que aborda o cultivo de laranjas deve ser indexado sob o termo “laranjas”, e não apenas sob “frutas cítricas” ou “frutas”.

Foskett (1973) apresenta a especificidade como uma função do sistema e a define: “[...] a extensão em que o sistema nos permite ser precisos ao especificarmos o assunto de um documento que estejamos processando” (FOSKETT, 1973, p. 12). Para o autor, essa característica está diretamente ligada à relevância dos documentos recuperados, e apresenta o exemplo do termo “gatos siameses”. Caso não fosse indexado com tamanha especificidade, seria necessário analisar todos os documentos sobre gatos para que se chegasse a esse assunto.

Tanto a exaustividade quanto a especificidade vão muito além de meras características do processo de indexação para revelarem toda uma dimensão axiológica, ligada aos valores que permeiam a indexação.

Em linhas gerais, Lancaster (1993) utiliza o termo revocação para designar a recuperação de documentos úteis, e o termo precisão como uma medida que evita a recuperação de documentos que não são pertinentes às necessidades do usuário. Essas duas medidas, segundo o autor, devem ser utilizadas para expressar os resultados de uma busca e encontram-se intrinsecamente relacionados com os conceitos de exaustividade e especificidade.

Carneiro (1985) apresenta a revocação como a medida do sistema de recuperar documentos relevantes, e precisão como a medida que impede a recuperação de documentos considerados não relevantes. Para a autora, há uma relação inversa entre essas duas medidas, pois quando se amplia a pesquisa para a recuperação de um maior número de documentos, deixa-se de utilizar o critério da precisão; e quando se restringe a pesquisa a termos específicos, diminui-se o nível de revocação. É uma decisão do sistema, no sentido de definir quais serão os níveis de revocação e precisão da representação, e devem-se considerar os interesses dos usuários.

Uma forma de se utilizar essas duas características é apresentada por Anderson (1983),

quando sugere que os termos devem estar conectados no momento da sua representação, a partir das relações de gênero/espécie, todo/parte, objeto/operação/agente, dentre outras relações, que são expressas principalmente por meio dos tesauros.

Para Cleveland e Cleveland (2001), exaustividade e especificidade compõem os dois lados de uma moeda, contribuindo as duas características para a profundidade da indexação. Para os autores, quando um documento é indexado de forma exaustiva e específica, o item será indexado de forma detalhada. Ainda segundo os autores, a exaustividade na indexação implica que os termos sejam atribuídos de forma pormenorizada, contribuindo para um maior número de termos de recuperação, os quais são um reflexo de uma análise específica. Quanto mais específicos forem os termos, maior há de ser a particularidade semântica dos documentos recuperados, fator que se encontra relacionado com a especificidade que o usuário busca nas pesquisas. Deste modo, a exaustividade refere-se à extensão com a qual um assunto será coberto e num dado documento; já a especificidade se prende com a precisão dos assuntos no mesmo.

É importante perceber que características como a exaustividade e a especificidade afetam diretamente a recuperação informacional, sendo consideradas não apenas de importância para o sistema, mas também como uma questão social, uma vez que um documento indexado em sua completude, de forma exaustiva, poderá ser recuperado pelo usuário por qualquer dos termos com ele relacionados, concorrendo para uma maior recuperação de um documento indexado a partir de todas as particularidades de seus assuntos. Quando o usuário buscar uma informação mais precisa, mais específica, poderá encontrá-la.

Dessa forma, fica evidente o caráter ético desses dois atributos da indexação, uma vez que o usuário é a razão de ser da unidade informacional. A exaustividade e a especificidade auxiliam diretamente na recuperação temática da informação e, conseqüentemente, no acesso à informação, na medida em que permitem ao usuário uma resposta rápida, pertinente e precisa à sua necessidade de informação.

Essas questões podem ser verificadas a partir do que Adler e Tennis (2013) denominam *harm* na organização do conhecimento. Para os autores, as ações dos profissionais da informação são verificadas proporcionalmente a partir de suas conseqüências boas ou danosas; neste sentido, as linguagens documentárias podem ser instrumentos de violência institucionalizados quando marginalizam as minorias e favorecem grupos maiores, além de reforçarem os preconceitos e marginalizações em relação às pessoas e instituições.

Os autores conceituam *harm* ou dano: “[...] quando se desvia de um acordo sobre os preceitos que ditam o que é ético” (ADLER; TENNIS, 2013, p. 268, tradução nossa). Num texto anterior, Tennis (2013) já havia previsto que evitar consequências como o *harm* é um dos nove preceitos para que haja uma organização do conhecimento comprometida, além da utilização de uma linguagem apropriada, como no momento da criação de termos, uma vez que ela possui o poder de prevenir conflitos. Ainda de acordo com os autores citados, há diversas formas de se identificar esses preceitos para se evitar o dano, questionando o que acontece ao redor, quem participa desses acontecimentos e como estes podem ser afetados. Dessa forma, é possível que esses danos sejam identificados e minimizados, levando em consideração que os indivíduos são responsáveis pelo que representam (ADLER, TENNIS, 2013).

Cabe, no entanto, verificar em que dimensões a literatura da Ciência da Informação concebe tais elementos, em especial no que se refere à dimensão ética.

5 METODOLOGIA

Levando em consideração que a investigação é uma “[...] atividade criativa e produtiva, assente em base ética, que procura a compreensão, em profundidade, de fenômenos relevantes” (OLIVEIRA, 2011, p. 2), é necessário utilizar dois procedimentos para se atingir os objetivos dessa atividade: a metodologia e o método.

Oliveira (2011) define essas duas ações de maneira simples, classificando a metodologia por ser um conjunto de princípios ou regras fundamentais a determinada teoria ou processo, que possui um caráter estratégico, pois, a partir de uma base de raciocínio, pode determinar certo plano de atuação para que sejam definidas respostas para questões formuladas anteriormente. No momento da implantação desse plano de atuação, são necessários alguns procedimentos denominados métodos, que possuem caráter funcional, pois são ferramentas utilizadas para executar o plano, a partir de um conjunto de formas e técnicas específicas.

Nessa mesma perspectiva, Quivy e Campenhoudt (2013) argumentam que os procedimentos de pesquisa apontam em direção a um objetivo. Quando esses procedimentos são expostos, descrevem-se os princípios que serão utilizados para colocar em prática o trabalho de investigação. Dessa forma, os métodos consistem em formalizações do procedimento, percursos configurados e adaptados à área estudada.

A pesquisa ora proposta se classifica como exploratória, já que possui por objetivo familiarizar-se com o problema, de maneira a evidenciá-lo, e construir algumas hipóteses, a partir da elaboração de algumas ideias sobre o tema. Como característica desse tipo de pesquisa, há o levantamento bibliográfico e a análise desse material. Este estudo apresenta um caráter bibliográfico, pois esse tipo de pesquisa tem sua base em materiais já publicados, especialmente em livros e artigos científicos (GIL, 2002).

Para responder às questões levantadas na primeira parte do referido estudo, conforme já demonstrado, realizou-se uma revisão de literatura, de forma a teorizar o campo da organização e representação do conhecimento no que diz respeito em particular aos aspectos éticos, e foi recolhida a bibliografia dos autores mais significativos que se tratam deste tema. Do mesmo modo, procedeu-se ao levantamento de bibliografia, que permitiu construir um arcabouço teórico relativamente à prática da indexação, em especial no que se refere às duas características predominantes nesta prática: a exaustividade e a especificidade.

Em um segundo momento, procurou-se realizar o estudo empírico, a partir da coleta e

análise de dados. Neste âmbito, as pesquisas foram realizadas exclusivamente em artigos científicos, a partir de um *corpus* advindo da área Ciência da Informação. A discussão teórica do tema partiu do levantamento bibliográfico e da análise do material selecionado, com o fim de se realizar uma fundamentação teórica.

A pesquisa ora desenvolvida foi realizada em dois momentos: em um primeiro, no Brasil, e em um segundo momento, na Universidade de Coimbra, financiada por uma Bolsa de Estágio Pesquisa no Exterior - FAPESP (Processo nº 2015/06278-6), orientada pela Profa. Dra. Maria da Graça de Melo Simões.

Dado o fato de a pesquisa se ter desenvolvido em diversas fases, entende-se oportuno registrar essas fases para possibilitar melhor entendimento.

5.1 CRITÉRIOS USADOS NA COLETA DE DADOS

As pesquisas foram efetuadas exclusivamente em artigos científicos, a partir de um *corpus* advindo da área da Ciência da Informação.

As buscas nas plataformas e repositórios de artigos científicos foram realizadas mediante a opção busca avançada, a única que permite a combinação de termos, além de permitir que os termos sejam buscados apenas em sua área de maior conteúdo informacional - título, resumo, palavras-chave e descritores de assunto. Para tanto, foram selecionados os termos exaustividade e especificidade, além de seus equivalentes em inglês - *exhaustivity* e *specificity* - combinados com os termos indexação/indexing, respectivamente, como termos de busca para o levantamento da literatura. Também é importante ressaltar que os artigos recuperados em mais de uma base de dados foram contabilizados e analisados apenas uma vez.

Para a seleção dos artigos, utilizou-se o critério de relevância orientada ao usuário, proposto por Abad García (2005), que possui uma natureza subjetiva e depende do âmbito do usuário. Para explicar esse critério, a autora apresenta as quatro categorias desse tipo de relevância, propostos por Saracevic (1996): a) a relevância do conteúdo, em que não só importa se o documento possui os termos de pesquisa, mas, também, o contexto em que os termos estão tratados; b) a pertinência, em que se busca estabelecer uma relação entre os conteúdos dos documentos recuperados e a necessidade informacional da busca; c) a relevância situacional, que relaciona o documento com o problema de pesquisa que motivou sua busca; d) e a relevância

motivacional, que relaciona o conteúdo do documento com o uso posterior deste.

Para escolha do *corpus*, optou-se por artigos que não apenas contivessem os termos de pesquisa, mas que também possuíssem os termos no âmbito da indexação, de forma a criar uma ponte entre os documentos recuperados e a pertinência destes para a presente pesquisa, identificando a relevância situacional e a relevância motivacional, uma vez que o corpus de pesquisa poderá ser utilizado em pesquisas posteriores sobre o tema.

5.2 FONTES DE OBTENÇÃO DE DADOS

As plataformas de bases de dados utilizadas para compor o corpus foram as seguintes: no âmbito da pesquisa realizada no Brasil, utilizaram-se as bases de dados Library and Information Science Abstracts (LISA), Scopus e Web of Science; de forma a apresentar o cenário brasileiro, os mesmos termos foram pesquisados na base de dados Scielo; outra busca paralela, e com o mesmo objetivo, foi feita nos códigos de ética profissional bibliotecária publicados na página da International Federation of Library Associations and Institutions (IFLA).

Já no âmbito português, a pesquisa foi realizada nas seguintes plataformas: B-on, uma plataforma de base de dados pública financiada pelo governo português; Ebsco (por meio da ferramenta de descobrimento EBSCO Discovery Search); e Elsevier (por meio da base Search Direct), que são plataformas de bases de dados privadas e financiadas pela Universidade de Coimbra. Consultou-se o Repositório Institucional Estudo Geral e o Repositório Científico de Acesso Aberto de Portugal (RCAAP), repositórios onde são depositadas as publicações científicas portuguesas, disponibilizadas na sua maioria em acesso aberto. Neste sentido, esta pesquisa integra-se numa abordagem metodológica qualitativa.

Fundada pela ProQuest - uma empresa fundada com o objetivo de desenvolver tecnologias informacionais -, a base de dados Library and Information Science Abstracts foi iniciada em 1969 e recupera resumos de artigos da área de Ciência da Informação, principalmente no que tange à biblioteconomia, recuperação informacional e tecnologia da informação. Por mês, a base indexa mais de mil registros, divididos em 440 periódicos de 45 países, em 20 diferentes idiomas. Por política de seleção, a LISA considera o conteúdo editorial, revisão por pares, o tema proposto, além da diversidade internacional dos autores e das citações (PROQUEST, 2015).

A editora Elsevier possui duas bases de dados utilizadas na pesquisa: a Scopus e a Science

Direct, descritas a seguir.

De acordo com o site da editora, a Scopus é uma grande base de dados de resumos e citações, incluindo periódicos científicos, livros e anais de congressos das mais diversas áreas científicas, como tecnologia, medicina, artes, humanidades e ciências sociais. Os anais de congressos, relacionado à presente pesquisa, corresponde a 24% do conteúdo da base. Seu público é distribuído em acadêmicos, agências de governo e corporações institucionais, dispostos a desenvolver pesquisas.

A Web of Science é a base de dados da editora Thomson Reuters, que, além de multidisciplinar, cruza referências de citações para uma busca mais completa. Com mais de 11 mil periódicos indexados em 55 diferentes disciplinas, ela possui o diferencial de indexar também patentes. Possui um público direcionado a acadêmicos, instituições governamentais e corporativas.

A plataforma Ebsco Discovery Search (2015) é um serviço privado, que agrega diversas bases de dados produzidas pela empresa, cujo objetivo é auxiliar a biblioteca e outras instituições a prover acesso informacional a seus usuários. Possui cerca de 70 anos de existência e relações com mais de 79 mil editores em todo o mundo, fornecendo materiais como periódicos, livros, ferramentas de gerenciamento de recursos eletrônicos, bases de dados secundárias, além de serviços relacionados a unidades informacionais.

Além de todas as bases de dados produzidas pela EBSCO - nas mais diversas áreas, como saúde, administração e novas tecnologias -, a plataforma possui um agregador de conteúdos, denominado EBSCO Discovery Search (EDS), que funciona como um polibuscador, indo a todas as bases de dados produzidas pela EBSCO e trazendo os recursos de acordo com a busca. Esse agregador de conteúdos pode ser entendido como uma ferramenta de descobrimento, que, de acordo com Rodriguez Bravo et al. (2014), é reconhecido por concentrar, em um único ponto de acesso, todos os materiais disponíveis no catálogo; além disso, combina diversas fontes de informação mesmo em uma busca simples, podendo ser considerado um grande avanço, pois combina o catálogo, o índice e diferentes fontes de informação em uma única ferramenta.

Além da base de dados Scopus, a Editora Elsevier possui também a base de dados Science Direct. De acordo com o site oficial da empresa (2015), a Science Direct é uma base utilizada por mais de 15 mil pesquisadores, que combina textos científicos, técnicos e publicações da área da saúde, de forma a tornar o trabalho desses pesquisadores mais eficiente. A base indexa mais de

13 milhões de publicações disponíveis em 2.500 periódicos científicos e 33 mil livros da editora.

De acordo com o site oficial (2015), o Repositório Institucional da Universidade de Coimbra, denominado Estudo Geral, surgiu no ano de 2008 e possui em sua base a produção científica desta universidade, em formato *open access*, tendo por objetivo marcar presença na rede mundial de informática e ser um emissor de cultura e conhecimento, além de disponibilizar e preservar a produção científica desta universidade.

Já o Repositório Científico de Acesso Aberto de Portugal - RCAAP (2015) – é uma iniciativa nacional, também do ano de 2008, que abrange toda a publicação portuguesa em acesso aberto, sendo o maior representante dos repositórios portugueses, por conta do número de acessos e downloads.

A plataforma B-on (Biblioteca do Conhecimento On-line) disponibiliza e dá acesso a textos integrais de diversos periódicos científicos, a partir de assinaturas de instituições de investigação. Teve início no ano de 2004, a partir de uma iniciativa pública, com a missão de garantir o acesso a publicações científicas à comunidade acadêmica. Tem por objetivos: a promoção, divulgação e o estímulo à produção e consulta de conteúdos científicos, além de contribuir para a construção da sociedade do conhecimento e facilitar a gestão de todo o conhecimento científico.

É importante ressaltar que a plataforma B-on e o Repositório Científico de Acesso Aberto de Portugal – RCAAP – constituem as maiores iniciativas portuguesas de agregação de conteúdo científico, visto que a primeira é afiliada por 34 universidades públicas do país e a segunda possui toda a produção científica nacional portuguesa (RODRIGUEZ BRAVO et al., 2014).

No contexto brasileiro, utilizou-se para a pesquisa a biblioteca online Scielo - Scientific Electronic Library Online -, que é resultado de uma parceria entre agências de fomento, como FAPESP e CNPq, e iniciativas governamentais para o desenvolvimento da ciência no Brasil, como a BIREME. Fundada há 16 anos, a biblioteca tem por objetivo disponibilizar informação científica de qualidade via acesso aberto, armazenando e disseminando as publicações, além de avaliá-las. Com uma ampla gama de periódicos disponíveis, as buscas podem ser realizadas por meio de índices e formulários de busca. Essas informações podem ser acessadas no site da biblioteca.

No que tange aos códigos de ética, definiu-se como material de coleta aqueles que estavam dispostos no site da International Federation of Librarys Associations (IFLA), o

principal órgão que respalda associações de bibliotecários no mundo, buscando representar os interesses das bibliotecas e dos usuários destas unidades informacionais. A federação foi criada em 30 de setembro de 1927 e, atualmente, possui membros de mais de 150 países. Como objetivos, a IFLA visa promover um padrão elevado de ofertas e prestação de serviços bibliotecários, incentivar a compreensão do valor de boas bibliotecas e representar os interesses dos membros em todo o mundo. O site da IFLA possui um total de 62 códigos de ética disponíveis, divididos em diferentes países, e um código de ética idealizado pela própria.

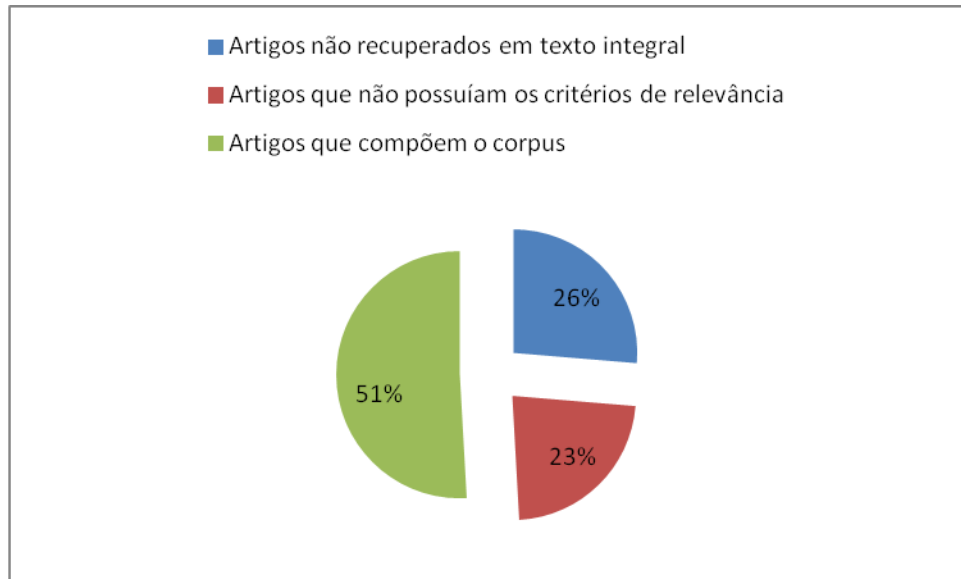
5.3 AMOSTRA

Da recolha de dados, obtiveram-se 57 textos. Este número resultou da pesquisa em dez fontes de dados: nas plataformas de dados B-on e EBSCO (por meio da ferramenta de descobrimento EBSCO *Discovery Search*); na base de dados Elsevier (por meio da base *Search Direct*); Library and Information Science Abstracts, Scopus e Web of Science; em dois repositórios institucionais portugueses: Estudo Geral e RCAAP; e nos códigos de ética profissional do bibliotecário, dispostos no site da IFLA .

Do total dos artigos recolhidos, 15 (26%) não foram encontrados na íntegra, pelo fato de as revistas não estarem incluídas no consórcio das bases de dados ou dada a circunstância de os autores terem solicitado o seu embargo. Dos 42 textos restantes, 13 (23%) não possuíam o critério de relevância proposto por Saracevic (1996) e foram excluídos da análise.

Registra-se também que não foram recuperados textos em nenhum dos dois repositórios, na base de dados brasileira Scielo e nem nos códigos de ética profissional, por meio dos termos de busca definidos. Dessa forma, o corpus analisado é composto por seis fontes de dados: B-On, EBSCO *Discovery Search*, LISA, Scopus e Web of Science.

Após a triagem referida anteriormente, a amostra é composta por 29 textos (51%). As porcentagens descritas podem ser verificadas na figura a seguir.

Figura 1 - Porcentagem do corpus

Fonte: elaborado pela autora.

Por uma questão de melhor entendimento do estudo e por questões metodológicas, os artigos não recuperados em texto integral, os que não possuíam os critérios de relevância e os artigos que compõem o corpus encontram-se registrados no capítulo de resultados (Quadros 2, 3 e 4).

5.4 Descrição do método usado na análise dos dados

Após a recolha e a análise da relevância dos dados, realizado com base nos critérios de Saracevic (1996), procedeu-se à interpretação dos textos recuperados e considerados pertinentes para pesquisa, recorrendo-se à análise de conteúdo (Bardin, 2008).

Esta opção prende-se ao fato de esse método possibilitar “[...] uma análise com base em inferências extraídas de conteúdos de documentos – a partir de uma interpretação controlada por meio de variáveis ou indicadores”. Parte-se de “[...] definições concebidas por pesquisadores, cujas características relativas à forma são homogêneas: argumentos descritivos registrados textualmente” (GUIMARÃES; SALES, 2010, p. 05).

O método foi aperfeiçoado por Bardin (2008), que o descreve como um conjunto de métodos aplicáveis a discursos, qualquer que seja sua natureza, possuindo um ponto principal: a inferência. A autora destaca que essa análise tem dois objetivos: a ultrapassagem da incerteza e

o enriquecimento da leitura. Ainda, possui duas funções: a função heurística e a administração da prova (hipóteses).

De acordo com Quivy e Campenhoudt (2013), no âmbito da investigação, ela é especialmente utilizada para

[...] análise das ideologias, dos sistemas de valores, das representações e das aspirações, bem como da sua transformação (...) análise de estratégias do que está em jogo num conflito, dos componentes de uma situação problemática, das interpretações de um acontecimento, das reações latentes a uma decisão [...]
(QUIVY; CAMPENHOUDT, 2011, p. 230)

De acordo com os autores, ela é ideal para análise do que não está exposto no texto, mas implícito, evitando que o investigador faça interpretações particulares, mas analisar o material a partir da ideologia em que foi produzido, muitas vezes em um suporte material, o que permite um maior controle por parte do pesquisador (QUIVY; CAMPENHOUDT, 2011).

Primeiramente, é necessário realizar uma fase de pré-análise, que, de acordo com Salles (2008), se caracteriza por ser “[...] a escolha dos documentos que serão submetidos à análise, [...] construção de um corpus de análise, a formulação de hipóteses e objetivos, e a elaboração de indicadores fundamentais à interpretação dos resultados” (SALLES, 2008, p. 100).

Para criação do corpus de pesquisa, Bardin (2003) sugere quatro regras principais: regra de exaustividade, regra de representatividade, regra da homogeneidade e regra da pertinência. Conforme realizado por Salles (2008), decidiu-se não utilizar a regra da exaustividade, visto que esta se define como a não seleção dos documentos encontrados.

Conforme recomendado por Bardin (2003), passa-se para a fase de elaboração de índices e indicadores. Para Sales (2008), esta fase tem “[...] a função de indicar quais são esses índices, e organizá-los sistematicamente com indicadores (frequência, em caso de análise quantitativa, e, presença, em caso de análise qualitativa) relativos a eles” (SALES, 2008, p. 106). Definiram-se, como variáveis de inferência, duas unidades de análise: **conceitos e conclusões dos termos**.

Para melhor identificar essas variáveis, foi elaborada uma ficha de análise para cada artigo recuperado, que se divide em sete colunas, a saber:

- Referência do artigo, para a identificação e localização do artigo objeto de análise;
- Resumo, cujo propósito consiste no reconhecimento do escopo do artigo;
- Palavras-chave do resumo;

- Registro dos extratos do texto nos quais se encontrem evidenciados os termos exaustividade e especificidade;
- Categorias de análise, que correspondem ao registro dos conceitos e das conclusões relativos às duas características consideradas.

Em seguida, apresenta-se uma ficha de análise.

Quadro 1 - Ficha de descrição do artigo

Referência: SPARCK-JONES, K. A statistical interpretation of term specificity and its application in retrieval. <i>Journal of Documentation</i> , v. 60, n. 05, p. 493 - 502, 2004.	
Resumo: The exhaustivity of document descriptions and the specificity of index terms are usually regarded as independent. It is suggested that specificity should be interpreted statistically, as a function of term use rather than of term meaning. The effects on retrieval of variations in term specificity are examined, experiments with three test collections showing, in particular, that frequently-occurring terms are required for good overall performance. It is argued that terms should be weighted according to collection frequency, so that matches on less frequent, more specific, terms are of greater value than matches on frequent terms. Results for the test collections show that considerable improvements in performance are obtained with this very simple procedure.	
Palavras-chave: Information research, Information retrieval, Information science and documentation	
CONCEITOS ENCONTRADOS	
Exaustividade	Especificidade
<p>“In this case the exhaustivity of a document description is the coverage of its various topics given by the terms assigned to it; and the specificity of an individual term is the level of detail at which a given concept is represented.” (p. 493)</p> <p>“the exhaustivity of a document description is the number of terms it contains, and the specificity of a term is the number of documents to which it pertains.” (p. 495)</p>	<p>“In this case the exhaustivity of a document description is the coverage of its various topics given by the terms assigned to it; and the specificity of an individual term is the level of detail at which a given concept is represented.” (p. 493)</p> <p>“(…) a term is more or less specific as its meaning is more or less detailed and precise.” (p. 494)</p> <p>“the exhaustivity of a document description is the number of terms it contains, and the specificity of a term is the number of documents to which it pertains.” (p. 495)</p>
CONCEITOS	CONCLUSÕES
<p>A exaustividade é cobertura dos vários tópicos de um documento, por meio dos termos atribuídos a ele (p.493)</p> <p>A exaustividade refere-se ao número de termos que a descrição de um documento contém (p. 495)</p>	<p>Um termo é mais ou menos específico, uma vez que é mais ou menos detalhado e preciso (p. 495).</p>

A especificidade pode ser considerada como o nível de detalhe que determinado conceito é representado (p. 493)	
--	--

Fonte: elaborado pela autora.

Com os dados recolhidos e apresentados de forma sistemática na ficha referida foi possível realizar a análise dos valores éticos, informação que se apresenta no próximo capítulo, em conjunto com o *corpus* recuperado.

