



## Indexação temática em textos de Rui Barbosa

*Thematic indexing in Rui Barbosa texts*

Dilza Ramos Bastos\*

Ana Paula Lima dos Santos\*\*

Lourdes Maria dos Santos\*\*\*

### RESUMO

Apresenta o desenvolvimento da primeira fase da pesquisa iniciada em 2016, que trata da aplicação de metodologia de indexação às Obras Completas de Rui de Barbosa. Aborda assim o estabelecimento do vocabulário sistematizado para representação dos acervos bibliográficos, da Fundação Casa de Rui Barbosa, empreendimento realizado a partir da constatação da carência de fundamentos teóricos que promovessem a melhoria do processo de elaboração das informações e da recuperação da informação, o que resultou na adoção de princípios metodológicos da teoria do conceito, de Ingetraut Dahlberg (1978). Tem como objetivo aprofundar a fundamentação teórica do processo de representação das informações oriundas dos acervos bibliográficos da instituição. A metodologia propõe identificar conceitos, a partir do estudo da metodologia de indexação que se fundamenta na teoria do conceito. Conclui que a indexação é uma prática que precisa ser sistematicamente discutida, revista e aprimorada.

### ABSTRACT

It presents to development the first phase research begun in 2016 about the indexing methodology application in the Rui de Barbosa Literary work. Describe the systematized vocabulary establishment for the representation of Casa de Rui Barbosa Foundation bibliographical collections, cause theoretical foundations don't improve the information elaboration process and retrieval therefore this study propose the Concept theory methodological by Ingetraut Dahlberg (1978) to solve this problem. Its purpose is deepen the foundation theoretical of the information representation process coming from the bibliographical collections. The methodology proposes to identify concepts, based on the study of the indexation methodology that is based on Concept Theory. It concludes that indexing is a practice that needs to be systematically discussed, reviewed and improved.

**Keywords:** Thematic Indexing; Bibliographic Collection; Concept Theory.

\* Mestre em Ciência da Informação pelo Ibict/UFRJ. Bibliotecária (coordenadora) da Fundação Casa de Rui Barbosa, Centro de Memória e Informação/Serviço de Biblioteca. Endereço: Rua São Clemente, 134 Botafogo, CEP 22260-000, Rio de Janeiro, RJ. Tel (21) 32898667. E-mail: dilzabastos@gmail.com.

\*\* Mestre em Ciência da Informação pela Universidade Federal Fluminense (UFF). Bibliotecária da Superintendência de Documentação. Sistema de Bibliotecas e Arquivos da UFF. Bolsista do Programa de Incentivo à Produção do Conhecimento Técnico e Científico na Área da Cultura da Fundação Casa de Rui Barbosa, CMI/Serviço de Biblioteca. Endereço: Avenida Rio Branco, s/n, São Domingos, Gragoatá, CEP 24020-006, Niterói, RJ. Tel (21) 2629-2781. E-mail: annalima27@gmail.com.

\*\*\* Mestre em Biblioteconomia pela Universidade Federal do Estado do Rio de Janeiro (Unirio). Bolsista do Programa de Incentivo à Produção do Conhecimento Técnico e Científico na Área da Cultura da Fundação Casa de Rui Barbosa, CMI/Serviço de Biblioteca. Endereço Postal: Rua São Clemente, 134 Botafogo, CEP 22260-000 - Rio de Janeiro, RJ. Tel (21) 32898667. E-mail: lourdes.maria@gmail.com.

**Palavras-chave:** Indexação Temática;  
Acervo Bibliográfico; Teoria do Conceito.

## INTRODUÇÃO

O artigo apresenta considerações sobre as ações de projeto de pesquisa no âmbito da indexação temática e, em especial, sobre a aplicação de metodologia fundamentada pela teoria do conceito de Ingetrud Dahlberg. Iniciado em outubro de 2016, como parte do Programa de Incentivo à Produção do Conhecimento Técnico e Científico na Área da Cultura, da Fundação da Casa de Rui Barbosa (FCRB), o projeto tem como objetivo aprofundar a fundamentação teórica do processo de representação das informações oriundas dos acervos bibliográficos da instituição, tendo como objeto empírico a coleção intitulada “Obras Completas de Rui Barbosa (OCRB)”. Desse modo, visa assim ampliar a aplicação da metodologia de indexação, fazendo com que a adoção de terminologia controlada seja utilizada nos sistemas de bases de dados referenciais e digitais. Tal empreendimento possibilita também investigar o uso da linguagem documentária e da linguagem natural nas obras digitalizadas constantes no repositório institucional. Logo, é uma revisão de literatura que dará o suporte teórico para o desenvolvimento do trabalho prático.

