

A EXAUSTIVIDADE E A ESPECIFICIDADE COMO VALORES ÉTICOS NO PROCESSO DE INDEXAÇÃO: uma análise baseada na literatura disponibilizada em Portugal

EXHAUSTIVITY AND SPECIFICITY AS ETHICAL VALUES IN THE PROCESS OF INDEXING: an analysis based in the literature available in Portugal

Isadora Victorino Evangelista | Maria da Graça Melo Simões | José Augusto Chaves Guimarães

Resumo: Considerando a natureza mediadora dos processos de organização da informação destaca-se a necessidade da abordagem da dimensão ética destes na recuperação da informação. Propõe-se analisar os conceitos de exaustividade e especificidade enquanto valores éticos na indexação com base na literatura da área disponibilizada em Portugal em bases de dados *online*. Parte-se de uma abordagem exploratória com o recurso ao levantamento bibliográfico e à análise de conteúdo baseada no método de Bardin. Dos resultados infere-se que os conceitos abordados podem ser valores quantitativos e qualitativos dos termos representativos de um documento, assumindo-se estes como valores éticos da indexação; afetam a recuperação da informação ao estabelecerem a quantidade/diversidade e a expressividade dos termos, dando resposta a uma sociedade culturalmente heterogénea; usados em simultâneo potenciando relações semânticas aumentam a probabilidade da recuperação de documentos, a literatura sobre o tema é escassa e dispersa, pelo que se recomenda um maior investimento no estudo do mesmo.

Palavras-chave: Indexação; Ética na Informação; Especificidade; Exaustividade

Abstract: The intermediate nature of the information organization process emphasizes ethical aspects of information retrieval. This study explores the completeness and specificity as concepts of indexing area, by moving them from simple information retrieval measures to ethical values of the indexer. The concepts are analyzed founded on the specific literature available in online databases and according to the content analysis based on Bardin method. The main results demonstrate that the concepts addressed could be considered as quantitative and qualitative values of the representative terms, accepting these as the ethical values of indexing; they affect the information retrieval by affecting the quantity and diversity of the terms, answering the cultural needs of a heterogeneous society; when used together, they increase the probability of information retrieval, potentiating the semantical relations; the literature on the subject is still scarce and disperse, suggesting the need of a greater investment in the study of this theme.

Keywords: Indexing; Information Ethics; Specificity; Exhaustivity

Introdução

De acordo com Guimarães (2008), o conhecimento assume um movimento helicoidal no contexto da Ciência da Informação, depois de produzido e registado é disponibilizado assumindo, deste modo, a sua função social. Para que possa ser utilizado pela sociedade e produzir novo conhecimento, a sua organização é realizada com base em instrumentos de representação documental que o permitem estruturar e desta forma poder tornar acessível a quem dele necessita.

No que respeita à representação do conhecimento, fase intermediária entre o documento e o utilizador, ela realiza-se com base em duas abordagens distintas mas complementares: uma descritiva, baseada essencialmente nos aspetos físicos do documento (catalogação) e

outra temática, que consiste na análise e representação do conteúdo do documento (indexação). Da última resultam, entre outros produtos, os índices e os catálogos de assunto, instrumentos de informação secundária que são pontes entre a informação contida num documento e o utilizador.

A indexação por assunto é uma operação documental milenar, da qual já existem testemunhos na Mesopotâmia. São vestígios desta remanescência temporal os envelopes de documentos encontrados nesta civilização, identificados com os assuntos de que tratavam, para preservar os documentos e economizar tempo a quem buscava a informação. Já mais perto da contemporaneidade, no século XIX, os índices assumem uma posição evidente, como pode observar-se nas duas situações que se apresentam. A primeira refere-se à sua introdução nos livros escolares britânicos, a segunda à autonomia que ganharam quando se formaliza a distinção entre os conceitos *table* (sumário) e *index* (índice), que os identifica inequivocamente. Ainda neste século, contribuíram para a sua afirmação as obras de William Frederick Poole - *Poole's Index to periodical literature*, e de John Shaw Billings, que elaborou o primeiro índice da literatura médica. No que respeita à elaboração dos catálogos de assuntos, destaca-se a obra de Charles Ammi Cutter - *Rules for a printed dictionary catalog*, que veio regular, de algum modo, esta prática (WITTY, 1973; PETTEE, 1945; BORKO e BERNIER, 1978; SILVA e FUJITA, 2004).

Esta operação documental consiste na seleção de conceitos de um documento e na sua tradução através de termos de uma linguagem controlada ou não controlada. Tal prática envolve processos documentais como a análise conceitual, pela qual se identificam e selecionam os conceitos mais representativos de um documento, e a representação dos mesmos através de termos de indexação (ROWLEY, 1982; CHAUMIER, 1986; ASSOCIAÇÃO, 1992).

Para uma recuperação da informação satisfatória, que atenda às carências informacionais do utilizador é necessária uma indexação eficaz, que possibilite a triangulação entre o conteúdo do documento, a sua representação e a necessidade informacional do utilizador.

No desenvolvimento deste processo há dois princípios que devem ser sempre considerados: a exaustividade e a especificidade. O primeiro relaciona-se com o número de assuntos de um documento que são selecionados, e que se traduz objetivamente na quantidade de termos representativos; o segundo tem a ver com o nível de pormenor com o qual um conceito é representado. De acordo com alguns teóricos da indexação, estes dois princípios são considerados como decisões administrativas, pelo facto de serem determinados no momento da indexação, baseados nas políticas de indexação das respetivas unidades de informação (FOSKETT, 1973; CHAUMIER, 1986; LANGRIDGE, 1989; ASSOCIAÇÃO, 1992; SIMÕES, 2011).

Devido a um conjunto de razões, de entre as quais se destaca a subjetividade do indexador, em geral a operação de indexação não é contextualizada no seu tempo e no seu espaço. Muitas vezes o profissional da informação não atende aos aspetos sociais, culturais, políticos, a todo um conjunto de valores, princípios, preconceitos e crenças que envolvem o contexto desta operação. A observância desta concorrerá, inevitavelmente para o não reconhecimento de uma dimensão ética inerente à organização e representação do conhecimento.