6 RESULTADOS E DISCUSSÃO

Embora no site da IFLA esteja listado um número significativo de códigos de ética do bibliotecário, em apenas dois deles – sendo o código de ética de Portugal e o idealizado pela IFLA - foi encontrado um dos termos (*specificity*/especificidade), embora em nenhum deles o termo estivesse no sentido utilizado pela pesquisa. Foi realizada, então, uma pesquisa pelo termo *indexing/indexación*/indexação, no entanto não foi recuperado nenhum código. Pode-se perceber, dessa forma, que há pouco interesse nos códigos de ética relativos às práticas de organização do conhecimento. Esses códigos preocupam-se apenas com a etapa final de acesso à informação, sem dar atenção à fase intermediária, relativa ao tratamento do documento a partir das práticas de organização e representação, que é uma prática inerentemente ética.

Interessa registrar, ainda, que nenhum dos artigos foi recuperado nos repositórios institucionais, independentemente das modalidades de pesquisa, inclusive quando a busca era feita por qualquer tipo de material. Quando foi realizada a pesquisa no repositório Estudo Geral, utilizando termos em inglês (*exhaustivity; specificity*) foram recuperados documentos, no entanto nenhum deles era da área da pesquisa - Ciência da Informação.

Dessa forma, o corpus constituiu-se apenas de artigos disponíveis nas bases de dados. Dos artigos recuperados por meio da busca descrita no capítulo anterior, não se encontrou um grupo deles em sua completude, devido ao fato de as revistas não estarem incluídas no consórcio fornecido pelas bases ou dada a circunstância de os autores terem solicitado o embargo. Por este motivo, esses artigos não foram considerados no estudo. Contudo, pela sua relevância, entendeu-se pertinente apresentar o quadro, onde foram registradas as referências dos artigos, a respectiva fonte e o total encontrado.

Quadro 2 - Artigos não recuperados em texto integral

REFERÊNCIA	FONTE					
	Scopus	EBSCO	Science Direct	B-On	LISA	Web of Science
BORKO, H. Toward a theory of indexing. Information processing and management , v. 13, n. 6, p. 355 - 365, 1977.		X	X	X		
BURGIN, Robert. The effect of indexing exhaustivity on retrieval performance. Information processing and management , v. 27, n. 6, p. 623 - 628, 1991.		X	X	X		

CHOI, Y. A complete assessment of tagging quality: A consolidated methodology. Journal of the Association for Information Science and Technology , v. 66, n. 4, 2015.					X	X
DILLON, M. Thesaurus-based automatic book indexing. Information processing and management , v. 18, n. 4, 1982.		X	X	X		
FUJIKAWA, T.; NAKAYAMA, K. Trends in biomedical-research and exhaustivity of animal indexing in view of laboratory-animal information in bibliographic databases. Library and Information Science , v. 19, p. 151-168, 1981.						X
GEDRIMOV, G. V.; ZILBERMI, L. V. Exhaustivity of coverage and criteria for selection of world literature for current bibliographies. Nauchno-tehnicheskaya Informatsiya Seriya 1-Organizatsiya i Metodika Informatsionnoi Raboty , v. 3, 1972.						X
HEINE, M. H. An investigation of the relative influences of database informativeness, query size and query term specificity on the effectiveness of Medline searching. Journal of Information Science , v. 21, n. 3, p. 173 – 185, 1995.						X
LYUBOVSKAYA, E. B. Determining exhaustivity of coverage of invention specifications in viniti abstract journal by means of an automated irs. Nauchno-tehnicheskaya Informatsiya Seriya 1-Organizatsiya i Metodika Informatsionnoi Raboty , v.1, p. 21-24, 1975.						X
RADECKI, T. On the inclusiveness of systems for retrieval documents indexed by unweighted descriptors. Information processing and management , v. 17, n. 5, p. 227 - 237, 1981.		X	X	X		
SHAW, W. M. Controlled and uncontrolled subject descriptions in the CF database: A comparison of optimal cluster-based retrieval results. Information processing and management , v. 29, n. 6, p. 751 - 763, 1993.		X	X	X		
SHAW, W. M. Retrieval expectations, cluster-based effectiveness, and performance standards in the CF database. Information processing and management , v. 30, n. 5, p. 711 - 723, 1994.		X	X	X		
SHAW, W. M. Subject indexing and citation indexing— part II: An evaluation and comparison.		X	X	X		

Information processing and management , 26, n. 6, p. 705 - 718, 1990.						
SPARCK-JONES, K. Term specificity – reply. Journal of Documentation , v. 28, n. 2, 1972.						X
STEINWEG, H. Specificity in subject-headings. Proceedings Of The American Society For Information Science , v. 23, n. 1, p. 55 – 68, 1979.						X
VALDMAN, R. V.; MAKHOTEN, Y. A.; KHROMCHE, L. G. Exhaustivity of selection, quality and time of delivery of contracted information in current information systems. Nauchno-tekhnicheskaya Informatsiya Seriya 1-Organizatsiya i Metodika Informatsionnoi Raboty , v. 4, 1970.						X
TOTAL	0	7	7	7	1	9

Fonte: elaborado pela autora.

Houve, ainda, um grupo de artigos que foi recuperado, mas sem o critério de relevância proposto por Saracevic (1996) e utilizado na presente pesquisa. Os artigos encontram-se no quadro a seguir, com as referências dos artigos e as fontes onde estes foram encontrados. É importante ressaltar que o único artigo encontrado na base de dados Scielo inclui-se neste quadro.

Quadro 3 - Artigos excluídos do corpus

REFERÊNCIA	FONTE						
	Scopus	EBSCO	Science Direct	B-On	LISA	Web of Science	Scielo
AJIFERUKE, I.; LU, K.; WOLFRAM, D. Who are the research disciples of an author? Examining publication recitation and oeuvre citation exhaustivity Journal of Informetrics , v. 5, p. 292 - 302, 2011.	X		X		X	X	
ANDONOV, Y. Some aspects of the specificity of vocabulary and terminology exercises. International Classification , v. 19, n. 2, p. 99, 1992						X	
BLAIR, D. C. The challenge of commercial document retrieval, part 1: major issues and a framework based on search exhaustivity, determinacy of representation and document collection size. Information Processing and Management , v. 38, p. 273 – 291, 2002.						X	
CHENGLIN, W. et al. A new DS _m T combination rule in open frame of discernment and its application. Science China Information Sciences ,	X						

v. 55, n. 3, p. 551 – 557, 2012							
DILLON, M. Thesaurus-based automatic book indexing. INFO. PROC. & MGMT , v. 18, n.4, p. 167-178, 1982.					X	X	
GONZÁLEZ, J. A. M. et al. Avaliação de repertórios brasileiros em agricultura, ciência da informação e direito: uma análise de conteúdo. Ciência da Informação , v. 27, n. 3, 1998.							X
LANCASTER, F. W.; MILLS, J. Testing Indexes and Index Language Devices: The ASLIB Cranfield Project. American Documentation , v. 15, n. 1, p. 4 – 14, 1964.					X	X	
LIOMA, C.; OUNIS, I. A syntactically-based query reformulation technique for information retrieval. Information Processing & Management , v. 44, p. 143 - 162, 2008.			X				
LOGAN, E. L. Subject specificity of co-author clusters. Proceedings Of The American Society For Information Science , v. 22, p. 124-126						X	
MANIEZ, J. Relationship in thesauri: some critical remarks. International Classification , v. 15, n. 3, 1988.						X	
NELSON, M. J.; TAGUE, J. M. Split size-rank models for the distribution of index terms. Journal of the American Society for Information Science , v. 36, n. 5, 1985.	X						
SCHABAS, A. Postcoordinate Retrieval: A Comparison of Two Indexing Languages. Journal of the American Society for Information Science , v. 33, n. 1, p. 32 – 36, 1982.	X						
SUMNER, R. G. Utilizing the age of references to control the exhaustivity of the reference representation in information retrieval. Information processing and management , v. 31, n. 01, p. 29 - 45, 1995.				X		X	
WILSON, M. D. Specificity, syndetic structure, and subject access to works about individual corporate bodies. Library resources & technical services , v. 42, n. 4, p. 272-281, 1998.						X	
ZELEPUKINA, V. A. et al. Particular experience in design and implementation of a Current Research Information System in Russia: national specificity. Procedia Computer Science , v. 33, p. 168 – 173, 2014.						X	
ZUNDE, P.; DEXTER, M. E. Indexing consistency and quality. American Documentation , v. 20, n. 3, p. 259 – 268, 1969.					X	X	

TOTAL	4	0	2	1	4	11	1
--------------	---	---	---	---	---	----	---

Fonte: elaborado pela autora.

Registram-se, no quadro a seguir, os artigos que compuseram o corpus do trabalho, portanto, objeto de análise da pesquisa. O quadro é composto pelas referências dos artigos e as fontes, além do total de artigos encontrados em cada uma delas.

Quadro 4 - Corpus do trabalho

REFERÊNCIA	FONTE					
	Scopus	EBSCO	Science Direct	B-On	LISA	Web of Science
BLAIR, D. The challenge of commercial document retrieval, part I: major issues and a framework based on search exhaustivity, determinacy of representation and document collection size. Information Processing & Management , v. 38, p. 273 - 291, 2002.		X	X	X		
BURGIN, R. The effect of indexing exhaustivity on retrieval performance. Information processing and management , v. 27, n. 6, p. 623 – 628, 1991.						X
BURGIN, R. The retrieval effectiveness of 5 clustering algorithms as a function of indexing exhaustivity. Journal of the American Society for Information Science , v. 46, n.8, p. 562-572, 1995.		X	X	X		X
CHEN, X. The influence of existing consistency measures on the relationship between indexing consistence and exhaustivity. In: ARSENAULT, C.; TENNIS, J. T. (Ed.). Cultural and Identity in Knowledge Organization . Würzburg: ERGON Verlag, 2008. p. 144 - 150. (Advances in Knowledge Organization, 11).	X					
CHUNG, E. K.; YOON, J. W. Categorical and specificity differences between user-supplied tags and search query terms for images: an analysis of Flickr tags and Web image search queries. Information Research , v. 14, n. 03, p. 01 – 22, 2009.						X
ROBERTSON, S. E. Documentation note. Journal of Documentation , v. 30, n. 1, p.41 – 46, 1974.						X
FERNANDÉZ, L. M. M.; ALONSO, M. I.; MONTÁVEZ, A. J. V. La consistencia como indicador de la fiabilidad (corrección) de la indización: los casos de LISA, PASCAL y VOCED. Scire , v. 12, n. 01, p. 85 - 98, 2006.					X	X
GIVEN, L. M; OLSON, H. A. Indexing and the "organized" researcher. Indexer , v. 23, n. 3, p. 129-133, 2003.					X	X
GIVEN, L. M; OLSON, H. A. Knowledge organization in research: a conceptual model for organizing data. Library & Information Science Research , v. 25, n. 2, p. 157-176, 2003.			X			
GÓMEZ, J. C. G.; OLIVARES, J. L. G. Aproximación a la evaluación cuantitativa de los sistemas de recuperación de información de la prensa en Internet: Exhaustividad y precisión. Scire , v. 7, n. 01, p. 143 -					X	X

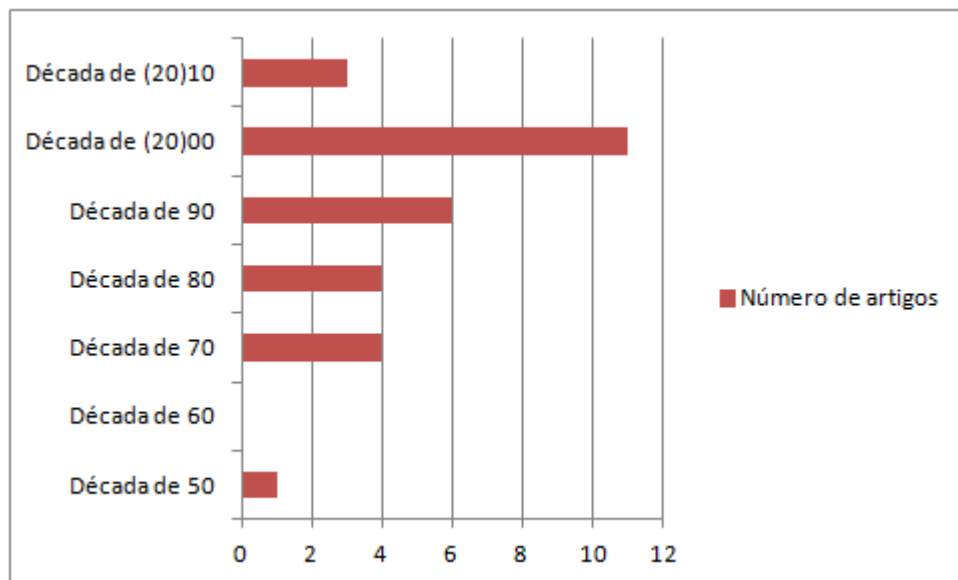
152, 2001.						
GUILLEMETTE, M. G.; LAROCHE, M.; CADIEUX, J. Defining decision making process performance: Conceptualization and validation of an index. Information & Management , v. 51, p. 618 - 626, 2014.			X			
GUIMARAES, J. A. C. Políticas de análisis y representación de contenido para la gestión del conocimiento en las organizaciones. Scire , v. 6, n. 02, p. 49 - 58, 2000.					X	X
HARDIN, A.; WILLET, P. Indexing exhaustivity and the computation of similarity matrices. Journal of the American Society for Information Science and Technology , v. 31, n. 4, p. 298 - 300, 1980.		X	X	X		X
HUGHES, A. V.; RAFFERTY, P. Inter-indexer consistency in graphic materials indexing at the National Library of Wales. Journal of documentation , v. 67, n. 1, 2011, p. 9 - 32.	X					
KIM, G. Relationship between index term specificity and relevance judgment. Information Processing and Management , v. 42, p. 1218 - 1229, 2006.		X	X	X		X
LYKKE, M.; PRICE, S.; DELCAMBRE, L. How doctors search: A study of query behaviour and the impact on search results. Information Processing & Management , v. 48, p. 1151 - 1170, 2012.			X			
PIRKOLA, A.; KESKUSTALO, H.; JÄRVELIN, K. The Effects of Conjunction, Facet Structure, and Dictionary Combinations in Concept-Based Cross-Language Retrieval. Information Retrieval , v. 1, n. 3, 1999.					X	X
HEINE, M. H. An investigation of the relative influences of database informativeness, query size and query term specificity on the effectiveness of Meadline searching. Journal of Information Science , v. 21, n. 03, p. 173 - 185, 1995.	X					
SHAW JR., W. M. Subject and Citation Indexing. Part I: The Clustering Structure of Composite Representations in the Cystic Fibrosis Document Collection. Journal of the American Society for Information Science , v. 42, n. 9, p. 669 - 675, 1991.			X			
SHAW, W M, JR. An investigation of document partitions. INFO. PROC. MANAGE , v. 22, n. 1, p. 19-28, 1986.					X	X
SPARCK-JONES, K A statistical interpretation of term specificity and its application in retrieval. Journal of documentation , v. 28, n. 1, 1972, p. 11 - 21.	X				X	X
SPARCK-JONES, K. Does indexing exhaustivity matter?. Journal of the American Society for Information Science and Technology , v. 24, p. 313 - 316, 1973.	X	X	X	X	X	X
TAUBE, M. Specificity in Subject Headings and coordinate indexing. Library Trends , v. 1, n. 2, p. 219 - 223, 1952.		X				
VAKKARI, P. et al. Query exhaustivity, relevance feedback and search success in automatic and interactive query expansion. Journal of documentation , v. 60, n. 02, 2004, p. 109 - 127.	X				X	X
WEINBERG, B. H.; CUNNINGHAM, J. A. Term specificity and online postings - inverse relationship. Proceedings Of The American Society For Information Science , v. 21, p. 144-147, 1984.						X
WILSON, P. End of specificity. Library Resources &						X

Technical Services, v. 23, n. 2, p. 116 – 122, 1979.						
WOLFRAM, D. Inter-Record Linkage Structure in a Hypertext Bibliographic Retrieval System. Journal of the American Society for Information Science , v. 46, n. 10, 1996, p. 765 – 774.	X					
WOLFRAM, D.; ZHANG, J. An investigation of the influence of indexing exhaustivity and term distributions on a document space. Journal of the American Society for Information Science and Technology , v. 53, n. 11, p. 943 - 952, 2002.	X	X	X	X	X	X
WONG, S. K. M.; YAO, Y. Y. An information-theoretic measure of term specificity. Journal of the American Society for Information Science , v. 43, n. 1, p. 54-61, 1992.						X
TOTAL	9	7	10	7	10	20

Fonte: elaborado pela autora.

Foi possível perceber que a maioria do corpus recuperado data de mais de dez anos. Apenas três dos artigos utilizados na pesquisa são mais recentes que esse período e somente um deles tem menos de três anos (KIM, 2006; LYKKE; PRICE; DELCAMBRE, 2012; GUILLEMETTE; LAROC; CADIEUX, 2014). Esse fator é representado no gráfico a seguir, que demonstra a contribuição de cada década para o *corpus* de pesquisa e se baseia nas informações dos artigos recuperados nas bases, portanto dependem do que é indexado por elas.

Figura 2 - Artigos publicados por década



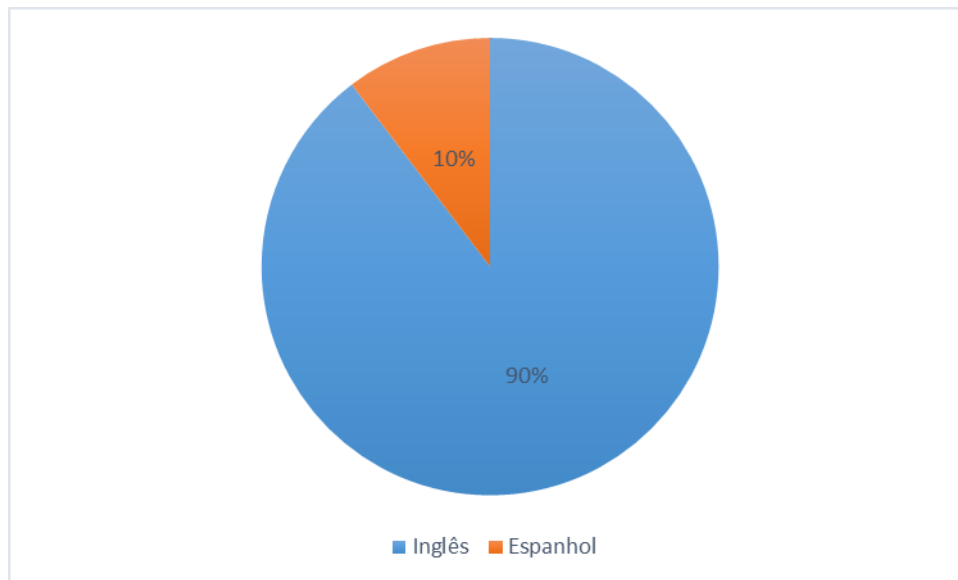
Fonte: elaborado pela autora

Por meio do gráfico, é possível perceber que, antes da década de 1990, a exaustividade e a especificidade ainda eram pouco retratadas. Já nos primeiros anos do milênio, elas foram

abordadas de maneira mais efusiva - o que pode ser justificado pelo avanço de novas tecnologias e a visão mais quantitativa dessas características -, mas voltou a decair a partir da segunda década do século XXI. Pode-se ressaltar que, com uma publicação mais efusiva no início dos anos 2000, o número de publicações no tema ainda é pouco significativo.

O idioma dos artigos também possui uma característica expressiva: uma maioria evidente dos artigos foi escrita em inglês, com uma minoria escrita em espanhol, conforme se verificará no próximo gráfico. O fato de os artigos terem sido escritos e divulgados nos dois principais idiomas mundiais demonstra que a barreira linguística é uma realidade e que, assim como pode trazer certa padronização para a ciência, pode também afetar a divulgação do tema, pois artigos escritos em diferentes idiomas acabam por não ser tão divulgados.

Figura 3 - Idioma dos artigos

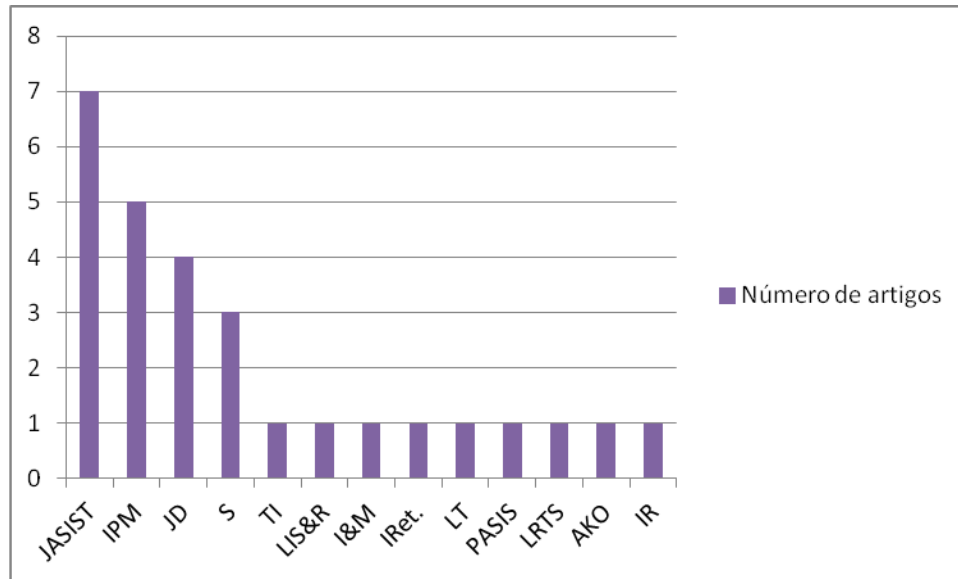


Fonte: elaborado pela autora.

No próximo gráfico, é possível perceber que, embora uma grande variedade de revistas tenha contribuído para o corpus, uma em particular se destaca - *Journal of American Society for Information Science and Technology* - representando 27% da composição do corpus. Diante disto, é possível inferir que essa publicação possui tradição no tema, e que isso talvez se deva ao fato de a revista trabalhar com publicações mais técnicas e referentes às novas tecnologias, visão muitas vezes compartilhada pelos teóricos que escrevem sobre os princípios de exaustividade e

especificidade. Os artigos publicados no periódico *Journal of American Society for Information Science* foram acrescidos na contagem do periódico *Journal of American Society for Information Science and Technology*, uma vez que os dois são o mesmo periódico, apenas com uma alteração na nomenclatura, a partir do ano 2000 (ASIS&T, 2015).

Figura 4 - Artigos por periódico¹



Fonte: elaborado pela autora.

Após a busca dos artigos e o preenchimento das fichas de registro (que se encontram no Apêndice 1), foi aplicada a análise de conteúdo, a qual permitiu a elaboração dos quadros seguintes. Algumas inferências estão repetidas nos dois quadros por se tratarem de ambos os termos. Os quadros são compostos pela referência dos artigos, que se encontram associados os extratos, e pelo registo da informação, que resultou na análise de conteúdo, tendo como base as respetivas categorias (conceitos e conclusões).

Segue a apresentação do quadro de análise de conteúdo relativo ao termo exaustividade.

¹ As publicações se referem respectivamente a: *Journal of American Society for Information Science and Technology*, *Information, Processing and Management*, *Journal of Documentation*, *Scire*, *The Indexer*, *Library Information Science & Retrieval*, *Information & Management*, *Information Retrieval*, *Library Trends*, *Proceedings of American Society for Information Science*, *Library Resources and Technical Services*, *Advances in Knowledge Organization* e *Information Retrieval*.