A coleção OCRB foi organizada com o objetivo de reunir a produção intelectual de Rui, a partir de um planejamento publicado em dois volumes como roteiro da coleção. Cada tomo pode corresponder a uma obra apenas ou se constituir um conjunto de trabalhos distintos. Além disso, alguns são de integral autoria de Rui Barbosa, outros de autoria partilhada ou mesmos apresentam apenas trechos de Rui, como no caso das transcrições de suas falas em assembleias, reuniões, etc. Essa é uma coleção viva, pois novos volumes ou novas edições continuam a ser publicados, após trabalho de pesquisa e de organização do Setor Ruiano, da FCRB.

Conhecer a atuação de Rui e sua influência política e social foi fundamental para identificar os assuntos e contextualizá-los no seu tempo, de modo a compreender situações e relações. De igual modo, era também importante conhecer como ele formou essa biblioteca ao longo de sua vida, considerando-se que as obras que a compõem constituíram rica fonte para suas análises, ponderações, doutrina e produção bibliográfica. Os relatos sobre o relacionamento com os seus livros evidenciam também a importância e a fundamentação dos trabalhos produzidos, fazendo ver como Rui embasava suas afirmações e mesmo seus pareceres, além dos seus discursos que tanto influenciaram os destinos do Brasil.

Outro aspecto significativo é que a Biblioteca de Rui Barbosa e sua produção intelectual são muito demandadas pelo público, requerendo assim estudar o perfil dos usuários e suas questões de pesquisa. Portanto, esse conhecimento propiciou melhorar a percepção do valor informacional do acervo e das necessidades dos usuários.

A revisão de literatura sobre indexação propiciou maior embasamento para a investigação dos registros já estabelecidos na base de dados referencial. A FCRB vem ao longo dos anos buscando aperfeiçoar as práticas de indexação com o objetivo de facilitar a comunicação entre o acervo e seu usuário. O estudo permanente e a avaliação dos catálogos são essenciais, visando manter a política de indexação e a coerência na criação e atribuição dos descritores para uma recuperação efetiva. (BRASIL et al., 2002).

De acordo com Bastos (2006), os estudos do processamento documental são de grande importância considerando-se o impacto das novas tecnologias da informação no processo da comunicação científica, pois surgem assim novas necessidades de pesquisa. Essa realidade impacta diretamente as atividades do indexador, fazendo com que reflita sobre o seu fazer e pense em novas formas de atender a essas demandas informacionais.

Entendemos que não basta criar uma política de indexação, é indispensável o aperfeiçoamento, treinamento e estudo constantes, pois com a frequente e necessária atualização dos sistemas de tecnologias da informação e comunicação (TICs), os gestores de unidades de informação devem acompanhar o desenvolvimento tecnológico, buscando aprimorar a organização das informações produzidas e os serviços oferecidos.

Recentemente a Fundação passou por migração do *software* gerenciador das bases de dados referenciais, o que ocasionou intercorrências durante o processo e que ocasionaram inconsistências que interferem diretamente na recuperação da informação. Tais problemas demandaram revisão das atividades técnicas, solucionando questões tais como duplicidade registros e informações deslocadas entre campos do Marc (Machine Readable Machine).

A constante atualização dos sistemas de recuperação da informação visa melhorar e ampliar as informações documentárias e o seu acesso, atendendo à crescente demanda dos usuários. Fazem-se assim necessários estudo e planejamento para adotar e aperfeiçoar as TICs. Portanto, a customização permanente busca aperfeiçoar os metadados, recursos de busca e produtos gerados. Apesar da existência de um planejamento estratégico, algumas questões são imprevisíveis devido à falta de compatibilidade técnica. Por isso, a migração de sistemas se torna imperativa para a melhoria da gestão da informação.

Ressaltamos assim que a migração de bases é necessária para atender as demandas informacionais. No entanto, toda mudança traz consigo alguns inconvenientes, tais como, adaptações de campos bibliográficos e controle de autoridades,<sup>1</sup> entre outros. As ações devem manter as bases de dados coesas em relação ao padrão definido pela instituição, para uma melhor recuperação da informação. Para tal, na próxima seção abordaremos os conceitos e algumas reflexões sobre a indexação.

## **INDEXAÇÃO: CONCEITOS E REFLEXÕES**

Em linhas gerais, Lancaster (2004) define a indexação como um processo que identifica do que trata o documento. Entende-se assim que a indexação é uma atividade que exige do indexador concentração e domínio do assunto. Portanto, há influência desse indexador no processo de indexação de modo significativo. Tal abordagem é também reforçada por Knight (1974, p. 21), quando afirma que o indexador “é a pessoa que realmente examina as páginas sob todos os pontos de vista”, e, desse modo, a indexação traz ao consulente maior relevância e precisão, evitando a alta revocação.

