

Análise de assunto a partir de uma perspectiva histórica do ARIST

Gercina Ângela Borém de Oliveira Lima

Pós-Doutorado pela Universidade de São Paulo (USP) - São Paulo, SP - Brasil. Pós-Doutorado pela Universidad Carlos III de Madrid (UC3M) - Espanha. Doutora em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil. Professora da Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/3183050056105009>

E-mail: limagercina@gmail.com

Benildes Coura Moreira dos Santos Maculan

Doutora em Ciência da Informação pela Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil. Professora da Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/5336218259257800>

E-mail: benildes@gmail.com

Recebido em: 10/07/2014. Aprovado em: 22/06/2015. Publicado em: 15/01/2016.

Resumo

A análise de assunto como primeira etapa da indexação tem recebido pouca atenção como potencial temática para pesquisas teóricas no campo da biblioteconomia e da ciência da informação. Este artigo apresenta um estudo bibliométrico, buscando analisar os fatores que influenciam a suposta carência. Também procura entender por que as pesquisas dão mais ênfase aos estudos relacionados à segunda etapa da indexação: a tradução. Foi utilizado o método da análise de conteúdo, com a técnica categorial temática. O recorte foi realizado dentro do universo das publicações do Annual Review of Information Science and Technology (ARIST), e é composto pelas revisões dos anos de 1977, 1982, 1986 e 1989. Os resultados encontrados apontam para a veracidade do pressuposto de que poucas pesquisas teóricas têm sido desenvolvidas sobre a primeira etapa da indexação – a análise de assunto – e isso pode ser explicado por meio da ideia de que a análise de assunto é considerada uma etapa mais intelectual, que sempre envolve subjetividade, o que pode dificultar o estudo e a sistematização de seus procedimentos.

Palavras-chave: Análise de assunto. Indexação. Estudo bibliométrico. ARIST.

Subject analysis from a historical perspective of ARIST

Abstract

Subject Analysis, the first step of Indexing, has received little attention as a potential subject for theoretical research in the field of Library and Information Science. Departing from this assumption, this paper presents a bibliometric study analyzing the factors related to this shortcoming. It also seeks to understand why research in the field emphasize more the second stage of indexing, that is the translation. The review of the literature encompasses the ARIST (Annual Review of Information Science and Technology) publications in the years 1977, 1982, 1986 and 1989. The results point to the fact that the Subject Analysis is considered a more intellectual and subjective phase what makes it more difficult to systematize its procedures.

Keywords: Subject analysis. Indexing. Bibliometric study. ARIST.

Análisis de asunto a partir de una perspectiva histórica del ARIST

Resumen

El análisis de asunto como primera etapa de la indización ha recibido poca atención como potencial temático para investigaciones teóricas en el campo de la bibliotecología y de la ciencia de la información. Este trabajo presenta un estudio bibliométrico, buscando analizar los factores que influyen la supuesta falta. También busca comprender porque las investigaciones dan mayor enfoque a los estudios relacionados a la segunda etapa de la indización: la traducción. Usa el método de análisis de contenido, con la técnica categorial temática. El recorte fue realizado dentro del universo de las publicaciones del Annual Review of Information Science and Technology (ARIST), compuesto por las revisiones de los años de 1977, 1982, 1986 y 1989. Los resultados encontrados apuntan para la veracidad de la suposición de que pocas investigaciones teóricas han sido desarrolladas sobre la primera etapa de la indización – el análisis de asunto – y eso puede ser explicado a través de la idea de que el análisis de asunto es considerada una etapa más intelectual, que siempre envuelve subjetividad, lo que puede dificultar el estudio y la sistematización de sus procedimientos.

Palabras clave: Análisis de asunto. Indización. Estudio bibliométrico. ARIST.

INTRODUÇÃO

A organização da informação supõe um processo desdobrado em dois subprocessos solidários: o armazenamento e a recuperação. Nesse sentido, compreende-se que a entrada de dados inconsistentes em um sistema de recuperação da informação (SRI) implicará a saída de dados também inconsistentes. No processo de entrada de informações, são ativados mecanismos cognitivos que influem na qualidade dos dados armazenados, porque são responsáveis pela maneira como usamos nossa mente para realizar abstrações. Como consequência, a qualidade das informações na saída do SRI também será influenciada.

A indexação é um importante processo dentro de um SRI, processo no qual se espera que o indexador leia o documento e faça a distinção entre informação relevante e periférica, para melhor representá-lo, para a posterior recuperação.

Nos últimos anos, muitas pesquisas têm sido orientadas para estudar e entender o processo de transferência da informação. Muitas dessas pesquisas têm sido relacionadas à interação entre o SRI e o usuário, talvez entendendo que esse é o ponto mais vulnerável de todo o processo.

A indexação é o processo dentro do SRI que tem a proposta de determinar o assunto dos documentos, traduzir esses assuntos e representá-los através de uma linguagem de indexação, tendo sempre em mente as necessidades do usuário

Soergel (1985) é um dos autores que mais defende o ponto de vista de que a indexação deve ser feita pensando-se na busca que pode ser realizada pelo usuário, modelo chamado de indexação orientada para o usuário. Nesse sentido, considera-se que a indexação, e os seus diferentes métodos, constituem a chave para a melhoria da eficácia e da eficiência do SRI.

Entretanto, como não há um método único que direcione a extração de conceitos relevantes de um documento, na prática, muitas vezes o indexador se orienta a partir dos termos que compõem o instrumento de tradução, que deveria ser utilizado somente na última etapa da indexação. Este procedimento, não cumprir a etapa de análise de assunto adequadamente, pode gerar ruído na recuperação, pois se pode afirmar que o processo de análise de assunto, primeira etapa da indexação, é o coração de um SRI.

Acreditando que poucas pesquisas teóricas têm sido desenvolvidas sobre a primeira etapa da indexação, a análise de assunto, este estudo objetiva lançar um olhar sobre a literatura existente. O propósito é procurar entender por que tal etapa é tão pouco estudada. Ao mesmo tempo, vamos analisar os fatores desta carência de pesquisas, e por que sempre dão maior ênfase aos estudos sobre a etapa da tradução e sobre linguagens de indexação. Para realizar este estudo sobre o processo análise de assunto, foi selecionado um corpus de quatro capítulos de revisões de literatura do *Annual Review of Information Science and Technology* (ARIST), dos anos de 1977, 1982, 1986 e 1989.