Quadro 5 - Análise de conteúdo do termo exaustividade

EXAUSTIVIDADE		
REFERÊNCIA	CONCEITO	CONCLUSÃO
BLAIR, D. The challenge of commercial document retrieval, part I: major issues and a framework based on search exhaustivity, determinancy of representation and document collection size. Information Processing & Management , v. 38, p. 273 - 291, 2002.	-----	Quando a busca é feita de maneira exaustiva pelo usuário, ela recupera uma maior parte de documentos úteis.
BURGIN, R. The effect of indexing exhaustivity on retrieval performance. Information processing and management , v. 27, n. 6, p. 623 – 628, 1991.	-----	Um compromisso efetivo entre exaustividade e especificidade, aumenta a recuperação informacional.
BURGIN, R. The retrieval effectiveness of 5 clustering algorithms as a function of indexing exhaustivity. Journal of the American Society for Information Science , v. 46, n.8, p. 562-572, 1995.	-----	Quando há um compromisso entre a exaustividade e a especificidade, há uma potencialização da recuperação informacional.
CHEN, X. The influence of existing consistency measures on the relationship between indexing consistence and exhaustivity. In: ARSENAULT, C.; TENNIS, J. T. (Ed.). Cultural and Identity in Knowledge Organization . Würzburg: ERGON Verlag, 2008. p. 144 - 150. (Advances in Knowledge Organization, 11).	A exaustividade diz respeito ao número de termos indexados.	Quando diferentes indexadores possuem o mesmo nível de exaustividade, a consistência da indexação não é prejudicada.
FERNANDEZ, L. M. M.; ALONSO, M. I.; MONTÁVEZ, A. J. V. La consistencia como indicador de la fiabilidad (corrección) de la indización: los casos de LISA, PASCAL y VOCED. Scire , v. 12, n. 01, p. 85 - 98, 2006.	-----	A exaustividade e a especificidade, aliada à consistência e à confiabilidade traz um indicador de qualidade à indexação.
GIVEN, L. M; OLSON, H. A. Indexing and the "organized" researcher. Indexer , v. 23, n. 3, p. 129-133, 2003.	Exaustividade é a extensão da representação – o número de fatores indexados.	A exaustividade e a especificidade são estratégias para uma efetiva organização e recuperação da informação. Embora a utilização de características como a exaustividade e a especificidade seja considerado o ideal, os indexadores consideram que há grandes problemas em sua aplicação. Um alto nível de exaustividade tende a um baixo nível de precisão, já que a adição de um grande número de termos traz a recuperação de informações consideradas irrelevantes. De forma recíproca, um alto nível de especificidade traz um baixo nível de revocação, já que categorias específicas recuperam menos dados que categorias mais gerais. De maneira prática, a aplicação do conceito exaustividade está relacionado a análise de quanto determinado tópico está incluído no livro, para posterior indexação do assunto analisado e ainda, análise de quantos termos são necessários para atender as necessidades informacionadis dos usuários.
GIVEN, L. M; OLSON, H. A. Knowledge organization in research: a conceptual model for organizing data. Library & Information Science Research , v. 25, n. 2, p. 157-176, 2003.	Exaustividade é a extensão da representação – o número de fatores indexados, uma característica do vocabulário e sua aplicação.	A exaustividade e a especificidade são estratégias para uma efetiva organização e recuperação da informação. Características que influenciam o vocabulário controlado. A exaustividade está diretamente ligada à

		<p>revocação, que calcula quanto da informação relevante é recuperada, a partir do total de informações relevantes disponíveis.</p> <p>Se o nível de exaustividade é alto, mais termos são utilizados, aumentando as possibilidades de recuperação.</p> <p>O nível de exaustividade é inversamente proporcional ao nível de precisão.</p> <p>Exaustividade está relacionada de quantas e quais variáveis são escolhidas na representação.</p> <p>Embora a utilização de características como a exaustividade e a especificidade seja considerado o ideal, os indexadores consideram que há grandes problemas em sua aplicação.</p> <p>Um alto nível de exaustividade tende a um baixo nível de precisão, já que a adição de um grande número de termos traz a recuperação de informações consideradas irrelevantes. De forma recíproca, um alto nível de especificidade traz um baixo nível de revocação, já que categorias específicas recuperam menos dados que categorias mais gerais.</p>
GÓMEZ, J. C. G.; OLIVARES, J. L. G. Aproximación a la evaluación cuantitativa de los sistemas de recuperación de información de la prensa en Internet: Exaustividad y precisión. Scire , v. 7, n. 01, p. 143 - 152, 2001.	-----	<p>A exaustividade está diretamente ligada à revocação do sistema.</p> <p>Uma melhor relação entre exaustividade e especificidade pode ser adquirida a partir de um critério de relevância dos documentos recuperados.</p>
GUILLEMETTE, M. G.; LAROCHE, M.; CADIEUX, J. Defining decision making process performance: Conceptualization and validation of an index. Information & Management , v. 51, p. 618 - 626, 2014.	-----	<p>Informações exaustivas, trazem um maior número de informações para auxiliar no momento de tomada de decisões.</p>
GUIMARÃES, J. A. C. Políticas de análisis y representación de contenido para la gestión del conocimiento en las organizaciones. Scire , v. 6, n. 02, p. 49 - 58, 2000.	-----	<p>Os níveis de exaustividade e especificidade devem estar de acordo com o perfil dos usuários, sempre considerando suas particularidades.</p> <p>Exaustividade e especificidade são características chave em um sistema de recuperação informacional, além de auxiliar a gestão do conhecimento nas organizações.</p>
HARDIN, A.; WILLET, P. Indexing exhaustivity and the computation of similarity matrices. Journal of the American Society for Information Science and Technology , v. 31, n. 4, p. 298 - 300, 1980.	Número de termos de indexação atribuídos a cada um dos documentos de uma coleção.	-----
HUGHES, A. V.; RAFFERTY, P. Inter-indexer consistency in graphic materials indexing at the National Library of Wales. Journal of Documentation , v. 67, n. 01, p. 09 - 32, 2011.	O número de termos atribuídos a um documento.	-----
LYKKE, M.; PRICE, S.; DELCAMBRE, L. How doctors search: A study of query behaviour and the impact on search results. Information Processing & Management , v. 48, p. 1151 - 1170, 2012.	-----	<p>O aumento da exaustividade leva também, a um aumento significativo de termos precisos.</p> <p>A escolha das facetas de busca e da terminologia são fatores importantes na exaustividade.</p>
PIRKOLA, A.; KESKUSTALO, H.;	O número de conceitos ou facetas	

JARVELIN, K. The Effects of Conjunction, Facet Structure, and Dictionary Combinations in Concept-Based Cross-Language Retrieval. Information Retrieval , v. 1, p. 217 - 250, 1999.	atribuídas no momento da busca informacional	-----
SHAW JR., W. M. Subject and Citation Indexing. Part I: The Clustering Structure of Composite Representations in the Cystic Fibrosis Document Collection. Journal of the American Society for Information Science , v. 42, n. 9, p. 669 - 675, 1991.	-----	A extensão com que um documento é representado o associa a outros documentos do mesmo assunto, evidenciando uma função da exaustividade.
SHAW, W. An investigation of document partitions. Information processing and management , v. 22, n. 01, p. 19 - 28, 1986.	-----	Os níveis de exaustividade e especificidade podem ser medidos por meio de fórmulas matemáticas.
SPARCK-JONES, K. A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation , v. 60, n. 05, p. 493 - 502, 2004.	A cobertura dos vários tópicos de um documento, por meio dos termos atribuídos a ele. O número de termos que a descrição de um documento contém.	-----
SPARCK-JONES, K. Does indexing exhaustivity matter?. Journal of the American Society for Information Science and Technology , v. 24, n. 5, p. 313 - 316, 1973.	O número de tópicos ou temas indexados de um documento.	Quando os termos são extraídos apenas como palavras-chave, sem a utilização de um vocabulário controlado, são observadas significativas diferenças no nível de exaustividade. Diferentes níveis de exaustividade são observados a partir de diferentes fontes de informação, sendo os resumos uma fonte de um alto nível de exaustividade. O nível de exaustividade e as requisições de um documento não são independentes, e uma boa recuperação informacional pode ser alcançada se um fator varia do outro. Nao é possível investigar a exaustividade de forma isolada, ela possui conexão com a característica especificidade.
VAKKARI, P. et al. Query exhaustivity, relevance feedback and search success in automatic and interactive query expansion. Journal of documentation , v. 60, n. 02, p. 109 - 127, 2004.	-----	Quanto mais exaustivos os usuários forem no momento da busca informacional, mais úteis serão as referências encontradas. A exaustividade no momento da busca, indica as facetes que os termos de busca possuem.
WILSON, P. End of specificity. Library Resources & Technical Services , v. 23, n. 2, p. 116 - 122, 1979.	Extensão da indexação.	A exaustividade é uma característica útil para o acesso por assunto.
WOLFRAM, D. Inter-Record Linkage Structure in a Hypertext Bibliographic Retrieval System. Journal of the American Society for Information Science and Technology , v. 46, n. 10, p. 764 - 774, 1996.	A exaustividade refere-se ao número de descritores selecionáveis de um recurso, além de dar mais opções de pesquisa ao recurso descrito.	Percebe-se que a partir de uma descrição exaustiva, é possível estabelecer uma maior articulação entre recursos, uma vez que documentos com mais termos possibilitam mais vias de exploração.
WOLFRAM, D., ZHANG, J. An Investigation of the Influence of Indexing Exhaustivity and Term Distributions on a Document Space. Journal of the American Society for Information Science and Technology , v. 53, n. 11, p. 943 - 952, 2002.	Quantos termos de indexação são utilizados para identificar um documento. O número de termos atribuídos a um documento.	Com um alto nível de exaustividade, uma menor densidade de documentos é encontrada, facilitando a distinção de documentos.

Fonte: Elaborado pela autora.

Pelo exposto, infere-se que a exaustividade apresenta-se conceitualmente como a extensão

com a qual um assunto é representado ou com a quantidade de termos usados para representar determinado documento. Ela tem a finalidade de auxiliar na recuperação informacional, além de sua correlação com a especificidade se apresentar como ideal. Segue, assim, a apresentação do quadro de análise de conteúdo relativo ao termo especificidade.

Quadro 6 - Análise de conteúdo do termo especificidade

ESPECIFICIDADE		
REFERÊNCIA	CONCEITO	CONCLUSÕES
BURGIN, R. The effect of indexing exhaustivity on retrieval performance. Information processing and management , v. 27, n. 6, p. 623 – 628, 1991.	-----	Um compromisso efetivo entre exaustividade e especificidade aumenta a recuperação informacional.
BURGIN, R. The retrieval effectiveness of 5 clustering algorithms as a function of indexing exhaustivity. Journal of the American Society for Information Science , v. 46, n.8, p. 562-572, 1995.	-----	Quando há um compromisso entre a exaustividade e a especificidade, há uma potencialização da recuperação informacional.
CHUNG, E. K.; YOON, J. W. Categorical and specificity differences between user-supplied tags and search query terms for images: an analysis of Flickr tags and Web image search queries. Information Research , v. 14, n. 03, p. 01 – 22, 2009.	-----	Um documento pode ser articulado de forma diferente dependendo do seu nível de especificidade - um mesmo documento pode ser descrito tematicamente como “animais”, “cachorros” ou “Chiuaua”.
FERNANDÉZ, L. M. M.; ALONSO, M. I.; MONTÁVEZ, A. J. V. La consistencia como indicador de la fiabilidad (corrección) de la indización: los casos de LISA, PASCAL y VOCED. Scire , v. 12, n. 01, p. 85 - 98, 2006.	-----	A exaustividade e a especificidade, aliada à consistência e à confiabilidade, traz um indicador de qualidade à indexação. Exaustividade e especificidade são características chave em um sistema de recuperação informacional, além de auxiliar a gestão do conhecimento nas organizações.
GIVEN, L. M; OLSON, H. A. Indexing and the "organized" researcher. Indexer , v. 23, n. 3, p. 129-133, 2003.	Especificidade é o nível de detalhamento do vocabulário, o número de níveis hierárquicos definidos.	A exaustividade e a especificidade são estratégias para uma efetiva organização e recuperação da informação. O nível de especificidade deve atender ao grupo de usuários da unidade. Embora a utilização de características como a exaustividade e a especificidade sejam consideradas o ideal, os indexadores consideram que há grandes problemas em sua aplicação. Um alto nível de exaustividade tende a um baixo nível de precisão, já que a adição de um grande número de termos traz a recuperação de informações consideradas irrelevantes. De forma recíproca, um alto nível de especificidade traz um baixo nível de revocação, já que categorias específicas recuperam menos dados que categorias mais gerais.
GIVEN, L. M; OLSON, H. A. Knowledge organization in research: a conceptual model for organizing data. Library & Information Science Research , v. 25, n. 2, p. 157-176, 2003.	Especificidade é o nível de detalhamento do vocabulário, o número de níveis hierárquicos definidos.	A exaustividade e a especificidade são estratégias para uma efetiva organização e recuperação da informação. Características que influenciam o vocabulário controlado. O nível de especificidade deve atender ao grupo de usuários da unidade. Especificidade está diretamente ligada à precisão. O nível de especificidade é inversamente proporcional ao nível de revocação. Embora a utilização de características como a exaustividade e a especificidade seja considerado o ideal, os indexadores consideram que há grandes problemas em sua aplicação. Um alto nível de exaustividade tende a um baixo nível de

		precisão, já que a adição de um grande número de termos traz a recuperação de informações consideradas irrelevantes. De forma recíproca, um alto nível de especificidade traz um baixo nível de revocação, já que categorias específicas recuperam menos dados que categorias mais gerais.
GÓMEZ, J. C. G.; OLIVARES, J. L. G. Aproximación a la evaluación cuantitativa de los sistemas de recuperación de información de la prensa en Internet: Exhaustividad y precisión. Scire , v. 7, n. 01, p. 143 - 152, 2001.	-----	Uma melhor relação entre exaustividade e especificidade pode ser adquirida a partir de um critério de relevância dos documentos recuperados.
GUIMARAES, J. A. C. Políticas de análisis y representación de contenido para la gestión del conocimiento en las organizaciones. Scire , v. 6, n. 02, p. 49 - 58, 2000.	-----	Os níveis de exaustividade e especificidade devem estar de acordo com o perfil dos usuários, sempre considerando suas particularidades.
HEINE, M. H. An investigation of the relative influences of database informativeness, query size and query term specificity on the effectiveness of Meadline searching. Journal of Information Science , v. 21, n. 03, p. 173 - 185, 1995.	-----	O nível de especificidade deve ser variável, dependendo da escolha do usuário para campos de registro acessíveis ao termo.
KIM, G. Relationship between index term specificity and relevance judgment. Information Processing and Management , v. 42, p. 1218 - 1229, 2006.	Atividade prática, que determina termos ou descritores para cada unidade documental. A relação do termo com o seu documento e quão preciso será esse termo em relação ao document representado. Uma importante característica da indexação para a recuperação de documentos relevantes Capacidade discriminativa, que distingue uma classe de documentos de outra ou uma subclasse de outra.	Influenciar em um julgamento relevante por parte do usuário juntamente com a linguagem, influenciando também, a recuperação informacional A especificidade pode relacionar o termo com o assunto do documento e entre a estrutura hierárquica dos termos. A especificidade também afirma que um grupo de termos é mais representativo que um único termo e existem termos mais específicos, que outros. A relação entre termo-documento ainda é inábil de ser medida por meios automáticos.
ROBERTSON, S. E. Documentation note. Journal of Documentation , v. 30, n. 1, p.41 - 46, 1974.	-----	Um termo menos frequente tende a ser um termo menos específico. Uma linguagem documentária é específica, caso contenha um grande número de termos específicos. A especificidade pode ser descrita por meio de fórmulas matemáticas. Questões relacionadas ao uso de métodos utilizando a especificidade merecem futuras pesquisas.
SHAW, W. An investigation of document partitions. Information processing and management , v. 22, n. 01, p. 19 - 28, 1986.	-----	Os níveis de exaustividade e especificidade podem ser medidos por meio de fórmulas matemáticas.
SPARCK-JONES, K. A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation , v. 60, n. 05, p. 493 - 502, 2004.	O nível de detalhe que determinado conceito é representado.	Um termo é mais ou menos específico, uma vez que é mais ou menos detalhado e preciso.
TAUBE, M. Specificity in Subject Headings and coordinate indexing. Library Trends , v. 1, n. 2, p. 219 - 223, 1952.	-----	De forma a obter uma maior especificidade, os termos representativos podem ser combinados, criando uma intersecção entre os documentos, que se diferem a partir do grau de especificidade utilizado.
WEINBERG, B. H.; CUNNINGHAM, J. A. Term specificity and online postings - inverse relationship. Proceedings Of The American Society For Information Science , v. 21, p. 144-147, 1984.	-----	O nível de especificidade deve aumentar conforme os termos tornam-se mais antigos e consolidados.
WONG, S. K. M.; YAO, Y. Y. Na information-theoretic measure of term specificity. Journal of the American	-----	O nível de especificidade pode ser medido por meio de fórmulas matemáticas.

Society for Information Science, v. 43, n. 1, p. 54-61, 1992.		
---	--	--

Fonte: Elaborado pela autora.

Face ao apresentado, intui-se que a especificidade pode ser definida como particularidade ou o nível de detalhamento que será dado no momento da representação, além de contribuir para o nível hierárquico da representação. A especificidade pode afetar a recuperação informacional e se correlacionar com a exaustividade para uma melhor indexação.

Com base nas categorias de análise consideradas (conceito e conclusão) e na análise de conteúdo dos artigos consultados, foi realizada uma sistematização dos valores éticos, registrados no quadro seguinte. Trata-se de um quadro constituído pelos seguintes elementos: a) valor ético inferido da análise de conteúdo; b) a explicação teórica sobre a qual assenta o valor ético; c) a referência do artigo analisado.

Quadro 7 - Sistematização de valores éticos a partir da análise de conteúdo do corpus

VALOR	EXPLICAÇÃO TEÓRICA	REFERÊNCIA
O número de conceitos extraídos de um documento a forma e o nível de especificidade que são representados são diretamente proporcionais à informação que se pretende representar	A exaustividade está ligada ao número de conceitos que podem ser representados a partir de determinado documento e a especificidade, é a particularidade com que se quer definir um documento, dessa forma, dependem sempre da análise conceitual do documento.	GIVEN, L. M; OLSON, H. A. Knowledge organization in research: a conceptual model for organizing data. Library & Information Science Research , v. 25, n. 2, p. 157-176, 2003. GUIMARÃES, J. A. C. Políticas de análisis y representación de contenido para la gestión del conocimiento en las organizaciones. Scire , v. 6, n. 02, p. 49 - 58, 2000. HARDIN, A.; WILLET, P. Indexing exhaustivity and the computation of similarity matrices. Journal of the American Society for Information Science and Technology , v. 31, n. 4, p. 298 - 300, 1980. HEINE, M. H. An investigation of the relative influences of database informativeness, query size and query term specificity on the effectiveness of Meadline searching. Journal of Information Science , v. 21, n. 03, p. 173 - 185, 1995. JONES, K. S. Does indexing exhaustivity matter?. Journal of the American Society for Information Science and Technology , v. 24, p. 313 - 316, 1973. WOLFRAM, D.; ZHANG, J. An investigation of the influence of indexing exhaustivity and term distributions on a document space. Journal of the American Society for Information Science and Technology , v. 53, n. 11, p. 943 - 952, 2002.
Os níveis de exaustividade e de especificidade concorrem para o impacto na recuperação informacional	Quando um documento é indexado a partir de várias facetas - de todos os seus conteúdos e particularidades - as chances de posteriormente ser recuperados, são maiores.	BURGIN, R. The retrieval effectiveness of 5 clustering algorithms as a function of indexing exhaustivity. Journal of the American Society for Information Science , v. 46, n.8, p. 562-572, 1995. GIVEN, L. M; OLSON, H. A. Knowledge organization in research: a conceptual model for organizing data. Library & Information Science Research , v. 25, n. 2, p. 157-176, 2003. LYKKE, M.; PRICE, S.; DELCAMBRE, L. How doctors search: A study of query behaviour and the impact on search results. Information Processing & Management , v. 48, p. 1151 - 1170, 2012. WOLFRAM, D.; ZHANG, J. An investigation of the

		influence of indexing exhaustivity and term distributions on a document space. Journal of the American Society for Information Science and Technology , v. 53, n. 11, p. 943 - 952, 2002.
A articulação entre a exaustividade e a especificidade concorre para uma maior recuperação informacional	Essas duas características não podem ser vistas de forma isolada, o compromisso entre as duas influencia na recuperação informacional, pois podem ser vistas enquanto estratégias para elevar a recuperação.	BURGIN, R. The effect of indexing exhaustivity on retrieval performance. Information processing and management , v. 27, n. 6, p. 623 – 628, 1991. BURGIN, R. The retrieval effectiveness of 5 clustering algorithms as a function of indexing exhaustivity. Journal of the American Society for Information Science , v. 46, n.8, p. 562-572, 1995. GIVEN, L. M; OLSON, H. A. Indexing and the "organized" researcher. Indexer , v. 23, n. 3, p. 129-133, 2003. GIVEN, L. M; OLSON, H. A. Knowledge organization in research: a conceptual model for organizing data. Library & Information Science Research , v. 25, n. 2, p. 157-176, 2003. GUIMARÃES, J. A. C. Políticas de análisis y representación de contenido para la gestión del conocimiento en las organizaciones. Scire , v. 6, n. 02, p. 49 - 58, 2000. JONES, K. S. Does indexing exhaustivity matter?. Journal of the American Society for Information Science and Technology , v. 24, p. 313 - 316, 1973.
Os níveis de exaustividade e especificidade qualificam a indexação	As duas características aliadas, trazem consistência no processo da indexação, sendo características chave em um sistema de recuperação informacional.	FERNANDEZ, L. M. M.; ALONSO, M. I.; MONTÁVEZ, A. J. V. La consistencia como indicador de la fiabilidad (corrección) de la indexación: los casos de LISA, PASCAL y VOCED. Scire , v. 12, n. 01, p. 85 - 98, 2006.
Elevados níveis de exaustividade concorrem para uma maior atribuição de termos a um documento, que se fará de acordo com o serviço e o utilizador	Quanto maior o número de termos representados no índice, maiores são as opções de recuperação. Há ainda, a facilidade de distinguir um termo do outro.	BLAIR, D. The challenge of commercial document retrieval, part I: major issues and a framework based on search exhaustivity, determinancy of representation and document collection size. Information Processing & Management , v. 38, p. 273 - 291, 2002. CHEN, X. The influence of existing consistency measures on the relationship between indexing consistence and exhaustivity. In: ARSENAULT, C.; TENNIS, J. T. (Ed.). Cultural and Identity in Knowledge Organization . Würzburg: ERGON Verlag, 2008. p. 144 - 150. (Advances in Knowledge Organization, 11). VAKKARI, P. et al. Query exhaustivity, relevance feedback and search success in automatic and interactive query expansion. Journal of documentation , v. 60, n. 02, p. 109 – 127, 2004. WOLFRAM, D. Inter-record linkage structure in a hypertext bibliographic retrieval system. Journal of the American Society for Information Science and Technology , v. 46, n. 10, p. 764 - 774, 1996. WOLFRAM, D.; ZHANG, J. An investigation of the influence of indexing exhaustivity and term distributions on a document space. Journal of the American Society for Information Science and Technology , v. 53, n. 11, p. 943 - 952, 2002.
Um elevado nível de especificidade concorre para uma maior precisão do documento	A especificidade está intrinsecamente relacionada com a precisão com que determinado conceito será representado, trazendo detalhamento à representação.	CHUNG, E. K.; YOON, J. W. Categorical and specificity differences between user-supplied tags and search query terms for images: an analysis of Flickr tags and Web image search queries. Information Research , v. 14, n. 03, p. 01 – 22, 2009. GIVEN, L. M; OLSON, H. A. Indexing and the "organized" researcher. Indexer , v. 23, n. 3, p. 129-133, 2003. GIVEN, L. M; OLSON, H. A. Knowledge organization in research: a conceptual model for organizing data. Library & Information Science Research , v. 25, n. 2, p. 157-176, 2003. KIM, G. Relationship between index term specificity

		and relevance judgment. Information Processing and Management , v. 42, p. 1218 - 1229, 2006. ROBERTSON, S. E. Documentation note. Journal of Documentation , v. 30, n. 1, p.41 – 46, 1974. SPARCK-JONES, K. A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation , v. 60, n. 05, p. 493 - 502, 2004.
A expansão do nível de especificidade é proporcional à expansão do vocabulário	Conforme o vocabulário vai se expandindo e tornando-se mais consolidado, é coerente que o nível de especificidade torne-se cada vez mais preciso.	WEINBERG, B. H.; CUNNINGHAM, J. A. Term specificity and online postings - inverse relationship. Proceedings Of The American Society For Information Science , v. 21, p. 144-147, 1984.
A especificidade pode contribuir para a elaboração de uma rede semântica consistente no que diz respeito a relações hierárquicas	Os níveis hierárquicos só podem ser construídos a partir de relações semânticas e termos específicos, já que a especificidade possui capacidade discriminativa, distinguindo classes ou subclasses de termos.	GIVEN, L. M; OLSON, H. A. Knowledge organization in research: a conceptual model for organizing data. Library & Information Science Research , v. 25, n. 2, p. 157-176, 2003. GIVEN, L. M; OLSON, H. A. Indexing and the "organized" researcher. Indexer , v. 23, n. 3, p. 129-133, 2003. KIM, G. Relationship between index term specificity and relevance judgment. Information Processing and Management , v. 42, p. 1218 - 1229, 2006.
Um conjunto significativo de termos específicos atribuídos a um documento expressa de forma mais precisa os conceitos	Um grupo de termos específicos combinados são mais representativos que um único termo geral, além de criar uma intersecção entre os documentos representados	KIM, G. Relationship between index term specificity and relevance judgment. Information Processing and Management , v. 42, p. 1218 - 1229, 2006. TAUBE, M. Specificity in Subject Headings and coordinate indexing. Library Trends , v. 1, n. 2, p. 219 - 223, 1952.
A exaustividade corresponde a um nível elevado de revocação	Representações exaustivas trazem um maior número de informações pertinentes, e é esse fator que tendenciona a exaustividade a estar diretamente ligada à revocação - de todas as informações relevantes, quanto é recuperado pelo sistema	GIVEN, L. M; OLSON, H. A. Knowledge organization in research: a conceptual model for organizing data. Library & Information Science Research , v. 25, n. 2, p. 157-176, 2003. GIVEN, L. M; OLSON, H. A. Indexing and the "organized" researcher. Indexer , v. 23, n. 3, p. 129-133, 2003. GUILLEMETTE, M. G.; LAROCHE, M.; CADIEUX, J. Defining decision making process performance: Conceptualization and validation of an index. Information & Management , v. 51, p. 618 - 626, 2014. GÓMEZ, J. C. G.; OLIVARES, J. L. G. Aproximación a la evaluación cuantitativa de los sistemas de recuperación de información de la prensa en Internet: Exaustividad y precisión. Scire , v. 7, n. 01, p. 143 - 152, 2001.
A especificidade corresponde a níveis elevados de precisão	Se uma informação é específica, ela é, portanto, mais detalhada e precisa possível, o que faz com que a especificidade esteja diretamente relacionada à precisão do sistema	GIVEN, L. M; OLSON, H. A. Knowledge organization in research: a conceptual model for organizing data. Library & Information Science Research , v. 25, n. 2, p. 157-176, 2003. GIVEN, L. M; OLSON, H. A. Indexing and the "organized" researcher. Indexer , v. 23, n. 3, p. 129-133, 2003.
A exaustividade traz regularidade e densidade à indexação	Uma indexação exaustividade traz uma proporção e quantidades consideráveis de documentos recuperados.	WOLFRAM, D.; ZHANG, J. An investigation of the influence of indexing exhaustivity and term distributions on a document space. Journal of the American Society for Information Science and Technology , v. 53, n. 11, p. 943 - 952, 2002.
As relações semânticas entre termos-conceitos são identificadas de forma mais precisa a partir de meios intelectuais	Embora a indexação automática seja uma realidade, estabelecer relações semânticas entre termos-conceitos são mais bem identificadas por humanos, o que ressalta ainda mais o papel da especificidade em definir essas relações.	KIM, G. Relationship between index term specificity and relevance judgment. Information Processing and Management , v. 42, p. 1218 - 1229, 2006.

Fonte: Elaborado pela autora.

A partir da análise de conteúdo do corpus recuperado e da sistematização dos valores

éticos foi possível chegar às conclusões que são apresentadas no próximo capítulo.

7 CONCLUSÕES

Considerando os resultados da pesquisa, pode-se concluir que foram atingidos os objetivos iniciais, que consistiam em analisar a construção teórica dos termos exaustividade e especificidade, a partir de seus conceitos e finalidades, identificando-os e sistematizando-os como valores éticos na representação informacional, de acordo com a literatura da área.