No que se refere ao conceito de indexação, Gil Leiva (1999, p.19-20) afirma que a maioria dos conceitos é incompleta por se referirem, muitas vezes, apenas aos

---

<sup>1</sup> Entende-se por “controle de autoridade” a padronização de termos como: termos tópicos, títulos uniformes, nomes próprios, nomes corporativos, áreas geográficas e outros.

documentos como fontes de análise, ignorando a pergunta do usuário – afinal, o usuário é o maior interessado nesse processo. Para o estudioso, a indexação ocorre em dois momentos: a “*indexação do documento*, para armazenamento”; e a *indexação da pergunta* do usuário, cujo objetivo é obter o que o autor chamou de “resposta documental”, ou seja, para recuperar documentos que atendam à necessidade do usuário, materializada na expressão de busca.

Gil Leiva (1999, p.20) divide a *indexação dos documentos* em duas etapas. A primeira refere-se à *leitura do documento*, que por sua vez se divide em uma “leitura horizontal”, na qual são analisados e selecionados os conceitos presentes no documento; e em uma “leitura vertical”, em que são identificados e atribuídos termos referentes aos conceitos implícitos no documento. Na segunda etapa, os conceitos em linguagem natural podem ser registrados na própria linguagem natural ou convertidos para os termos de uma linguagem documentária.

No entendimento de Lancaster (2004, p. 81-82), uma boa indexação implica na existência de fatores que influenciam “na qualidade da indexação”, tanto referente ao indexador, ao vocabulário, ao documento, quanto ao “processo”<sup>2</sup> em si e aos fatores ambientais. No que se refere ao *indexador*, pode-se elencar o conhecimento que ele tem do assunto, a percepção das necessidades dos usuários, da capacidade de compreensão de leitura, da experiência profissional, entre outros fatores. Quanto ao *processo de indexação*, observa-se o fator ligado ao vocabulário que se refere à especificidade/sintaxe, ambiguidade ou imprecisão, na especificidade de uso dos termos, pois quanto mais específico for o vocabulário, mais minucioso será o seu significado, bem como na qualidade da estrutura e disponibilidade de instrumentos auxiliares afins. No que se referem ao “processo” propriamente dito, os fatores estão relacionados ao tipo de indexação; se é uma indexação exaustiva ou específica (existem também os aspectos referentes a regras e instruções que variam de acordo com a instituição); a questão da produtividade exigida; e a exaustividade da indexação. Em relação ao *documento*, pode-se considerar: conteúdo temático, complexidade, língua, linguagem, extensão, apresentação e sumarização. E, ainda, outros fatores como os *ambientais*, que dizem respeito à calefação, refrigeração, iluminação e ruído, que podem afetar a atuação do indexador.

No processo de recuperação da informação, a relevância de acordo com o pensamento de Greisdorf (2000) é um conceito difícil de construir e, para alguns, de quantificar. O estudioso ressalta que no momento da busca em um sistema de informação, o usuário empregou seus conhecimentos intrínsecos e extrínsecos para decidir quais palavras utilizar, tendo em vista recuperar informações importantes, ou seja, relevantes. Ao separar o que é importante do que é “lixo”, muitas informações podem ser aproveitadas para outros problemas que não necessariamente aquele para qual efetuou a pesquisa. Desse modo, para o usuário que fez a busca, todas as informações passam a ser relevantes. Observamos também que Saracevic (1970, p. 112) define a relevância ou precisão<sup>3</sup> como “a medida de contato efetivo entre a fonte e o destinatário”.

---

<sup>2</sup> “tipo de indexação, regras e instruções, produtividade exigida, exaustividade da indexação” (LANCASTER, 2004, p. 81).

<sup>3</sup> Relevância e precisão são consideradas por Saracevic como sinônimos por serem vistas por ele como uma propriedade em um processo de comunicação e que pode ser quantificada, tal como a revocação, por meio das fórmulas matemáticas. São tópicos amplamente discutidos na literatura referente à disseminação seletiva da informação (DSI) e estão diretamente relacionados aos sistemas de recuperação da informação (SRI).

Portanto, os resultados obtidos na busca, nos sistemas de recuperação da informação, precisam ser avaliados. De acordo com Lancaster (2004, p. 4), a medida denominada “revocação” é a extensão com que os itens ou informações são recuperados, ou seja, é tudo o que é recuperado. Assim, no sentido de evitar a alta revocação, que resulta em baixa precisão, Knight (1974) e Lancaster (2004) concordam que o indexador deve trabalhar com instrumentos auxiliares, tais como os dicionários especializados e atlas.

Todavia, o grande problema é que, na realidade diária de trabalho, o indexador se depara com outros fatores também mencionados por Lancaster (2004). Isto é, a produtividade exigida agregada à falta de profissionais qualificados para fazê-lo e tempo suficiente para a alta demanda de afazeres que se tem na rotina diária são fatores relacionados ao tipo de indexação. Logo, o domínio dos fundamentos teóricos é essencial para tal amadurecimento, como veremos na seção abaixo.