O PROCESSO DE INDEXAÇÃO DE ASSUNTO

Jens-Eric Mai (2001) ressalta que o processo de indexação é algumas vezes apontado, na literatura, como tendo duas, três ou mesmo quatro etapas. Para autores como Lancaster (1998) e Fugmann (1993), esse processo é realizado em duas etapas: análise de assunto e tradução. Já os autores Farrow (1991), Taylor (1994) e Petersen (1994) adicionam mais uma etapa: apesar de também considerarem como primeira a análise de assunto, incluem uma etapa intermediária, que reformula o assunto em uma frase de indexação na linguagem natural, para depois fazer a tradução. Os autores Hutchins (1978), Frohmann (1990) e Blair (1990) também comungam da sugestão das três etapas no processo de indexação e salientam, ainda, que a literatura sobre indexação se concentra mais na terceira etapa, a da tradução, sendo incipiente a literatura referente à primeira e à segunda etapas. Isso pode ser explicado por meio da ideia de que a análise de assunto é considerada uma etapa mais intelectual, que envolve analisar e extrair, dos documentos, termos que possam representar os assuntos mais relevantes. Essa atividade sempre envolve subjetividade, o que pode dificultar a sistematização de seus procedimentos.

A Norma ABNT 12676, de 1992, também sugere que o processo de indexação seja realizado em três etapas: a primeira é a da análise do documento para identificação de seu conteúdo informacional; a segunda a de seleção dos termos que possam representar a informação; e a terceira a de tradução dos conceitos nos termos de uma linguagem de indexação.

Diferentemente das sugestões anteriores, Chu e O'Brien (1993) e Langridge (1989) consideram que esse processo possui quatro etapas, e apontam duas para a tradução. Para as autoras, a indexação se expressa assim:

1. determinação do assunto do documento;
2. reformulação do assunto em frases de indexação;
3. tradução dessas frases de indexação de acordo com o vocabulário da linguagem de indexação;
4. tradução do assunto de acordo com a linguagem de indexação.

No entanto, o processo dividido em quatro etapas não é o que a literatura aponta e estuda comumente.

A indexação envolve ações interpretativas complexas, o que justifica, de certo modo, a ausência de metodologias que possam dar consistência ao processo. Para solucionar esse problema, algumas pesquisas em outras áreas têm surgido, desde a década de 1970.

Análise de assunto, como primeira etapa indexação do processo, é um tema importante nos estudos sobre a indexação, pois os SRIs sempre procuram melhorar o acesso à informação. Entretanto, apesar de ser um tema central dentro da área da biblioteconomia e ciência da informação, nota-se que tem sido difícil encontrar revisões de literatura que cubram extensivamente as discussões sobre as teorias de indexação, principalmente sobre a primeira etapa do processo, a análise de assunto. Isso pode ser justificado pela complexidade que essa atividade exige do ser humano, pois requer conhecimentos cognitivos, lógicos e linguísticos, e, conseqüentemente, exige do professor/pesquisador conhecimentos interdisciplinares para avançar em suas pesquisas.

Nesse sentido, Nohr (1991) sugere que se busque suporte em três áreas: na filosofia, para responder às questões sobre a organização do conhecimento; em fundamentos conceituais, para representar, ordenar e relacionar semanticamente os conceitos dentro da proposta da análise conceitual; e na linguística, para tratar os aspectos sintáticos e morfológicos da linguagem. Os referenciais linguísticos como suportes teóricos para a indexação também são sugeridos por Navarro (1988) e Blair (1990). Além da linguística, a abordagem dos aspectos cognitivos e lógicos dos processos de representação é sugerida por Daniels (1986), Pinto Molina (1994) e Allen (1991).

Esta temática foi recorrente em quase todos os volumes do ARIST. Esse *Annual* foi uma publicação de referência dentro da comunidade científica da ciência da informação, fornecendo revisões analíticas de temas significativos dentro da área de ciência da informação e tecnologia, tendo como principal objetivo tornar acessíveis à comunidade científica as principais tendências e avanços do campo.

O ARIST foi publicado anualmente, em nome da Associação para a Ciência da Informação e Tecnologia (ASIS&T), de 1966 a 2011.

O TEMA DA ANÁLISE DE ASSUNTO

Para Foskett (1973, p. 34), “a análise de assunto é a operação chave da indexação, que é a decisão sobre o que o documento é, e ainda é menos discutida e a menos reduzível a regras”.

Análise de assunto como uma atividade intelectual tem como objetivo responder à pergunta “do que trata?”, a respeito de determinado documento. É um processo intelectual que envolve atividades cognitivas na compreensão do texto e a composição da representação do documento. Durante o processo de análise de assunto, a memória é ativada constantemente, buscando o aprendizado adquirido na experiência e no conhecimento prévio do leitor indexador; da mesma forma ocorre no momento de realização da estratégia de busca, seja ela realizada pelo bibliotecário de referência ou pelo usuário. Além disso, o bibliotecário indexador, na realização dessa atividade intelectual, é influenciado pela sua formação e experiência. Sendo o indexador um ser humano, a subjetividade é inerente a ele, por isto, não se pode ter ideias de senso comum no momento da interpretação e identificação do assunto no texto, pois é necessário que se tenha sempre em mente a responsabilidade de mediador das ideias do autor/criador para o usuário, sem ignorar a política de indexação adotada pelo sistema.

Sabe-se que o processo cognitivo está intrinsecamente ligado ao processo de identificação da ideia principal apontada em um texto. Para Fujita,

[...] o processo de análise de assunto reveste-se (*sic*) de uma subjetividade característica, dadas as circunstâncias e elementos envolvidos, pois, a partir da leitura do documento pelo indexador, é realizado um processo de comunicação interativo entre três variáveis: leitor, texto e contexto. Cada uma dessas variáveis estará sujeita a diferentes condições, mas é o indexador como leitor a variável mais influente nessa interação para análise de assunto, porque precisa realizar a compreensão da leitura mediante sua cognição (FUJITA, 2003, p. 69).