O principal foco da pesquisa foi, justamente, demonstrar que características como a exaustividade e a especificidade devem ser entendidas além de medidas, ou seja, como valores éticos. Esse fato já havia sido evidenciado em pesquisas anteriores (GUIMARÃES et al., 2008; GUIMARÃES; MILANI; EVANGELISTA, 2015) e agora comprovado com o presente estudo, que teve como base a literatura do tema. Foi possível perceber como se vêm construindo os conceitos e as conclusões desses dois termos, a partir de um corpus que data desde a década de 1950 até os dias atuais (tendo uma amostra de 51% do total de textos recuperados, ou seja, 29 artigos de um total de 57), e, dessa forma, analisar e sistematizar os traços que os comprovem enquanto valores éticos indo ao encontro da proposição do presente estudo.

As características podem ser assim entendidas porque confirmaram ser mais que meras medidas, pois, quando levadas em consideração na escolha dos termos, afetam de forma direta a recuperação informacional. Dessa forma, a exaustividade e a especificidade atuam de maneira a reger a prática da escolha dos termos, em busca do bem-comum, que é o acesso à informação, característica própria de um valor ético, aquilo que possui um caráter positivo, que é bom e útil para a sociedade.

Foi possível concluir, ainda, que o presente estudo se integra em três princípios que cercam a organização do conhecimento, propostos por Hjørland (1994), conforme referido na revisão de literatura. O primeiro princípio estabelece que percepções ingênuas em organização do conhecimento não são possíveis em ciências mais avançadas. Não é possível que a indexação ocorra de forma ingênuo, deve-se ter em mente do papel de agente da informação que o profissional possui nesse momento, e que a escolha de termos pertinentes, precisos e que aborem a totalidade do documento faz-se necessário. O segundo princípio refere-se às categorizações, que devem refletir seus objetivos. No momento da escolha dos termos, os objetivos da instituição e do usuário devem sempre ser levados em consideração, além de estarem de acordo com os objetivos do documento que se quer representar. O terceiro princípio relaciona-se ao conceito de “polirrepresentação”. O autor ressalta que, em alguns momentos, é necessário

que o termo seja delimitado a uma especificidade de determinada disciplina; já em outras situações, que haja cruzamento entre diferentes ciências. Esse fator evidencia a importância de uma representação ser exaustiva e específica.

Do presente estudo, e de acordo com o que Barité (2001) determina como as premissas da organização do conhecimento, infere-se que: a) O conhecimento deve ser organizado para sua melhor utilização. A forma como se organiza e representa o documento deve sempre levar em consideração a potencialização de seu acesso. Assim, como foi demonstrado a partir dos resultados da pesquisa, um documento representado de forma exaustiva e específica, tem maior possibilidade de recuperação. b) O conhecimento é transmitido por meio de conceitos, que são traduzidos por termos, sendo a tradução uma tarefa inerente à organização do conhecimento, que deve ser realizada levando em consideração sua importância.

No âmbito ético da organização do conhecimento, retratado na revisão bibliográfica, foi possível concluir que o estudo vai ao encontro do que Dalhberg (1992) já havia alertado, há mais de duas décadas, sobre a atitude antiética de se elaborarem índices sem sua devida reflexão, não tendo zelo e compromisso com o documento representado. Ressalta-se, também, a relação do estudo com o que Olson (2002) destaca como o “poder de nomear”, ou seja, o poder do indexador no momento da escolha dos termos, podendo dar acesso ou censurar a informação. Assim, o presente estudo buscou dar continuidade a uma trajetória de pesquisa iniciada por Guimarães et al. (2008), sobre o campo dos valores éticos em organização do conhecimento.

A partir da análise de conteúdo, foi possível concluir que a exaustividade é conceituada enquanto ocorrência ou quantidade de termos designados para representar o conteúdo de um documento, abrangendo, de maneira extensa, os temas de um recurso e as diferentes facetas que este possui, além de seu valor ético. Quando relacionada com a especificidade, elas representam estratégias efetivas para a organização e recuperação da informação.

Já no que tange à especificidade, característica bem menos tratada no âmbito da indexação, de acordo com o *corpus* recuperado, ela pode ser definida como o nível de detalhamento com que um conceito será representado, além de ser um elemento necessário no momento da construção de hierarquias. Ela ainda é responsável por aprimorar a precisão, possuir capacidade discriminativa e auxiliar na intersecção de documentos, pois, dependendo do seu nível, permite combinar termos.

Essas duas características se aproximam quando fica evidente que, relacionadas,

aperfeiçoam de forma ainda mais positiva a recuperação e o acesso à informação. Quando há o compromisso entre elas, o nível de recuperação é consideravelmente maior, além de otimizar a construção de categorizações e a abordagem dos conceitos. A exaustividade e a especificidade se aproximam por se tratar de atividades práticas e de contexto intelectual, considerando sua abordagem semântica, responsáveis por estabelecer relações entre os documentos.

A natureza que a indexação possui em direcionar e guiar o usuário à informação, auxiliando-o a encontrar a informação desejada, já havia sido ressaltada por importantes teóricos da área (BORKO; BERNIER, 1978; ANDERSON, 1983; CLEVELAND, CLEVELAND, 2001; FETTERS, 2008). Diante do exposto na análise de conteúdo, aspectos como a exaustividade e a especificidade auxiliam e afetam positivamente essas questões, o que os torna inerentes a esse processo.

A análise de conteúdo realizada vai ao encontro daquilo que teóricos fundamentais do campo da indexação (FOSKETT, 1973; BORKO, BERNIER, 1978; LANCASTER, 1993; BONURA, 1994; CLEVELAND, CLEVELAND, 2001; SIMÕES, 2011) tratam - a exaustividade e especificidade -, características entendidas como a extensão de cobertura de um documento ou a exatidão e particularidade da representação desses conceitos, respectivamente. A análise também demonstrou que essas características se relacionam diretamente com a revocação e precisão do Sistema, sendo que, em alguns momentos, a literatura do corpus de pesquisa utiliza os dois conceitos como sinônimos de exaustividade e especificidade, respectivamente.

A título de conclusão, essas duas características revelam-se enquanto valores éticos, pois, a partir da análise dos termos, foi possível chegar a um grupo de ações: considerar o documento como fonte para os níveis de exaustividade e especificidade; estabelecer níveis que condigam com as necessidades dos usuários; e reconhecer que características como a exaustividade e a especificidade afetam positivamente a recuperação informacional, para auxiliar na prática indexadora, de forma a atingir o bem-comum da comunidade em que a unidade informacional está inserida. Dessa forma, é possível afirmar que a exaustividade e a especificidade se encontram diretamente proporcionais ao documento que se quer representar, o que revela seu compromisso com a fonte de informação. Elas ainda concorrem, efetivamente, para o aumento da recuperação informacional, já que um elevado nível de exaustividade contribuirá para um elevado número de termos e um elevado nível de especificidade, assegurando a presença de termos precisos e detalhados e potencializando o acesso à informação.

Além disso, características como a especificidade ajudam a estabelecer uma rede hierárquica de alto nível semântico, pois contribui para a precisão da representação dos conceitos de um documento, criando um conjunto significativo de termos. Já a exaustividade é diretamente proporcional a um alto nível de revocação, que é a medida responsável por calcular, de todo o universo de documentos pertinentes, quais são recuperados pelo sistema; além de contribuir sistematicamente para a regularidade e densidade da indexação, permitindo uma prática mais efetiva.

Por meio dos valores éticos, é possível concluir que a relação entre essas duas práticas não é só ideal, mas, também, necessária para auxiliar na recuperação informacional, pois um documento representado de forma exaustiva e específica tem maior probabilidade de ser recuperado e utilizado para a construção de novos conhecimentos, contribuindo, portanto, para o movimento de produção e organização do conhecimento. São atividades intrinsecamente intelectuais, pois, embora a indexação automática seja uma realidade, relações semânticas precisas entre termo-conceito são mais efetivamente estabelecidas por humanos.

Teóricos seminais na área da indexação (BORKO; BERNIER, 1978; BELL, 1991; CLEVELAND; CLEVELAND, 2001; FETTERS, 2008) já vinham alertando sobre as questões éticas nessa atividade, ressaltando o traço subjetivo da indexação e a necessidade em ser pautada por normas éticas. O presente estudo ressalta a relevância dessas questões, destacando o inerente papel da exaustividade e da especificidade para se atingir uma indexação realizada de forma ética.

Como referido nos resultados, a maioria do corpus recuperado possui mais de dez anos de publicação (90% do corpus), e apenas três artigos (10% do corpus) são mais recentes que esse período, dando uma regularidade de menos um artigo a cada três anos. Esse fator também é refletido no corpo teórico utilizado para tratar sobre o processo de indexação, que não registrou muitos avanços nas últimas décadas. Isso evidencia que os temas tratados na pesquisa são pouco abordados, o que pode acabar refletindo em pouca teorização sobre o assunto.

Outro agravante é o fato de os teóricos da indexação nem sempre considerarem as duas características estudadas pela pesquisa, e, quando fazem referência a elas, desenvolvem pouco este tema e raramente as abordam na perspectiva ética, que é o foco da pesquisa. Isso também fica evidente no corpus recuperado, que, mesmo não se delimitando a um período temporal, recuperou um número pouco relevante de artigos, sendo que nenhum deles foi encontrado em

repositórios institucionais. A falta de publicação sobre o tema nos repositórios é preocupante, uma vez que atesta a falta de tradição de pesquisa no tema, embora sua necessidade seja inerente.

A lacuna de um número expressivo de publicações na área pode ser refletida nos aspectos deontológicos do campo. Embora exista um número considerável de códigos de ética (63) dispostos no site de uma das mais influentes federações de bibliotecários (IFLA), em nenhum deles foram encontrados os termos exaustividade e especificidade, assim como nenhum deles fazia referência ao processo de indexação. Deste modo, fica evidente que a pouca pesquisa na área pode afetar a forma com que a prática profissional é realizada, causando o desinteresse por parte do profissional no bem fazer dos processos de representação.

No âmbito profissional, é importante ressaltar que, na maioria das vezes, o profissional se encontra dependente hierarquicamente de um superior agindo sob suas decisões. Muitas vezes, a decisão antiética de privar ou censurar a informação, por meio da não precisão dos termos utilizados ou da economia dos termos, é uma decisão que não diz respeito ao profissional que efetua a indexação, mas sim a quem o supervisiona. Cabe então ao profissional demonstrar os malefícios que uma má representação pode causar não só ao usuário, mas à própria instituição, que deve ter por objetivos principais o acesso à informação. É nesse sentido, ainda, que fica evidente a necessidade de uma política de indexação, estabelecendo diretrizes que abarquem essas duas características, para que o trabalho do indexador seja pautado em normas e assegurado por elas.

Outra conclusão está relacionada com a publicação da literatura, que, além de pouco atualizada e escassa, é predominantemente em língua inglesa, não tendo sido encontrado nenhum artigo que trate essas duas características e que seja pertinente ao corpus em língua portuguesa. Foi encontrado um reduzido número (três artigos) em espanhol, nas bases e repositórios utilizados para se estabelecer o corpus. Dada essa circunstância, intui-se que possam existir trabalhos e publicações sobre o tema em outros idiomas, mas que não são publicados por uma questão de barreira linguística, sendo esse um dos impasses na divulgação do tema.

A exaustividade e a especificidade como valores éticos são fruto de uma reflexão de um grupo mais voltado para questões culturais de forte presença latina (GARCIA GUTIÉRREZ, 2002; GUIMARÃES, FERNÁNDEZ-MOLINA, 2003; GUIMARÃES et al. 2008; LÓPEZ-HUERTAS, 2008), cuja produção se veicula principalmente nos anais de eventos (como os anais da *International Society for Knowledge Organization*), que não foram objeto de sistematização e

análise desta pesquisa.

Concluiu-se também que a maioria das publicações centra-se em uma revista em especial - *Journal of American Society for Information Science (and Technology)*- inferindo-se que esta revista é a que mais atende às questões relacionadas com exaustividade e especificidade. As publicações sobre o tema, na sua maioria, são de natureza técnica e aliadas às tecnologias informacionais, constituindo-se estes assuntos o foco das publicações neste periódico.

Ainda de acordo com a pesquisa, intui-se que as características exaustividade e especificidade são, efetivamente, essenciais na recuperação informacional, pois agregam *valor* ao documento, facilitando a sua recuperação, levando-se em consideração que um número maior de termos significativos e a escolha de termos mais precisos e particulares aumentam a probabilidade de recuperação. Além disso, essas duas características fazem com que a ponte entre o documento e o usuário, traço inerente ao processo da indexação, seja realizado de forma mais eficaz.

A pesquisa enfrentou algumas dificuldades durante sua elaboração, como a não recuperação de todos os textos completos encontrados por meio dos termos definidos, já que nem todas as publicações estão disponíveis em acesso aberto ou financiadas pelas universidades em que a pesquisa foi desenvolvida – Universidade Estadual Paulista e Universidade de Coimbra – e, ainda, o fato de haver uma publicação reduzida, por vezes inexistente, sobre o tema tratado. Acredita-se, porém, que essas limitações não afetaram o resultado da pesquisa, considerando, como já mencionado, que foi possível atingir os objetivos propostos.

Dada a relevância do tema, sugere-se que outras pesquisas sejam realizadas nesse sentido, buscando-se ampliar o idioma utilizado para a busca dos termos, para que se verifique se há uma tradição americana no tema ou que se comprove que, de fato, há uma barreira linguística no momento das publicações. Sugere-se, ainda, um estudo prático na área, para analisar se os profissionais indexadores conseguem identificar a importância das características estudadas e se é dada a devida relevância ao estudo das mesmas durante a formação dos profissionais.

Reforça-se a conclusão geral deste estudo, afirmando que a exaustividade e a especificidade não podem ser entendidas apenas como meras medidas, mas devem ser consideradas, no processo de indexação, como características essenciais que acrescentam valor à pesquisa.

Inferese que essas características afetam, de forma direta, a recuperação informacional,

visto que estabelecem a quantidade e a expressividade dos termos, podendo ser encaradas enquanto valores éticos ideais para o bem fazer na prática indexadora. Mudando de perspectiva, observa-se que, embora sejam inversamente proporcionais quando encaradas como medidas, podem ser combinadas quando consideradas como valores.

Ao final desta trajetória investigativa, em que a exaustividade e a especificidade foram abordadas sob o prisma da ética, uma indagação mais específica se coloca, no sentido de melhor conhecer esse domínio do conhecimento e como ele se articula por meio de comunidades epistêmicas, aspecto que se almeja investigar em continuidade, possivelmente no âmbito de um doutorado.

REFERÊNCIAS

- ABAD GARCÍA, M. F. **Evaluación de la calidad de los sistemas de información**. Madrid: Editorial Síntesis, 2005.
- ADLER, M.; TENNIS, J. Toward a taxonomy of harm in knowledge organization systems. **Knowledge Organization**, Würzburg, v. 40, n. 4, p. 266 – 272, 2013.
- ANDERSON, J. D. Essential decisions in indexing systems design. FEINBERG, H. (ed). **Indexing specialized formats and subjects**. Metuchen: Scarecrow Press, 1983.
- BARDIN, L. **Análise de conteúdo**. Tradução de Luiz Antero Reto e Augusto Pinheiro. Edições 70: Lisboa, 2009.
- BARITÉ, M. Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA, K. (Org.). **Educação, universidade e pesquisa**. Marília, 2001. p.35-50.
- BEGHTOL, C. A proposed ethical warrant for global knowledge representation and organization systems. **Journal of Documentation**, London, v. 58, n. 5, p. 507-532, 2002.
- BEGHTOL, C. Ethical decision-making for knowledge representation and organization systems for global use. **Journal of the American Society for Information Science and Technology**, New York, v. 56, n. 9, p. 903-912, 2005.
- BELL, H. K. Bias in indexing and loaded language. **The Indexer**, v. 17, n. 3, p. 173 – 177, 1991.
- BLACKBURN, S. **Dicionário Oxford de filosofia**. Rio de Janeiro: Jorge Zahar, 1997.
- BOCCATO, V. **Avaliação do uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo sociocognitivo com protocolo verbal**. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista Júlio de Mesquita Filho, 2009.
- BOCCATO, V. R. C.; FUJITA, M. S. L. O uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo de avaliação sociocognitiva com protocolo verbal. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 15, n. 3, p. 23-51, 2010.
- BONURA, L. S. **The art of indexing**. New York: Wiley, 1994.
- BORGES, G. S. B. Indexação automática de documentos textuais: proposta de critérios essenciais.
In: **Encontro nacional de pesquisa em ciência da informação, XI.**, 2010, Rio de Janeiro, RJ: ENANCIB, 2010.
- BORKO, H.; BERNIER, G. L. **Indexing concepts and methods**. New York: Academic Press,

1978.

BORKO, H. Information science: what is it? **American Documentation**, v.19, n.1, p.3-5, 1968.

BUCKLAND, M. Information as thing. **Journal of the American Society for Information Science**, v. 42, n. 5, p. 351-360, 1991.

CARNEIRO, M. V. Diretrizes para uma política de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 15, n. 02, 1985.

CLEVELAND, D.; CLEVELAND, A. **Introduction to indexing and abstracting**. Estados Unidos: Libraries Unlimited, 2001.

COMPARATO, F. K. **Ética: direito, moral e religião no mundo moderno**. São Paulo: Companhia das Letras, 2006.

DAHLBERG, I. Ethics and knowledge organization: in memory of Dr. S. R. Ranganathan in his centenary year. **International Classification**, Frankfurt, v. 19, n. 1, p. 1-2, 1992.

DICIONÁRIO DE ÉTICA E FILOSOFIA MORAL. São Leopoldo: Universidade do Vale do Rio dos Sinos, 2007.

ELSEVIER. **Scopus**. Disponível em < <https://www.elsevier.com/solutions/scopus>>. Acesso em 05.02.2015.

FEINBERG, H. The thesaurus in indexing and searching: a review. In: FEINBERG, H. (ed). **Indexing specialized formats and subjects**. Metuchen: Scarecrow Press, 1983.

FETTERS, L. K. **Handbook of indexing techniques: a guide for beginning indexers**. Estados Unidos: FIM-Books, 2001.

FOSKETT, A. C. **A abordagem temática da informação**. Tradução de Antonio Agenor Briquet de Lemos. São Paulo: Editora Polígono, 1973.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 1, n.1,p.60-90,jul./dez.2003.

FUJITA, M. S. L. (Org). **A indexação de livros a partir da percepção de catalogadores e usuários de bibliotecas universitárias: um estudo de observação do contexto sociocognitivo com protocolos verbais**. São Paulo : Cultura Acadêmica, 2009.

FUJITA, M. S. L.; RUBI, M. P.; BOCCATO, V. R. C. O contexto sociocognitivo do catalogador em bibliotecas universitárias: perspectivas para uma política de tratamento da informação documentária. **DataGramaZero - Revista de Ciência da Informação**, v.10, n.2, 2009.

GARCÍA GUTIÉRREZ, A. Knowledge organization from a “culture of the border” towards

a transcultural ethics of mediation. In: LÓPEZ-HUERTAS, M. M. (Ed.). **Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries**. Wüzburg: ERGON, 2002. p. 516-522. (Advances in Knowledge Organization, v. 8).

GLOSSÁRIO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO. Rio de Janeiro: Sociedade da Informação Brasil, 2002.

GOMES, H. F. Comportamento ético: fundamentos e orientações normativas ao exercício profissional do bibliotecário. In: GOMES, H. F.; BOTTENTUIT, A. M.; OLIVEIRA, M. O. E. (org). **A ética na sociedade, na área da informação e da atuação profissional: o olhar da Filosofia, da Sociologia, da Ciência da Informação e da Formação e do Exercício Profissional do Bibliotecário no Brasil**. Brasília: Conselho Federal de Biblioteconomia, 2009, 187p.

GUIMARÃES, J.A.C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). **Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 1, p. 77-99, 2008a.

GUIMARÃES, J.A.C. A análise documentária no âmbito do tratamento temático da informação: elementos históricos e conceituais. In: RODRIGUES, Georgete Medleg; LOPES, Ilza Leite. (Org.). **Organização e representação do conhecimento na perspectiva da Ciência da Informação**. Brasília: Thesaurus, 2003, v. 2, p. 100-117.

GUIMARÃES, J. A. C. **Aspectos éticos do tratamento temático da informação (TTI): elementos para sua caracterização a partir da interface das dimensões profissional, pedagógica e investigativa na área de Biblioteconomia no Mercosul**. Marília: UNESP, 2007. Relatório de Produtividade do Projeto Integrado de Pesquisa referente ao período 2003-2006 apresentado ao CNPq.

GUIMARÃES, J. A. C. Aspectos éticos em organização e representação do conhecimento (ORC): uma reflexão preliminar. In: GONZÁLEZ DE GÓMEZ, M. N.; ORRICO, E. G. D. (Org.). **Políticas de memória e informação: reflexos na organização do conhecimento**. Natal: EDUFRN, 2006. p. 237-264.

GUIMARÃES, J. A. C. As políticas de indexação como elementos para a gestão do conhecimento nas organizações. In: VIDOTTI, Silvana Aparecida Borsetti Gregorio. (Org.). **Tecnologia e conteúdos informacionais**. São Paulo: Polis, p. 43-52, 2004.

GUIMARÃES, J.A.C. Abordagens teóricas em tratamento temático da informação: catalogação de assunto, indexação e análise documental. In: GARCÍA MARCO, Francisco Javier. (Org.). **Avances y perspectivas en sistemas de información y de documentación**. Zaragoza: Prensas Universitarias de Zaragoza, p. 105-117, 2009a.

GUIMARÃES, J. A. C. et al. Aspectos éticos da organização da informação: abordagens teóricas acerca da questão dos valores. In: GOMES, H. F.; BOTTENTUIT, A. M.; OLIVEIRA, M. O. E.

(org). **A ética na sociedade, na área da informação e da atuação profissional: o olhar da Filosofia, da Sociologia, da Ciência da Informação e da Formação e do Exercício Profissional do Bibliotecário no Brasil**. Brasília: Conselho Federal de Biblioteconomia, 2009, 187p.

GUIMARÃES, J. A. C. et al. Ética nas atividades informativas: aspectos teóricos. **PontodeAcesso**, Salvador, v. 2, n. 1, p. 137 – 152, 2009b.

GUIMARÃES, J. A. C. et al. Ethics in the knowledge organization environment: an overview of values and problems in the LIS literature. In: ARSENAULT, C.; TENNIS, J. T. (Ed.). **Cultural and Identity in Knowledge Organization**. Würzburg: ERGON Verlag, 2008b. p. 340-346. (Advances in Knowledge Organization, 11).

GUIMARAES, J. A. C.; FERNÁNDEZ-MOLINA, J. C. Los aspectos éticos de la organización y representación del conocimiento en la revista Knowledge Organization. In: FRÍAS, J. A.; TRAVIESO, C. (Org.). **Tendencias de investigación en organización del conocimiento**. Salamanca: Universidad de Salamanca, 2003. p. 809-816.

GUIMARÃES, J.A.C. . O profissional da informação sob o prisma de sua formação. In: VALENTIN, M. L. P. (Org.). **Profissionais da informação: formação, perfil e atuação profissional**. São Paulo: Polis, 2000a, p. 53-70.

GUIMARÃES, J. A. C. ; MILANI, S. ; EVANGELISTA, I. V. Valores éticos na organização e representação do conhecimento. **Encontros Bibli**, Florianópolis, v. 20, n. 1, p. 19-32, 2015.

GUIMARÃES, J. A. C.; PINHO, F. A. Desafios da representação do conhecimento: uma abordagem ética. **Informação e Informação**, v. 12, n. 1, 2007.

GUIMARÃES, J. A. C. Políticas de análisis y representación de contenido para la gestión del conocimiento en las organizaciones. **Scire**, Zaragoza, v. 6, n. 2, p. 49-58, 2000b.

GUIMARÃES, J. A. C.; SALES, R. Análise documental: concepções do universo acadêmico brasileiro em Ciência da Informação. **Datagrama**, v. 11, n. 1, 2010.

HJØRLAND, B. **Information seeking and subject representation: an activity-theoretical approach to Information Science**. New York: Greenwood Press, 1997

HJØRLAND, B. Nine principles of knowledge organization. In: ALBRECHTSEN, H.; ORNAGER, S. (ed.) **Knowledge organization and quality management**. Frankfurt/Main: Indeks, 1994. p. 91-100. (Advances in Knowledge Organization, v. 4).

HJØRLAND, B. The concept of “subject” in information science. **Journal of Documentation**, Londres, v. 48, n. 2, p. 172 – 200, 1992.

HJØRLAND, B. What is knowledge organization (KO)?. **Knowledge Organization**, Würzburg, v. 35, n. 2/3, 2008, p. 86 – 101.

HUDON, M. Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts. **Knowledge Organization**, Würzburg, v. 24, n. 2, p. 84-91, 1997.

IFLA. About IFLA. Disponível em < <http://www.ifla.org/about>>. Acessado em 24.01.2016.

KEYSER, P. **Indexing: from thesauri to the semantic web**. Oxford: Chandos Publishing, 2012.

LANGRIDGE, D. Subject headings. In: KNIGHT, G. N. **Indexing, The Art of**. Boston : Allen & Unwin, 1979.

LANCASTER, F. W. **Indexing and abstracting in theory and practice**. Arlington: Information Resources Press, 1993.

LÓPEZ-HUERTAS, M. J. Cultural impact on Knowledge Representation and Organization in a Subject Domain. In: ARSENAULT, C.; TENNIS, J. T. (Ed.). **Cultural and Identity in Knowledge Organization**. Würzburg: ERGON Verlag, 2008a. p. 340-346. (Advances in Knowledge Organization, 11).

LÓPEZ-HUERTAS, M. J. Some current research questions in the field of knowledge organization. **Knowledge Organization**, Würzburg, v. 35, n. 2/3, p. 113-136, 2008b.

LEIVA, I. G. **Manual de indización: teoría y práctica**. Asturias: Ediciones Treas, 2008.

MALIANDI, R. **Ética: conceptos y problemas**. Buenos Aires: Biblos, 2004.

MARCHIONI, A. **Ética: a arte do bom**. Rio de Janeiro: Vozes, 2008.

MENDES, M. T. P.; SIMÕES, M. G. **Indexação por assuntos: princípios gerais e normas**. Lisboa: Gabinete de estudos, 2002.

MILANI, S. O. **Aspectos éticos em organização e representação do conhecimento (ORC): uma análise preliminar de valores e problemas a partir da literatura internacional da área**. 2007. 115 f. Trabalho Conclusão de Curso (Bacharel em Biblioteconomia) – Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2007.

MILANI, S. O. **Bias na representação de assunto: uma discussão de oposições binárias nos functional requirements for subject authority data (FRSAD)**. 2014. 134 f. Tese (Doutora em Ciência da Informação) – Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2014.