## FUNDAMENTOS TEÓRICOS DA REPRESENTAÇÃO DO CONHECIMENTO E DA INFORMAÇÃO

Para Naves (2001, p. 189), o aumento da produção científica aliado à crescente interdisciplinaridade entre áreas do conhecimento, que vem ocorrendo nas últimas décadas, tem tornado o trabalho de organização e tratamento da informação cada vez mais árduo e complexo. “No campo da ciência da informação, a ação de identificar e descrever o conteúdo de um documento” ocorre em diversos processos de análise e de descrição dos aspectos físicos e temáticos. Para a autora, “a indexação é considerada como um processo básico na recuperação da informação e, dentro desse processo ocorre” o ato de compreender e interpretar o conteúdo informativo do documento, por meio da atividade denominada análise de assunto.

De acordo com Feitosa (2006), o advento da internet proporcionou grande crescimento na quantidade de informações, de usuários e de computadores ligados à rede (*notebooks*, celulares, *tablets*). Com isso, torna-se imprescindível a padronização de metodologias para estruturar toda essa imensa rede de informação que é disponibilizada e acessada. A correria do dia a dia leva o usuário a buscar suas informações cada vez mais na internet por ser atualmente a forma mais cômoda e rápida. Por isso, também a importância dessas informações estarem bem representadas para a sua recuperação.

Brascher e Café (2008) em trabalho apresentado pelo GT2, no Enancib, ressaltam uma distinção entre:

[...] dados, informação e conhecimento: os dados são informação potencial, que somente são percebidos por um receptor se forem convertidos em informação e esta passa a converter-se em conhecimento no momento em que produz uma modificação na estrutura do conhecimento do receptor. (FERNANDEZ-MOLINA, 1994, p. 328 apud BRASCHER; CAFÉ, 2008, p. 3).

Para Vickery (1986), no campo da ciência da informação (CI), a representação documentária em seu aspecto de representação é distinguida em dois tipos: conhecimento pessoal e conhecimento público. Esses conhecimentos são transformados em memória documentária e passam a ser representações presentes na dinâmica social. Todos esses estudos teóricos relativos à organização e representação da informação são importantes para estabelecer padrões e normas visando representar o conhecimento e informação para que estes sejam recuperados

pelo usuário. Alguns instrumentos decorrentes dessas teorias são as linguagens documentárias (LD), que para Cintra (2002) são: os tesouros e os sistemas de classificação. Os mais conhecidos são de natureza enciclopédica, Classificação Decimal Universal (CDU), Classificação Decimal de Dewey (CDD), Classificação da Biblioteca do Congresso (LC) e posteriormente a Classificação Facetada de Ranganathan que visa atender particularidades não contempladas nas anteriores (CDD, CDU, LC). Surge então o tesouro oriundo das classificações facetadas com o objetivo específico de controle de vocabulário.

Vickery (1986), ao abordar a estruturação da representação do conhecimento, traz-nos uma reflexão enquanto profissionais da informação. Até que ponto nossos sistemas, catálogos e termos que usamos para traduzir ou representar um conteúdo estão sendo relevantes e consistentes? Será que o nosso usuário tem encontrado tudo que procura ou apenas uma quantidade razoável que lhe permita encontrar o que procura? É claro que essas indagações só poderão ser respondidas com um estudo e avaliação dos serviços prestados ao usuário, porém não podemos perder de vista a complexidade que envolve todo esse processo de representação do conhecimento. É provável que uma das causas da ineficiência dos serviços de informação seja a falta de compromisso de se criar sistemas especiais voltados aos tipos de necessidades informacionais e que atendam às complexidades da informação tratada e ao objetivo institucional. Além disso, a ordem de arquivamento deveria ser eficiente para aquela instituição, atendendo aos propósitos para que foi criada.

A representação do conhecimento é a organização do cognitivo expressa pela fala ou escrita (registro) de determinado fato e/ou objeto. Sendo assim, cabe ao profissional da informação estar atento à melhor forma de representar essa informação de modo consistente e relevante. Logo, enquanto profissionais da área da informação, é nosso papel escolher os métodos mais adequados e viáveis dentro de cada instituição.

O trabalho de Vickery mostra o crescimento dos estudos na área da representação da informação e ressalta também a relevância e a importância da continuidade do registro e da disseminação das pesquisas nessa área.

Para se representar a informação, é necessário que se conheça os conceitos que originam cada assunto, para fazer suas associações e subordinações adequadas, sempre refletindo a aplicação dessas divisões e métodos orientados na literatura e nos esquemas de classificação. Os profissionais encarregados dessa tarefa devem ter em conta não somente seus valores internos e expectativas pessoais, mas, sobretudo, as necessidades dos usuários. Se não for assim, para que organizamos o conhecimento? Ou para quem? Portanto, não podemos perder de vista a responsabilidade e o compromisso com a informação e com o usuário – ator principal e norteador de todo esse processo. Por isso, é premissa para que o conhecimento se perpetue e continue no topo do campo científico que é o lugar aonde nós, pesquisadores, iremos sempre buscá-lo.

