

---

# Exaustividade e especificidade na indexação: uma análise de conteúdo como perspectiva de investigação sobre o tema

*Exhaustivity and specificity in indexing: a content analysis as a research perspective on the topic.*

---

**José Augusto Chaves Guimarães (1), Isadora Victorino Evangelista (2)**

(1) Universidade Estadual Paulista, Av. Hygino Muzzi Filho, 737 – Marília/SP, guima@marilia.unesp.br.

(2) isadora.biblio@marilia.unesp.br

## Resumen

A indexação, enquanto processo da representação temática na organização do conhecimento, possui outros dois micro-processos, responsáveis para o aumento de sua eficiência – a exaustividade e a especificidade. O presente artigo tem por objetivo identificar as abordagens conceituais acerca de exaustividade e especificidade na literatura internacional da Ciência da Informação. Para tanto, utilizou-se como método a análise de conteúdo de artigos recuperados pela base de dados SCOPUS, SCIELO, Web of Science e LISA, que continham os termos em suas áreas de maior conteúdo informacional (título, resumo e palavras-chave) publicados em periódicos científicos do campo da Ciência da Informação, sem limite temporal. Foram recuperados um total de 51 artigos, os quais 10 não foram encontrados para leitura na íntegra e 14 foram excluídos por não conterem as variáveis de inferência. Foram estabelecidas como categorias de inferência Conceitos, Características e Objetivo. Os resultados revelam uma preponderância de abordagens conceituais genéricas e "matemáticas" na medida em que se referem a questões de mensuração, sem que se revelem preocupações mais efetivas com aspectos ligados às necessidades dos usuários, objetivos da instituição e a questões éticas. Conclui-se, portanto, que existe a necessidade de maiores estudos na área, para que o tema seja visto de maneira mais ampla

**Palabras clave:** 1. Exaustividade. 2. Especificidade. 3. Indexação. 4. Ética Informacional. 5. Análise de Conteúdo.

## 1. Introdução

O campo da organização e representação do conhecimento materializa-se em duas dimensões complementares – descritiva e temática, a primeira tendo como resultado os catálogos que serão representados em bases de dados, de maneira que a representação chegue o mais próximo possível do documento físico e, a segunda, voltada ao conteúdo do recurso, envolvendo processos como a classificação e a in-

## Abstract

Indexing, as a process of subject representation in knowledge organization, presents two more micro-processes that are responsible for the improvement of its efficiency –exhaustivity and specificity. This paper aims to identify the conceptual approaches to exhaustivity and specificity in the international literature on Information Science. For this, we used the method of content analysis for the articles retrieved from the databases SCOPUS, Scielo, Web of Science and LISA, that include the terms in their main areas of information content (title, abstract and keywords), published in scientific journals in the area of Information Science without time limit. A total of 51 articles were retrieved, where 10 of them could not be accessed and 14 were excluded for not including the inference variables. These variables included: Concepts, Characteristics and Objective. Results show a predominance of generic and "mathematical" conceptual approaches that refer to questions of measuring, without revealing more effective concerns on aspects related to users' needs, goals of the institution and ethical aspects. We conclude that there is a need of more studies in the area.

**Keywords:** Exhaustivity. Specificity. Indexing. Information ethics. Content analysis.

dexação, que visam a resgatar os principais assuntos do documento, traduzindo-os por meio de termos que serão utilizados para futura busca na base de dados.

No caso da indexação, dois procedimentos encontram-se diretamente voltados à questão da eficiência: exaustividade, que diz respeito à amplitude da cobertura da temática do documento e a especificidade, entendida como a precisão e particularidade que será dada à re-

apresentação. Esses dois procedimentos possuem poucos estudos na área, entendidos muitas vezes apenas como questões matemáticas relacionadas à indexação, embora estudos recentes (Milani, 2007; Guimarães et al., 2008; Evangelista, 2013) demonstraram ser entendidos também como valores éticos.

O presente artigo tem por objetivo demonstrar como a literatura internacional da área de Ciência da Informação vem abordando conceitualmente a exaustividade e a especificidade, procedimentos intrínsecos à indexação valendo-se, para tanto, da metodologia de análise de conteúdo (Bardin, 2008).