A Ética é um domínio no campo da Filosofia que visa a procura do bem-estar na convivência em sociedade, ideia que se encontra ancorada nos conceitos de *bem* e de *mal*, e que tem como objetivo uma harmoniosa coexistência social. É seu objeto a moral, que trata das ações relacionadas com o binómio bem/mal e que a partir de regras baseadas numa delimitação espaço-temporal, determina o que se deve ou não se deve fazer em sociedade (MALIANDI, 2004; JAPIASSÚ, 2008; GUIMARÃES *et al.* 2008).

O campo da organização do conhecimento, por sua vez, encontra-se respaldado num conjunto de valores que são os fundamentos da moral, que se destacam em determinada sociedade, num dado espaço e num determinado tempo, com o objetivo de disciplinar a convivência dos membros que a constituem, buscando o bem-comum (JAPIASSÚ, 2008; GUIMARÃES *et al.* 2008).

Dada a sua complexidade e afirmação na sociedade contemporânea, as questões relacionadas com a privacidade, a equidade, o acesso à informação, o respeito pelo utilizador e o princípio pela garantia cultural, têm vindo a ser destacadas como objeto de discussão na literatura da área da Ciência da Informação, em trabalhos académicos e científicos, desde meados do século XX. É o caso do que foi realizado por Guimarães *et al.* (2008), no qual, entre outras questões relacionadas com a ética na indexação, se apontam a especificidade e a exaustividade como valores éticos a serem considerados nesta operação. Esta nova abordagem leva a que conceitos direta ou indiretamente associados aos valores éticos, deixem de ser considerados apenas como simples medidas matemáticas usadas para determinar o sucesso ou insucesso de um sistema de recuperação de informação, no que se refere ao desempenho nos aspetos relacionados com a precisão e/ou revocação dos resultados de uma pesquisa, e passem a ser considerados também como medidas de natureza qualitativa que, na prática, concorrem para uma maior abrangência em termos de heterogeneidade cultural na recuperação da informação.

Deste modo, pode inferir-se que a presença da exaustividade e da especificidade na representação da informação vai ao encontro de um compromisso ético com o utilizador (GUIMARÃES, 2000), que está patente no modo como o catálogo de assuntos ou o índice evidencia o mais adequadamente possível. A diversidade temática do documento (exaustividade) e a profundidade com a qual é abordada a realidade que manifesta (especificidade).

Entre outros fatores que determinam as tomadas de decisão sobre o uso destes dois princípios da indexação destaca-se a ética do indexador. O modo como este aspeto, extremamente subjetivo, influencia os produtos da indexação e conseqüente recuperação da informação faz dele, desde sempre, objeto de estudo de vários autores da área. Entre outros, salientam-se: (DALHBERG, 1992; GUIMARÃES, 2000, 2006; GUIMARÃES e FERNÁNDEZ-MOLINA, 2003; GUIMARÃES *et al.*, 2008; BEGHTOL, 2002, 2005; OLSON, 2002; PINHO, 2006; MILANI, 2007, LÓPEZ-HUERTAS, 2008).

Partindo-se da ideia de que a exaustividade e a especificidade são dois primados da indexação por assunto, este estudo apresenta como objetivo geral analisar estes conceitos enquanto valores éticos desta prática, no âmbito da literatura da área, disponibilizada em Portugal em bases de dados *online*. Como objetivos específicos propõe-se: i) identificar e sistematizar os elementos que caracterizam a construção concetual dos dois princípios considerados na literatura na área da Ciência da Informação; ii) analisar como se colocam

estes dois princípios na representação da informação, verificando a partir da literatura se são considerados valores éticos; iii) sistematizar um conjunto de argumentos extraídos da literatura consultada que corroboram o valor ético destes dois princípios no processo de indexação.

Como pode observar-se no ponto que segue, para dar cumprimento aos objetivos enunciados optou-se por uma abordagem metodológica de âmbito exploratório, com o intuito de contribuir para a familiarização do problema proposto, de forma a evidenciar essa questão e construir algumas hipóteses para futuras soluções.

A revisão bibliográfica efetuada indicia que o compromisso entre exaustividade e especificidade se encontra estritamente relacionado com a recuperação da informação. Ao estabelecerem valores quantitativos e qualitativos dos termos representativos podem ser, neste contexto, considerados como valores éticos da atividade de indexação.

Metodologia

A pesquisa proposta insere-se numa abordagem metodológica exploratória, que tem como objetivo familiarizar-se com o problema focado no estudo, de modo a evidenciá-lo e a construir algumas hipóteses com base na elaboração de algumas ideias sobre o tema. Para isso fez-se o levantamento bibliográfico e a análise de conteúdo desse material. Deste modo, este estudo apresenta um caráter bibliográfico, na medida em que recorre a materiais já publicados, especialmente livros e artigos científicos (GIL, 2002).

Num primeiro momento, para construir um *corpus* teórico sólido sobre o assunto realizou-se uma revisão de literatura na qual foi privilegiada a bibliografia de autores que abordam os aspetos éticos relacionados com a organização e representação do conhecimento. Procedeu-se ao levantamento de um conjunto de bibliografia que permitisse construir um arcabouço teórico sobre a prática da indexação, em especial no que se refere aos dois princípios considerados. Esta contextualização teórica, além de permitir um enquadramento concetualmente robusto ao estudo, permitiu suportar teoricamente também o estudo empírico, que consistiu na recolha e na análise do *corpus* da pesquisa.

Num segundo momento procedeu-se à recolha e à análise dos dados, sendo a análise de conteúdo feita de acordo com o método proposto por Bardin (2003), o qual consiste num conjunto de procedimentos aplicáveis a diversos tipos de discurso e possui como ponto principal da análise a inferência.