MILANI, S. O. **Estudos éticos em representação do conhecimento: uma análise da questão feminine em linguagens documentárias brasileiras**. 2010. 140 f. Dissertação (Mestre em Ciência da Informação) – Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2010

NALINI, J. R. **Ética geral e profissional**. Revista dos Tribunais: São Paulo, 2004.

NELSON, M. J.; TAGUE, J. M. Split size-rank models for the distribution of index terms.

Journal of the american society for information Science, v. 36, n. 5, p.283-296,1985.

NOVELLINO, M. S. F. Instrumentos e metodologias de representação da informação. **Informação & Informação**, v. 1, n. 2, 1996, p. 37 – 45.

OLIVEIRA, L. A. **Dissertação e tese em ciência e tecnologia**. Lisboa: Lidel, 2011.

OLSON, H.A. Exclusivity, teleology and hierarchy: our aristotelan legacy. **Knowledge Organization**, Wuzburg, v. 26, n. 2, p. 62-72, 1999.

OLSON, H.A. How we construct subjects: a feminist analysis. **Library Trends**, Baltimore, v.56, n.2, p.509-542, 2007.

OLSON, H. A. Patriarcal structures of subject access and subversive techniques for change. **The Canadian Journal of Information and Library Science**, Toronto, v. 26, n. 2/3, 2001.

OLSON, H. A. **The power to name: locating the limits or subject representation in libraries**. Dordrecht: Kluwer Academic Publisher, 2002.

OLSON, H. A. Wind and rain and dark of night: classification in scientific discourse communities. In: Arsenault, C.; Tennis, J. T. (Ed.). **Cultural and Identity in Knowledge Organization**. Würzburg: Ergon, 2008.

ORTEGA, C. D. Contexto de desenvolvimento da Organização da Informação, com enfoque para a Catalogação, na Escola de Ciência da Informação da UFMG. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.18, n.2, 2013, p. 182-215.

PINHO, F. A. **Aspectos éticos em representação do conhecimento: em busca do diálogo entre Antonio García Gutierrez, Michele Hudon e Clare Beghtol**. 2006. 132 f. Dissertação (Mestre em Ciência da Informação) – Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2006.

PROQUEST. **Library and information science abstracts (LISA)**: about. Disponível em <<http://proquest.libguides.com/lisa>>. Acessado em 05.01.2015.

QUIVY, R.; CAMPENHOUDT, R. **Manual de investigação em ciências sociais**. Lisboa: Gradiva, 2013. Tradução de João Minhoto Marques, Maria Amália Mendes e Maria Carvalho.

ROTHMAN, J. Is indexing obsolete? Keyword indexing and free-text searching. In: FEINBERG, H. (ed). **Indexing specialized formats and subjects**. Metuchen: Scarecrow Press, 1983.

RUBI, M. P. **A política da indexação na perspectiva do conhecimento organizacional**. 2004. 135 f. Dissertação (Mestrado em Ciência da Informação) – UNESP, Marília, 2004.

SALES, R. **Tesauros e ontologias sob a luz da teoria comunicativa da terminologia**. 164 f. Dissertação (Mestrado em Ciência da Informação) - PGCIN/ UFSC, Florianópolis, 2008.

SARACEVIC, T. Relevance reconsidered “96”. In: INGWERSEN, P.; PORS, N. O. (eds). **Proceedings of CoLIS 2, Second International Conference on Conceptions of Library and Information Science: integration in perspective**. Copenhagen: Royal School of Librarianship, 1996.

SCIELO. Disponível em <<http://www.scielo.br/?lng=pt>>. Acessado em 18.01.2016.

SEARIGHT, B. E. E. G. L. Name headings: practical exercise. KNIGHT, G. N. **Indexing, The Art of**. Boston : Allen & Unwin, 1979.

SIMÕES, M. G. **Classificações bibliográficas: percurso de uma teoria**. Coimbra: Almedina, 2011.

SOUSA, B. P. Representação Temática da Informação Documentária e sua Contextualização em Biblioteca. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 9, n. 2, 2013, p.132-146.

SOUZA, F. C. Dos deveres profissionais ou a deontologia. In: GOMES, H. F.; BOTTENTUIT, A. M.; OLIVEIRA, M. O. E. (org). **A ética na sociedade, na área da informação e da atuação profissional: o olhar da Filosofia, da Sociologia, da Ciência da Informação e da Formação e do Exercício Profissional do Bibliotecário no Brasil**. Brasília: Conselho Federal de Biblioteconomia, 2009, 187 p.

TENNIS, J. Ethos and ideology of knowledge organization: toward precepts for an engaged knowledge organization. **Knowledge Organization**, Wuzburg, v. 40, n. 1, p. 42 – 49, 2013.

WEB OF SCIENCE. Disponível em <<http://wokinfo.com/citationconnection/>>. Acesso em 05.01.2015.

WITTY, F. J. The beginnings of indexing and abstracting: some notes towards a history of indexing and abstracting in antiquity and the middle ages. **The Indexer**, v. 8, n. 4, p. 193 – 198, 1973.

APÊNDICE 01 - FICHAS DE ANÁLISE DE CONTEÚDO DOS TERMOS

FICHA 01

Referência: VAKKARI, P. et al. Query exhaustivity, relevance feedback and search success in automatic and interactive query expansion. *Journal of documentation*, v. 60, n. 02, p. 109 – 127, 2004.

Resumo: This study explored how the expression of search facets and relevance feedback (RF) by users was related to search success in interactive and automatic query expansion in the course of the search process. Search success was measured both in the number of relevant documents retrieved, whether identified by users or not. Research design consisted of 26 users searching for four TREC topics in Okapi IR system, half of the searchers using interactive and half automatic query expansion based on RF. The search logs were recorded, and the users filled in questionnaires for each topic concerning various features of searching. The results showed that the exhaustivity of the query was the most significant predictor of search success. Interactive expansion led to better search success than automatic expansion if all retrieved relevant items were counted, but there was no difference between the methods if only those items recognised relevant by users were observed. The analysis showed that the difference was facilitated by the liberal relevance criterion used in TREC not favouring highly relevant documents in evaluation.

Palavras-chave: Searching, Query languages.

CONCEITOS ENCONTRADOS

Exaustividade

“Pennanen and Vakkari (2003) have shown that the more exhaustively the users were able to articulate their information problems in query terms in a Boolean system, the more useful references they retrieved in a search.” (p. 110)

“The exhaustivity of a query denotes to the number of facets covered by query terms (Sormunen, 2002a).” (p. 114)

“a marginal document only points to the topic, a fairly relevant document contains more information but the presentation is not exhaustive, and a highly relevant document discusses the themes of the topic exhaustively.” (p. 115)

“Query exhaustivity correlates most strongly with the number of relevant documents ($r = \frac{1}{4} 0:37$), be they identified ($r = \frac{1}{4} 0:34$) or not ($r = \frac{1}{4} 0:28$) and relevance scores ($r = \frac{1}{4} 0:32$). Thus, users who generated most terms for the initial query were able to cover most extensively the facets by query terms, which was vital for the search success.” (p. 117)

Especificidade

Nada consta.

CONCEITOS

CONCLUSÕES

Quanto mais exaustivos os usuários forem no

	<p>momento da busca informacional, mais úteis serão as referências encontradas (p. 110).</p> <p>A exaustividade no momento da busca, indica as facetras que os termos de busca possuem (p. 115).</p>
--	--

FICHA 02

Referência: WOLFRAM, D. Inter-Record Linkage Structure in a Hypertext Bibliographic Retrieval System. **JASIST**, v. 46, n. 10, p. 764 - 774, 1996.

Resumo: The author explores inter-record linkage relationships of a bibliographic hypertext system through the use of descriptor term co-occurrences. Using term distribution and term exhaustivity data for an existing system, three models of term co-occurrence are developed and tested against the observed data. The observed distribution of term co-occurrences using term sizes follows a complex, but regular relationship. The developed models, although matching the observed behavior for much of the distribution, do not adequately model the observed co-occurrence patterns for select parts of the distribution using chi-square values. With knowledge of the structure of such a hypertext system, an appropriate model may be constructed and used as the basis for studying systems design of inter-record linkages and system navigation by users in such a hypertext system.

Palavras-chave:

CONCEITOS ENCONTRADOS

Exaustividade

“Nelson and Tague (1985) examined the fits of different mathematical distributions to empirical data sets from several databases, using data for the distribution of index terms over documents, the distribution of exhaustivity, and the distribution of term co-occurrences.” (p. 766)

“Distribution of descriptor term exhaustivity: This distribution describes the number of selectable descriptors within a record. These define the number of alternate search path options for further exploration of other search terms within each record. This is also represented in size-frequency form.” (p. 768)

“Data were collected by parsing through each record within the HyperLynx system and recording the descriptor term and a record identifier. They were stored in a database management system for easy searching and retrieval. From this database, term distributions,

Especificidade

Nada consta.

descriptor exhaustivity, and term co-occurrences were tabulated. The empirical distributions were then used to help generate hypothetical systems using simulation models.” (p. 768)

“Data from the HyperLynx system were collected to determine the probabilities of assignment for different exhaustivity values.” (p. 768)

“For the distribution of term exhaustivity, the number of search terms per page follows a Poisson-like pattern (Table 2), as was observed by Jacquesson and Schieber (1973), and Nelson and Tague (1985)” (p. 768)

“The enhanced simulation model required extraction of co-occurrence data for records containing the different exhaustivity values. Because of the distribution of descriptor exhaustivity, resulting in mid-range values for descriptors occurring more frequently (Table 2), the data were normalized to represent the relative inclusion of each term size for each exhaustivity value.” (p. 769)

“The data show that terms occurring only once are selected more frequently than other sizes of terms, and increasingly so as the descriptor exhaustivity values increase. Terms of greatest size are selected most frequently with lower exhaustivity values, but account for a progressively lower percentage of term assignments as the descriptor exhaustivity increases.”(p. 769)

“The basic simulation model adds to the model complexity by introducing the descriptor exhaustivity data.” (p. 770)

“Even the enhanced model, which attempts to take into account exhaustivity and term size frequency characteristics on an individual level does not provide a complete representation of the behavior.” (p. 772)

“A further investigation of the frequency of occurrence of singly occurring terms reveals that they are unequally distributed among the records, producing a skewed distribution of exhaustivity.” (p. 772)

<p>“This skewness produces more combinations of term co-occurrences for higher exhaustivity values than if terms were distributed more equally among the records.” (p. 773)</p> <p>“An appropriate simulation model could also be used to predict the number of inter-record linkages based on term distribution and term exhaustivity data.” (p. 773)</p> <p>“The Poisson relationship for term exhaustivity will also play a role. Records with only one term provide no linkages to other records. Records with large numbers of terms may present too many avenues for exploration.”(p. 774)</p>	
<p style="text-align: center;">CONCEITOS</p> <p>A exhaustividade refere-se ao número de descritores selecionáveis de um recurso, além de dar mais opções de pesquisa ao recurso descrito (p. 768).</p>	<p style="text-align: center;">CONCLUSÕES</p> <p>Percebe-se que a partir de uma descrição exaustiva, é possível estabelecer uma maior articulação entre recursos, uma vez que documentos com mais termos possibilitam mais vias de exploração (p. 769).</p>

FICHA 03

<p>Referência: WOLFRAM, D., ZHANG, J. An Investigation of the Influence of Indexing Exhaustivity and Term Distributions on a Document Space. JASIST, v. 53, n. 11, p. 943 - 952, 2002.</p>	
<p>Resumo: The authors investigate the influence of index term distributions, and indexing exhaustivity levels on the document space within a visual information retrieval environment called DARE. Using combinations of three levels of term distributions (shallow, observed, steep) and indexing exhaustivity (low, observed, high), hypothetical document sets were generated and projected onto the DARE environment. The results from the simulated document sets demonstrate the importance of term distribution and exhaustivity characteristics on the density of document spaces and their implications for retrieval, particularly when different term weighting schemes are used. The results also demonstrate how different combinations of exhaustivity and term distributions may result in similar document space density characteristics.</p>	
<p>Palavras-chave:</p>	
<p style="text-align: center;">CONCEITOS ENCONTRADOS</p>	
<p>Exaustividade</p>	<p>Especificidade</p>

<p>“These rules will also ultimately shape the observed regularities that result from indexing, namely how frequently specific terms are included in the index, as well as how many index terms are used to identify a given document (indexing exhaustivity)..” (p. 943)</p> <p>“Another important feature of IR systems important in modeling system content is the document indexing exhaustivity, or the number of index terms assigned to a document.” (p. 944)</p> <p>“With high exhaustivity to more completely describe a document and steeper term distributions, defining more unique sets of terms, the lowest document densities are encountered, making it easier to distinguish documents from one another.” (p. 951)</p>	<p>Nada consta.</p>
<p style="text-align: center;">CONCEITOS</p> <p>A exaustividade refere-se a quantos termos de indexação são utilizados para identificar um documento (p. 943).</p> <p>A exaustividade refere-se ao número de termos atribuídos a um documento (p. 944)</p>	<p style="text-align: center;">CONCLUSÕES</p> <p>Com um alto nível de exaustividade, uma menor densidade de documentos é encontrada, facilitando a distinção de documentos (p. 951).</p>

FICHA 04

<p>Referência: SPARCK-JONES, K. A statistical interpretation of term specificity and its application in retrieval. <i>Journal of Documentation</i>, v. 60, n. 05, p. 493 - 502, 2004.</p>	
<p>Resumo: The exhaustivity of document descriptions and the specificity of index terms are usually regarded as independent. It is suggested that specificity should be interpreted statistically, as a function of term use rather than of term meaning. The effects on retrieval of variations in term specificity are examined, experiments with three test collections showing, in particular, that frequently-occurring terms are required for good overall performance. It is argued that terms should be weighted according to collection frequency, so that matches on less frequent, more specific, terms are of greater value than matches on frequent terms. Results for the test collections show that considerable improvements in performance are obtained with this very simple procedure.</p>	
<p>Palavras-chave: Information research, Information retrieval, Information science and documentation</p>	
<p>CONCEITOS ENCONTRADOS</p>	
<p>Exaustividade</p>	<p>Especificidade</p>

<p>“In this case the exhaustivity of a document description is the coverage of its various topics given by the terms assigned to it; and the specificity of an individual term is the level of detail at which a given concept is represented.” (p. 493)</p> <p>“the exhaustivity of a document description is the number of terms it contains, and the specificity of a term is the number of documents to which it pertains.” (p. 495)</p>	<p>“In this case the exhaustivity of a document description is the coverage of its various topics given by the terms assigned to it; and the specificity of an individual term is the level of detail at which a given concept is represented.” (p. 493)</p> <p>“(…) a term is more or less specific as its meaning is more or less detailed and precise.” (p. 494)</p> <p>“the exhaustivity of a document description is the number of terms it contains, and the specificity of a term is the number of documents to which it pertains.” (p. 495)</p>
CONCEITOS	CONCLUSÕES
<p>A exaustividade é cobertura dos vários tópicos de um documento, por meio dos termos atribuídos a ele (p.493)</p> <p>A exaustividade refere-se ao número de termos que a descrição de um documento contém (p. 495)</p> <p>A especificidade pode ser considerada como o nível de detalhe que determinado conceito é representado (p. 493)</p>	<p>Um termo é mais ou menos específico, uma vez que é mais ou menos detalhado e preciso (p. 495).</p>

FICHA 05

Referência: HUGHES, A. V.; RAFFERTY, P. Inter-indexer consistency in graphic materials indexing at the National Library of Wales. *Journal of Documentation*, v. 67, n. 01, p. 09 - 32, 2011.

Resumo: Purpose – This paper seeks to report a project to investigate the degree of inter-indexer consistency in the assignment of controlled vocabulary topical subject index terms to identical graphical images by different indexers at the National Library of Wales (NLW). Design/methodology/approach – An experimental quantitative methodology was devised to investigate inter-indexer consistency. Additionally, the project investigated the relationship, if any, between indexing exhaustivity and consistency, and the relationship, if any, between indexing consistency/exhaustivity and broad category of graphic format. Findings – Inter-indexer consistency in the assignment of topical subject index terms to graphic materials at the NLW was found to be generally low and highly variable. Inter-indexer consistency fell within the range 10.8 per cent to 48.0 per cent. Indexing exhaustivity varied substantially from indexer to indexer, with a mean assignment of 3.8 terms by each indexer to each image, falling within the range 2.5 to 4.7 terms. The broad category of graphic format, whether photographic or non-photographic, was found to have little influence on either inter-indexer consistency or indexing exhaustivity. Indexing exhaustivity and inter-indexer consistency exhibited a tendency toward a direct, positive relationship. The findings are necessarily limited as this is a small-scale study within a single institution.

Originality/value – Previous consistency studies have almost exclusively investigated the indexing of print materials, with very little research published for non-print media. With the literature also rich in discussion of the added complexities of subjectively representing the intellectual content of visual media, this study attempts to enrich existing knowledge on indexing consistency for graphic materials and to address a noticeable gap in information theory.

Palavras-chave: National libraries, Indexing, Wales

CONCEITOS ENCONTRADOS

Exaustividade

“For each image, indexing efforts were examined and individual indexer exhaustivity (the number of assigned descriptors) and the number of terms in agreement between all 28 indexer pairs were transcribed and recorded.” (p. 17)

CONCEITOS

A exaustividade é o número de termos atribuídos a um documento (p. 17)

Especificidade

Nada consta.

CONCLUSÕES

FICHA 06

Referência: SPARCK-JONES, K. Does indexing exhaustivity matter?. JASIST, v. 24, n. 5, p. 313 - 316, 1973.

Resumo: Indexing exhaustivity, which may be broadly defined as the numbers of terms assigned to a document, is thought to be of some importance in retrieval, and it has been suggested that there may be an optimal level of exhaustivity for a particular collection. Experiments with two distinct collections, using three levels of indexing exhaustivity for both documents and requests, show that substantially the same performance is obtained for very different levels of document indexing, if suitable choices are made of request level.

Palavras-chave:

CONCEITOS ENCONTRADOS

Exaustividade

Especificidade

“A good deal has been said about indexing depth or exhaustivity. There have been tests to compare the effects of different levels of exhaustivity on retrieval performance, and it has been asserted that a given level of exhaustivity may be optimal for a particular document collection. The term 'exhaustivity' may refer to the number of topics or themes indexed for a document, or the level of detail with which a given topic is treated.” (p. 313)

“(…) greater exhaustivity will be defined as the assignment of a greater number of index terms to a document.” (p. 313)

“When the index terms are simple extracted keywords systems, from an unstructured indexing vocabulary, differences in exhaustivity are clearly observed, and it is convenient to study their effects in this context.” (p. 313)

“In some experiments, for example Salton's, different indexing levels are associated with different sources, so less exhaustive indexing is derived from titles, say, and more exhaustive indexing from abstracts.” (p. 313)

“In retrieval experiments, the consequences of indexing at different levels of exhaustivity are usually examined in relation to a fixed set of requests. Request exhaustivity is treated as a separate matter, alterations being primarily intended to raise precision or recall in searching. The object of this paper is to show that request and document exhaustivity are not so independent, and that the same retrieval performance may be achieved if either is varied in relation to the other; there may be therefore be no optimum level of document exhaustivity” (p. 313)

“Some options are open all the time so that what may turn out, as a collections grows, to have been a wrong decision about document exhaustivity, can be counteracted without re-indexing, by suitable treatment of requests.”

“Unfortunately, it is impossible to investigate exhaustively as a retrieval variable in isolation. There is a connection between indexing exhaustivity and term specificity, in that changes

“Unfortunately, it is impossible to investigate exhaustively as a retrieval variable in isolation. There is a connection between indexing exhaustivity and term specificity, in that changes in the numbers of terms per document are correlated with changes in the numbers of document per term.” (p. 313)

<p>in the numbers of terms per document are correlated with changes in the numbers of document per term.” (p. 313)</p>	
CONCEITOS	CONCLUSÕES
<p>A exaustividade está relacionada ao número de tópicos ou temas indexados de um documento (p. 313)</p>	<p>Quando os termos são extraídos apenas como palavras-chave, sem a utilização de um vocabulário controlado, são observadas significativas diferenças no nível de exaustividade (p. 313)</p> <p>Diferentes níveis de exaustividade são observados a partir de diferentes fontes de informação, sendo os resumos uma fonte de um alto nível de exaustividade (p. 313)</p> <p>O nível de exaustividade e as requisições de um documento não são independentes, e uma boa recuperação informacional pode ser alcançada se um fator varia do outro (p. 313)</p> <p>Não é possível investigar a exaustividade de forma isolada, ela possui conexão com a característica especificidade. (p. 313)</p>

FICHA 07

Referência: CHUNG, E. K.; YOON, J. W. Categorical and specificity differences between user-supplied tags and search query terms for images: an analysis of Flickr tags and Web image search queries. *Information Research*, v. 14, n. 03, p. 01 – 22, 2009.

Resumo: Introduction. The purpose of this study is to compare characteristics and features of user supplied tags and search query terms for images on the Flickr Website in terms of categories of pictorial meanings and level of term specificity. Method. This study focuses on comparisons between tags and search queries using Shatford's categorization schemes and the level of specificity based on the basic level theory. Analysis. Frequency distributions and chi-squared analyses were

performed on the data. The results of statistical analyses demonstrated that there were significant differences in categories among tags and different stages of search query terms. The overall distributions of the levels of term specificity also had similar pattern in both tags and search query terms, but statistically significant differences were found between tags and search query terms. Results. The results of this study demonstrated that Flickr tags have their own unique features compared to users' queries for image searching. The findings suggest that image tags should not be generally applied to other image collections although they have been considered as useful data in developing a user-centered indexing system. Conclusions. When utilizing user-supplied tags for user-centered indexing systems, it is desirable to consider the functions and users' tasks applied in tags rather than depending on statistical features merely obtained from the tag analysis.

Palavras-chave:

CONCEITOS ENCONTRADOS

Exaustividade	Especificidade
Nada consta.	<p>“More specifically, this study focuses on comparisons between tags and search queries using Shatford’s (1986) categorization schemes and the level of specificity based on the basic level theory. These two methods have been adopted as tools in related studies (Armitage and Enser 1997; Choi and Rasmussen 2003; Jorgensen 2003; Rorrisa 2008) because they provide a clear understanding of an image's semantic content.” (p. 02)</p> <p>“In addition, an object can be articulated differently depending on term specificity (e.g., the same object can be described as 'animal', 'dog', or 'Chihuahua').” (p. 02)</p> <p>“Following two methods, Shatford's categorization and specificity level, the characteristics and features of image tags are elucidated through comparison to search queries.” (p. 02)</p> <p>“In this context, the goals of this study are as follows: To determine to what extent user supplied tags and search query terms for images are different in the level of specificity.” (p. 02)</p> <p>“However, an analysis demonstrated no differences between tags generated in 2004 and in 2007 in terms of categorization distribution and specificity levels.”(p. 06)</p> <p>“For comparing the level of term specificity, the basic level theory was adopted. The basic theory explains that concepts can be categorized into one of three levels, the superordinate level, the basic level or the subordinate level.” (p. 07)</p>

	<p>“Since it has been found that a set of basic level terms are dominantly used and commonly shared by general users, researchers in library and information science assumed that basic level terms should be the level of specificity for concepts, and should receive focus during the indexing process (Bates 1998; Green 2006). By following this assumption, this study examined the level of specificity of user-supplied tags and search terms by applying the basic level theory.” (p.07)</p>
CONCEITOS	CONCLUSÕES
	<p>Um documento pode ser articulado de forma diferente dependendo do seu nível de especificidade - um mesmo documento pode ser descrito tematicamente como “animais”, “cachorros” ou “Chiuaua” (p. 02).</p>

FICHA 08

<p>Referência: ROBERTSON, S. E. Documentation note. Journal of Documentation, v. 30, n. 1, p.41 – 46, 1974.</p> <p>Resumo: THE CONCEPT OF the 'specificity' of the terms used in an information retrieval system has come up several times in the last few years. In fact, there are two different concepts which have been described by this term. It is perhaps helpful to distinguish between them.</p> <p>Palavras-chave: National libraries, Indexing, Wales</p>	
CONCEITOS ENCONTRADOS	
Exaustividade	Especificidade
Nada consta.	<p>“This is concerned with how specific a term is within the context of na indexed collection of documents, irrespective of any questions that might be put to the retrieval system. It can be measured by the frequency of occurrence of the term (a less frequent term is more specific). Specificity in this sense has been used by Svenonius and Sparck Jones; the latter has demonstrated how useful it can be in a retrospective retrieval system. 'Index language specificity', as used by Cleverdon et al. and Keen, can be regarded as the generalization of</p>

	<p>this concept to the whole index language—a language is specific if it has a high proportion of specific terms.” (p. 41)</p> <p>“Both Sparck Jones and Barkla use specificity (one in each sense) in retrieval systems based on weighted retrieval.” (p. 42)</p> <p>“In the discussion below, the phrase 'the term specificity (TS) model' will be used to describe the pure information-theoretic model without Sparck Jones's approximations.” (p. 43)</p> <p>“Barkla⁵ defines the specificity of a term t_i in relation to a question as $S = \frac{r}{R} \frac{n}{N}$” (p. 44)</p> <p>“Hence the log of the specificity forms a good basis for an additive weighting system of the type described above, and the matching value assigned to a given document can be used directly to predict the probability of the document being relevant.” (p. 45)</p> <p>“Miller⁸ develops an identical model, without however describing the measure as 'specificity'. In this case the weighting is intended for a retrospective search system, so the numerator (r_i/R) cannot be estimated from previous runs of the profile. Miller suggests that the user who makes the request can estimate this proportion.” (p. 46)</p> <p>“A term-question specificity of one represents a term that occurs no more frequently in relevant documents than in the rest of the collection, and is therefore of no value (and indeed the weighting in the TQS model is zero).” (p. 47)</p> <p>“The questions posed in this paper, and other questions concerning methods of using specificity information in a weighting scheme, deserve further investigation” (p. 47)</p>
CONCEITOS	<p style="text-align: center;">CONCLUSÕES</p> <p>Um termo menos frequente tende a ser um termo menos específico (p. 41).</p> <p>Uma linguagem documentária é específica, caso contenha um grande número de termos específicos (p. 41).</p>

	<p>A especificidade pode ser descrita por meio de fórmulas matemáticas (p. 44)</p> <p>Questões relacionadas ao uso de métodos utilizando a especificidade merecem futuras pesquisas (p. 47)</p>
--	---

FICHA 09

<p>Referência: KIM, G. Relationship between index term specificity and relevance judgment. <i>Information Processing Management</i>, v. 42, p. 1218 - 1229, 2006.</p>	
<p>Resumo: Concurrent concepts of specificity are discussed and differentiated from each other to investigate the relationship between index term specificity and users' relevance judgments. The identified concepts are term-document specificity, hierarchical specificity, statement specificity, and posting specificity. Among them, term-document specificity, which is a relationship between an index term and the document indexed with the term, is regarded as a fruitful research area. In an experiment involving three searches with 175 retrieved documents from 356 matched index terms, the impact of specificity on relevance judgments.</p>	
<p>Palavras-chave:</p>	
<p>CONCEITOS ENCONTRADOS</p>	
<p>Exaustividade</p>	<p>Especificidade</p>
<p>Nada consta.</p>	<p>“Specificity has been one of the core concepts in librarianship and information science from its earliest days. The concern about specificity emerged as a practical issue for cataloguing, especially in allocating subject headings or descriptors to each documentary unit. In 1876, Cutter wrote, “Enter a work under its subject heading, not under the headings of a class, which includes that subject” (p. 37). Since then, specificity has been discussed usually as a criterion for identifying index terms or descriptors. A common focus of the previous research on specificity is on its discriminating ability, which is the ability to distinguish a class of documents from other classes of documents, or one subclass from another.” (p. 1218)</p> <p>“What is the impact of specificity on a user’s relevance judgment? Because relevance judgment is a basis for evaluation of effectiveness of an information retrieval system, the impact of specificity, if it exists, is also a factor for effective information retrieval. Although some previous research has focused on the impact of specificity on some notions based on relevance judgment, such as recall and</p>

precision (Soergel, 1994; Sparck Jones, 1972; Svenonius, 1971), no research has been identified which addresses the impact of specificity on relevance judgment” (p. 1218)

“However, several concepts of specificity have been used in research, but these have not been consistent as the list below indicates:

(1) ‘Indexing’ is more specific than ‘information retrieval’ to a book titled ‘indexing’, because the book is about indexing (term-document specificity).