Desse modo, parte-se de uma abordagem conceitual sobre indexação para se chegar especificamente à exaustividade e à especificidade. Em seguida, adentra-se nas questões éticas da organização do conhecimento como subsídio para a análise que se realiza no material explorado, de modo a identificar e discutir as dimensões conceituais almeçadas.

## 2. Indexação

A organização e representação do conhecimento é uma área de atuação da Ciência da Informação, centrada em tratar o documento para posterior recuperação e disseminação. Para que essa recuperação ocorra, têm-se a necessidade de alguns aportes teóricos que contribuam para tanto, um deles é a indexação.

A principal tarefa da indexação é a descrição temática do documento, empregando termos representativos geralmente selecionados a partir de um vocabulário controlado. De acordo com Lancaster (2004), a indexação é composta por duas etapas: análise conceitual, onde o assunto do documento é definido e a tradução, onde os assuntos são traduzidos para termos, de acordo com uma linguagem documentária. Ainda segundo o autor "(...) os termos atribuídos pelo indexador servem como pontos de acesso mediante os quais um item é localizado e recuperado, durante uma busca por assunto." (Lancaster, 2004, p. 06).

Hjørland define assunto como sendo "potenciais epistemológicos ou informativos dos documentos" e ainda ressalta que "(...) a melhor análise temática é a que faz o melhor prognóstico quanto ao uso futuro do documento" (Hjørland, 2011, p. 776).

Para Leiva (2012), "(...) a indexação gera palavras-chave, índices ou os cabeçalhos de assunto de um documento" (Leiva, 2012, p. 31). O autor ainda sistematiza algumas definições da

literatura para o termo indexação, como pode ser visto a seguir.

Analisar o conteúdo informacional dos registros do conhecimento e expressar o conteúdo informacional na linguagem do sistema de indexação. (Borko; Bernier, 1978, p. 8).

Representação pelos elementos de uma linguagem documentária ou natural, das noções resultantes da análise do conteúdo de um documento para facilitar sua localização. (NF Z 47 – 102 1978)

Descrever o conteúdo de documentos ou demandas documentárias para possibilitar a elaboração de estratégias de recuperação mediante conceitos ou assuntos. (García Gutiérrez (1984, p. 105)

Ação de descrever ou identificar um documento em relação ao conteúdo. (ISO 5963 – 1985)

Identificar informação numa entidade de conhecimento (que seja texto ou não) e organizá-la para que esteja disponível num sistema de recuperação. (Cleveland, D. B.; Cleveland, A. D., 2001, p. 97)

Determinar o assunto temático dos documentos e expressar em índices (por exemplo, descritores, cabeçalhos de assunto, números de chamada, códigos de classificação ou índice) para tornar possível a recuperação temática. (Mai, 2005, p. 599)." (Leiva, 2012, p.68).

Ou seja, a indexação visa a reduzir o documento a um grupo de termos, para que o armazenamento e a busca em bases de dados sejam facilitados e deve satisfazer as necessidades informacionais dos usuários, o que é buscado deve corresponder com o que for encontrado.

Considerando que os documentos irão responder a questões internas dos usuários, é essencial que a análise de assunto esteja preparada para as perguntas da comunidade em que a unidade informacional está inserida. Ou seja, o assunto do documento deve estar diretamente ligado ao objetivo daquela unidade (Hjørland, 2011).

Para Lancaster (2004), não há uma indexação "correta", uma indexação eficiente é aquela que além de levar em conta os assuntos de determinado documento, também reflete sobre o porquê dele ser importante para determinada comunidade. Comunidades com interesses diferentes buscarão através de termos diferentes e, portanto, não há apenas um conjunto de termos para aquele documento.

A indexação utiliza-se de processos que contribuem para melhorar sua eficácia, no presente artigo, busca-se focar em dois deles: a exaustividade e a especificidade.