Dada a importância que a metodologia assume neste estudo entendeu-se oportuno apresentar as fases do seu desenvolvimento:

Critérios usados na recolha de dados

A recolha dos dados foi realizada no dia 13 de outubro de 2015. As pesquisas foram efetuadas exclusivamente em artigos científicos, a partir de um *corpus* advindo da área da Ciência da Informação. Na pesquisa foram considerados os termos **exaustividade** e **especificidade** e os seus equivalentes em inglês - **exhaustivity** e **specificity**, combinados com os termos **indexação** e **indexing**.

Para a seleção dos artigos utilizou-se o critério proposto por Abad García (2005); é caracterizado por apresentar uma natureza subjetiva orientada para o utilizador, dependendo por isso, do seu âmbito. Assenta nas quatro categorias que caracterizam o tipo de relevância proposto por Saracevic (1996): a) a relevância do conteúdo, onde não só se considera se o documento possui os termos de pesquisa, mas também o contexto em que eles são tratados; b) a pertinência, onde se procura estabelecer uma relação entre os conteúdos dos documentos recuperados e a necessidade de informação; c) a relevância situacional, que relaciona o conteúdo do documento com o problema que motivou a pesquisa, e ainda, d) a relevância motivacional, que relaciona o conteúdo do documento com o seu uso.

Fontes de obtenção de dados

As plataformas de bases de dados utilizadas na pesquisa para constituir o *corpus* de dados foram as seguintes: a B-on, enquanto plataforma agregadora de conteúdos de distintas bases de dados, acessível por subscrição na Universidade de Coimbra e financiada pelo governo português; a EBSCO (por meio da ferramenta de descobrimento EBSCO *Discovery Search*) e a Elsevier (por meio da base *Search Direct*), enquanto plataformas de bases de dados privadas igualmente subscritas pela Universidade de Coimbra.

A plataforma EBSCO (2015) é um serviço privado que agrega diversas bases de dados desenvolvidas pela empresa comercial. É seu objetivo auxiliar as bibliotecas e outras instituições de investigação a prover acesso informacional aos seus utilizadores. Possui cerca de 70 anos de existência e relações com mais de 79.000 editores no mundo fornecendo materiais como periódicos, livros, ferramentas de gestão de recursos eletrónicos, bases de dados secundárias, e serviços relacionados com unidades de informação.

Além das bases de dados, nas mais diversas áreas do conhecimento, das quais a EBSCO é produtora, como a saúde, a administração e as novas tecnologias, a plataforma possui também um agregador de conteúdos, denominado EBSCO *Discovery Search* (EDS), que funciona como um polibuscador, que pesquisa em todas as bases de dados produzidas pela EBSCO¹, tornando assim a pesquisa da informação mais rápida e eficaz.

A editora Elsevier (2015) surgida em 1880, evoluiu hoje para uma grande editora científica. Entre outras bases, é detentora da *Science Direct*, que possui segmentos em mais de 17 áreas do conhecimento e disponibiliza 50 periódicos da área de Ciência da Informação. Além dos periódicos que faculta em acesso *online*, a plataforma oferece também livros digitalizados e bases de dados secundárias que disponibilizam documentos da área específica do conhecimento.

A base de dados da Elsevier utilizada para a presente pesquisa, foi a *Science Direct*. De acordo com o *site* oficial da empresa (2015) é uma base utilizada por mais de 15.000 pesquisadores, que combina textos científicos, técnicos e publicações da área da saúde, para tornar o trabalho desses pesquisadores mais eficientes. A base indexa mais de 13

¹ Este agregador de conteúdos pode ser entendido como uma ferramenta de descoberta que, de acordo com Rodriguez Bravo *et al.* (2014) é reconhecida por concentrar num único ponto de acesso todos os materiais disponíveis no catálogo, além de combinar diversas fontes de informação mesmo numa busca simples, podendo assim, ser considerada um grande avanço, pois combina o catálogo, o índice e diferentes fontes de informação numa única ferramenta.

milhões de publicações, que se encontram disponíveis em 2.500 periódicos científicos e 33.000 livros da editora.

A plataforma B-on (Biblioteca do Conhecimento *online*) é uma plataforma que disponibiliza e dá acesso a textos integrais de diversos periódicos científicos a partir das assinaturas feitas pelas instituições subscritoras. Data de 2004, a partir de uma iniciativa pública, e tem como missão garantir à comunidade acadêmica o acesso a diversas publicações científicas através do acesso a uma única plataforma *online*. São seus objetivos: a promoção, a divulgação e o estímulo à produção e consulta de conteúdos científicos, a contribuição para a construção da sociedade do conhecimento e facilitar a gestão de todo esse conhecimento científico² (PORTUGAL, 2015).

Foi também feita a busca em dois repositórios institucionais portugueses: Repositório Científico de Acesso Aberto de Portugal (RCAAP) e Repositório Institucional Estudo Geral (Universidade de Coimbra). Não se apresenta a sua caracterização, pelo facto, de neles não se ter recuperado informação relevante para este estudo.