(2) ‘Cat’ is more specific than ‘domestic animal’, because ‘cat’ is semantically narrower than ‘domestic animal’ (hierarchical specificity).

(3) ‘Cancer-Chemotherapy-Congresses’ is more specific than ‘Cancer’, because the longer statement describes more closely the desired document (statement specificity).

(4) ‘Thesaurus’ is more specific than ‘indexing’ in a bibliographic database named Library and Information Science Abstracts, because ‘thesaurus’ has fewer postings than ‘indexing’ (posting specificity).” (p. 1219)

“This diversity in the meaning of the term has affected research on specificity, and shows different characteristics between the concepts. For instance, Weinberg and Cunningham (1984, 1985) point out that hierarchical specificity is not always positively correlated with posting specificity when using two groups of terms: core

terms and peripheral terms in a subject area. Also, Khosh-Khui (1987) indicates that statement specificity is not related to posting specificity at all. The existence of various concepts of index term specificity allows us to refine and expand the research question, as follows (1) Do the above concepts of specificity affect users’ relevance judgments? (2) What are the relationships among the concepts? (3) Does a combination of the concepts affect relevance judgments more than any single concept?” (p. 1219)

“In spite of the diversity, researchers generally agree on the characteristics of specificity as a relationship in terms of subject, topic, meaning, and/or theme, and as a degree. This means that a

specific term indicates that the term is involved in a topical relationship with a document, another term, or a set of documents, and that one term may be more or less specific than another term. This helps to set a framework for the discussion of specificity.” (p. 1219)

“Term-document specificity focuses on the topical relationship between an index term and a document indexed with the term. The question regarding this specificity is how accurately the term represents the topic of the document which is indexed with the term. Due to the semantic characteristics of term-document specificity, this is quite complicated to estimate, and automatic methods to assess this, beyond human intelligence, have not been successful.” (p. 1219)

“Hierarchical specificity is based on the term’s own meaning and is defined in a generic-specific hierarchical structure of terms, such as a thesaurus, by a human being. It is a relationship between terms, and it is easily measured in the hierarchy. The assumption for this specificity is that a narrower term is more specific than a broader term in a term hierarchy. Human knowledge can be organized into hierarchies of terms. In the hierarchies, a child or leaf item (narrower term) can be defined as a more specific one than its parent or root item (broader term), and this relationship is fixed. In this sense, the opposite concept of specificity would be generalization, with specificity’s synonym being specialization.” (p. 1219)

“Statement specificity is based on the assumption that a string of terms, as in a subject heading or in a search statement, is more specific than a single term. This concept has usually been discussed in cataloguing, in which subject headings can be modified with various tools, such as modified terminology, modifiers, parentheses, and subdivisions (Steinweg, 1979). It can be easily measured by the number of terms in a statement.” (p. 1220)

“Posting specificity is based on the assumption that if the number of documents which are indexed with term A is smaller than the number of documents indexed with term B in a set of documents, then term A is more specific than

term B. Its meaning is similar to the uniqueness of a term in a set of documents so that it is a relationship between a term and a set of documents. It can easily be measured with the number of documents indexed with a term (document frequency).” (p. 1220)

“Term-document specificity is the innate definition of specificity. When Cutter (1876) emphasized filing and locating a document under its specific subject heading, his concept of specificity was term-document specificity, because he focused on appropriate representation of a document using a specific subject heading. Balnaves (1976) also discussed term-document specificity with other definitions. He concluded that this specificity is the most promising one among the definitions according to the general theory of indexing, and that specificity deals with the accuracy of document representation. When Anderson and Pe´rez-Carballo (2000) discussed specificity, they defined it as “a degree of closeness between the meanings of an index term and the topics that are discussed in messages, or between the meanings and the features that a message, a text, or a document exhibits or possesses”. (p. 1220)

“Among the four types of specificity, statement specificity is not applied to index terms, but to compound subject headings and search statements” (p. 1220)

“Here, the focus is not to identify almost all factors affecting user’s relevance judgment, not to predict the group membership of cases (terms) as well as possible, but to show what type of specificity affects more significantly on the user’s relevance judgment. In other words, the focus is rather on specificity than on relevance.”(p. 1225)

“According to the results of correlation analysis, the three types of specificity have their own characteristics and it is potentially useful if these are distinguished from each other. The results of ANOVA, LOGIT, and DA show that t-spec is most significant in relation to relevance, and h-spec is not significant at all. t-spec can identify non-relevant documents from partially-relevant and relevant ones, but it does not distinguish

between relevant and partially-relevant documents. It shows that partially-relevant and relevant documents can occur in one category in terms of term-document specificity.” (p. 1227)

“Though the results are significant, the effect size is not very large. In other words, there are features other than specificity which affect the relevance judgments. An obvious feature discovered is language.” (p. 1227)

“Even though specificity was not expected as the only feature linked to users’ relevance judgment, specificity was proven to be one of the significant features of an index term for retrieving relevant documents, especially term-document specificity (t-spec).” (p. 1228)

“In order to apply the results to automatic indexing, the operationalization of term-document specificity should be investigated. As pointed out previously, the only drawback of term-document specificity is its inability to be measured by automatic means” (p. 1228)

“The results of the experiment show that each of the three types of specificity, term-document specificity, posting specificity, and hierarchical specificity, has its own characteristics, and each is independent from the others. Term-document specificity is shown to affect relevance judgments. Moreover, hierarchical specificity is not shown as a significant variable related to relevance judgments. Interestingly, despite its curvilinear characteristics, posting specificity affects relevance judgments significantly, albeit reflecting a non-normal distribution of relevance judgments.” (p. 1228)

“In future studies, it is suggested that work be undertaken to develop a way to measure term-document specificity automatically, in order to improve effective information retrieval. Because specificity has been differentiated from the other related concepts, and because it is still measurable by human beings, studies could explore the indexing phenomena with attention on the evidence available to measure specificity.” (p. 1229)

CONCEITOS	CONCLUSÕES
<p>A especificidade é um dos conceitos núcleo da Ciência da Informação, emergindo a partir de uma questão prática na catalogação: como alocar cabeçalhos de assunto ou descritores para cada um dos registros informacionais (p. 1218)</p>	<p>Um dos principais focos da especificidade é sua habilidade discriminatória: distinguindo classes e subclasses de documentos (p. 1218).</p>
<p>A diversidade de significados do termo especificidade acaba afetando as pesquisas sobre o tema (p. 1219)</p>	<p>Um termo específico indica que este termo está relacionado com um documento, outro termo ou um grupo de documentos e que este termo pode ser mais ou menos específico que outro termo (p. 1219).</p>
	<p>A especificidade dos termos afetam a relevância dos dados recuperados, já que podem auxiliar na identificação de documentos relevantes e não relevantes para a busca do usuário (p. 1221)</p>
	<p>A especificidade de um termo ainda encontra dificuldades para ser medida por meios automáticos (p. 1219)</p>
	<p>Temas relacionados à especificidade ainda merecem ser melhor explorados (p. 1228)</p>

FICHA 10

<p>Referência: BURGIN, R. The effect of indexing exhaustivity on retrieval performance. Information processing and management, v. 27, n. 6, p. 623 – 628, 1991.</p>	
<p>Resumo: effect of variations in indexing exhaustivity on retrieval performance in a vector space retrieval system was investigated by using a term weight threshold to construct different document representations for a test collection. Retrieval results showed that retrieval performance, as measured by the mean optimal E measure for all queries at a term weight threshold, was highest at the most exhaustive representation, and decreased slightly as terms were eliminated and the indexing representation became less exhaustive. These findings, coupled with those of Shaw for a retrieval system based on single-link clustering, suggest that the vector space model is more robust against variations in indexing exhaustivity than is the single-link clustering model.</p>	
<p>Palavras-chave: National libraries, Indexing, Wales</p>	
<p>CONCEITOS ENCONTRADOS</p>	
<p>Exaustividade</p>	<p>Especificidade</p>

“Shaw (1990a, 1990c, 1986a) has observed that the performance of a retrieval system based on single-link clustering varies as a function of indexing exhaustivity. Document representations produced by varying the levels of indexing exhaustivity were tested by Shaw, and retrieval performance was found to improve as the exhaustivity of representations decreased from the most exhaustive representation to an optimal representation based on an intermediate level of indexing exhaustivity. In some cases, retrieval performance was also found to decrease as the exhaustivity of representations decreased from that optimal representation of less exhaustive representations. In all cases, the most successful retrieval results were found at a point where an optimal compromise between indexing exhaustivity and specificity was produced by the document representation.” (p. 623)

“Sparck Jones (1973) varied the indexing exhaustivity of both queries and documents to show that the two are not independent, and that the same level of retrieval performance may be obtained by different combinations of the two.” (p. 623)

“Harding and Willett (1980) examined the effects of indexing exhaustivity on clustering efficiency by showing that the number of interdocument comparisons varies with the mean number of index terms per document in such a way that the number of comparisons for an inverted file approach will be greater than with conventional approaches for exhaustive indexing. Finally, El-Hamdouchi and Willett (1987) suggested that variations in indexing exhaustivity may affect clustering tendency and cluster search performance, and recommend this question as an area for future research.” (p. 624)

“As in Shaw (1990a, 1990b, 1990c, 1986a), these thresholds were used to control the exhaustivity of document representations by eliminating terms whose weights fell below the threshold” (p. 624)

“The effect of indexing exhaustivity on retrieval performance, reported by Shaw (1990a, 1990c, 1986a) for a retrieval system

“In all cases, the most successful retrieval results were found at a point where an optimal compromise between indexing exhaustivity and specificity was produced by the document representation.” (p. 623)

based on single-link clustering, was not observed for a vector space retrieval system, and would appear to be an artifact of single-link cluster-based retrieval. This finding suggests that the vector space model is more robust against variations in indexing exhaustivity than is the single-link clustering model.” (p. 625)

“Although this feature of cluster-based strategies is sometimes considered an advantage over conventional, linear associative retrieval models (Griffiths, Luckhurst, & Willett, 1986; Jardine & Van Rijsbergen, 1971), the clustering structures generated from these interdocument similarities would appear to be sensitive to variations in the indexing exhaustivity, and to contribute thereby to suboptimal retrieval effectiveness in many situations.” (p. 627)

“Shaw’s (1990a, 1990b) finding that the evidence of clustering structure was strongest for the most exhaustive document representations and weakest for intermediate values of term exhaustivity, coupled with his finding that retrieval effectiveness was worst for the most exhaustive document representations and optimal for an intermediate value of term exhaustivity, provides confirmation of the potentially adverse effects of these clustering structures on retrieval effectiveness.” (p. 627)

“This investigation further indicates that, unless individuals who make use of single-link clustering algorithms take care to determine the level of indexing exhaustivity that produces the best performance in a specific experimental environment, they may not obtain optimal performance.” (p. 627)

“Since the vector space model is more robust against variations in indexing exhaustivity, that method would appear to be preferable, unless one is willing to take the effects of variations in indexing exhaustivity into consideration.” (p. 627)

“Other clustering algorithms have been shown to be superior to single-link clustering in terms of retrieval performance (Willett, 1988), and these algorithms may also prove to be more robust against variations in indexing exhaustivity than

<p>single-link clustering.” (p. 627)</p> <p>“Finally, although vector-based retrieval appears to be more robust against variations in indexing exhaustivity than is single-link cluster-based retrieval, the retrieval effectiveness results reported in Table 2 are nevertheless comparable to the retrieval effectiveness results obtained by Shaw and characterized as “low performance levels.” (Shaw 1990a, p. 347)” (p. 627)</p> <p style="text-align: center;">CONCEITOS</p>	<p style="text-align: center;">CONCLUSÕES</p>
	<p>Um compromisso efetivo entre exaustividade e especificidade, aumenta a recuperação informacional (p. 623)</p>

FICHA 11

<p>Referência: GÓMEZ, J. C. G.; OLIVARES, J. L. G. Aproximación a la evaluación cuantitativa de los sistemas de recuperación de información de la prensa en Internet: Exhaustividad y precisión. Scire, v. 7, n. 01, p. 143 - 152, 2001.</p>	
<p>Resumo: Las ediciones de los medios de comunicación escrita en Internet son cada vez más diferentes de las tradicionales, gracias a la incorporación de nuevos servicios que aprovechan las posibilidades que la red les ofrece: inmediatez, alcance mundial, hipertextualidad, multimedia, integración de bases de datos, flexibilidad de formatos, etc. Entre estos servicios destacan la personalización de la página principal de acceso, el envío personalizable de noticias por correo electrónico y los sistemas de recuperación de noticias, cuyas características se evalúan tanto cuantitativamente mediante el modelo de evaluación de recuperación de la información, como cualitativamente en relación a aspectos como la facilidad de consulta, la ayuda disponible, etc.</p>	
<p>Palavras-chave: Prensa. Ediciones electrónicas. Recuperación de información. Evaluación. Internet.</p>	
<p>CONCEITOS ENCONTRADOS</p>	
<p>Exaustividade</p>	<p>Especificidade</p>
<p>“Se trata, por tanto, de hacer una primera aproximación a las posibilidades reales de recuperación de información que ofrece cada uno de dichos sistemas, así como tratar de valorar de la forma más objetiva posible su calidad basándose en la pertinencia de los resultados obtenidos tras la búsqueda, a partir de lo que se obtienen una serie de datos cuantitativos sobre la precisión y exhaustividad.” (p. 145)</p> <p>“Se ha partido de la adaptación y simplificación de algunas metodologías existentes para la evaluación de la recuperación de información — fundamentalmente en lo que a términos de exhaustividad y precisión se refiere—,</p>	<p>Nada consta.</p>

concretamente respecto a los planteamientos metodológicos recogidos en los experimentos de Cranfield que propusiera Cleverdon (1972), posteriormente modificados por Salton y McGill (1983) (Hipola, Vargas-Quesada y Montes, 1999).” (p. 145)

“Se construyó una tabla (tabla 1), donde se incorporó este dato junto con el cálculo de la precisión y la exhaustividad. La exhaustividad para cada document de la lista de documentos recuperados viene dada por el resultado de la division del número de documentos relevantes recuperados hasta ese momento entre el total de documentos relevantes recuperados para esa consulta ($E = \text{drr} / \text{Tdrr}$).” (p. 147)

“Con estos datos se fue construyendo una matriz de resultados y se calculó la media de la matriz de resultados de exhaustividad y la correspondiente a la de precisión con el objeto de obtener una columna de datos que representara la exhaustividad y otra que representara la precisión, y así poder plasmar el resultado en una representación gráfica, a la que a continuación se le calculó la correspondiente curva de regresión logarítmica, con la que se pudo representar el funcionamiento general del ingenio de búsqueda (1) en cuestión.” (p. 147)

“Tal y como hemos calculado los valores de exhaustividad y precisión, ese sistema tendría asociada una representación gráfica consistente en una línea totalmente horizontal a la altura del máximo valor de precision” (p. 147)

“Así, un sistema que ordene los resultados por orden de relevancia tenderá a ofrecer una curva de relación precisión/exhaustividad decreciente, y un sistema que no ordene por relevancia tenderá a ofrecer como resultado una línea plana por debajo del valor máximo de precisión.” (p. 147)

“Para comparar varios sistemas de recuperación de información a partir de su representación gráfica tendremos en cuenta ante todo el hecho de que para iguales valores de exhaustividad se obtengan mayores valores de precisión, así como el hecho de obtener una curva decreciente frente a una horizontal.” (p. 147)

“Así se podían observar dos cualidades en la misma gráfica: por un lado la relación precisión/exhaustividad —que se habría de manifestar en la pendiente de la curva—, y, por otro, la profundidad del número de resultados que devuelve cada sistema de recuperación —lo que se observaría a partir de la longitud de la curva.” (p. 148)

“En cuanto a la relación entre exhaustividad y precisión obtuvimos como resultado cuatro matrices de datos, que representamos gráficamente mediante sus cuatro líneas de regresión correspondientes, con las que tratamos de mostrar visualmente el rendimiento de los ingenios de búsqueda.” (p. 148)

“El hecho de que el SRI sea capaz de ordenar los resultados por orden de relevancia será un factor clave para conseguir obtener mejor relación precisión/ exhaustividad que aquellos que no lo hacen, aún cuando los que ordenan los resultados cuenten en su base documental con un número sustancialmente menor de documentos.” (p. 150)

CONCEITOS

CONCLUSÕES

A exaustividade está diretamente ligada à revocação do sistema (p. 147)

Uma melhor relação entre exaustividade e especificidade pode ser adquirida a partir de um critério de relevância dos documentos recuperados (p. 150)

FICHA 12

Referência: FERNANDÉZ, L. M. M.; ALONSO, M. I.; MONTÁVEZ, A. J. V. La consistencia como indicador de la fiabilidad (corrección) de la indización: los casos de LISA, PASCAL y VOCED. Scire, v. 12, n. 01, p. 85 - 98, 2006.

Resumo: Se analizan las posibilidades de la consistencia como indicador de calidad de la indización mediante un doble procedimiento. En primer lugar, se calcula la coherencia habida en las bases de datos de LISA y PASCAL aplicando solo las fórmulas de Hooper y Rolling. Esto permite demostrar que la consistencia no necesariamente muestra que estemos ante una indización de calidad, si no tenemos en cuenta el context en el que se lleva a cabo esa descripción de los documentos. Después se procede a investigar en la base de datos VOCED la vinculación entre consistencia y calidad de la indización, estudiando la interrelación entre aquella y los restantes factores que modulan la indización: corrección, exhaustividad y especificidad. Para ello se compara la indización hecha en VOCED con una indización efectuada como “modelo” o “patrón oro”, lo que permite asimismo añadir a las dos fórmulas citadas la del índice de Kappa. Este procedimiento se ha concretado en un

modelo que contempla toda la casuística que incide en una correcta indización (semántica, pragmática y ambiental), a partir de un patrón de referencia. El método revela que la consistencia por sí sola no constituye un indicador de calidad, pero que sí es condición necesaria para alcanzarla, correlacionando la consistencia con las demás características de la indización. Asimismo determina las causas de variabilidad en las ratios de consistencia y detecta los errores u omisiones que se produzcan en la indización.

Palavras-chave: Consistencia en la indización. Bases de datos. LISA. PASCAL. VOCED. Calidad de la indización.

CONCEITOS ENCONTRADOS	
<p>Exaustividade</p> <p>“Posteriormente, en esta segunda fase, investigamos la relación entre consistencia y calidad tratando de analizar empíricamente la vinculación entre la coherencia y los restantes factores que caracterizan la indización: consistencia o coherencia, corrección o exactitud, exhaustividad y especificidad.” (p. 93)</p> <p>“Para que eso no suceda, por ejemplo, la corrección de la indización requiere la adecuada identificación de los conceptos, y, además, la asignación de terminus con el nivel preciso de exhaustividad y especificidad; la coherencia requiere que el indizador escoja unos descriptores que se ajusten a las normas de corrección, exhaustividad y especificidad, si no queremos contar con una indización coherentemente mala. Y, en fin, una elevada exhaustividad comporta, si bien de manera indirecta, cierto nivel de especificidad (Moreno, 2003a).” (p. 94)</p> <p>“Por lo tanto, si convenimos que la indización adecuada de un documento debe tener cierto grado de corrección y especificidad, podremos aceptar también que la consistencia, cuando aparece imbricada con la corrección, la exhaustividad y la especificidad, puede constituir un indicador de la “calidad” de la indización.” (p. 94)</p>	<p>Especificidade</p> <p>“Posteriormente, en esta segunda fase, investigamos la relación entre consistencia y calidad tratando de analizar empíricamente la vinculación entre la coherencia y los restantes factores que caracterizan la indización: consistencia o coherencia, corrección o exactitud, exhaustividad y especificidad.” (p. 93)</p> <p>“Para que eso no suceda, por ejemplo, la corrección de la indización requiere la adecuada identificación de los conceptos, y, además, la asignación de terminus con el nivel preciso de exhaustividad y especificidad; la coherencia requiere que el indizador escoja unos descriptores que se ajusten a las normas de corrección, exhaustividad y especificidad, si no queremos contar con una indización coherentemente mala. Y, en fin, una elevada exhaustividad comporta, si bien de manera indirecta, cierto nivel de especificidad (Moreno, 2003a).” (p. 94)</p> <p>“Por lo tanto, si convenimos que la indización adecuada de un documento debe tener cierto grado de corrección y especificidad, podremos aceptar también que la consistencia, cuando aparece imbricada con la corrección, la exhaustividad y la especificidad, puede constituir un indicador de la “calidad” de la indización.” (p. 94)</p>
CONCEITOS	CONCLUSÕES
	A exaustividade e a especificidade, aliada à consistência e à confiabilidade traz um indicador de qualidade à indexação (p. 94)

FICHA 13

Referência: GUIMARÃES, J. A. C. Políticas de análisis y representación de contenido para la gestión del conocimiento en las organizaciones. Scire, v. 6, n. 02, p. 49 - 58, 2000.

Resumo: Teniendo en cuenta que las unidades y sistemas de información se constituyen por definición en organizaciones de gestión de conocimiento, se observa que las actividades de análisis y representación de contenido asumen un rol estratégico en cuanto que puentes para la construcción de nuevo conocimiento. En este sentido, la indización deja de ser meramente una actividad de naturaleza eminentemente técnica para integrar el universo gerencial de tales organizaciones. Por tanto, se analiza el diseño de políticas de análisis y representación de contenido, partiendo de elementos contextuales (la organización, los clientes, los recursos) para llegar a elementos instrumentales, que llevarán a su vez a la toma de decisiones sobre aspectos tales como la cobertura temática, el universo de documentos fuente, los niveles de exhaustividad y especificidad deseados y el perfil del indizador, del sistema de indización y del lenguaje documental. Finalmente, se analiza la importancia de tales políticas, en tanto instrumentos para la gestión del conocimiento por parte de la organización en cuestiones tales como el conocimiento propio y estratégico allí generado, el que se puede obtener externamente, su carácter público o estratégico, así como el costo del proceso.

Palavras-chave: Análisis documental de contenido. Organización del conocimiento. Políticas de indización.

CONCEITOS ENCONTRADOS

Exhaustividade	Especificidade
<p>“Con respecto al estudio del tipo de organización, Carneiro (1985) resalta la importancia de determinar los objetivos y actividades de la misma, de modo que sirvan de base a decisiones relacionadas con las áreas temáticas prioritarias, con el sistema de indización a adoptar y también con el nivel de exhaustividad y especificidad” (p. 54)</p>	<p>“Con respecto al estudio del tipo de organización, Carneiro (1985) resalta la importancia de determinar los objetivos y actividades de la misma, de modo que sirvan de base a decisiones relacionadas con las áreas temáticas prioritarias, con el sistema de indización a adoptar y también con el nivel de exhaustividad y especificidad” (p. 54)</p>
<p>“La cuestión financiera influye —y mucho— en la implementación y en el desarrollo de una política de este tipo. En este sentido, Carneiro (1985), propone el cotejo entre tales recursos y las decisiones relativas a la creación, adaptación o adopción de un lenguaje de indización, y también entre exhaustividad, especificidad, exigencias de los usuarios, inversión en equipos y programas de indización, inversión en la formación de indizadores, etc.” (p. 55)</p>	<p>“La cuestión financiera influye —y mucho— en la implementación y en el desarrollo de una política de este tipo. En este sentido, Carneiro (1985), propone el cotejo entre tales recursos y las decisiones relativas a la creación, adaptación o adopción de un lenguaje de indización, y también entre exhaustividad, especificidad, exigencias de los usuarios, inversión en equipos y programas de indización, inversión en la formación de indizadores, etc.” (p. 55)</p>
<p>“Nivel de exhaustividad y de especificidad de indización, en directa conexión con el perfil del usuario. De ese modo, se pueden diferenciar los sistemas mas orientados a la exhaustividad — como ocurre en el caso de las búsquedas de naturaleza histórica o legislativa— y otros en los que la precisión —y, por consiguiente, la especificidad— asume el caracter determinante, como ocurre en los centros de documentación empresarial, donde se solicitan productos documentales más orientados a la toma de decisiones gerenciales rápidas que suponen la asunción de riesgos financieros.” (p. 55)</p>	<p>“Nivel de exhaustividad y de especificidad de indización, en directa conexión con el perfil del usuario. De ese modo, se pueden diferenciar los sistemas mas orientados a la exhaustividad — como ocurre en el caso de las búsquedas de naturaleza histórica o legislativa— y otros en los que la precisión —y, por consiguiente, la especificidad— asume el caracter determinante, como ocurre en los centros de documentación empresarial, donde se solicitan productos documentales más orientados a la toma de decisiones gerenciales rápidas que suponen la asunción de riesgos financieros.” (p. 55)</p>

“Elección del lenguaje de indización, sea libre —principalmente áreas muy especializadas y con terminologías consolidadas o en casos en los que se dispone para la indización de recursos humanos escasos—, sea controlado —en sistemas que exigen alta exhaustividad y precisión temática.” (p. 55)

“Es interesante señalar que los elementos sistematizados por Carneiro (1985) coinciden en muchos aspectos con los que Foskett (1982) enunció como las características claves de un SRI: llamada y precisión, exhaustividad y pertinencia, tiempo, modo de búsqueda y coste.” (p. 55)

“En cuanto al proceso: existencia o no de reglas o instrucciones, productividad requerida a los indizadores y exhaustividad de la indización.” (p. 56)

CONCEITOS

“Selección y adquisición de documentos fuente, abarcando aspectos relativos a la cobertura y a la calidad de los documentos, a los que se puede modernamente añadir cuestiones como la procedencia —sobre todo en el análisis de paginas Web—, costos, lengua, orientación técnica (especificidades temáticas de un SRI), etc.” (p. 55)

“Tiempo de respuesta del sistema, en relación con el grado de recuperación y de precisión exigidos y también del perfil de la organización. Ciertamente, el número de organizaciones en las cuales el refrán “información retrasada es información negada” se sobrepone a cuestiones como las especificidad o la precisión no deja de crecer.” (p. 55)

“En cuanto al vocabulario de indización: especificidad, sintaxis, grado de ambigüedad y de imprecisión, calidad del vocabulario de entrada y de la estructura del lenguaje de indización y también la disponibilidad de instrumentos auxiliares (por ejemplo, diccionarios terminológicos).” (p. 56)

CONCLUSÕES

Os níveis de exaustividade e especificidade devem estar de acordo com o perfil dos usuários, sempre considerando suas particularidades (p. 55)

Exaustividade e especificidade são características chave em um sistema de recuperação informacional, além de auxiliar a gestão do conhecimento nas organizações (p. 55-56).