### 3. Exaustividade e especificidade

A exaustividade caracteriza-se, basicamente, pelo nível de abrangência que a indexação vai optar por atribuir. A exaustividade concerne ao número de termos que serão utilizados para representar um documento e qual a extensão do conteúdo que esses termos irão cobrir.

Para Lancaster (2004), a quantidade de termos escolhidos para serem representativos daquele documento influencia o quanto o documento tornar-se-á acessível, e como resultado, recuperado, considerando que no momento da busca, a proporção de termos indexados/termos buscados será proporcionalmente maior. O autor ainda ressalta que ainda irá ocorrer o aumento de características que distingam aquele documento de outros.

Fosket (1973) divide a exaustividade em duas partes, a primeira dizendo respeito ao conteúdo global de um documento e a segunda, aos subtemas que correspondem a ele.

De acordo com Fujita (2004), quando optar pela exaustividade, o indexador deverá atentar-se ao número de termos que irá extrair de cada recurso, os selecionando baseados sempre no critério da abrangência.

Leiva (2012) diz que, embora não se tenha claro qual a quantidade exata de termos a serem atribuídos aos documentos, parece lógico que documentos como resumos gerem menos termos que textos completos, por exemplo.

Já a especificidade, consolida-se por ser o grau de particularidade dado à representação ou o nível de precisão que iremos utilizar na extensão de cobertura dos documentos. Decide-se quão específico serão os termos atribuídos no momento da representação.

A especificidade é atribuída prezando pelo interesse dos grupos que utilizam daquele sistema de busca e depende do nível de especificidade dos usuários (Lancaster, 2004). Ainda de acordo com o autor, enquanto a exaustividade diz respeito ao âmbito de cobertura, a especificidade refere-se à profundidade de tratamento do conteúdo. Estabelece-se a exaustividade como uma decisão da política de indexação, enquanto a especificidade é uma propriedade do vocabulário adotado na indexação (Lancaster, 2004, p. 203).

Foskett (1973), define a especificidade como sendo também a extensão dada ao documento, mas nesse caso, sendo relativa à extensão da precisão ao se especificar um assunto. O autor ainda ressalta que, quanto maior a especificidade, maior será a possibilidade de documentos

relevantes serem encontrados. Com uma precisão limitada, conseqüentemente, a relevância dos documentos será baixa.

Segundo Fujita (2004), a especificidade depende de uma decisão estabelecida pelo sistema: se essa será ou não uma recomendação dada ao indexador; e caso seja, o indexador deve realizar a leitura documental tendo em mente o nível de especificidade exigida pelo sistema. Ainda segundo a autora, “a especificidade está relacionada ao nível de abrangência que a biblioteca e a linguagem documentária permitem especificar os conceitos identificados documento.” (Fujita, 2004, p. 85).

Fujita (2004) ainda demonstra, com o seguinte exemplo, como os dois processos – exaustividade e especificidade – podem trabalhar de maneira conjunta adequadamente: caso um livro trate especificamente do tema “tilápia, pacus e lambaris”, e o indexador optar também pela exaustividade, o usuário que buscar pelo termo “peixes” também encontrará o mesmo documento.

Tem-se, portanto, que esses são dois processos complementares, que dizem respeito diretamente à abrangência e à profundidade. Estudos realizados na área (Milani, 2007; Guimarães et al, 2008; Evangelista, 2013), demonstram que esses dois aspectos da indexação, podem ser vistos como valores éticos.

### 4. Ética em representação do conhecimento

Essa dimensão das atividades de indexação faz com que se remonte à Ética, enquanto ramo da Filosofia, nascida na Grécia Antiga, cuja reflexão está centrada no estudo das normas, princípios e condutas que orientam o homem na vida em sociedade, a partir da busca por um bem-estar individual e social. Dela decorre a Moral, que situa as questões éticas a uma dada realidade de costumes e valores, no tempo e no espaço (Guimarães, 2007).

A questão ética, apenas mais recentemente vem sendo objeto de pesquisa em organização do conhecimento, no campo da Ciência da Informação, ainda que, há mais de duas décadas, Ingetraut Dalghberg (1992), fundadora da ISKO, já enunciava fatores alarmantes que levavam a tal necessidade. Passada mais de uma década, López-Huertas (2008), ao analisar as tendências de pesquisa em organização do conhecimento na atualidade, ainda apontava a falta de suficiente zelo com o assunto.