Vickery (1986), ao analisar a estrutura semântica, afirma que nos exemplos dos tesouros as declarações de assuntos e índices são usadas de certa forma como uma informação adquirida, ou seja, a experiência do indexador é o que determinará o entendimento dos assuntos para ampliar as possibilidades de recuperação da informação pelos usuários. Outra abordagem conferida à representação do conhecimento é a redução da palavra dada ao significado mais original, ou seja, que origina a palavra. Vickery (1986) afirma que um exemplo disso em ciência da informação é o “fatoramento semântico”, que foi um sistema desenvolvido por Perry

e Kent. Eles estabeleceram uma lista contendo em torno de 200 termos primitivos que foram utilizados para representar aproximadamente 800 termos em Inglês, especificamente na área de ciência e tecnologia, “Cada termo era representado por um código de quatro letras” e as relações analíticas representadas pelos códigos eram: categórico, intrínseco, incluso, abrangente, produtivo, afetivo, instrumental, negativo, atributivo, simulador (VICKERY, 1986, p. 151).

Pela abordagem do autor, podemos perceber a preocupação dos estudiosos em criar formas de se representar e estruturar técnicas e padrões de se representar a informação, preparando-a para o sistema informatizado. Apesar de auxiliar em muito na disseminação da informação, o computador não faz tudo sozinho, sendo assim necessários especialistas em estruturação de conceitos e ordenação de termos para facilitar a representação do conhecimento. A tecnologia e seus efeitos sociais como agente de mudança na sociedade surgiu para suprir necessidades humanas, tendo imenso impacto, e suas contribuições vão além do computador, como é o caso dos telefones.

Essa necessidade tecnológica muda padrões comportamentais e cria novos padrões de comportamentos sociais ao incluir a sociedade na era da informação. Ao mesmo tempo exclui, por vários motivos, muitos usuários que poderiam ser reais em um sistema informacional, por desistirem de usar o sistema devido à complexidade que ele impõe. Assim, faltam pessoas qualificadas e preparadas para auxiliar um pesquisador, estudante ou qualquer que seja o indivíduo que busca informação por meio de catálogos eletrônicos. As informações não estão bem estruturadas e organizadas, e conseqüentemente bem representadas. Faltam nas instituições responsáveis pelo armazenamento e pela disseminação da informação, o que se constitui especial atenção em padronizar termos, estruturar assuntos de forma sistematizada e eficaz. A teoria do conceito é um exemplo de metodologia a ser seguida no que diz respeito à criação de termos padronizados necessários à recuperação dos assuntos abrangidos pela coleção OCRB, considerando-se também o potencial informativo identificado pelos indexadores.

### **Dahlberg: teoria do conceito**

A teoria do conceito veio possibilitar uma contribuição científica aos estudos terminológicos relacionados aos requisitos de definição, “e para seu posicionamento em um sistema de conceitos. O conceito não é mais apenas um elemento de significação do termo: o termo acaba sendo um elemento do próprio conceito” (CAMPOS, 2001, p. 100). Assim, já não importa se o termo é composto por um termo ou mais, se é composto por substantivos ou adjetivos, mas sim o que importa é o referente, ou seja, “tratar o termo como representante de um referente, com suas características, é dar a ele um tratamento terminológico” (CAMPOS, 2001, p. 100).

Um sistema de conceito revela sua relação com outros conceitos por meio das características próprias do referente, suas relações lógicas e ônticas (o superficial que fundamenta o senso comum, ou seja, o que todo mundo vê). Por essa razão, Dahlberg (1978) considera que a comunicação necessita de conceitos unívocos, e para tal é preciso avaliar os termos, seus significados, e designar graus de relevância.

A autora nos esclarece que com a ajuda das linguagens naturais, podemos formular enunciados sobre conceitos individuais e conceitos gerais. Os conceitos individuais são como o próprio nome diz, individuais (únicos). Um exemplo: Casa de Rui Barbosa, Rui Barbosa, etc. Já os conceitos gerais são os gerais, o que parece óbvio, mas se faz

necessário o esclarecimento na formação de conceitos. Como exemplo, temos: as instituições, os diplomatas brasileiros, etc.

A teoria do conceito revolucionou, simplificou e tem contribuído para a evolução das linguagens documentárias, consolidando sua estruturação teórica específica para a sua elaboração, e hoje é um instrumento fundamental na representação do conhecimento.

Para Dahlberg (1980 apud CAMPOS, 2001, p. 87), os fundamentos da teoria do conceito têm sido ferramentas úteis para se elaborar um tesouro, pois eles proporcionam segurança para se formar as relações de assuntos e também “para sua realização do plano verbal, ou seja, a determinação do que se denomina termo. Segundo esta teoria, as soluções para o termo e sua forma não são mais o ponto de partida, mas sim o ponto de chegada [...]”. Sem dúvida nenhuma, o tesouro é uma ferramenta eficaz e prática para o controle de vocabulários e representação do conhecimento, e, conseqüentemente, para a recuperação da informação. Os tesouros foram desenvolvidos para representar áreas específicas do conhecimento, logo são indicados para bibliotecas especializadas, e no caso do tratamento das obras de Rui, na área jurídica é o mais funcional.