FICHA 14

Referência: SHAW, W. An investigation of document partitions. Information processing and management, v. 22, n. 01, p. 19 - 28, 1986.

Resumo: In this paper, the empirical significance of document partitions is investigated as a function of term-weight and similarity thresholds. The term-weight threshold selects a particular level of indexing exhaustivity and specificity for the document representation and the similarity threshold selects a specific level of the associated single-link hierarchy. The results show that the same empirically “preferred” partitions can be detected by two independent strategies: an analysis of cluster-based retrieval effectiveness and an analysis of regularities in the underlying structure of the document graph. These results represent the first step in an investigation designed to determine

if the statistical significance of document partitions can explain the empirical significance of the same partitions.

Palavras-chave:

CONCEITOS ENCONTRADOS

Exaustividade	Especificidade
<p>“The term-weight threshold can be used to control the exhaustivity and specificity of document representations] 13, 141.” (p. 20)</p> <p>“The structure of a document graph and the resulting partition of the document collection are dictated by the magnitudes of tw, and ts. The value of tw selects a particular level of indexing exhaustivity and specificity for the document representation, and the value of ts, selects a particular level of the associated single-link hierarchy.” (p. 22)</p> <p>“The term-weight threshold ($tw = 30$) produces an optimal compromise between indexing exhaustivity and specificity for the document collection and the associated document representation.” (p. 25)</p> <p>“Thus, a statistical regularity in the structures of document graphs may indicate a desirable compromise between indexing exhaustivity and specificity. This regularity can be understood by reconsidering the properties of the similarity coefficient.” (p. 26)</p> <p>“The result suggests that for a sufficiently high fixed value of ts, the magnitude of tw (which controls the exhaustivity and specificity of document representations) can appear in one of three regions.” (p. 26)</p> <p>“The term-weight threshold establishes the degree of specificity a weighted index term must possess to describe a particular document and selects a particular level of indexing exhaustivity and specificity for the document representation.” (p. 27)</p>	<p>“The term-weight threshold can be used to control the exhaustivity and specificity of document representations] 13, 141. For the present document collection, the specificity weights vary from N^{-1}; = 0 to $\xi = 200$.” (p. 20)</p> <p>“The structure of a document graph and the resulting partition of the document collection are dictated by the magnitudes of tw, and ts. The value of tw selects a particular level of indexing exhaustivity and specificity for the document representation, and the value of ts, selects a particular level of the associated single-link hierarchy.” (p. 22)</p> <p>“The term-weight threshold ($tw = 30$) produces an optimal compromise between indexing exhaustivity and specificity for the document collection and the associated document representation.” (p. 25)</p> <p>“Thus, a statistical regularity in the structures of document graphs may indicate a desirable compromise between indexing exhaustivity and specificity. This regularity can be understood by reconsidering the properties of the similarity coefficient.” (p. 26)</p> <p>“The result suggests that for a sufficiently high fixed value of ts, the magnitude of tw (which controls the exhaustivity and specificity of document representations) can appear in one of three regions.” (p. 26)</p> <p>“The term-weight threshold establishes the degree of specificity a weighted index term must possess to describe a particular document and selects a particular level of indexing exhaustivity and specificity for the document representation.” (p. 27)</p>
CONCEITOS	CONCLUSÕES
	Os níveis de exaustividade e especificidade podem ser medidos por meio de fórmulas matemáticas (p. 20)

FICHA 15

Referência: PIRKOLA, A.; KESKUSTALO, H.; JARVELIN, K. The Effects of Conjunction, Facet Structure, and Dictionary Combinations in Concept-Based Cross-Language Retrieval. *Information Retrieval*, v. 1, p. 217 - 250, 1999.

Resumo: The paper studies concept-based cross-language information retrieval (CLIR). The document collection was a subset of the TREC collection. The test requests were formed from TREC's health related topics. As translation dictionaries the study used a general dictionary and a domain-specific (=medical) dictionary. The effects of translation method, conjunction, and facet order on the effectiveness of concept-based cross-language queries were studied, and concept-based structuring of cross-language queries was compared to mechanical structuring based on the output of dictionaries. The performance of translated Finnish queries against English documents was compared to the performance of original English queries against the English documents, and the performance of different CLIR query types was compared with one another. No major difference was found between concept-based and mechanical structuring. The best translation method was a simultaneous look-up in the medical dictionary and the general dictionary, in which case cross-language queries performed as well as the original English queries. The results showed that especially at high exhaustivity (the number of mutually restrictive concepts in a request) levels cross-language queries perform well in relation to monolingual queries. This suggests that conjunction disambiguates cross-language queries. An extensive study was made of the relative importance of the concepts of requests. On the basis of the classification data of request concepts it was shown how the order of facets in a query affects cross-language as well as monolingual queries

Palavras-chave:**CONCEITOS ENCONTRADOS****Exaustividade**

“The term exhaustivity is defined as the number of conjunctive concepts identified in a request (the number of facets in a query). For instance, the exhaustivity of the request “vitamins” above and the query vitamin AND human AND ailment AND (cure OR cause) is four.” (p. 221)

“Both types included queries whose exhaustivity ranged from exhaustivity one (one facet) to the highest exhaustivity allowed by a request. In the first type, each query had the best order of facets in terms of precision. The best order was determined a posteriori. The queries of the second type had a completely reverse of facet order to that of the queries of the first type.” (p. 221)

“We hypothesize further that the fuzziness introduced by dictionaries causes the fewer retrieval failures the higher the query exhaustivity. In other words, the likelihood of having, for example, four incorrect translation equivalents (for four different conjunctive

Especificidade

Nada consta.

concepts) in the same document is less than that of having just two. We argue that the hypothesis is supported, if the performance of high exhaustivity CLIR queries in relation to high exhaustivity monolingual queries is better than the performance of low exhaustivity CLIR queries in relation to low exhaustivity monolingual queries.” (p. 221)

“Obviously, the best operator to test the disambiguation effect of conjunction (research problem four) would be the Boolean and-operator (band in INQUERY). However, the pilot tests which were done using the strict Boolean constraint showed that in many cases high exhaustivity queries gave zero results.” (p. 222)

“No differences were found between the results of the present study and that of Pirkola and Keskustalo (1999) which considered query exhaustivity as an additional variable. The best translation method was the sd+gd method at all exhaustivity levels. Exhaustivity did not affect the relative effectiveness of the and and sum operators nor did it have a remarkable effect on the relative performance of CA-CLIR and MS-CLIR queries. In the study by Pirkola and Keskustalo (1999), the same $P(\text{CLIR},n)/P(\text{Eng},n)$ pattern for conjunction effect as in the present study was found when the number of queries was maximized at each exhaustivity level. The figures at the exhaustivity levels 1, 2, 3, 4, and 5 were 1.01; 0.89; 0.90; 1.03; 1.38, respectively.” (p. 233)

“For the CA-CLIR and CA-Eng queries with the worst facet order, precision improves systematically as exhaustivity is increased from exhaustivity 1 to the highest exhaustivity (figures 8–11).” (p. 234)

“There are two counteracting factors that affect the performance of CA-CLIR and CA-Eng queries with the best facet order: conjunction (high exhaustivity) and the bad concepts, i.e., high exhaustivity concepts of requests. The former factor has positive effects and the latter negative effects. The bad concepts had a strong influence on the performance of CA-Eng queries. For CA-CLIR queries the positive

<p>effects of conjunction resisted effectively the negative effects of the bad concepts.” (p. 240)</p> <p>“When facets were in the worst order, CA-CLIR queries benefited from high (3–5) exhaustivity; in each case the highest precision was obtained at the highest exhaustivity level” (p. 240)</p> <p>“The results show that mechanical structuring is as good a structuring method as structuring based on a concept analysis. Query exhaustivity does not affect this. Therefore it seems that the concept analysis is not worth the effort in CLIR queries provided that mechanical structuring is carried out. However, query expansion extent may affect concept-based CLIR queries as well as mechanically structured queries. The effect of query expansion was not examined in this study.” (p. 240)</p>	
CONCEITOS	CONCLUSÕES
A exaustividade diz respeito ao número de conceitos ou facetas atribuídas no momento da busca informacional (p. 221)	

FICHA 16

Referência: GIVEN, L. M; OLSON, H. A. Knowledge organization in research: a conceptual model for organizing data. Library & Information Science Research , v. 25, n. 2, p. 157-176, 2003.	
Organizing research data for effective analysis has been insufficiently addressed in the methodological literature. This article proposes that concepts of knowledge organization relating to relevance, precision, recall, coextensiveness, exhaustivity, specificity, and consistency offer an already-made model that can be applied to research data. The knowledge organization (KO) model is reinterpreted for transferability to quantitative, qualitative, and textual research. In each instance, the model's applicability is illustrated with examples from the authors' research. This exploration demonstrates the model's resiliency in organizing numeric data, coding transcripts, and marking up textual statements. The limitations of the model are noted and compromises are described, providing a valuable approach to meaningful data preparation for researchers, educators, students, and reviewers of research across disciplines. The article concludes that the KO model contributes significantly to the ability of researchers to collect and organize data in a manner most likely to shed light on research problems they address.	
Palavras-chave:	
CONCEITOS ENCONTRADOS	
<p>Exaustividade</p> <p>“In information science, strategies for effective organization and retrieval of information have been honed by examining specificity, exhaustivity, relevance, and other elements</p>	<p>Especificidade</p> <p>“In information science, strategies for effective organization and retrieval of information have been honed by examining specificity, exhaustivity, relevance, and other elements</p>

central to the process of subject analysis.” (p. 158)

“Barzun and Graff (1985) alluded to the relative exhaustivity and specificity applied in selecting statements and categorizing them (pp. 30–31) but do not explore the ramifications of those actions and how to adjust for desired results.” (p. 159)

“Two other factors govern the vocabulary or encoding scheme and its application: specificity and exhaustivity” (p. 163)

“Exhaustivity is defined as the breadth of representation—the number of factors indexed or encoded. As with specificity, exhaustivity is a characteristic of both the vocabulary and its application.” (p. 164)

“In application, exhaustivity is related to the level of indexable matter; that is, how much of a given topic must be covered by an information-bearing document before it is represented in the scheme (Milstead, 1984).” (p. 164)

“Exhaustivity is closely related to recall, the measure of how much of available relevant information or data is retrieved from the total available relevant information.” (p. 165)

“If exhaustivity is high, more codes are used, which allows more data to be retrieved and analyzed.” (p. 165)

“The notions of relevance, precision, recall, specificity, and exhaustivity seem like they could produce perfect categories and codes—the data rainbow—but as indexers know, there are several potential problems in applying these organizational concepts.” (p. 165)

“High exhaustivity tends to lower precision, because the addition of more and more codes results in the retrieval of irrelevant data along side the relevant. Conversely, high specificity results in low recall. Since high specificity uses narrower categories, it produces fewer data in each category than does low specificity.” (p. 166)

“If indexers of a document use different levels of

central to the process of subject analysis.” (p. 158)

“Barzun and Graff (1985) alluded to the relative exhaustivity and specificity applied in selecting statements and categorizing them (pp. 30–31) but do not explore the ramifications of those actions and how to adjust for desired results.” (p. 159)

“Two other factors govern the vocabulary or encoding scheme and its application: specificity and exhaustivity” (p. 163)

“Specificity can be viewed in at least two ways: as a characteristic of the vocabulary and as a factor in the application of a vocabulary (the process of indexing). In the first sense, specificity is the relative detail within the vocabulary—the number of hierarchical levels defined.” (p. 163)

“As with coextensiveness, the level of specificity should serve the needs of users of that organizational scheme.” (p. 163)

“Specificity of application is standard indexing practice; each item must be coded at the most specific level available” (p. 163)

“It is important to remember that precision is enhanced by specificity: generally, the High specificity means that data are indexed (or categorized or coded) at a very precise level—the selected terms are finely grained, using detailed levels of categorization. Low specificity tends to result in high recall (Wellisch, 1991).” (p. 164)

“The notions of relevance, precision, recall, specificity, and exhaustivity seem like they could produce perfect categories and codes—the data rainbow—but as indexers know, there are several potential problems in applying these organizational concepts.” (p. 165)

“High exhaustivity tends to lower precision, because the addition of more and more codes results in the retrieval of irrelevant data along side the relevant. Conversely, high specificity results in low recall. Since high specificity uses narrower categories, it produces fewer data in

specificity or exhaustivity when creating and/or applying the vocabulary scheme, the end result will be documents that are difficult or impossible for searchers to retrieve (i.e., low recall and/or precision). If different categorizers or coders (or the same categorizers or coders at different times) use different levels of specificity or exhaustivity when preparing research data, the analysis will produce potentially misleading results.” (p. 166)

“The three factors that most often increase consistency are as follows: documentation to aid application of vocabularies or encoding schemes, low specificity, and low exhaustivity. Obviously, the solution of consistency problems may create other inadequacies. Using low specificity and exhaustivity to achieve consistency will lead to low precision and low recall.” (p. 166)

“Exhaustivity relates to what and how many variables are chosen.” (p. 167)

each category than does low specificity.” (p. 166)

“If indexers of a document use different levels of specificity or exhaustivity when creating and/or applying the vocabulary scheme, the end result will be documents that are difficult or impossible for searchers to retrieve (i.e., low recall and/or precision). If different categorizers or coders (or the same categorizers or coders at different times) use different levels of specificity or exhaustivity when preparing research data, the analysis will produce potentially misleading results.” (p. 166)

“The three factors that most often increase consistency are as follows: documentation to aid application of vocabularies or encoding schemes, low specificity, and low exhaustivity. Obviously, the solution of consistency problems may create other inadequacies. Using low specificity and exhaustivity to achieve consistency will lead to low precision and low recall.” (p. 166)

CONCEITOS

Especificidade é o nível de detalhamento do vocabulário, o número de níveis hierárquicos definidos (p. 163)

Exaustividade é a extensão da representação – o número de fatores indexados, uma característica do vocabulário e sua aplicação (p. 165)

CONCLUSÕES

A exaustividade e a especificidade são estratégias para uma efetiva organização e recuperação da informação (p. 158)

Características que influenciam o vocabulário controlado (p. 163)

A exaustividade está diretamente ligada à revocação, que calcula quanto da informação relevante é recuperada, a partir do total de informações relevantes disponíveis (p. 165).

Se o nível de exaustividade é alto, mais termos são utilizados, aumentando as possibilidades de recuperação (p. 165).

O nível de exaustividade é inversamente proporcional ao nível de precisão (p. 166).

Exaustividade está relacionada de quantas e quais variáveis são escolhidas na representação (p. 167)

O nível de especificidade deve atender ao grupo de usuários da unidade (p. 163)

	<p>Especificidade está diretamente ligada à precisão (p. 164)</p> <p>O nível de especificidade é inversamente proporcional ao nível de revocação (p. 164)</p> <p>Embora a utilização de características como a exaustividade e a especificidade seja considerado o ideal, os indexadores consideram que há grandes problemas em sua aplicação (p. 165)</p> <p>Um alto nível de exaustividade tende a um baixo nível de precisão, já que a adição de um grande número de termos traz a recuperação de informações consideradas irrelevantes. De forma recíproca, um alto nível de especificidade traz um baixo nível de revocação, já que categorias específicas recuperam menos dados que categorias mais gerais (p. 166)</p>
--	---

FICHA 17

Referência: WONG, S. K. M.; YAO, Y. Y. Na information-theoretic measure of term specificity. *Journal of the American Society for Information Science*, v. 43, n. 1, p. 54-61, 1992.

Resumo: The inverse document frequency (IDF) and signal-noise ratio (S/N) approaches are two well known term weighting schemes based on term specificity. However, the existing justifications for these methods are still somewhat inconclusive and sometimes even based on incompatible assumptions. Although both methods are related to term specificity, their relationship has not been thoroughly investigated. An information-theoretic measure for term specificity is introduced in this study. It is explicitly shown that the IDF weighting scheme can be derived from the proposed approach by assuming that the frequency of occurrence of each index term is uniform within the set of documents containing the term. The information-theoretic interpretation of term specificity also establishes the relationship between the IDF and S/N methods.

Palavras-chave:

CONCEITOS ENCONTRADOS

Exaustividade
Nada consta.

Especificidade
“From the precision point of view in particular, index terms of this type should be assigned higher weights. In other words, index terms should be weighted according to their specificity (Robertson, 1974). Sparck Jones (1972) proposed that the value of a term may be measured by the logarithm function of the ratio of the document frequency of the term over the total number of documents in the collection, which reflects in some sense the specificity of the term.” (p. 54)

“Many existing justifications for IDF do not discuss the relationship between the IDF and S/N weighting schemes, although both methods measure the value of a term based on its specificity. If we accept the assumption that the value of a term depends on its specificity, then the problem is reduced to choosing a suitable measure of term specificity” (p. 55)

“Based on the analogue between coding in information theory and indexing in information retrieval (Chen, 1988), we introduce in this paper an information-theoretic measure of term specificity.” (p. 55)

“The information-theoretic interpretation of term specificity also clearly establishes the relationship between the IDF and S/N methods.” (p. 55)

“We then develop a measure of the quality of an evidence within the framework of information theory and apply such a measure to compute the specificity of an index term.” (p. 55)

“In fact, the entropy function has been used as a measure of the diversity of a probability distribution (Rao, 1982). Thus, the evaluation of evidence based on the entropy function essentially amounts to the measurement of specificity of the evidence.” (p. 58)

“Based on such interpretations, the measure of the value of evidence defined by equation can be used as a measure of term specificity, and hence a measure of term value. In fact, Salton and Yang (1973) investigated the relationship between the entropy function and term specificity.” (p. 58)

“In the above analysis, we have clearly shown that both the IDF and S/N weighting schemes can be considered as approximations of the proposed information theoretic measure for term specificity.” (p. 61)

“It is perhaps worth mentioning that although the information-theoretic measure uses the information about term distribution, the statistical data alone may not be sufficient to characterize term specificity.” (p. 61)

CONCEITOS	<p>“Term specificity was also studied extensively with the two-Poisson model, in which specific words are characterized by Poisson distributions (Srinivasan, 1990a; Harter, 1975; Bookstein & Swanson, 1974).” (p. 61)</p> <p style="text-align: center;">CONCLUSÕES</p> <p>O nível de especificidade pode ser medido por meio de fórmulas matemáticas (p. 55).</p>
------------------	---

FICHA 18

Referência: BURGIN, R. The retrieval effectiveness of 5 clustering algorithms as a function of indexing exhaustivity. <i>Journal of the American Society for Information Science</i> , v. 46, n.8, p. 562-572, 1995.	
Resumo: The retrieval effectiveness of five hierarchical clustering methods (single link, complete link, group average, Ward’s method, and wighted average) is examined as a function of indexing exhaustivity with four test collections (CF, Cranfield, Medlars and Time). Evaluations of retrieval effectiveness, based on three measures of optimal retrieval performance, confirm earlier findings that the performance of a retrieval system based on a single link clustering varies as a function of indexing exhaustivity but fail to find similar patterns for other clustering methods. The data also confirm earlier findings the regarding the poor performance of single-link clustering in a retrieval environment. The poor performance of single-link clustering appears to derive from from that method’s tendency to produce a small number of large, ilt-defined document clusters. By contrast, the data examined here found the retrieval performance of the other clustering methods to be generally comparable. The data present here also provide an opportunity to examine the theoretical limits of cluster-based retrieval and to compare these theoretical limits to the effectiveness of operational implementations. Performance standardsfor the four document collections examined here were found to vary widely, and the effectiveness of operational implementations were found to be in the range defined as “unacceptable”. Further improvements in search strategies and document representations warrant investigation.	
Palavras-chave:	
CONCEITOS ENCONTRADOS	
<p>Exaustividade</p> <p>“The results suggests that the performance of a retrieval system based on a single-link clustering varies as a function of indexing exhaustivity. In particular, retrieval performance in such systems appears to improve as the exhaustivity of representations decreases from the most exhaustive representation to an optimal representation based on a intermediate level of indexing exhaustivity where a compromise between indexing exhaustivity and specificity is produced by the document representations. The most exhaustive representations produce low levels of performance, while the highest performance levels are found at relatively low levels of exhaustivity.” (p. 562)</p>	<p>Especificidade</p> <p>“The results suggests that the performance of a retrieval system based on a single-link clustering varies as a function of indexing exhaustivity. In particular, retrieval performance in such systems appears to improve as the exhaustivity of representations decreases from the most exhaustive representation to na optimal representation based on a intermediate level of indexing exhaustivity where a compromise between indexing exhaustivity and specificity is produced by the document representations.” (p. 562)</p>

<p>“It should be noted, of course, that these performance differences have been found only for the most exhaustive representations, which are those typically studied. Whether such differences in retrieval performance will persist across a range of exhaustivity levels remains to be tested.” (p. 562)</p> <p>“The presente study extends Shaw’s work by examining the effectiveness of retrieval as a function of indexing exhaustivity with five hierarquical clustering methods and by examining the effectiveness of theses methods with four tests collections. It also provides na opportunity to determine the best clustering method across a wider range of indexing exhaustivity levels than is tipically investigated.” (p. 563)</p> <p>“The thresholds were used to control exhaustivity of document representations by eliminating terms whose weights fell bellow the thresholds.” (p. 564)</p> <p>“If a cluster is defined as any available component with two or more documents, then these measures are based on the selection of clusters from the set of available clusters at three levels: partition, hierarchy, and exhaustivity space, respectively.</p>	
CONCEITOS	CONCLUSÕES
	Quando há um compromisso entre a exaustividade e a especificidade, há uma potencialização da recuperação informacional (p. 562)

FICHA 19

Referência: CHEN, X. The influence of existing consistency measures on the relationship between indexing consistence and exhaustivity. In: ARSENAULT, C.; TENNIS, J. T. (Ed.). Cultural and Identity in Knowledge Organization. Würzburg: ERGON Verlag, 2008. p. 144 - 150. (Advances in Knowledge Organization, 11).

Resumo: Consistency studies have discussed the relationship between indexing consistency and exhaustivity, and it is commonly accepted that higher exhaustivity results in lower indexing consistency. However, this issue has been oversimplified, and previous studies contain significant misinterpretations. The aim of this study is, to investigate the relationship between consistency and exhaustivity based on a large sample to analyse the misinterpretations in earlier studies. A sample of 3,307 monographs i. e. 6,614 records was drawn from two Chinese bibliographic catalogues. Indexing consistency was measured using two formula which were popular in previous consistency

studies. A relatively high level of consistency was found (64.21% according to the first formula, 70.71% according to the second). Regarding the relationship between consistency and exhaustivity, it was found that when two indexers has identical exhaustivity, indexing consistency was substantially high. On the contrary, when they had different levels of exhaustivity, consistency was significantly low. It was inevitable with the use of two formula. Moreover, a detailed discussion was conducted to analyse the misinterpretations in previous studies.

Palavras-chave:

CONCEITOS ENCONTRADOS

Exaustividade

“Many factors can affect indexing consistency, among which indexing exhaustivity is the most important one and had been most extensively explored in former consistency studies. Considering the relationship between indexing exhasutivity and consistency, it is widely believed that higher exhaustivity leads to lower indexing consistency. However, the relationship between indexing consistency and exhaustivity is not as straightforward as it might seem.” (p. 144)

“A t-test shows no signifcant difference between the two catalogues with regard to the number of index terms. That is to say, the two catalogues have identical level of indexing exhaustivity.” (p. 145)

“Therefore, we should discuss the relationship between consistency and exhaustivity under two different conditions (...) Under condition 1, there is a positive correlation between consistency and exhaustivity. It indicates that the more two indexers totally assign to a book, the more likely they assign the same terms. Under condition 2, the relationship between consistency and the total number of terms assigned is a little more complicated.” (p. 145)

“The experimental data has shown that when two indexers have identical level of exhaustivity, the indexing consistency is considerably high.” (p. 146).

“There is a common accepted statement about the relationship between indexing consistency and exhaustivity in this field: consistency drops as indexing exhaustivity increases.” (p. 148)

CONCEITOS

A exaustividade diz respeito ao número de

Especificidade

Nada consta.

CONCLUSÕES

Quando diferentes indexadores possuem o

termos indexados (p. 145).

mesmo nível de exaustividade, a consistência da indexação não é prejudicada (p. 145).

FICHA 20

Referência: HEINE, M. H. An investigation of the relative influences of database informativeness, query size and query term specificity on the effectiveness of Medline searching. *Journal of Information Science*, v. 21, n. 03, p. 173 - 185, 1995.