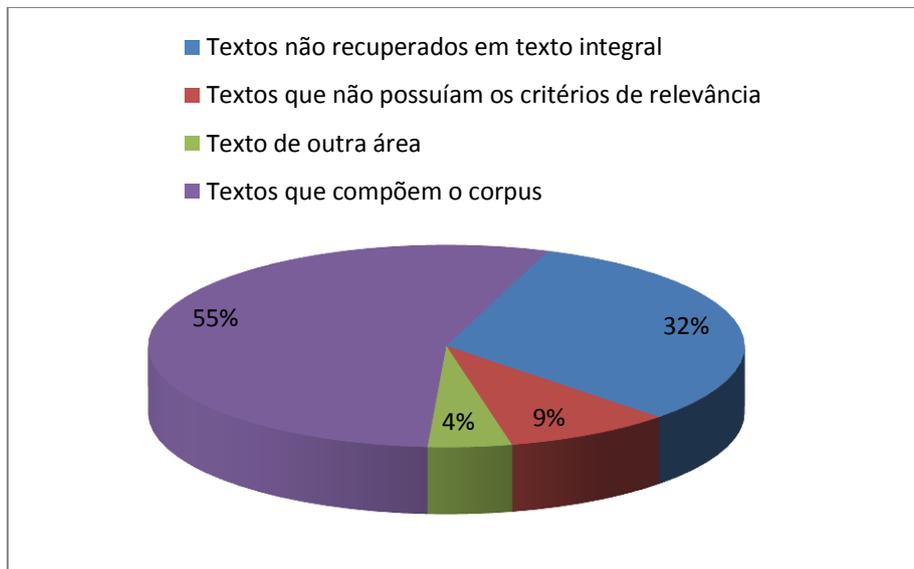
Amostra

Da recolha de dados obtiveram-se 22 textos. Como foi referido, este número resultou da pesquisa em 5 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*), e ainda em dois repositórios institucionais. Do total dos artigos recolhidos, 7 (32%) não foram encontrados na íntegra, pelo facto de as revistas não estarem incluídas no consórcio ou dada a circunstância de os autores terem solicitado o seu embargo. Dos 14 textos restantes, 2 (9%) não possuíam o critério de relevância proposto por Saracevic (1996 *apud* ABAD GARCIA, 2005) e também foram excluídos da análise. No repositório Institucional Estudo Geral foi recuperado um texto sobre o assunto abordado, mas não se classifica como sendo da área, por ser uma tese de doutoramento do campo da Engenharia Informática (4%). No Repositório Científico de Acesso Aberto de Portugal (RCAAP) não foi encontrado qualquer registo. Após esta triagem, a amostra é composta por 12 textos (55%).

O motivo pelo qual não se individualiza o número de artigos encontrados em cada fonte prende-se com o facto de, na maioria, eles se repetirem nas três fontes consideradas. O seu registo, além de não proporcionar uma ideia clara da sua ocorrência concorreria ainda para a entropia na leitura dos dados.

² A realçar que a plataforma B-on agrega também os conteúdos da plataforma EBSCO, pelo que, em certos casos, pode haver uma duplicação de resultados nas pesquisas realizadas em ambos os buscadores.

Fig. 1 – Percentagem do corpus



Descrição do método usado na análise de dados

Após a recolha e a análise da relevância dos dados procedeu-se à interpretação dos textos julgados pertinentes para o estudo, recorrendo-se para tal à análise de conteúdo, segundo a metodologia proposta por Bardin (2008).

Esta opção prende-se com o fato de este 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” partindo de “definições concebidas por pesquisadores, cujas características relativas à forma são homogêneas: argumentos descritivos registados textualmente” (GUIMARÃES e SALES, 2010:5).

O método foi aperfeiçoado por Bardin (2008), que o descreve como um conjunto de métodos aplicáveis a discursos, qualquer que seja a sua natureza, possuindo um ponto principal: a inferência. A autora ressalta ainda que os objetivos dessa análise são dois: a ultrapassagem da incerteza e o enriquecimento da leitura. Refere também que este método apresenta duas funções: a heurística e a administração da prova (hipóteses).

Conforme recomenda por Bardin (2008), é necessário que haja a definição de índices e indicadores, para organizar sistematicamente a análise. De acordo com o contexto metodológico, foram definidas como variáveis de inferência duas unidades de análise: a definição e a função/conclusão dos dois princípios considerados: a exaustividade e a especificidade.

Para melhor identificar essas variáveis foi elaborada uma ficha de análise para cada artigo recuperado, que se dividiu em i) referência do artigo para a sua identificação e localização; ii) resumo e palavras-chave, cujo propósito visa o reconhecimento do escopo do artigo; iii) registo dos extratos do texto nos quais se encontrem evidenciados os termos exaustividade e especificidade; iv) categorias de análise que correspondem ao registo dos

conceitos e das conclusões relativos aos dois princípios considerados, tal como se observa no exemplo que se segue.

Quadro 1: Ficha de análise de conteúdo

Referência: SPARCK-JONES, K. Does indexing exhaustivity matter? <i>JASIST</i> . 24:5 (1973) 313-316.	
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: Não constam	
CONCEITOS ENCONTRADOS	
Exaustividade	Especificidade
<p>“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).</p> <p>“(…) greater exhaustivity will be defined as the assignment of a greater number of index terms to a document” (p. 313).</p> <p>“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).</p> <p>“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).</p> <p>“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).</p> <p>“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”.</p> <p>“Unfortunately, it is impossible to investigate exhaustivity 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>	<p>“Unfortunately, it is impossible to investigate exhaustivity 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>

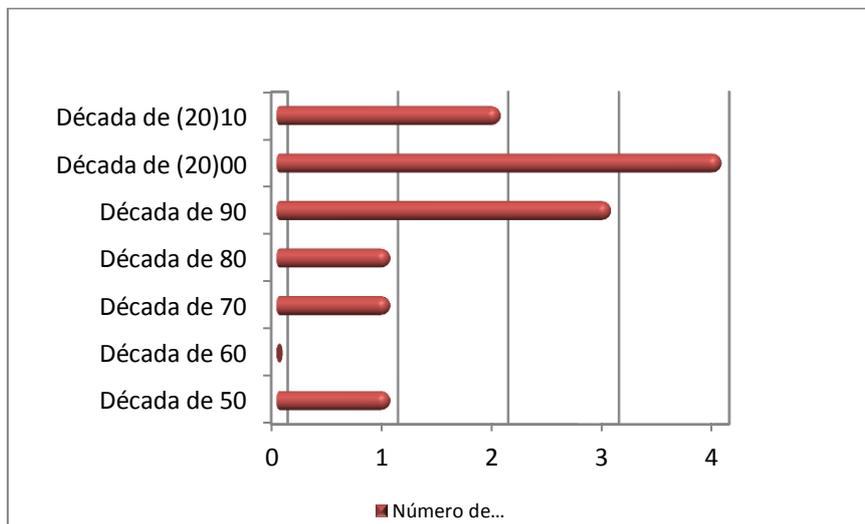
CONCEITOS	Conclusão
Refere-se ao número de temas indexados relativos um documento (p. 313)	Uma boa indexação é aquela que atribui um alto nível de termos ao documento (p. 313). Diferentes fontes de informação - como títulos ou resumos - afetam a exaustividade. A exaustividade não pode ser estudada de maneira isolada, estando correlacionada com a especificidade (p. 313).