Guimarães (2000), referindo-se ao universo da organização e representação do conhecimento,

mais especificamente no que se refere ao aspecto temático destaca cinco níveis de compromissos éticos: com as necessidades informacionais do usuário, com os objetivos da organização, com a especificidade e a natureza da própria informação disponibilizada, com o bom nome da profissão e com o próprio universo de valores do indexador. Observam-se, assim, cinco níveis de valores a serem considerados: os do usuário, os da organização, os inerentes ao conteúdo informacional, os da profissão coletivamente considerada e os do profissional indexador em sua individualidade. Assim, a questão dos valores é central na discussão das questões éticas, pois revela um universo de normas de conduta que uma dada sociedade considera importantes para atingir o bem-comum. Da negativa desses valores nascem os denominados problemas éticos, que a sociedade considera como impeditivos para que se atinja o bem-comum.

Ao analisarem a literatura internacional da área no período de 1995-2004, relativamente às questões éticas em organização e representação do conhecimento, Guimarães et al. (2008) depararam-se com um núcleo axiológico composto por três níveis de valores: a) valores maiores (ou supervalores) que permeiam toda a atividade informacional, b) valores antes havidos como requisitos profissionais, na medida em que integram a essência do fazer profissional na área; c) valores antes havidos como meras medidas de ORC, mas que hoje se integram ao universo axiológico da área, como Precisão (também entendido como Especificidade). Decorrendo da negativa dos valores, surgem os problemas que, de acordo com Guimarães et al. (2008), podem ser categorizados em dois contextos: os Problemas que permeiam o mundo atual; e os Problemas que afetam diretamente as atividades de ORC, quando do exercício profissional.

Milani (2007) e Evangelista (2013) analisaram, respectivamente aos períodos de 1995- 2004, e de 1990-1994; 2005-2010, a presença de valores e problemas éticos na literatura internacional de organização e representação do conhecimento, identificando um conjunto de 54 valores éticos, com a preponderância de Privacidade, Garantia cultural, Precisão (relativo à especificidade), Autoria e Exaustividade. Relativamente aos problemas, obteve-se um conjunto de 44, sendo predominantes: Vigilância, Negligência, Censura Direcionamento informacional, Falta de garantia cultural e Má representação.

Viu-se, portanto, a necessidade de realizar a análise de conteúdo de maneira que se demonstre o que a literatura internacional da área

vem dizendo a respeito desses dois processos e em que âmbito eles são entendidos.

## 5. Material e métodos

A pesquisa demonstra-se ser qualitativa, uma vez que se utiliza inferências e interpretações para demonstrar como se estabelecem dois conceitos, de acordo com a literatura.

A análise de conteúdo caracteriza-se por 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 registrados textualmente” (Guimarães & Sales, 2010, p. 05).

O método foi aperfeiçoado por Bardin (2003), que o descreve como um conjunto de métodos aplicáveis a discursos, qualquer que seja sua natureza, possuindo um ponto principal: a inferência. A autora ainda ressalta, que o objetivo dessa análise são dois: a ultrapassagem da incerteza e o enriquecimento da leitura, e ainda possui duas funções: a função heurística e a administração da prova (hipóteses).

Primeiramente, é necessário que se realize uma fase de pré-análise, que de acordo com Salles (2008) caracteriza-se 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).

Foi definido como material para posterior aplicação da análise de conteúdo, artigos que contenham os termos exaustividade e especificidade em inglês – exhaustivity e specificity, respectivamente -, recuperados pelas bases de dados Scopus, SCIELO, Web of Science e LISA, através do campo de busca que compreende o título do artigo, o resumo e as palavras-chave. Os termos foram procurados separadamente e não se estabeleceu limite de tempo de publicação, mas foram selecionados apenas os artigos publicados em periódicos da área de Ciência da Informação.