Por isso, se um SRI quiser aprimorar seus serviços, é necessário dispor de profissionais com conhecimentos teóricos e práticos de tratamento da informação. Porém, o processo de análise de assunto, primeira etapa da indexação, é feito, muitas vezes, sem o conhecimento de regras ou diretrizes, por falta de conhecimento do próprio indexador, ou por falta de propostas metodológicas que possam ser testadas e entendidas. Além disso, a análise de assunto não pode ser feita se desconsiderando a cultura, o ambiente e o contexto no qual ela é realizada, incluindo o conhecimento individual do indexador e os interesses coletivos do SRI como um todo. Essa análise deve ser alcançada de maneira precisa e objetiva, sempre se considerando a necessidade da comunidade de usuários. Tais requisitos tornam esse processo desafiador para os indexadores, exigindo deles procedimentos subjetivos.

Outra variável que influencia a análise de assunto é a tipologia de textual. Segundo Kobashi (1994, p. 112), o texto, em sentido mais amplo, designa uma unidade de comunicação organizada sintagmaticamente e dotada de coesão e coerência. Quanto mais estruturado for um texto, mais fácil será sua representação semântica. O texto é constituído de uma estrutura global, composto de macroestruturas parciais que se organizam em macroestruturas mais específicas, que, quando organizadas, evidenciam a organização temática (superestrutura) do texto. Essa organização é feita através da sintaxe, da semântica e da pragmática, que auxiliam na análise das partes mais relevantes do texto.

Deve-se esclarecer que mesmo o termo “análise de assunto” pode ser usado por abordagens diferentes dentro da área de biblioteconomia e ciência da informação. Mai (2000) apresenta, em seu artigo intitulado *Deconstructing the indexing process*, duas diferentes abordagens: uma como uma área do conhecimento que estuda a construção e uso das linguagens de indexação e sistemas de classificação; outra como o processo de análise conceitual, primeira etapa da indexação de um documento.

O autor salienta, ainda, que o estudo da área tem recebido mais atenção dos pesquisadores do que o processo da análise conceitual. Isso pode ser

justificado pelo fato de tal etapa ser considerada a mais difícil de ser entendida e realizada, por necessitar dos conhecimentos do indexador sobre o domínio no qual seu sistema está inserido, sobre a tipologia textual e sobre o usuário do sistema.

Outra observação que deve ser feita se refere às diversas denominações sob as quais o processo de análise de assunto pode ser apresentado. Naves (1996) afirma que o processo de abstrair a essência de um documento é conhecido como “análise de assunto para alguns, análise temática para outros e ainda como análise documentária ou análise de conteúdo” (NAVES, 1996, p. 215). Já Hjørland (1997) aponta outros termos que têm sido utilizados, tais como análise de conteúdo, análise conceitual, análise da informação, *aboutness* e análise textual.

A partir do que foi colocado até aqui, este estudo tem como objetivo fazer a análise da revisão de literatura que trata da primeira etapa da indexação, a que denominamos análise de assunto.

O MÉTODO: A ANÁLISE DE CONTEÚDO

O método da Análise de Conteúdo (AC) vem sendo utilizado com destaque desde o século XX e designa um conjunto de técnicas empregadas para a sistematização de análises realizadas em textos. Do ponto de vista histórico, Bardin (2009) expõe que:

[...] descrever a história da “análise de conteúdo” é essencialmente referenciar as diligências que nos Estados Unidos marcaram o desenvolvimento de um instrumento de análise de comunicações, é seguir passo a passo o crescimento quantitativo e a diversificação qualitativa dos estudos empíricos apoiados na utilização de uma das técnicas classificadas sob a designação genérica de análise de conteúdo; é observar *a posteriori* os aperfeiçoamentos materiais e as aplicações abusivas de uma prática que funciona há mais de meio século (BARDIN, 2009, p. 15).

Assim, percebemos que, inicialmente considerado como um método cartesiano e positivista de pesquisa, na contemporaneidade, o método AC já une análises de cunho quantitativo e qualitativo de dados, para estudos históricos ou bibliométricos, por exemplo. “Na análise quantitativa, o que serve de informação

é a frequência com que surgem certas características do conteúdo. Na análise qualitativa é a presença ou a ausência de uma dada característica de conteúdo ou de um conjunto de características” (GEORGE, 1959 apud BARDIN, 1977, p. 21).

O método AC é usado com intensidade em pesquisas empíricas realizadas nas ciências humanas e sociais. Bardin (2009) define esse método como

[...] um conjunto de técnicas de análise das comunicações visando (*sic*) obter (*sic*) por procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos às condições de produção/ recepção (variáveis inferidas) destas mensagens (BARDIN, 2009, p. 44).

As condições de produção/recepção citadas por Bardin são as “variáveis inferidas” de distintas maneiras, dentre as quais o autor destaca as “variáveis psicológicas do indivíduo emissor, variáveis sociológicas e culturais, variáveis relativas à situação de comunicação ou do contexto de produção da mensagem” (BARDIN, 2009, p. 42).

Bardin (2009) aponta que, no método AC, alguns conceitos são considerados relevantes, a saber:

- (a) Objetividade: definição de categorias que possam ser replicadas por outro pesquisador, desde que seguindo os mesmos critérios;
- (b) Sistemática: conservação do objetivo proposto na análise, visando a minimizar a influência da subjetividade e conhecimentos pessoais do pesquisador e a evitar escolhas arbitrárias de categorias;
- (c) Conteúdo Manifesto: limitação da análise ao que está explícito no texto;
- (d) Unidades de Registro (UR): menor fragmento do texto (uma palavra, um termo, um parágrafo do texto ou parte de uma frase) a ser selecionado, de acordo com as categorias escolhidas, que devem fazer sentido no contexto. A UR também pode ser denominada unidade de análise ou unidade de significado;

- (e) Unidades de Contexto (UC): orientação do contexto mais amplo das unidades de registro (UR), sendo que cada UC pode conter diversas URs;
- (f) Construção de Categorias: classificação do conteúdo do texto em conformidade com o critério e objetivos estabelecidos. Os critérios podem ser léxicos (as palavras e seus sentidos), semânticos (temáticos), expressivos (problemas de linguagem) ou criados a partir de verbos, adjetivos, substantivos, entre outros. O conjunto de categorias deve obedecer a um único critério, em uma organização intencional do conteúdo, condensando o conteúdo do texto.

O conjunto de “Categorias” pode ser considerado o mais importante no contexto do método AC, uma vez que são as variáveis a ser estudadas e têm a função de resumir o conteúdo do texto. É a partir da observação das relações entre as “Categorias” que as inferências e interpretações são obtidas. Bardin afirma que o método AC se “assenta implicitamente na crença de que a categorização (passagem de dados brutos a dados organizados) não introduz desvios (por excesso ou por recusa) no material, mas que dá a conhecer índices invisíveis, ao nível dos dados em bruto” (BARDIN, 2009, p. 147).