Resultados e discussão

As datas de publicação dos artigos revelam que a maioria do *corpus* recuperado (num total de 8 artigos, 67%) foi publicado há mais de dez anos (1952-2003) e a minoria (num total de 4 artigos, 33%) há menos de 10 anos, tendo sido três dos artigos utilizados na pesquisa publicados em 2006, 2008 e 2011, e um em 2012, o que revela uma amplitude e disparidade considerável na publicação sobre o tema (Fig. 2).

Tal situação revela não só obsolescência destes tema na literatura da área, mas também a falta de renovação teórica da mesma.

Fig. 2 - Artigos publicados por década

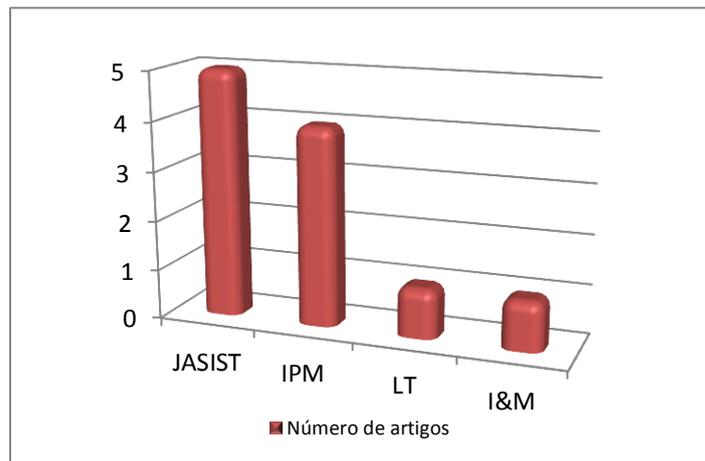


Com base no gráfico apresentado (Fig. 2) verifica-se que na década de 1990 os conceitos de exaustividade e especificidade são pouco expressivos na literatura. Já nos primeiros anos do milênio (2002), ela é mais representativa, situação que pode justificar-se pelo avanço de novas tecnologias e pela visão mais quantitativa destes princípios contudo; observa-se um abrandamento a partir da segunda década do século XXI. Os resultados

gerais mostram que, apesar de se observar um número considerável de artigos sobre este tema nos inícios dos anos 2000, a sua representatividade nas bases consultadas é pouco significativa.

De entre os periódicos que constituem o *corpus* de estudo: *Journal of American Society for Information Science* (JASIST); *Information, Processing and Management*; *Library Trends*; *Information & Management* salientam-se pela predominância destas matérias o *Journal of American Society for Information Science and Technology* e o *Information, Processing and Management*, o que pode revelar a tradição destas publicações relativamente ao tema. Os artigos publicados no *Journal of American Society for Information Science* foram acrescidos na contagem do periódico *Journal of American Society for Information Science and Technology* dado serem o mesmo periódico, apenas com uma alteração na nomenclatura a partir do ano 2000 (JASIS&T) (Fig. 3).

Fig. 3 – Artigos recuperados por publicação



As definições do conceito *exaustividade* expressas nos artigos consultados, referem que este se relaciona com o nível de extensão com o qual um assunto é representado e com a quantidade de termos que são usados para representar os assuntos de um determinado documento. Nas definições e nas conclusões apresentadas observa-se que é objetivo deste princípio facilitar a recuperação da informação e que é desejável a sua correlação com o princípio da especificidade, sendo entendida esta como uma situação ideal.

No que respeita ao conceito *especificidade*, esta noção encontra-se relacionada com o nível de maior ou menor pormenor que é atribuído à representação de um assunto. Pode ainda ler-se que este princípio contribuiu para o nível hierárquico da representação dos assuntos expressos nos documentos e, tal como acontece com o princípio da exaustividade, também ele poderá afetar a recuperação da informação. Seria também uma mais-valia para a indexação se estes dois princípios estivessem correlacionados (Quadro 2).

A partir da reflexão sobre o conceito de ética realizada com base nos autores estudados e da análise de conteúdo dos princípios considerados, efetuada segundo as categorias de

inferência *conceito e conclusão*, foi possível sistematizar um conjunto de argumentos extraídos da literatura consultada que corroboram o valor ético que assumem estes dois princípios no processo de indexação (Quadro 2).

Quadro 2 - Sistematização dos valores éticos a partir da análise de conteúdo do corpus

VALOR	EXPLICAÇÃO TEÓRICA	REFERÊNCIAS
O número de conceitos extraídos de um documento, a forma e o nível de especificidade que são representados são diretamente proporcionais à informação que pretende representar-se.	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; desta forma, dependem sempre da análise conceitual do documento.	JONES, K. S., 1973; HARDIN, A.; WILLET, P., 1980; WOLFRAM, D.; ZHANG, J., 2002; GIVEN, L. M.; OLSON, H. A., 2003.
Os níveis de exaustividade e de especificidade concorrem para o impacto na recuperação informacional.	Quando um documento é indexado a partir de várias facetras - de todos os seus conteúdos e particularidades - as chances de posteriormente ser recuperados são maiores.	SUMNER, R. G., 1995; BURGIN, R., 1995; WOLFRAM, D.; ZHANG, J., 2002; GIVEN, L. M.; OLSON, H. A., 2003.
Elevados níveis de exaustividade concorrem uma maior atribuição de termos a um documento, que se fará de acordo com o serviço e o utilizador.	Se se têm um número maior de termos representados no índice as opções de recuperação serão maiores e há ainda a facilidade de distinguir um termo do outro	WOLFRAM, D.; ZHANG, J., 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 rigor à representação.	GIVEN, L. M.; OLSON, H. A., 2003; KIM, G., 2006.
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, porque a especificidade possui capacidade discriminativa, distinguindo classes ou subclasses de termos.	GIVEN, L. M.; OLSON, H. A., 2003; KIM, G., 2006.
Um conjunto significativo de termos específicos atribuídos a um documento expressa de forma mais precisa os conceitos.	Um grupo de termos específicos combinados é mais representativo que um único termo geral, cria uma intersecção entre os documentos representados	KIM, G., 2006; TAUBE, M., 1952.
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.	JONES, K. S., 1973; BURGIN, R., 1995; GIVEN, L. M.; OLSON, H. A., 2003.
A exaustividade corresponde a um nível elevado de revocação.	Representações exaustivas trazem um maior número de informações pertinentes, fator que concorre para a revocação.	GIVEN, L. M.; OLSON, H. A., 2003; GUILLEMETTE, M. G.; LAROCHE, M.; CADIEUX, J., 2014.