Para criação do corpus de pesquisa, Bardin (2003) sugere quatro regras principais: regra de exaustividade, regra de representividade, regra da homogeneidade e regra da pertinência. Decidiu-se não utilizar-se da regra da exaustividade, visto que essa define-se como a não-seleção dos documentos encontrados.

O processo de coleta de dados gerou dois corpora de pesquisa, a saber:

Corpus 1: conjunto de todos os artigos recuperados pela base de dados. Foram recuperados 51 artigos, os quais 10 não foram encontrados na íntegra. Realizou-se a leitura dos resumos dos 41 recuperados, para saber se constituiriam o Corpus 2.

Corpus 2: textos selecionados por tratarem do assunto no âmbito da Ciência da Informação e ainda, conterem as variáveis de inferência definidas, resultando em um corpus de 27 artigos. Para tanto, leu-se os textos na íntegra.

Conforme recomendado por Bardin (2003), passa-se para a fase de elaboração de índices e indicadores. Para Salles (2008) essa é 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” (Salles, 2008, p. 106). Foi definido como variáveis de inferência três concepções: conceitos (as definições ou as funções dos termos), características (quais os atributos que caracterizam cada um dos termos) e objetivos (qual a intenção de cada um dos termos).

Posteriormente, passa-se a fase das inferências, onde se procura sintetizar em sentenças, o conteúdo dos documentos. A análise gerou dois quadros – um correspondente ao termo exaustividade e outro, correspondente ao segundo termo, especificidade – que serão explorados na seção a seguir.

## 6. Análise dos resultados

A seguir, procurou-se definir de forma resumida, o que os autores dos artigos do corpus de pesquisa, interpretam sobre esses dois processos da indexação. Em um primeiro momento, procura-se demonstrar as principais inferências dadas ao termo exaustividade:

No que se refere a categoria *Conceito*, duas dimensões se colocam: uma ligada à análise (matemática) de distribuição de ocorrência de dados e outra ligada à cobertura de termos utilizados na indexação.

Relativamente às *Catacterísticas*, foram identificados os seguintes atributos: Seletividade, Representatividade, Abrangência, Relação direta com a revocação, Relação inversa com a precisão, Gradatividade, Consistência, e Quantificação.

Em termos de *Objetivos*, destacam-se o estabelecimento de relações entre os documentos, a redução da densidade dos documentos encon-

trados, o fornecimento de subsídios à distinção entre documentos na recuperação e a promoção da relevância.

O que pode se perceber, dessa análise, é que, de maneira geral, a exaustividade é definida como a abrangência que o sistema permite indexar um documento e que está diretamente ligada à revocação (maior recuperação de documentos) do sistema. Também devemos chamar a atenção que, muitas vezes, ela é vista apenas como uma expressão matemática, não tomando consideração por elementos como as necessidades dos usuários, os objetivos da instituição e questões éticas. Outro fator importante é a constatação por alguns atores de que ela está inversamente relacionada à precisão e especificidade do sistema.

O resultado da análise de conteúdo da literatura sobre especificidade revela, em termos de *Conceito*, uma referência aos documentos diretamente ligados àquela temática e ao nível de minúcia dada a um documento, quando representado por meio de termos.

As *Características* da especificidade se revelam por meio do detalhamento, da profundidade, e de sua relação direta com a precisão e inversa à revocação.

No que se refere aos *Objetivos*, a especificidade se revela como uma estratégia para uma organização e representação do conhecimento efetiva por meio da análise e seleção de sentenças para posterior categorização.

De maneira geral, podemos identificar, na especificidade, a particularidade que será dada à indexação, na hora da escolha dos termos.

Podemos estabelecer algumas semelhanças e algumas diferenças entre os termos. Por semelhanças, podemos perceber que fica claro que os dois tratam de características do sistema. A questão matemática também está sempre presente, visto que muitas vezes, os processos são vistos apenas pensando na quantidade de termos que eles irão gerar. Os autores também evidenciam que os dois métodos ainda causam potenciais problemas em sua aplicação, fator que pode se dever aos poucos estudos na área.