De acordo com Brasil et al. (2002), com os catálogos *onlines* das bibliotecas digitais, virtuais e outros, o uso de instrumentos de controle terminológico para apoiar a indexação são fundamentais para a recuperação da informação, pois possuem mais eficiência para filtrar essas informações.

Brasil et al. (2002) nos esclarece que com o surgimento das novas TICs, os cabeçalhos de assunto ganharam relevância em relação às formas tradicionais de representação da informação, que ainda são muito utilizadas pela maioria das bibliotecas brasileiras. Porém, estudos como de Dahlberg (1978), Wuester (1981), Campos (2002), entre outros, vêm apontando que a unidade de uma representação temática não necessariamente pode ser feita pelo assunto, mas sim pela reunião de conceitos.

A teoria do conceito de Dahlberg (1978) é a base para o desenvolvimento deste trabalho e tem como princípio básico a representação do conhecimento. Dessa forma, partimos sempre do específico para o genérico, e quanto mais específico for o termo mais fácil será a recuperação do documento. Para tal, a FCRB aprimorou as técnicas de sistematização do seu vocabulário com estudos e pesquisas que norteiem um melhor caminho para a recuperação da informação. Vejamos a seguir a metodologia utilizada para elaboração desta pesquisa.

## **METODOLOGIA**

A coleção OCRB apresenta a compilação dos escritos de Rui Barbosa referentes ao período de 1865 a 1922. Abrange assuntos que ainda permeiam o atual contexto político-social do país e que são de grande interesse de pesquisadores e do público em geral, razão pela qual, faz-se necessário representar tematicamente os conteúdos mesmo diante do acesso aos documentos digitais. Os recursos de busca no texto podem não atender às necessidades informacionais, pois nem sempre a linguagem natural representará claramente os assuntos demandados pelos usuários do sistema de recuperação da informação. Nesse aspecto, enquadram-se, por exemplo, as figuras de linguagem, tais como a metáfora e a metonímia. Desse modo, também no sentido de melhor representar o conteúdo dos documentos, Lancaster (2004) menciona a importância do saber do indexador, ou seja, seu conhecimento de mundo, vivências e experiências que influenciarão diretamente no estabelecimento

da linguagem documentária. Ao representar tematicamente a coleção, o indexador precisa ser cuidadoso ao estabelecer as pontes entre o passado e o presente, de forma mais precisa possível, e mesmo com termos “atuais” retratando uma história em um século passado sem fazer um julgamento sobre o que se indexa, até porque, como afirma Vickery (1986), o que se passa em seu conhecimento pessoal se tornará um registro público.

A caracterização dos documentos pode ser feita em grau de maior ou menor importância, de acordo com o propósito que se tem. Contudo, as propriedades se tornam visíveis no uso do documento, não sendo, entretanto, possível uma profunda descrição do conteúdo somente pela formalização da linguagem. Para identificarmos as propriedades, deve haver uma argumentação explícita, prova ou probabilidade, não sendo passível de automatização, e sim altamente dependente de condições particulares (HJORLAND, 1992, p. 181-187).

Os assuntos são os potenciais epistemológicos de documentos, e um potencial é uma possibilidade objetiva. A descrição de assunto é um prognóstico de potenciais futuros, que pode ser baseado tanto nos julgamentos positivos como também nos julgamentos negativos, e, assim, o pré-requisito relevante na descrição de assunto é a maturidade no julgamento, e não um tipo especial de método. Podemos também afirmar que o usuário deve entrar no universo do sistema de recuperação da informação e entender a sua lógica. Compreender e interpretar requer capacidade de se colocar no lugar do outro, esforçando-se de forma sistemática e metódica, no sentido de conhecer a natureza do documento.

Portanto, o projeto propõe identificar conceitos, a partir do estudo da metodologia de indexação que se fundamenta na teoria do conceito, de Ingetraut Dahlberg (1978). Representar o conhecimento na forma simbólica é uma questão antiga no campo da documentação, e, portanto, é necessário decidir como o conhecimento pode ser representado de forma que as representações possam ser manipuladas (VICKEREY, 1986, p. 145). Desse modo, após os primeiros meses de estudo da fundamentação teórica, foi iniciada a fase empírica, com a revisão da descrição física e temática dos registros estabelecidos no sistema de recuperação da informação, referentes à coleção OCRB. Isto é, os trabalhos passaram a ser objeto de leitura, de revisão das informações sobre os aspectos físicos e de revisão da representação temática utilizando descritores já existentes ou estabelecendo novos termos no vocabulário sistematizado das bases bibliográficas. Por conseguinte, foram estabelecidos: forma verbal, definição, fontes, notas necessárias e relacionamentos, constituindo assim uma linguagem documentária que espelha a tematicidade da coleção, isto é, “um conjunto controlado de termos dotados de regras sintáticas e semânticas cujo objetivo” é representar os conceitos significativos dos assuntos dos documentos (FUJITA, 2004, p. 15).