<p>A especificidade corresponde a níveis elevados de precisão.</p>	<p>Se uma informação é específica, ela é o mais detalhada e precisa possível, o que faz com que a especificidade esteja diretamente relacionada com a precisão do sistema</p>	<p>GIVEN, L. M.; OLSON, H. A., 2003.</p>
<p>A exaustividade traz regularidade e densidade à indexação.</p>	<p>Uma indexação exaustiva traz consigo uma quantidade considerável de documentos recuperados.</p>	<p>WOLFRAM, D.; ZHANG, J., 2002.</p>
<p>As relações semânticas entre termos-conceitos são identificadas de forma mais precisa a partir de meios intelectuais.</p>	<p>Embora a indexação automática seja uma realidade, o estabelecimento de relações semânticas entre termos-conceitos é uma operação mais consistente quando é feita por humanos o que destaca ainda mais o papel da especificidade no definir dessas relações.</p>	<p>KIM, G., 2006.</p>

Como foi referido, o principal objetivo deste estudo é demonstrar que os princípios de exaustividade e de especificidade devem ser entendidos não só como valores mensuráveis, mas também como valores éticos, facto que já havia sido evidenciado em pesquisas anteriores como as realizadas por (GUIMARÃES *et al.*, 2008; GUIMARÃES, MILANI e EVANGELISTA, 2015), ideia corroborada no presente estudo, na medida em que ele atesta essa afirmação. Estes princípios podem ser entendidos como valores éticos, porque quando são considerados na seleção e na escolha dos termos, a boa ou a má prática do seu uso podem afetar a recuperação da informação. Deste modo, intuiu-se que a exaustividade e a especificidade atuam de maneira a regular a prática da indexação, com o intuito de buscar o bem-comum, característica intrínseca a um valor ético que se traduz no que é bom e útil para a sociedade.

A partir da análise de conteúdo dos artigos analisados é possível inferir que a exaustividade é definida com base na ocorrência ou quantidade de termos selecionados para representar o conteúdo de um documento, abrangendo de maneira extensa os temas de um documento e as diferentes facetas em que estes se manifestam, concorrendo neste contexto para um conjunto significativo de termos (SUMNER, 1995; BURGIN, 1995; WOLFRAM e ZHANG, 2002; GIVEN e OLSON, 2003).

No que diz respeito à especificidade, o princípio que se encontra menos presente no *corpus* recuperado, quando comparado com a exaustividade, pode ser definido como o nível de detalhe usado na representação de um conceito. Alguns dos artigos referem que a especificidade é necessária no momento da construção das hierarquias conceituais, na medida em que pode estabelecer-se uma rede hierárquica de alto nível semântico, contribuindo deste modo para a precisão da representação dos conceitos de um documento. Ele aprimora a precisão na pesquisa por possuir uma capacidade discriminativa ao distinguir classes de termos e subclasses (GIVEN e OLSON, 2003; KIM, 2006), auxiliando ainda na interceção de documentos, na medida em que a combinação de termos específicos é mais representativa conceitualmente do que um termo geral (KIM, 2006; TAUBE, 1952).

Os resultados deste estudo atestam que a combinação destes dois princípios contribui para o aumento efetivo das pesquisas. A exaustividade contribuirá para um elevado

número de termos (WOLFRAM e ZHANG, 2002), situação que aumentará as probabilidades de pesquisa; a especificidade contribuirá com um elevado nível de especificidade, assegurando, deste modo, a presença de termos precisos e detalhados e potencializando, desta forma, o acesso à informação pertinente, através de um termo específico. A importância da sua complementaridade é por si evidente, devem estes dois princípios ser considerados de um modo isolado, pois o número significativo de termos pertinentes num sistema (exaustividade) concorrerá para taxas expressivas de revocação (informação relevante), se essa informação for específica convergirá indubitavelmente para uma elevada taxa de precisão no momento da recuperação (GIVEN e OLSON, 2003).

Considerações finais

As considerações gerais que se apresentam tiveram como base a análise, a definição e o desenvolvimento do conceito de especificidade e de exaustividade, assim como a sua finalidade. Para tal, foram analisados 12 artigos publicados entre a década de 1950 e a atualidade. A sua análise permitiu examinar e sistematizar os traços que comprovam estes princípios enquanto valores éticos, indo ao encontro da proposição deste estudo.

As ideias apresentadas, mais do que conclusões são apenas inferências, situação que tem a ver com o tema objeto de estudo, caracterizado por um elevado nível de subjetividade e pela aplicação de um método cujo objetivo é chegar a inferências e não a conclusões precisas.

Inferre-se que a exaustividade e a especificidade são essenciais na prática de indexação, na medida em que agregam *valor acrescentado* ao documento, o que concorre para a existência de pesquisas mais consistentes e precisas, contribuindo para a construção de uma ponte sólida entre o documento e o utilizador, que se reflete em resultados da pesquisa mais eficazes.