Dentro desse propósito, o conjunto de “Categorias” deve atender às seguintes características:

- (a) Adequação (válidas e pertinentes): ser útil para o objetivo proposto, ao problema e à fundamentação teórica utilizada na pesquisa. Se criado *a priori*, deve estar respaldado por base teórico-metodológica, com critérios de classificação preestabelecidos;
- (b) Exaustividade (ou inclusividade): ser capaz de acolher a todas as unidades de registro (UR) sem a exclusão de qualquer dado significativo;
- (c) Homogeneidade: obedecer a um único critério de categorização, para que atenda a uma única variável;
- (d) Exclusividade (ou exclusão mútua): aceitar o pertencimento de uma unidade de registro (UR) a uma única categoria, a partir de regras precisas e claras;

- (e) Objetividade (ou consistência e fidedignidade): permitir que outro pesquisador chegue a resultados semelhantes, desde que aplicando os mesmos critérios de análise.

De acordo Bardin (2009), os procedimentos para o método de AC são organizados a partir de três etapas: (1) a pré-análise; (2) a exploração do material; (3) e o tratamento dos dados, inferência e interpretação.

A primeira etapa consiste em escolher uma unidade ou um conjunto de textos para a análise, que ocorrerá por meio da organização e sistematização dos assuntos a partir de quatro fases: (a) leitura flutuante, para conhecimento inicial do texto; (b) escolha dos documentos, que é a delimitação do que será analisado; (c) determinação dos pressupostos e objetivos; (d) elaboração de indicadores e/ou recortes de texto para análise. Na segunda etapa, há a exploração do material e definição de categorias (que pode ser feita *a priori*), a classificação e a criação de sistemas de codificação. Ressaltamos que a categorização é um procedimento contínuo e cíclico de formulação até que se chega ao conjunto final de “Categorias” a ser analisado, que deve obedecer às características já discriminadas anteriormente (adequação, exaustividade, homogeneidade, exclusividade e objetividade).

Nesse processo, determinam-se ou se identificam as unidades de registro (UR), que serão a unidade base para a contagem da frequência, e as unidades de contexto (UC), que têm o papel de significação que irá permitir a compreensão de uma UR. Essa é uma fase fundamental no método AC, pois possibilitará, ou não, melhores interpretações e inferências, dependendo da maneira como for realizada.

Finalmente, a terceira etapa consiste no tratamento, inferência e interpretação dos resultados, quando se faz um resumo e se destacam as informações para análise, elementos que possibilitam interpretações inferenciais.

Os resultados são descritos por meio de gráficos ou tabelas, em análises quantitativas (frequências e percentuais) ou qualitativas (definições para as categorias, preferencialmente com citações diretas).

É uma fase destinada a uma análise reflexiva e crítica (BARDIN, 2009) dos resultados, visando a saber se são passíveis de generalização, usando basicamente duas formas:

- (a) comparação entre os resultados e a literatura;
- (b) e construção de uma hipótese a partir da análise dos documentos.

Bardin (1977) esclarece que, na literatura, em algumas situações, o método AC e a análise documental ou documentária (AD) são tratados como sinônimos. De fato, o autor aponta que os dois métodos possuem semelhanças em relação aos processos intelectuais de análises empregados, tais como recortes de informação, analogias ou representação sumária das informações (BARDIN, 2009, p. 47). Todavia, são métodos diferentes, que têm objetivos distintos, com a básica diferença expressa pelo fato de que o método AC se ocupa do conteúdo da mensagem, enquanto o método AD manipula os assuntos dos documentos (BARDIN, 2009, p. 47). Reafirmando isso, Tálamo et al. (1992, *on-line*) assinalam que o método AD “tem como objetivo propor princípios para a elaboração de informações documentárias, definidas como múltiplas representações de textos para fins de recuperação em face das diferentes demandas de informação”. Já o método AC “consiste em extrair sentido dos dados de texto e imagem” (CRESWELL, 2007, p. 194). Assim, é uma análise que enriquece a leitura do texto, uma vez que dá maior significado, permitindo a compreensão e interpretação dos dados que foram quantificados.

Dentre as técnicas de análise que o método AC abriga, podem-se destacar: análise da anunciação (processo de comunicação), análise léxica (vocabulário/ideias), análise de avaliação (de atitudes), análise proposicional do discurso (significado dos enunciados), análise da expressão (indicadores formais), análise das relações (intratextual) e dedução frequencial ou análise categorial (temática ou semântica). Dessa maneira, ele pode ser aplicado a diversos contextos, podendo os procedimentos empregados variar conforme

os objetivos de análise, desde que se submetam a regras precisas e não intuitivas (OLIVEIRA, 2008).

Neste artigo, aplicamos a técnica de dedução frequencial ou análise por categorias, pois com ela é possível localizar

[...] uma série de significações que o codificador detecta por meio de indicadores que lhe estão ligados; [...] [pois] codificar ou caracterizar um segmento é colocá-lo em uma das classes de equivalências definidas, a partir das significações, [...] em função do julgamento do codificador. (PÊCHEUX, 1993, p. 65).

Diante do exposto, percebemos que essa técnica permite ao pesquisador (codificador) fragmentar, sintetizar e manipular o conteúdo do texto, a partir de categorias de informação, atendendo a critérios, objetivos e recorte de pesquisa. Isso é realizado de tal maneira que evidencia e comporta inferências sobre uma realidade que não a mesma da mensagem do texto (BARDIN, 2009).

Para estabelecer o conjunto de “Categorias” para a coleta de dados, é preciso determinar o que os elementos do conteúdo do texto têm em comum, que permita o seu agrupamento. Nesse processo, as temáticas vão emergindo do próprio texto analisado.

Esse foi o procedimento adotado para este estudo, e será apresentado na próxima seção.

APLICAÇÃO DA ANÁLISE DE CONTEÚDO NA AMOSTRAGEM

Nosso universo de estudos é o *Annual Review Of Information Science and Technology* (ARIST). Ele é publicado em língua inglesa, portanto, reflete mais a literatura produzida nesse idioma. Contudo, não há, no Brasil, qualquer publicação com o perfil do ARIST, que compile revisões sobre temas de interesse da ciência da informação, razão que motivou nossa decisão por sua escolha.