Como diferenças, o principal aspecto é a desigualdade entre um termo estar totalmente relacionado à alta revocação enquanto outro, completamente relacionado à alta precisão.

## 7. Conclusões

A abordagem do processo de indexação a partir de dois dos micro-processos que a compõem -

a exaustividade e a especificidade – e a análise dessa temática a partir do aspecto conceitual na literatura especializada, com aporte da análise de conteúdo evidenciou que, exaustividade e especificidade são entendidos como atributos do sistema, no mais das vezes ligados à questão quantitativa (quantidade de termos que deve ser atribuído ao documento) sem maiores reflexões sobre as consequências dessas ações ou, em última instância, os valores envolvidos que integram a sua dimensão ética.

Isso leva à necessidade de uma ampliação de horizontes rumo a uma abordagem mais social (Hjørland, 2002) que vá além dos estritos limites do sistema para ir ao encontro do papel mediador que a indexação deve realizar entre o contexto de produção da informação (o conteúdo do documento), e o contexto de uso dessa informação (necessidades do usuário), aproximando-se do que Antonio García Gutiérrez (2002) denomina como “ética transcultural de mediação”.

**Agradecimentos:** Agradece-se o apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP (proc. 2015/06278-6).

## Referências

- Bardin, L. (2009) *Análise de conteúdo*. Tradução de Luiz Antero Reto e Augusto Pinheiro. Lisboa: Edições 70..
- Dahlberg, I. (1992). Ethics and knowledge organization: in memory of Dr. S. R. Ranganathan in his centenary year. *International Classification*, 19:1, p. 1-2.
- Foskett, A. C. (1973) *A abordagem temática da informação*. Tradução de Antonio Agenor Briquet de Lemos. São Paulo: Polígono.
- Evangelista, I. V. (2013) Ética em representação do conhecimento: uma análise de valores e problemas a partir da literatura internacional da área (1991-2010). Trabalho de Conclusão de Curso (Bacharelado em Biblioteconomia) - Universidade Estadual Paulista Júlio de Mesquita Filho, Marília.
- García Gutiérrez, A. (2002) Knowledge organization from a “culture of the border” towards a transcultural ethics of mediation. En: López-Huertas, M. J. (ed.). (2002) *Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries*. Würzburg: ERGON, p. 516-522.
- Guimarães, J. A. C. et al. (2008) Ethics in the Knowledge Organization Environment: an overview of values and problems in the LIS literature. En: Arsenault, C.; Tennis, J. T. (eds.). (2008) *Cultural and Identity in Knowledge Organization*. Würzburg: ERGON, p. 340-346
- Guimarães, J. A. C. (2000) O profissional da informação sob o prisma de sua formação. En: VALENTIM, M. L. P. (ed.). *Profissionais da Informação: formação, perfil e atuação profissional*. São Paulo: Polis, p. 53-70.
- Guimarães, J. A. C.; Sales, R. (2010) Análise documental: concepções do universo acadêmico brasileiro em Ciência da Informação. *Datagrama*, 11:1.
- Hjørland, B. (2002) Domain analysis in information science. Eleven approaches - traditional as well as innovative. *Journal of the American Society for Documentation*, 58:4, p. 422-462.
- Hjørland, B. (2001) Toward a theory of aboutness, subject, topicality, theme, domain, field, content...and relevance. *Journal of the American Society for Information Science and Technology*, 52: 09, p. 724 – 778.
- Fujita, M. S. L. (2009). *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.
- Lancaster, F. W. (2004) *Indexação e resumos: teoria e prática*. Tradução de Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos.
- Leiva, I. G. (2012) Aspectos conceituais da indexação. In: LEIVA, I. G.; FUJITA, M. S. L. *Política de indexação*. São Paulo: Cultura Acadêmica, p. 31 – 106.
- López-huertas, M. J. (2008) Some current research questions in the field of knowledge organization. *Knowledge Organization*, 35: 2/3, p. 113-136.
- Sales, R. (2008) *Tesauros e ontologias sob a luz da teoria comunicativa da terminologia*. Dissertação Mestrado em Ciência da Informação. Forianópolis: PGCIN/ UFSC.