Outro aspeto relevante do estudo prende-se com o facto de se observar uma proximidade entre a exaustividade e a especificidade motivada, por um lado pelo facto de ambas se encontrarem no contexto intelectual, quando consideradas na abordagem semântica, por outro por se situarem no contexto prático, quando são consideradas na prática da indexação.

Identificou-se também uma relação estreita entre estes dois princípios e os valores éticos. Partindo deste pressuposto e da análise de conteúdo, foi possível chegar a um conjunto de princípios a ter presentes na prática da indexação, para atingir o bem-comum. Também foi possível intuir que a exaustividade e a especificidade concorrem para uma representação efetiva do documento, revelando um compromisso efetivo com a fonte de informação.

Refere-se ainda que os dois conceitos estimulam as relações semânticas entre os conteúdos dos documentos a exaustividade ao contribuir com um conjunto significativo de termos e a especificidade por potenciar e otimizar a categorização. Assim, estes dois princípios relacionados aperfeiçoam de forma muito positiva a recuperação e o acesso à informação. Quando há um compromisso entre os dois o nível de recuperação é consideravelmente maior demonstrando ser, neste sentido, uma estratégia muito eficaz para a organização e recuperação da informação. Tal situação decorre do facto de um

documento representado de forma exaustiva e específica ter uma maior probabilidade de ser recuperado, sendo posteriormente utilizado para na construção de novos conhecimentos e contribuindo para o movimento de produção e organização do conhecimento.

Conclui-se este ponto com um conjunto de inferências sobre a bibliografia consultada que nos parece significativa para um melhor entendimento do estado da questão.

Como foi referido, a maioria do *corpus* deste estudo (67%) teve como base artigos publicados entre os anos de 1952-2003, e uma menor parte (33%) do *corpus* em artigos publicados entre 2006-2012, o que reflete, em termos aritméticos, uma produção de menos de um artigo em cada três anos. Situação idêntica é observada na bibliografia usada para a construção do *corpus* teórico deste estudo, onde a presença destes tem uma abordagem pouco significativa, o que leva a que se infira que há um escasso nível de teorização sobre os mesmos. Os princípios da especificidade, da exaustividade e a questão dos valores éticos em indexação encontram-se pouco desenvolvidos nas obras analisadas; elas revelam também uma certa redundância de ideias nos autores.

A inexistência de bibliografia sobre este tema nos repositórios institucionais portugueses (RCAAP e EG) revela o escasso interesse que ele tem na investigação científica da área, apesar de lhe ser reconhecida importância e necessidade na prática da indexação.

Outro aspeto que vem reforçar o facto de ele ser tão pouco considerado prende-se com a circunstância de a literatura além de escassa, não ser suficientemente atualizada, ser predominantemente em língua inglesa. Comprova-o o facto de não ter sido encontrado nenhum artigo em língua portuguesa, nas bases de dados e repositórios que foram utilizados. Esta observação leva a intuir que possam existir trabalhos e publicações sobre o tema em outros idiomas, mas que não sejam publicados nestas bases por uma questão de barreira linguística, contribuindo assim para a frágil divulgação do mesmo.

Considerando a parca quantidade de artigos recuperados e a fugaz referência feita pelos teóricos da área a este assunto, recomenda-se a elaboração de estudos académicos e científicos sobre este tema, para torná-lo presente e sustentável na prática da indexação.

Referências bibliográficas

ABAD GARCÍA, Maria Francisca
2005 *Evaluación de la calidad de los sistemas de información*. Madrid: Editorial Síntesis, 2005.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS
1992 *NBR 12676, 1992: Documentação: Métodos para a análise de documentos*. Rio de Janeiro: ABNT, 1992.

BARDIN, Laurence
2008 *Análise de conteúdo*. Edições 70: Lisboa, 2008.

BEGHTOL, Clare

2002 A Proposed ethical warrant for global knowledge representation and organization systems. *Journal of Documentation*. London. 58:5 (2002) 507-532.

BEGHTOL, Clare

2005 Ethical decision-making for knowledge representation and organization systems for global use. *Journal of the American Society for Information Science and Technology*. New York. 56:9 (2005) 903-912.

BORKO, Harold; BERNIER, Charles L.

1978 *Indexing concepts and methods*. New York: Academic Press, 1978.

BURGIN, Robert

1995 The Retrieval effectiveness of 5 clustering algorithms as a function of indexing exhaustivity. *Journal of the American Society for Information Science*. New York. 46:8 (1995) 562-572.

CHAUMIER, Jacques

1986 *Análisis y lenguajes documentales: el tratamiento lingüístico de la información documental*. Barcelona: Editorial Mitre, 1986.

DAHLBERG, Ingetraut

1992 Ethics and knowledge organization: in memory of Dr. S. R. Ranganathan in his centenary year. *International Classification*. Frankfurt. 19:1 (1992) 1-2.

EBSCO DISCOVERY SERVICES

2015 *EBSCO Discovery Services* [Em linha]. Massachusetts: EBSCO, 2015. [Consult. 20 out. 2015].

Disponível em: <https://www.ebsco.com/about>

ELSEVIER

2015 *Research platforms* [Em linha]. Amsterdam: ELSEVIER, 2015. [Consult. 20 out. 2015].

Disponível em: <http://www.elsevier.com/about/company-information>

FOSKETT, A. C.

1976 *A Abordagem temática da informação*. São Paulo: Editora Polígono, 1976.

GIL, António Carlos

2002 *Como elaborar projetos de pesquisa*. São Paulo: Atlas, 2002.

GIVEN, Lisa M.; OLSON, Hope A.

2003 Knowledge organization in research: a conceptual model for organizing data. *Library & Information Science Research*. 25:2 (2003) 157-176.

GUILLEMETTE, Manon G.; LAROCHE, Maxime; CADIEUX, Jean

2014 Defining decision making process performance: conceptualization and validation of an index. *Information & Management*. 51 (2014) 618-626.