Ressaltamos que este trabalho é um recorte de uma pesquisa maior que pretende, através da literatura, avaliar estudos feitos sobre o processo da etapa de

análise de assunto, para traçar, através do tempo, seu desenvolvimento teórico e experimental.

Para analisar o conjunto de documentos, aplicamos a técnica da Análise Categórica Temática em três etapas.

PRIMEIRA ETAPA

Compôs-se de uma pré-análise, com a escolha dos documentos que fossem representativos em relação ao objetivo predefinido; de leitura flutuante, para conhecimento do texto; e de formulação de pressupostos. Como afirma Bardin (2009, p. 121), “a escolha dos documentos depende dos objetivos, ou, inversamente, o objetivo só é possível em função dos documentos disponíveis”. Assim, foi realizada uma primeira organização do material, de forma analítica, verificando se ele atendia ao objetivo proposto. Todos os capítulos já publicados do ARIST foram analisados e realizamos um primeiro recorte. Para esse recorte, escolhemos quatro capítulos do ARIST, tomando como primeiro critério selecionar aqueles capítulos, dentre todos os já publicados, que traziam no título “*Subject Analysis*”. Cabe registrar que esses capítulos foram os últimos publicados pelo ARIST, dentro dessa temática, desde seu primeiro volume, em 1966. Logo, nossa amostra é composta por: (1) o capítulo 4, do volume 12, de 1977; (2) o capítulo 5, do volume 17, de 1982; (3) o capítulo 2, do volume 17, de 1986; (4) o capítulo 2, do volume 24, de 1989.

Após uma leitura flutuante dos capítulos, formulamos alguns pressupostos:

- (1) pouca pesquisa tem sido realizada sobre o tema análise de assunto;
- (2) a publicação sobre este tema ainda é incipiente;
- (3) existe necessidade de ampliar os estudos metodológicos que facilitem a estratégia de determinação de assuntos de um documento textual.

SEGUNDA ETAPA

Caracterizou-se pela exploração do material para criação de um conjunto de categorias que são

classificadas e codificadas, com a identificação das unidades de registro (UR), que são os conceitos relevantes no texto; e as unidades de contexto (UC), que são as categorias mais gerais que contextualizam e dão significado às URs. Para isso, foi realizada uma leitura nos capítulos selecionados, e foram marcados nos quatro capítulos do *corpus* os segmentos em que aparecia a palavra “*Subject Analysis*”, quando houve a segunda organização, pelo critério de temas ou elementos conceituais dos textos, buscando e identificando eles para relacionar um texto ao outro, em conformidade com os objetivos propostos. Foi o próprio material que nos conduziu à determinação final do conjunto de categorias. Esse conjunto foi sendo continuamente refinado até que chegamos ao conjunto final:

- (a) unidades de contexto (UC): processo de indexação, etapa da análise de assunto, etapa da tradução, recuperação da informação, classificação e outros temas;
- (b) unidades de registro (UR): indexação intelectual, indexação automática, abordagem cognitiva, abordagem linguística, abordagem lógica, abordagem probabilística, abordagem semiótica, ensino, processos, teoria e prática, instrumentos, aspectos diferentes, avaliação.

Depois, partimos para a última etapa da técnica, com o tratamento dos resultados, inferência e interpretação, que será apresentada a seguir.

TERCEIRA ETAPA

Constituiu-se do tratamento dos resultados, inferência e interpretação, com a validação do conjunto de “Categorias” e a apreciação e análise dos resultados. Para orientar a análise, foram construídos quadros de autores e de temas e subtemas, sobre os quais ia sendo elaborado um fichamento com as observações acerca das relações entre os textos, com os pressupostos e os objetivos desejados. Através da análise, foi possível inferir significados importantes para o objetivo do estudo, criando novos conhecimentos sobre o campo de investigação, cujas análises estão expostas na próxima subseção.

INFERÊNCIAS E INTERPRETAÇÕES DOS DADOS

Inicialmente, foram analisados todos os volumes já publicados do ARIST, cuja coleção totaliza 45 volumes, dentre os quais fizemos nosso recorte de pesquisa: o *corpus*. A coleção ARIST trouxe publicações que citaram a temática análise de assunto nos seguintes volumes: (a) volume 1, 1966; (b) volume 2, 1967; (c) volume 3, 1968; e (d) volume 4, 1969, que foram intitulados “*Content Analysis, Specification, and Control*”. Porém, os volumes de 5 a 9 (1970-1974) foram publicados sob o título “*Document Description and Representation*”. Nos anos de 1975 e 1976, os ARIST não trouxeram qualquer

capítulo que cobrisse essa temática especificamente; já os volumes 12 (1977), 17 (1982), 21 (1986) e 24 (1989) trouxeram capítulos intitulados “*Subject Analysis*”. Como registrado anteriormente, a seleção do *corpus* se baseou na escolha dos capítulos por eles trazerem o título de análise de assunto.

Ressalta-se que o ARIST é uma publicação que traz revisões bibliográficas publicadas somente em língua inglesa. Portanto, autores que publicaram em outros idiomas não serão citados nesta amostra. A tabela 1 apresenta o conjunto de categorias com as ocorrências das temáticas e subtemáticas extraídas do *corpus*, em uma relação de temáticas/subtemáticas *versus* ano.

Tabela 1 – Ocorrências de temáticas e subtemáticas no *corpus*

Temáticas	Subtemáticas	1977	1982	1986	1989
Processo de Indexação	Intelectual	X	X	X	X
	Automática	X	X	X	X
Etapa da Análise de Assunto	Abordagem cognitiva	X	X		X
	Abordagem linguística	X	X	X	X
	Abordagem lógica	X			
	Abordagem probabilística		X		
	Abordagem semiótica		X		
	Ensino			X	
	Processos	X	X		X
Etapa da Tradução	Teoria e prática	X	X	X	X
	Instrumentos	X	X	X	X
Recuperação da Informação	Aspectos diferentes	X	X	X	X
	Avaliação			X	X
Classificação		X		X	X
Outros temas		X	X	X	X

Fonte: elaborada pelas autoras, a partir dos dados coletados.