Esperamos ter como resultados: o aperfeiçoamento profissional, no âmbito da indexação temática para acervos bibliográficos; o desenvolvimento dos descritores aplicados à indexação das obras que abordam temas diversos; o aprimoramento da indexação do acervo bibliográfico no âmbito das Obras Completas de Rui Barbosa; e a melhoria da recuperação da informação nos sistemas referenciais e digitais.

Assim sendo, para atingir o objetivo inicial, foi estudada uma metodologia que se constituiu nas etapas e atividades descritas a seguir:

## 1ª Etapa

- a) Contextualização: quem foi Rui Barbosa; como foi formada a coleção de trabalhos que compõem a produção intelectual de Rui; qual o contexto histórico de sua época; qual seria a importância dessas obras para a instituição; quem são os usuários dos acervos bibliográficos da FCRB. E a realização de estudo teórico de trabalhos que versem sobre a teoria do conceito, de Dahlberg (1978).
- b) Análise dos catálogos do acervo bibliográfico, da FCRB, para verificar algumas inconsistências já percebidas e mais bem visualizadas após a migração para o novo sistema de gerenciamento de bases de dados.
- c) Verificação quanto à correta descrição das informações documentárias, incluindo sua adequação aos campos da estrutura da base de dados.
- d) Realização de treinamento das bolsistas participantes, capacitando-as a efetuar as alterações necessárias e a inserção de novos registros no sistema de recuperação da informação.
- e) Nos registros bibliográficos referentes à coleção OCRB, análise dos descritores temáticos aplicados, conforme as concepções estudadas e referentes à metodologia adotada (teoria do conceito).

## 2ª Etapa

- f) Aspectos que envolvem o indexador: ao analisar a obra a ser indexada, levando em consideração o fato de se ter várias obras em um volume, ou seja, fazer os devidos desdobramentos de títulos e atribuir descritores conforme a temática abordada.
- g) Especificamente quanto à tematicidade, a análise minuciosa e crítica se faz necessária para identificar assuntos “camuflados”, como, por exemplo, um livro que trata da ditadura, mas está de forma velada na obra. Tal processo envolve uma leitura técnica das seguintes partes: título, resumo, sumário, introdução, conclusão. Além disso, o foco deve estar na especificidade, para evitar o alto índice de revocação.
- h) No processo de análise da obra, são atribuídos termos que representam os conceitos identificados. Portanto, serão utilizados termos já existentes no vocabulário estabelecido no sistema ou serão criados novos termos segundo a metodologia fundamentada na teoria do conceito.
- i) Quanto aos novos termos criados para atribuição às obras analisadas, esses serão incluídos na estrutura do vocabulário que apresenta como principais elementos os termos autorizados (Campo 150); tradução dos termos autorizados para o inglês, espanhol e francês (Campo 750); relações entre os termos (Campo 450 e 550); notas que apresentam a definição e, caso necessário, a aplicação do termo (Campo 680); e notas referentes às fontes pesquisadas na constituição do termo e demais informações (670 e 675).

## CONSIDERAÇÕES FINAIS

O potencial informativo de um documento é analisado em dois aspectos: os formais e os temáticos. Quanto ao aspecto temático, o indexador deve prezar pelo resgate de todo o potencial de assuntos constituindo pontos de acesso possíveis, ou seja, o conteúdo do documento deve ser analisado e desdobrado por meio do assunto/conteúdo, apresentando os aspectos específicos que visam possibilitar seu real significado, considerando representar o mais próximo do pensamento do autor, bem como a possibilidade de prever aplicações nas pesquisas realizadas pelos usuários de um sistema. Quanto ao aspecto formal, o indexador deverá analisar a informação com o objetivo de investigar tanto o objeto de análise quanto o processo de produção de um item informativo; observar a definição contextual; a intenção do documento levando em conta a coleção na qual estão inseridos, os objetivos organizacionais do serviço de recuperação da informação, bem como as necessidades de informação dos usuários. Vale ressaltar que uma coleção merece atenção especial na sua análise e descrição, por possuir características, como, por exemplo, uma nota; uma série. Essas informações irão individualizá-las, facilitando a busca realizada pelo usuário e, posteriormente, a recuperação.

O acervo bibliográfico da Fundação Casa de Rui Barbosa (FCRB) possui diversas obras nos assuntos diversos, em especial as que fazem parte do acervo da Biblioteca Rui Barbosa e a própria produção intelectual de Rui. Faz-se necessário aprimorar a indexação desse rico acervo, visando ampliar as possibilidades de busca das informações e, desse modo, melhorar os recursos de recuperação da informação nos sistemas de bases de dados referenciais e digitais.