GUIMARÃES, José Augusto Chaves

2000 O Profissional da informação sob o prisma de sua formação. In VALENTIM, Marta L. P., ed. - *Profissionais da informação: formação, perfil e atuação profissional*. São Paulo: Polis, 2000, p. 53-70.

GUIMARÃES, José Augusto Chaves

2006 Aspectos éticos em organização e representação do conhecimento (ORC): uma reflexão preliminar. In GONZÁLEZ DE GÓMEZ, Maria Nelida; ORRICO, Evelyn, org. - *Políticas de memória e informação: reflexos na organização do conhecimento*. Natal: EDUFRN, 2006, p. 237-264.

GUIMARÃES, José Augusto Chaves

2008 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* [Em linha]. 1:1 (2008) 7-99. [Consult. 12 out. 2015]. Disponível em: <http://eprints.rclis.org/19114/>.

GUIMARÃES, José Augusto Chaves

2009 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, 2009, p. 105-117.

GUIMARAES, José Augusto Chaves; FERNÁNDEZ-MOLINA, Juan Carlos

2003 Los Aspectos éticos de la organización y representación del conocimiento en la revista Knowledge Organization. In FRIAS MONTOYA, J. A.; TRAVIESO, C., org. - *Tendencias de investigación en organización del conocimiento*. Salamanca: Universidad de Salamanca, 2003, p. 809-816.

GUIMARÃES, José Augusto Chaves; MILANI, Suellen; EVANGELISTA, Isadora Victorino

2015 Valores éticos na organização e representação do conhecimento. *Encontros Bibli*. Florianópolis. 20:1 (2015) 19-32.

GUIMARÃES, José Augusto Chaves; SALES, Rodrigo

2010 Análise documental: concepções do universo acadêmico brasileiro em Ciência da Informação. *Datagramazero*. [Rio de Janeiro]. 11:1 (2010).

GUIMARÃES, José Augusto Chaves [et al.]

2008 Ethics in the Knowledge Organization Environment: an overview of values and problems in the LIS literature. In ARSENAULT, C.; TENNIS, Joe T., ed. - *Culture and identity in Knowledge Organization*. Würzburg: Ergon Verlag, 2008, p. 340-346.

HARDIN, Alan. ; WILLET, Peter

1980 Indexing exhaustivity and the computation of similarity matrices. *Journal of the American Society for Information Science*. New York. 31:4 (1980) 298-300.

JONES, Karen Sparck

1973 Does indexing exhaustivity matter? *Journal of the American Society for Information Science*. New York. 24 (1973) 313-316.

KIM, Giyeoung

2006 Relationship between index term specificity and relevance judgment. *Information Processing and Management*. 42 (2006) 1.218-1.229.

LANGRIDGE, Derek Wilton

1989 *Subject analysis: principles and procedures*. London: Bowker-Saur, 1989.

LÓPEZ-HUERTAS, Maria José

2008 Some current research questions in the field of knowledge organization. *Knowledge Organization*. Frankfurt. 35:2/3 (2008) 113-136.

MALIANDI, Ricardo

2004 *Ética: conceptos y problemas*. 3.^a ed. Buenos Aires: Biblos, 2004.

MILANI, Suellen Oliveira

2007 *Aspectos éticos em Organização e Representação do Conhecimento (ORC): uma análise preliminar de valores e problemas a partir da literatura internacional da área* [Texto policopiado]. Marília: [s. n.], 2007.

Trabalho de conclusão de curso - Faculdade de Filosofia da Universidade Estadual Paulista (UNESP), Marília.

OLSON, Hope A.

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

PETTEE, J.

1945 *Subject headings: the history and theory of the alphabetical subject approach to books*. New York: The H. W. Wilson Company, 1945.

PINHO, Fábio Assis

2006 *Aspectos éticos em representação do conhecimento: em busca do diálogo entre Antonio García Gutierrez, Michele Hudon e Clare Beghtol* [Texto policopiado]. Marília: [s. n.], 2006.

Dissertação de mestrado – Universidade Estadual Paulista (UNESP), Marília.

PORTUGAL. Fundação para Ciência e Tecnologia

2009 *B-on: Biblioteca do Conhecimento Online* [Em linha]. [Lisboa]: FCT; FCCN, 2009. [Consult. 12 out. 2015].

Disponível em: <http://www.b-on.pt>.

RODRIGUEZ BRAVO, Blanca [et al.]

2014 Evaluating discovery tools in Portuguese and Spanish academic libraries. In BABIK, W., ed. - *Knowledge organization in the 21st century: between historical patterns and future prospects*. Würzburg: Ergon-Verlag GmbH, 2014, p. 516-523.

ROWLEY, Jennifer

1982 *Abstracting and indexing*. London: Clive Bingley, 1982.

SARACEVIC, Tefko

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

SILVA, Maria dos Remédios; FUJITA, Mariângela Spotti Lopez

2004 A Prática da indexação: análise da evolução de tendências teóricas e metodológicas. *Transinformação*. Campinas. 16:2 (2004) 133-161.

SIMÕES, Maria da Graça

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

SUMNER, Robert G.

1995 Utilizing the age of references to control the exhaustivity of the reference representation in information retrieval. *Information processing and management*. 31:1 (1995) 29-45.

TAUBE, Mortimer

1952 Specificity in subject headings and coordinate indexing. *Library Trends*. Baltimore. 1:2 (1952) 219-223.

WITTY, Francis J.

1973 The Beginnings of indexing and abstracting: some notes towards a history of indexing and abstracting in antiquity and the middle ages. *The Indexer*. Sheffield. 8:4 (1973) 193-198.

WOLFRAM, Dietmar; ZHANG, Jin

2002 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*. New York. 53:11 (2002) 943-952.

Isadora Victorino Evangelista | isadora.biblio@marilia.unesp.br

Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Brasil

Maria da Graça Melo Simões | gsimoes@fl.uc.pt

Universidade de Coimbra – Faculdade de Letras, Portugal

José Augusto Chaves Guimarães | guima@marilia.unesp.br

Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Brasil