Observando a tabela 1, percebemos que, mesmo que o *corpus* apresente, no título dos capítulos, o termo “análise de assunto”, nenhuma das revisões tratou, especificamente, sobre a primeira etapa da indexação, a análise de assunto. A despeito dessa não exclusividade, esse tema sempre esteve associado ao processo de indexação como um todo, que apareceu em todas as revisões. Nesses casos, a maior parte das revisões analisadas tratou da temática de forma abrangente,

confirmando o apontamento de MAI (2000) quando ressalta que a análise de assunto é, algumas vezes, abordada como uma área do conhecimento que investiga a construção e o uso das linguagens de indexação e dos sistemas de classificação bibliográficos. Ao examinarmos o *corpus* para a criação do conjunto de categorias a serem analisadas, foi preciso elencar outras temáticas fronteiriças relacionadas ao tema principal da análise de assunto.

A temática da etapa da análise de assunto recebeu referências em todos os anos, assinalando estudos relativos às teorias, práticas e abordagens envolvidas nos seus procedimentos, os processos incluídos, assim como questões relativas ao ensino dessa disciplina nos cursos de graduação em biblioteconomia. Notamos que, algumas vezes, não foi possível discernir, inicialmente, quando a análise de assunto era estudada como área ou como processo. Somente uma exploração mais aprofundada permitiu distinguir tais situações, quando então emergiram do texto as subtemáticas.

Concomitantemente à coleta das temáticas e subtemáticas, elencamos os autores, referenciados nas revisões do *corpus*, que tinham relação com a temática e, assim, com as subtemáticas da etapa da análise de assunto, que é o nosso objetivo principal. O resultado desse levantamento está apresentado na figura 1, e relaciona AUTORES *versus* ANOS, AUTORES *versus* SUBTEMÁTICAS e, ainda, SUBTEMÁTICAS *versus* ANOS.

A figura 1 é muito importante, pois demonstra o total de 47 autores envolvidos em estudos sobre a temática da etapa da análise de assunto, fornecendo, também, uma visão temporal das subtemáticas abordadas.

Desse total, destacamos o autor Lancaster, que aparece cobrindo a subtemática da teoria e prática nas revisões dos anos 1986 e 1989. Lancaster também aparece em 1986 tratando do ensino da análise de assunto. Da mesma maneira, a autora Svenonious merece destaque, pois aparece nas revisões dos anos 1982 (com duas publicações) e de 1989, dedicando-se à subtemática teoria e prática.

Dentro da subtemática do ensino, além do autor Lancaster, salientamos a contribuição do autor Dagobert Soergel, cujo livro *Organizing information: principles of data base and retrieval systems*, de 1985, é amplamente utilizado em escolas de biblioteconomia e ciência da informação em língua inglesa (ARIST, 1986).

Ainda sobre a subtemática das abordagens, no ARIST de 1977 aparece o autor K. Jones, com estudos teóricos acerca das abordagens cognitiva, lógica e linguística. Sobre a última, no ARIST de 1982, ressaltamos a ocorrência de citação da autora Dahlberg, contribuindo com os estudos sobre abordagem linguística, trazendo a Teoria do Conceito como uma técnica auxiliar na determinação do assunto de um documento.

No gráfico 1, temos os dados demonstrados na forma de um gráfico, o que facilita a percepção da incidência das subtemáticas em cada ARIST do *corpus*, em relação ao valor absoluto de ocorrências.

Analisando o gráfico 1, é possível apreender o perfil dos ARIST do *corpus*. No volume 12 do ano de 1977, curiosamente, verificamos homogeneidade de ocorrência de todas as subtemáticas dentro da temática da etapa da análise de assunto. Esse capítulo pode ser considerado, na amostra, o que mais apontou trabalhos relacionados à atividade de análise de assunto como a primeira etapa da indexação.

Nota-se que, no capítulo do ano de 1982, houve ampliação das subtemáticas relacionadas a outras áreas de conhecimento. Isso se justifica pela necessidade de buscar contribuições em áreas afins, com o intuito de auxiliar na compreensão da complexidade e da sofisticação da atividade da análise de assunto. Especificamente na revisão de 1982, aparecem todas as subtemáticas de abordagens. Esse fato coincide com apontamentos posteriores, na literatura, sobre a necessidade de avaliar e melhorar o ensino da etapa de análise de assunto nas escolas de biblioteconomia e ciência da informação. Antes dessa época, o ensino era bastante focado na indexação pré-coordenada, controle de vocabulário e construção de tesouros.

Por meio do gráfico 1, observa-se que no capítulo de 1986 houve um registro maior da subtemática processos. Neste período, houve aumento muito grande de pesquisas trazendo questionamentos

Figura 1 - Autores referenciados na amostra

Autores da temática Análise de assunto	1977	1982	1986	1989	Subtemáticas
BRENNER, E. H. & SARACEVI, T. (1985)			○		○ Ensino
CLEVELAND, D. & CLEVELAND, A. (1983)			○		● Processos
FOSKET, A. C. (1982)			○		● Teoria e prática
LANCASTER, F. W. (1980)			○		□ Abordagem linguística
SOERGEL, D. (1985)			○		▣ Abordagem cognitiva
TARR, D. & BORKO, H. (1974)	●				▣ Abordagem lógica
TANGUE, J. M. (1981)		●			▣ Abordagem probabilística
AZUBUIKE, A. A. & UMOH, J. S. (1988)				●	■ Abordagem semiótica
BAKEWELL, K. G. B. (1987)				●	
BEGHTOL, C (1986a)				●	
BLAIR, D. C. (1986)				●	
CHANG, L. M. (1986a e b)				●	
VICKERY, B. C. (1986)				●	
WHITE, H. D. & GRIFFITH, B. C. (1987)				●	
HICKEY, D. J. (1976)	●				
BORKO, H. (1977)		●			
COOPER, W. S. (1978)		●			
FUGMANN, R. (1979,1980)		●			
HUCTCHIS, W. J. (1978)		●			
SETTEL, B. (1977)		●			
SPARCK JONES, K. (1979b)		●			
SVENONIUS, E. & CHMIERER, H. F. (1977)		●			
SVENONIUS, E. (1981)		●	●		
GRUNBERG, M. W. (1986)			●		
LANCASTER, F. W. (1986)			●		
CROWE, J. D. (1986)				●	
KHANZHIN, A. G. (1986)				●	
LANCASTER, W. J. (1968)				●	
NELL, S. D. (1987)				●	
WEINBERG, B. H. (1988)				●	
JONES, K. (1976)		□			
CLEVERDON, C. W. (1979)			□		
DAHLBERG, I (1978)			□		
DAHLBERG, I (1981a)			□		
DOVBENKO, M. A. & UMANSKI, A. (1980)			□		
KAPLUN, L. M. & AZGALDOV, G. G. (1979)			□		
RIGGS, F. W. (1979)			□		
VASARHELYI, P. (1980)			□		
CRISTAL, D. (1984)				□	
SALTON, G. (1988b)				□	
JONES, K. (1974)	▣▣				
KOBLITZ, J. (1979)		▣			
DANIELS, P. J. (1986)				▣	
MARON, M.E. & HUHNS, J. L. (1960)		▣			
ROBERTSON, S. et al (1981)		▣			
ROBERTSON, S. (1977)		▣			
HENRIKSEN, T. (1979)		■			