Resumo: An investigation of search effectiveness in Medline used regression analysis to compare the relative influences on search performance (characterised by recall, precision and the MZ-metric) of the three variables: database informativeness, and size and mean specificity of the query (a "query" is defined as an unstructured set of search terms). A reduction in random and/or systematic error was sought by means of three experimental devices: error arising from the subjectivity of third party relevance judges was minimised by using behaviourally-generated target sets of records, errors associated with the definitions of queries were avoided (or at least standardised) by using algorithmically derived queries and error associated with uncontrolled variation in the logical structure of the search expression was minimised by using search expressions that had been optimised separately against each of the three performance criteria referred to. It was found that the dominant variable influencing search performance was the number of query terms. Database informativeness had a statistically significant, but small, influence on search performance in almost all data-sets, and the influence could be in either direction, depending on the data-set. The influence of term specificity was in general not statistically significant. Overall, the three predicting variables were able to explain up to 50% of the variation in search performance.

Palavras-chave: Searching, Query languages.

CONCEITOS ENCONTRADOS

Exaustividade

Especificidade

CONCEITOS

CONCLUSÕES

"However, there appears to have been on prior study of the relative influences on retrieval effectiveness of term specificity, database generality and the key related variables of query-set size and individuation." (p. 175)

"The assessment they offer of the relative influences of the search variables 'query size' and 'query term specificity' on retrieval effectiveness." (p. 175)

"It is instructive to note that 'specificity' should properly be regarded as a variable, in that, even when the database is fixed, its value will depend on the user's choice of record fields accessible to the term." (p. 177)

O nível de especificidade deve ser variável, dependendo da escolha do usuário para campos de registro acessíveis ao termo (p. 177).

FICHA 21

Referência: WEINBERG, B. H.; CUNNINGHAM, J. A. Term specificity and online postings - inverse relationship. Proceedings Of The American Society For Information Science, v. 21, p. 144-147, 1984.

Resumo: The assumption that there is an inverse relationship between the specificity of a descriptor and that term's postings in an on-line database is widespread in the literature. As part of a National Science Foundation grant project, this assumption was tested empirically. Our hypothesis was that the inverse relationship holds only for peripheral terms of a discipline rather than for central terms. The data included hierarchical levels of selected terms from several thesauri that cover the field of engineering/technology either as a major focus or as peripheral subject, including TEST, INSPEC, ERIC and MESH, as well as the online postings of those terms. Results weakly support the peripheral hypothesis. The implication for online search strategy and training is that increasing term specificity is not a guaranteed method of lowering recall, i. e., that the theoretical inverse relationship is not borne out by the empirical evidence.

Palavras-chave: Specificity. Thesauri. Term hierarchy. Inverse relationships. Online search strategy. Postings.

CONCEITOS ENCONTRADOS

Exaustividade

Especificidade

“The assumption there is an inverse relationship between the specificity of an index term (descriptor) and the number of postings to that term in online database is widespread in the literature.” (p. 144)

“Donald T. Hawkins, in an article on online search strategy intimating equivalence between specificity and low postings, suggested “search the most specific aspect of the subject (or the lowest posted facet)” (p. 144)

“If upposting were automatically done, the inverse relationship would always obtain; however, in most online database documents are indexed to the appropriate level of specificity only.” (p. 145)

“It is widely believed that specific terms are added to thesauri when broad terms get too many postings. If so, the level of specificity should increase as the dates of terms get later.” (p. 145)

“We used the various hierarchical structures of thesauri to assign numeric levels to terms indicative of increasing specificity, with 1 indicating the most general term, 2 the next most specific term, etc.” (p. 145)

“Our evidence from online database indicates that not only is there a very wide range in the number of postings for index terms, but that

CONCEITOS	<p>disparity is not explained by levels of specificity.” (p. 147)</p> <p style="text-align: center;">CONCLUSÕES</p> <p>O nível de especificidade deve aumentar conforme os termos tornam-se mais antigos e consolidados (p. 145)</p>
------------------	---

FICHA 22

Referência: WILSON, P. End of specificity. Library Resources & Technical Services, v. 23, n. 2, p. 116 – 122, 1979.	
Resumo: Recently announced subject cataloging practitioners at the Library of Congress, calling for systematic duplication of entries at specific and generic leads, ate in direct violation of the Rule Exclusively Specific Entry, accepted by LC. It is argued that if there practices are justified, consistency calls for their general application, which results in abandonment of the rule. But the practices do not accomplish their goals, do not reveal the content of LC's collections, do introduce new inconveniences, do constitute a pointless enlargement of catalog and hence should be abandoned.	
Palavras-chave:	
CONCEITOS ENCONTRADOS	
<p>Exaustividade</p> <p>“It can be argued that an increase of the depth or exhaustivity of indexing would be a useful improvement in subject access.” (p. 122)</p>	<p>Especificidade</p> <p>Nada consta.</p>
A exaustividade pode ser considerada a extensão da indexação (p. 122)	A exaustividade é uma característica útil para o acesso por assunto (p. 122).

FICHA 23

Referência: GIVEN, L. M; OLSON, H. A. Indexing and the "organized" researcher. Indexer, v. 23, n. 3, p. 129-133, 2003.	
This article proposes that indexing concepts relating to relevance, precision, recall, coextensiveness, exhaustivity, specificity and consistency offer a ready-made model that can be applied to the organization of research data. This knowledge organization model contributes significantly to the ability of researchers to collect and organize data in a manner most likely to shed light on the research problems they address.	
Palavras-chave:	
CONCEITOS ENCONTRADOS	
<p>Exaustividade</p> <p>“In indexing, strategies for effective organization and retrieval of such different types of information have been honed by examining specificity, exhaustivity, relevance, and other elements central to the process of subject analysis.” (p. 129)</p>	<p>Especificidade</p> <p>“In indexing, strategies for effective organization and retrieval of such different types of information have been honed by examining specificity, exhaustivity, relevance, and other elements central to the process of subject analysis.” (p. 129)</p>

“A related question is how many concepts will be represented? Here, exhaustivity meets specificity. If a discussion on pets includes cats and dogs, it seems reasonable to be both specific and exhaustive in categorizing those concepts. However, if the discussion includes cats, dogs, fish, birds, rodents, rabbits and llamas it may not include enough information about any one type of animal to justify highly specific representations. In this case, 'pets' might be the more appropriate choice.” (p. 131)

“It would seem that the notions of relevance, precision, recall, specificity and exhaustivity could produce perfect categories and codes, but as indexers know, there are several potential problems in applying these principles.” (p. 131)

“High exhaustivity tends to lower precision: adding more and more codes results in the retrieval of irrelevant data alongside the relevant. Conversely, high specificity tends to lower recall. Since high specificity uses narrower categories it will produce fewer data in each category than will low specificity.” (p. 131)

““The three factors that most often increase consistency are:

1. documentation to aid application of vocabularies or encoding schemes as stressed in indexing standards (Calvert, 1996: 74);
2. low specificity;
3. low exhaustivity.” (p. 131)

“Exhaustivity is also an issue in both the development and application of the vocabulary. Exhaustivity is defined as the breadth of representation - the number of factors indexed. At the level of the vocabulary, exhaustivity is concerned with the different aspects or facets included.” (p. 131)

“The application of the concept of exhaustivity is related to the level of indexable matter: how much of a particular topic must be covered by a book before it is represented in the index? Does it need to be particularly insightful or will a passing mention suffice? How is it related to the core themes of the book?” (p. 131)

“Exhaustivity is closely related to recall - how

“Specificity is the relative detail within the vocabulary – the number of hierarchical levels defined. In organizing animals, a scheme that stops at differentiating only domesticated from wild animals would have very low specificity. Specificity increases with each level as the hierarchy becomes deeper, extending to a level that differentiates particular breeds of particular animals (such as 'sealpoint Siamese cats'). As with coextensiveness, the level of specificity should serve the needs of users of that organizational scheme; an index in veterinary science will probably need much higher specificity in regard to animals than an index used by elementary schoolteachers. In approaching research data, the research questions and/or hypotheses determine the level of specificity required for organizing the categories and codes assigned by the researcher. Once the level of specificity is defined, it must be applied to take full advantage of that specificity, which is, of course, standard indexing practice. Each item must be coded at the most specific level available. This application of specificity would dictate that Siamese cats be indexed as 'Siamese' and not as 'cats' or 'pets' or 'domesticated animals.'” (p. 130)

“To approach this ideal it is important to remember that precision is enhanced by high specificity, meaning that data are indexed at a very precise level - the selected terms are finely grained, using very detailed levels of categorization.” (p. 130)

“A related question is how many concepts will be represented? Here, exhaustivity meets specificity. If a discussion on pets includes cats and dogs, it seems reasonable to be both specific and exhaustive in categorizing those concepts. However, if the discussion includes cats, dogs, fish, birds, rodents, rabbits and llamas it may not include enough information about any one type of animal to justify highly specific representations. In this case, 'pets' might be the more appropriate choice.” (p. 131)

“It would seem that the notions of relevance, precision, recall, specificity and exhaustivity could produce perfect categories and codes, but as indexers know, there are several potential problems in applying these principles.” (p. 131)

much of the available relevant information or data is actually retrieved. Maximum recall means retrieving every last instance of a theme or variable.” (p. 131)

“One way to enhance recall is to seek high exhaustivity. If exhaustivity is high more codes are used, which will allow more data to be retrieved and analysed.” (p. 131)

“It is important to remember, however, that over-coding leads to extreme levels of exhaustivity and specificity and thus to low precision and low recall. The problem of overcoding occurs when researchers code beyond the research questions, including interesting themes that are simply not relevant - a challenge that all researchers must address in managing data.” (p. 132)

“The other half of the model relates to exhaustivity: how many themes are needed to address the research questions? A completely exhaustive codebook (i.e. a list of all relevant thematic terms) is just not practical as it would take too long to develop all the themes and search for them across all transcripts.” (p. 132)

CONCEITOS

Especificidade é o nível de detalhamento do vocabulário, o número de níveis hierárquicos definidos (p. 130)

Exaustividade é a extensão da representação – o número de fatores indexados (p. 131)

“High exhaustivity tends to lower precision: adding more and more codes results in the retrieval of irrelevant data alongside the relevant. Conversely, high specificity tends to lower recall. Since high specificity uses narrower categories it will produce fewer data in each category than will low specificity.” (p. 131)

“The three factors that most often increase consistency are:

1. documentation to aid application of vocabularies or encoding schemes as stressed in indexing standards (Calvert, 1996: 74);
2. low specificity;
3. low exhaustivity.” (p. 131)

“The key was to make choices about the level of specificity to ensure optimal precision in code assignment.” (p. 132)

“It is important to remember, however, that over-coding leads to extreme levels of exhaustivity and specificity and thus to low precision and low recall. The problem of overcoding occurs when researchers code beyond the research questions, including interesting themes that are simply not relevant - a challenge that all researchers must address in managing data.” (p. 132)

CONCLUSÕES

A exaustividade e a especificidade são estratégias para uma efetiva organização e recuperação da informação (p. 129)

O nível de especificidade deve atender ao grupo de usuários da unidade (p. 130)

Embora a utilização de características como a exaustividade e a especificidade seja considerado o ideal, os indexadores consideram que há grandes problemas em sua aplicação (p. 131)

Um alto nível de exaustividade tende a um baixo nível de precisão, já que a adição de um grande número de termos traz a recuperação de informações consideradas irrelevantes. De forma recíproca, um alto nível de especificidade traz um baixo nível de revocação, já que categorias específicas recuperam menos dados que

	<p>categorias mais gerais (p. 166)</p> <p>De maneira prática, a aplicação do conceito exaustividade está relacionado a análise de quanto determinado tópico está incluído no livro, para posterior indexação do assunto analisado e ainda, análise de quantos termos são necessários para atender as necessidades informacionadís dos usuários (p. 131-132)</p>
--	---

FICHA 24

Referência: HARDING, A. F.; WILLETT, P. Indexing exhaustivity and the computation of similarity matrices. JASIST, v. 31, n. 4, p. 298 - 300, 1980.

Resumo: Some of the automatic classification procedures used in information retrieval derive clusters of documents from an intermediate similarity matrix, the computation of which involves comparing each of the documents in the collection with all of the others. It has recently been suggested that many of these comparisons, specifically those between documents having no terms in common, maybe avoided by means of the use of an inverted file to the document collection. This communication shows that the approach will effect reductions in the number of inter document comparisons only if the documents are each indexed by a limited number of indexing terms; if exhaustive indexing is used, many document pairs will be compared several times over and the computation will be greater than when conventional approaches are used to generate the similarity matrix.

Palavras-chave:

CONCEITOS ENCONTRADOS

Exaustividade

“Thus, while neither expression is likely to give an accurate value for C, they both suggest that C will be strongly dependent on the exhaustivity of indexing,i.e., on the numbers of index terms assigned to each of the documents in the collection.The exhaustivity will also affect T in the denominator of both expressions.Thus, if k is high, it is possible that the procedure will require more computation than the normal $N(N-1)/2$ approach, the difference being duet o the large numbers of document pairs which are compared several times over.” (p. 299)

“It is likely that the term deletion methods used result in document descriptions different from those that would be obtained if the collection was processed by trained indexers or by an automatic indexing procedure to comparable degrees of exhaustivity.” (p. 299)

“As a partial check on the results above, the titles of the documents were stop listed,

Especificidade

Nada consta.

<p>stemmed, and used to produce an inverted file in the same manner as the abstracts had been processed: the number of comparisons required was found to be little dissimilar to the numbers resulting from abstracts processed to a similar level of exhaustivity.” (p. 299)</p> <p>“As the indexing exhaustivity increases and the lists in the inverted file lengthen, the neglect of the most frequent terms is likely to have less and less effect on the structure of the final hierarchy which is obtained after clustering” (p. 300)</p> <p style="text-align: center;">CONCEITOS</p>	RESULTADOS
<p>Número de termos de indexação atribuídos a cada um dos documentos de uma coleção (p. 299)</p>	

FICHA 25

<p>Referência: BLAIR, D. The challenge of commercial document retrieval, part I: major issues and a framework based on search exhaustivity, determinancy of representation and document collection size. <i>Information Processing & Management</i>, v. 38, p. 273 - 291, 2002.</p>	
<p>Resumo: With the growing focus on what is collectively known as “knowledge management”, a shift continues to take place in commercial information system development: a shift away from the well-understood data retrieval/database model, to the more complex and challenging development of commercial document/retrieve models. While document retrieval has had a long and rich legacy of research, its impact on commercial applications has been modest. At the enterprise level most large organizations have little understanding of, or commitment to, high quality access and management. Part of the reason for this is what we still do not have a good framework for understanding the major factors which affect the performance of large-scale corporate document retrieval systems. The thesis of this discussion is that document retrieval - specifically, access to intellectual content - is a complex process which is most strongly influenced by three factors: the size of the document collection; the type of search (exhaustive, existence or sample); and, the determinancy of document representation. Collectively, these factors can be used to provide a framework for, or taxonomy, document retrieval, and highlight some of the fundamental issues facing the design and development of commercial document retrieval systems. This is the first of a series of three articles. Part II (D. C. Blair, The challenge of commercial document retrieval. Part II A strategy for document searching based on identifiable document partitions, <i>Information Processing and Management</i>, 2001b, this issue) will discuss the implications of this framework for search strategy, and Part III (D. C. Blair, Some thoughts on the reported results of text retrieval Conference TREC, <i>Information Processing Management</i>, 2002, forthcoming) will consider the importance of the TREC results for our understanding of operating information retrieval system.</p>	
<p>Palavras-chave:</p>	
CONCEITOS ENCONTRADOS	
Exaustividade	Especificidade
<p>“In short, there is no single type of document retrieval situation, but a family of similar types of searches which can be classified along three dimensions: search exhaustivity, database size</p>	<p>Nada consta.</p>

and the determinancy of document representation.” (p. 283).

“We already discussed how database size can significantly affect the retrieval of intellectual content. Search exhaustivity has a similar important influence. An exhaustive search on a document retrieval system is one in which the inquirer needs to see all, or nearly all, of the documents which are usefull for him.” (p. 283)

CONCEITOS

RESULTADOS

Quando a busca é feita de maneira exaustiva pelo usuário, ela recupera uma maior parte de documentos úteis (p. 283)

FICHA 26

Referência: GUILLEMETTE, M. G.; LAROCHE, M.; CADIEUX, J. Defining decision making process performance: Conceptualization and validation of an index. *Information & management*, v. 51, p. 618 - 626, 2014.

Resumo: Many studies have demonstrated the impact of information technology (IT) on decision making but few have used decision making process performance (DMPP) as a dependent variable. Our study proposes a rich formative conceptualization of DMPP, a valid and reliable measure for this construct, and studies its influence on the quality of decision making. The results show that DMPP is a formative second-order aggregate construct composed of procedural rationality, exhaustivity of the information analyzed, openness of spirit, and effort. This study illustrates the importance of building proper definitions of constructs and contributes to the development of shared meaning in IS.

Palavras-chave: Query reformulation; Pseudo-relevance feedback; Part of speech tagging; Part of speech blocks (POS blocks).

CONCEITOS ENCONTRADOS

Exaustividade

“Consequently, this paper proposes a rich formative conceptualization of decision making process performance (DMPP) and studies its influence on the quality of decision making. First, we conceptualize DMPP as a multidimensional construct in the form of a formative second-order aggregate construct [26] composed, at the first-order level, of four reflective dimensions: procedural rationality, exhaustivity of the information consulted, openness of spirit, and effort.” (p. 619)

“Finally, we wanted the dimensions to be theoretically consistent with our conceptualization of the decision making process as one that is largely rational and

Especificidade

Nada consta.

structured to ensure the internal coherence of the DMPP construct. Four dimensions met these criteria: procedural rationality, exhaustivity of the information analyzed, effort and openness of spirit.” (p. 620).

“The exhaustivity of the information taken into account by the decision maker is another important factor in the decision making process. Different types of information are required to make business decisions. The more information that is available to and used by a decision maker, and the more detailed this information is, the better the quality of the decision making process [21]. Exhaustivity of the information refers specifically to the “completeness” of the information available to the decision maker when following each step of the decision making process. In other words: Does the decision maker have enough information at the time that the decision must be made?” (p. 620)

“Procedural rationality, exhaustivity of the information analyzed, effort and openness of spirit may be conceptualized as the underlying dimensions of DMPP, the dimensions that are conceptually distinct” (p. 620)

“We have shown that DMPP should be conceptualized as a formative construct, consisting of four dimensions that capture its different characteristics (i.e. procedural rationality, the exhaustivity of the information analyzed, effort, and openness of spirit) that ensure better coverage of the construct.” (p. 625).

CONCEITOS

RESULTADOS

Informações exaustivas trazem um maior número de informações para auxiliar no momento de tomada de decisões (p. 620).

FICHA 27

Referência: LYKKE, M.; PRICE, S.; DELCAMBRE, L. How doctors search: A study of query behaviour and the impact on search results. Information, processing and management, v. 48, p. 1151 - 1170, 2012.

Resumo: Professional, workplace searching is different from general searching, because it is typically limited to specific facets and targeted to a single answer. We have developed the semantic component (SC) model, which is a search feature that allows searchers to structure and specify the search to context-specific aspects of the main topic of the documents. We have tested the model in an interactive searching study with family doctors with the purpose to explore doctors' querying behaviour, how they applied the means for specifying a search, and how these features contributed to the search outcome. In general, the doctors were capable of exploiting system features and search tactics during the searching. Most searchers produced well-structured queries that contained appropriate search facets. When searches failed it was not due to query structure or query length. Failures were mostly caused by the well-known vocabulary problem. The problem was exacerbated by using certain filters as Boolean filters. The best working queries were structured into 2–3 main facets out of 3–5 possible search facets, and expressed with terms reflecting the focal view of the search task. The findings at the same time support and extend previous results about query structure and exhaustivity showing the importance of selecting central search facets and express them from the perspective of search task. The SC model was applied in the highest performing queries except one. The findings suggest that the model might be a helpful feature to structure queries into central, appropriate facets, and in returning highly relevant documents

Palavras-chave: Query reformulation; Pseudo-relevance feedback; Part of speech tagging; Part of speech blocks (POS blocks).

CONCEITOS ENCONTRADOS

Exaustividade

“Sormunen (2002a) studied query exhaustivity and extent for 18 search tasks using a Finnish newspaper text database as the test collection. The study showed that increasing the exhaustivity of optimized best-match queries led to a significant increase in precision, especially for tasks that require high precision.” (p. 1154)

“Selecting the right set of key search facets seems to have more impact on search success compared to a high degree of exhaustivity.” (p. 1668)

“The findings support and extend previous results about query structure and exhaustivity that indicate the importance of determining the crucial search facets and right perspective and terminology.” (p. 1169).

CONCEITOS

Especificidade

“Since Semantic Component model are intended to facilitate retrieval, and not intended to describe the domain; knowing the contents of a particular document collection, or the common information needs among users of the collection, may lead to selecting SCs with varying degrees of specificity to represent document content.” (p. 1155)

“Neither broader keys (BT) nor narrower keys (NT) were used; the searchers chose search keys at the same level of specificity as the task description.” (p. 1164).

RESULTADOS

O aumento da exaustividade leva também, um aumento significativo de termos precisos (p. 1154)

A escolha das facetes de busca e da terminologia são fatores importantes na exaustividade.

Referência: SHAW JR., W. M. Subject and Citation Indexing. Part I: The Clustering Structure of Composite Representations in the Cystic Fibrosis Document Collection. *JASIST*, v. 42, n. 9, p. 669 - 675, 1991.

Resumo: The presence of clustering structure in the cystic fibrosis (CF) Document Collection is evaluated as a function of the exhaustivity of five composite representations. The composite representations are constructed from two subject descriptions, based on Medical Subject Headings and subheadings, and two citation indexes, based on the complete set of references in and a comprehensive set of citations to each document. Experimental results reveal observable evidence for clustering structure at all exhaustivity levels of all composite representations but also show that the evidence for clustering structure diminishes as the exhaustivity of each representation is decreased. The representation composed of references and citations shows less evidence of clustering structure at the exhaustive level but more uniform evidence of clustering structure over a wide range of exhaustivity levels than composite representations that include subject descriptions. The structures imposed on the CF Document Collection by all composite representations satisfy the necessary condition for a meaningful clustering outcome.

Palavras-chave:

CONCEITOS ENCONTRADOS

Exaustividade

Especificidade

“In a previous investigation, the presence of clustering structure in the cystic fibrosis (CF) Document Collection (Shaw, Wood, Wood, & Tibbo, in press) was investigated as a function of the exhaustivity of four subject representations, based on Medical Subject Headings (MeSH) and subheadings, and two citation representations, derived from the complete set of references in and a comprehensive set of citations to each document (Shaw, 1990b). Experimental results show that:

- (1) the strongest evidence of clustering structure is associated with exhaustive subject representations,
- (2) the weakest evidence of clustering structure is associated with the least exhaustive subject representations, and
- (3) the evidence for clustering structure associated with the citation representations is more uniform than that associated with the subject representations as a function of the exhaustivity of the representations.

Despite characteristics that distinguish subject and citation representations, evidence for the presence of clustering structure is shown to diminish as the exhaustivity of each representation is decreased (Shaw, 1990b).” (p. 669).

“A specificity weight based on the conventional inverse document frequency formulation has been computed for each descriptor assigned to the papers (Salton & McGill, 1983, p. 63).” (p. 670)

“Results show how the evidence for clustering structure varies as a function of the exhaustivity of each composite representation.” (p. 669)

“A descriptor-weight threshold, denoted by DW, is used to control the exhaustivity and specificity of the composite document representations; descriptor k is retained if $w_k \geq DW$. When $DW = 0$, all descriptors are retained.” (p. 670)

“For each composite representation, the number of documents with one or more descriptor, the maximum number of descriptors assigned to a document, and the average number of descriptors per document decrease as the exhaustivity of the representation is decreased.” (p. 670)

“The analyses presented here focus on systematic patterns in the evidence for clustering structure as a function of exhaustivity and similarity thresholds.” (p. 672)

“Large values of DS over a wide range of TS values provide strong evidence for the presence of clustering structure at the selected exhaustivity level.” (p. 672)

“For each composite representation, differences between the observed and expected number of components have been computed for levels of the single-link hierarchies associated with exhaustivity levels in the range $0 < DW < 900$.” (p. 673)

“Despite exceptions, a maximum value of DS can be detected for all exhaustivity levels for which $DW < 800$.” (p. 673)

“For each composite representation, the range of values for which $DS > 0$ and the maximum values of DS decrease as DW is increased and the exhaustivity of the representation is decreased.” (p. 673)

“Like the previous study, results for the composite representations presented here show the strongest evidence of clustering structure at the exhaustive representations, and like the

<p>previous study, the evidence for clustering structure diminishes as the exhaustivity of the representation is decreased.” (p. 674)</p> <p>“Like the constituent subject representations, the evidence for clustering structure, based on composite representations that include subject descriptions, diminishes as the exhaustivity of the representation is decreased.” (p. 675)</p> <p>“Specifically, the extent to which documents relevant to the same query are associated and documents relevant to different queries are segregated in single-link clustering outcomes is investigated as a function of the exhaustivity of the five composite representations, and the optimal effectiveness of cluster-based retrieval is related to the evidence for clustering structure presented here.” (p. 675)</p>	
CONCEITOS	RESULTADOS
	A extensão com que um documento é representado o associa a outros documentos do mesmo assunto, evidenciando uma função da exaustividade (p. 675)

FICHA 29

Referência: TAUBE, M. Specificity in Subject Headings and coordinate indexing. Library Trends, v. 1, n. 2, p. 219 - 223, 1952.	
Resumo:	
Palavras-chave: CONCEITOS ENCONTRADOS	
Exaustividade	Especificidade

<p>Não se aplica.</p>	<p>“In previous papers on coordinate indexing it has been indicated that any desired degree of specificity could be attained by "coordinating" as many general terms as were desired or necessary. What was not realized was that the specificity achieved by the intersection, coordination, or logical product of terms of equal generality differs in principle from the specificity of a specific word or phrase and whatever degree of subdivision is allowed. Gull and Taube knew and had participated in establishing the rules and practices set forth and described by Haykin and Aronson. But it was only through a recent intensive study of the relation of specificity to subdivision on well over a thousand catalog cards, that the difference in kind between the two methods of attaining specificity was clearly recognized.” (p. 222)</p>
<p>CONCEITOS</p>	<p>RESULTADOS</p>
	<p>De forma a obter uma maior especificidade, os termos representativos podem ser combinados, criando uma intersecção entre os documentos, que se diferem a partir do grau de especificidade utilizado. (p. 222)</p>