Diante do exposto, a indexação não deve ser vista como uma prática mecânica, ela exige do indexador domínio do assunto da obra e maturidade para representá-lo, constituindo-se um conceito ou um conjunto de conceitos representados, refletindo de fato o conteúdo. Por isso, é fundamental o estudo permanente e a atualização constante desse profissional. A rotatividade de indexadores prejudica a qualidade da indexação, pois o compromisso e a técnica acabam ficando em segundo plano, tendo em vista as características profissionais de cada indivíduo. O treinamento deve ser uma prática contínua e ininterrupta para a qualidade da indexação.

Outra questão que deve ser observada é que o avanço tecnológico trás muitos benefícios, principalmente ao profissional da informação, porém essa mesma tecnologia torna o pesquisador mais exigente e com necessidades informacionais mais precisas, e isso impacta diretamente a comunicação científica. Ao impactar a comunicação científica, toda a estrutura de serviços oferecidos aos usuários de informação deve ser revista e estudada para que se encontre a melhor maneira de atender a esses usuários. Por isso, procurar rever a metodologia de indexação utilizada para o tratamento técnico dos materiais bibliográficos é uma atitude imperativa e de grande importância para o aperfeiçoamento dos sistemas de recuperação da informação.

Artigo recebido em 14/06/2018 e aprovado em 05/10/2018.

## REFERÊNCIAS

BASTOS, Dilza Ramos. *Em busca de uma metodologia de análise documental para as crônicas jornalísticas de Carlos Drummond de Andrade*. 2006. Dissertação (Mestrado

em Ciência da Informação) – Instituto Brasileiro de Informação e Tecnologia/ Universidade Federal Fluminense, 2006.

BRASIL, Maria Irene et al. Vocabulário sistematizado: a experiência da Fundação Casa de Rui Barbosa. In: CONGRESSO INTERNACIONAL DE ARQUIVOS, BIBLIOTECAS, CENTROS DE DOCUMENTAÇÃO E MUSEUS, 1., 2002, São Paulo. *Anais...* São Paulo: Imprensa Oficial do Estado, 2002. v. 1, p. 81-93.

BRASCHER, Marisa; CAFÉ, Lígia. Organização da informação ou organização do conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO: diversidade cultural e políticas de informação, 2008, São Paulo. *Anais...* São Paulo: Ancib, 2008. Disponível em: < <http://enancib.ibict.br/index.php/enancib/ixenancib/paper/viewFile/3016/2142>>. Acesso em: 4 nov. 2008.

CAMPOS, Maria Luiza de Almeida. *Linguagem documentária: teorias que fundamentam sua elaboração*. Niterói: EdUFF, 2001.

CINTRA, Anna Maria Marques et al. *Para entender as linguagens documentárias*. São Paulo: Pólis, 2002.

DAHLBERG, Ingetraut. Teoria do conceito. *Ciência da Informação*, Rio de Janeiro, v. 7, n. 2, p.101-107, 1978.

FEITOSA, Ailton. *Organização da informação na web: das tags à web semântica*. Brasília: Thesaurus, 2006.

FUJITA, Mariângela Spotti Lopes. A leitura documentária na perspectiva de suas variáveis: leitor-texto-contexto. *DataGramaZero: revista de ciência da informação*, v. 5, n. 4, art. 1, ago. 2004.

GIL LEIVA, Isidoro. *La automatización de la indización de documentos*. Gijón [Astúrias]: Eciciones Trea, 1999.

GREISDORF, Howard. Relevance: an interdisciplinary and information science perspective. *Information Science*, v. 3, n. 2, p. 67-71, 2000. Special issue on information science research.

HJØRLAND, B. The concept of “subject” in information science. *Journal of Documentation*, v. 48, n. 2, p. 172-200, 1992. DOI: <https://doi.org/10.1108/ebo26805>.

KNIGHT, Norman G. *Treinamento em indexação*. Rio de Janeiro: Fundação Getúlio Vargas, 1974.

LANCASTER, F. W. *Indexação e resumo: teoria e prática*. 2. ed. rev., ampl. e atual. Tradução de Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos/Livros, 2004.

NAVES, Madalena Martins Lopes. Estudos de fatores interferentes no processo de análise de assunto. *Perspectivas em ciência da informação*, Belo Horizonte, v. 6, n. 2, p. 189-203, jul./dez. 2001.

SARACEVIC, T. The concept of “relevance” in information science: a historical review. In: \_\_\_\_\_. *Introduction to information science*. New York: R. R. Bowker, 1970. p. 111-151.

VICKERY, B.C. Knowledge representation: a brief review. *Journal of Documentation*, v. 42, n. 3, p. 145-159, Sept.1986.

WÜESTER, E. L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses. In:

RONDEAU, G.; FELBER, F. (Org.). *Textes choisis de terminologie I: fondements théoriques de la terminologie*. Québec: Girsterm, 1981. p. 57-114.