Fonte: elaborada pelas autoras, a partir dos dados coletados.

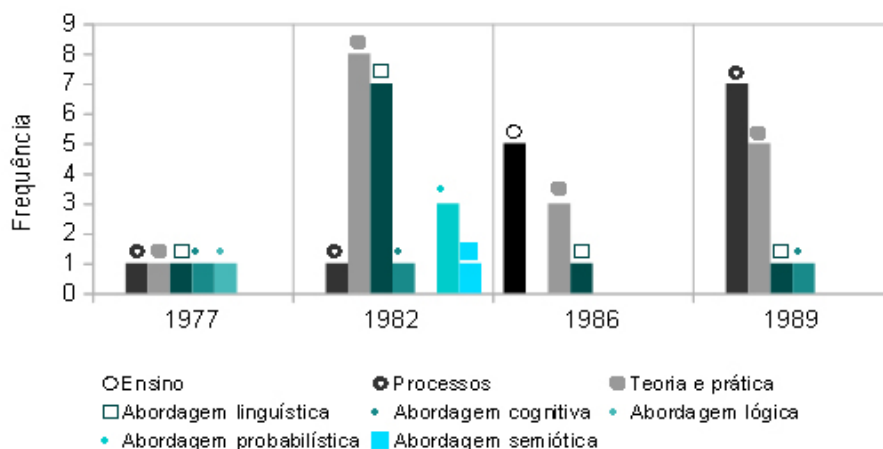
sobre o acesso à informação. Com isso, as publicações citadas nesse capítulo evidenciaram um cruzamento de ideias, teorias e modelos que pudessem melhorar a eficiência dos procedimentos de entrada e saída em um sistema de recuperação da informação (SRI), já apontando fortemente para as aplicações da tecnologia como uma das soluções.

O último capítulo analisado, de 1989, retrata fielmente a ampliação do uso das novas tecnologias para auxiliar na atividade de análise de assunto, com o uso de processamento da linguagem natural (PNL). A área da PNL ou linguística computacional trabalha com a automação da interpretação e da geração da língua humana em aplicações como sumarização automática de textos, recuperação e extração de informação, categorização textual, entre outros, visando a minimizar a morosidade da atividade de análise de assunto realizada por humanos. Com isso, notamos aumento da ocorrência da subtemática de abordagem linguística, com estudos abordando a fonologia, a estrutura da língua (morfologia e sintaxe) e o significado (semântica e pragmática), usando um princípio da tradição gerativista segundo o qual o processamento das línguas é modular.

Também observamos maior incidência de estudos da subtemática da abordagem cognitiva, trazendo uma visão mais holística desse processamento, afirmando que a língua humana somente pode ser explicada por princípios cognitivos, através da análise do todo e não de elementos segmentados.

Notamos, também, uma preocupação maior com a interação do usuário com o sistema e, como consequência dessa maior preocupação, iniciaram-se os estudos sobre a interdependência entre os processos na análise de assunto e a área de recuperação da informação. Isso sugere que as abordagens cognitiva e linguística tiveram participação na expansão desses estudos, contribuindo para os avanços das pesquisas sobre os métodos semiautomatizados e automatizados, tanto no processo de analisar o assunto, como nas técnicas usadas nas estratégias de busca. A subtemática Teoria e Prática aparece em todos os capítulos dos ARIST estudados, demonstrando, mais uma vez, que as pesquisas e estudos expressos na literatura tratam a análise de assunto, na maior parte das vezes, como uma área do conhecimento e, poucas vezes, como um processo dentro da indexação, sempre abordada de modo mais abrangente.

Gráfico 1 – Ocorrência das subtemáticas por ano (absoluta)



Fonte: elaborado pelas autoras, a partir dos dados coletados.

CONSIDERAÇÕES FINAIS

O conteúdo das revisões nos capítulos relacionados à análise de assunto dos ARIST veio confirmar os pressupostos apontados na metodologia deste artigo. Como exemplo, podemos citar o pouco que se tem pesquisado sobre esse tema, havendo raras publicações sobre isso na literatura da área.

Em razão de a atividade de análise de assunto envolver procedimentos interpretativos complexos, existe uma tendência de pesquisas buscando respaldo teórico nas áreas das ciências cognitivas e da linguística, para entender como a mente processa a informação e, também, para buscar aportes no entendimento das estruturas e tipologias textuais. Podemos perceber que a análise de assunto continua sendo uma área fértil e com grande potencial para o desenvolvimento de pesquisas.

A deficiência em se entender como o ser humano pensa, processa e comunica a informação durante a atividade da análise de assunto continua preocupando os pesquisadores. Para aprimorar essa atividade é preciso dominar os aspectos relativos ao processamento da linguagem na mente humana, para apreender como o fenômeno linguístico se forma em nosso cérebro.

Notamos também que o tema análise de assunto foi apontado como o principal nos livros de autores que discorrem sobre a subtemática teoria e prática para o ensino na biblioteconomia. Entretanto, o foco utilizado é, principalmente, estudá-la como uma área do conhecimento que investiga a construção e o uso das linguagens de indexação e dos sistemas de classificação bibliográficos.

A prática da indexação tem sido impulsionada e influenciada em maior grau pela automação dos procedimentos, enquanto a teoria tem se desenvolvido lentamente. A indexação automática e a semiautomática têm sido apontadas como uma solução para melhorar a morosidade da atividade de análise de assunto, gerando protótipos que utilizam a inteligência artificial e embasamento linguístico.

Alguns resultados já trouxeram progressos nesses estudos, tendo em vista a quantidade de *softwares* que podem ser adquiridos no mercado. Entretanto, ainda existe necessidade de ampliar os estudos metodológicos de análise de assunto, a fim de facilitar o uso de estratégias na determinação de assuntos de um documento textual.

Finalmente, podemos ressaltar que o principal problema no aprimoramento da atividade de análise de assunto é o entendimento de como pensamos, organizamos e integramos a informação em um SRI. Pretende-se, futuramente, dar continuidade à presente análise, completando-a com os outros oito capítulos que tratam dessa temática no ARIST.

AGRADECIMENTOS

Agradecimento ao apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e ao CNPq.

REFERÊNCIAS

- ALLEN, B. Cognitive research in information science: implications for design. *Annual Review of Information Science and Technology*, v. 26, p. 3-37, 1991.
- ARIST – The Annual Review of Information Science and Technology. Silver Spring: ASIS&T, 1966-2011. Disponível em: <http://www.asis.org/Publications/ARIST/statement.php>. Acesso em: 13 ago. 2014.
- BARDIN, L. *Análise de conteúdo*. 4. ed. rev. e atual. Lisboa: Edições 70, 2009.
- BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, 1977.
- BLAIR, D.C. *Language and representation in information retrieval*. New York: Elsevier Science, 1990.
- CHU, C.; O'BRIEN, A. Subject analysis: the critical first stage in indexing. *Journal of Information Science*, v. 19 n. 6, p. 439-454, 1993.
- CRESWELL, J. W. *Projeto de pesquisa: métodos qualitativo, quantitativo e misto*. 2. ed. Porto Alegre: Artmed, 2007.
- DANIELS, P. J. Cognitive models in information retrieval: an evaluation review. *Journal of Documentation*, v. 42, n. 4, p. 275-305, Dec. 1986.
- FARROW, J. F. *A cognitive process model of document indexing*. *Journal of Documentation*, London, v. 47, n. 2, p. 149-166, 1991.
- FOSKETT, A. C. *A abordagem temática da informação*. São Paulo: Polígono, 1973.
- FROHMANN, B. Rules of indexing: a critique of mentalism in information retrieval theory. *Journal of Documentation*, v. 46, n. 2, p. 81-101, June 1990.
- FUGMANN, R. *Subject analysis and indexing: theoretical foundation and practical advice*. Frankfurt/Main: Indeks Verlag, 1993.
- FUJITA, M. S. L. *A leitura documentária do indexador: aspectos cognitivos e lingüísticos influentes na formação do leitor profissional*. 2003. 321f. Tese (Livre-Docência em Análise Documentária e Linguagens Documentárias Alfabéticas) - Faculdade de Filosofia e Ciências, UNESP, 2003.
- GEORGE, A. L. *Quantitative and qualitative approaches to content analysis*. New York: Rand, 1959.
- HJORLAND, B. *Information seeking and subject representation: an activity-theoretical approach to Information Science*. Westport & London: Greenwood Press, 1997.
- HUTCHINS, W. J. The concept of «aboutness» in subject indexing. *Aslib Proceedings*, London, v. 30, n. 5, p. 172-181, May 1978.
- KOBASHI, N. Y. *A elaboração de informações documentárias: em busca de uma metodologia*. 1994. 195f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 1994.
- LANCASTER, F. W. *Indexação e resumos: teoria e prática*. 2. ed. Brasília: Briquet de Lemos, 1998.
- LANCASTER, F. W.; ELLIKER, C.; CONNELL, T.H. Subject analysis. *ARIST*, v. 24, p. 35-61, 1989.
- LANGRIDGE, D. W. *Subject analysis: principles and procedures*. London: Bowker-Saur, 1989.
- LISTON JR, D.M.; HOWDER, M.L. Subject analysis. *ARIST*, v. 12, p. 37-61, 1977.
- MAI, J-E. Deconstructing the indexing process. *Advances in Librarianship*, v. 23, p. 269-298, 2000.
- MAI, J-E. Semiotics and indexing: an analysis of the subject indexing process. *Journal of documentation*, v. 57, n. 5, p. 591-622, 2001.
- NAVES, M. M. L. Análise de assunto: concepções. *Revista de Biblioteconomia de Brasília*, Brasília, v. 20, n. 2, p. 215-226, jul./dez. 1996.
- NAVARRO, S. Interface entre linguística e Indexação: uma revisão de literatura. *Rev. Bras. Biblio. Doc.*, São Paulo, v. 21, n. 1/2, p. 46-62, jan./jun. 1988.
- NOHR, H. The training of librarians in content analysis: some thoughts on future necessities. *Int. Classif.*, v. 18, n. 3, p. 153-157, 1991.
- OLIVEIRA, D. C. de. Análise de conteúdo temático-categorial: uma proposta de sistematização. *Revista de Enfermagem*, Rio de Janeiro, v. 16, n. 4, p. 569-576, out./dez. 2008.
- PÊCHEUX, M. Análise automática do discurso. In: GADET, F.; HAK, T. (Org.). *Por uma análise automática do discurso: uma introdução à obra de Michel Pêcheux*. 2. ed. Campinas (SP): Unicamp, 1993. p. 61-105.
- PETERSEN, T. Introduction. In: GUIDE to indexing and cataloging with the Arts and Architecture Thesaurus. New York: Oxford University Press, 1994.
- PINTO MOLINA, M. Interdisciplinary approaches to the concept and practice of written text documentary content analysis (WTDC). *Journal of Documentation*, London, v. 50, n. 2, p. 111-133, jun. 1994.
- SCHWARTZ, C.; EISENMANN, L.M. Subject analysis. *ARIST*, v. 21, p. 37-61, 1986.
- SOERTEL, D. *Organizing information: principles of data base and retrieval systems*. Orlando: Academic Press, 1985.
- TÁLAMO, M. de F.G.M. et al. A interface análise documentária, linguística documental e terminologia. In: SIMPÓSIO DE TERMINOLOGIA, 3., 1992, Millán de La Cogolla. *Anais...* San Millán de la Cogolla: RITerm, 1992.
- TAYLOR, A.G. *The organization of information*. Englewood, CO: Libraries Unlimited, 1994.
- TRAVIS, I. L.; FIDEL, R. Subject analysis. *ARIST*, v. 17, p. 123-157, 1982.