

# Indexação em repositórios digitais: uma abordagem sobre o metadado assunto da Biblioteca Digital de Monografias da UFRN

Raimunda Fernanda dos Santos

Mestra em Ciência da Informação pela UFPE

E-mail: nandaflorania@gmail.com

## RESUMO:

Estuda aspectos relativos à representação temática da informação no contexto dos Repositórios Digitais. Objetiva analisar a representação da informação em tais ambientes. Especificamente, visa analisar as implicações da indexação por extração no processo de representação do conteúdo dos documentos na Biblioteca Digital de Monografias da Universidade Federal do Rio Grande do Norte (BDM/UFRN); analisar os descritores utilizados na indexação das monografias do curso de Biblioteconomia no metadado assunto da BDM/UFRN. Utiliza como metodologia a pesquisa bibliográfica pautada em revisão de literatura sobre indexação e representação da informação em repositórios digitais e pesquisa exploratória com abordagem qualitativa mediante a análise da indexação realizada no metadado assunto da BDM/UFRN. Destaca que os termos atribuídos em linguagem natural, ainda que elencados por autores da área de biblioteconomia viabiliza implicações negativas no processo de representação dos documentos em Repositórios Digitais. Identifica a presença de algumas implicações mediante a análise qualitativa do metadado "assunto" do Repositório Digital, dentre elas: emprego de termos com polissemia, ocorrência de erros ortográficos, presença de descritores abrangentes e com multiplicidade de sentidos, abreviações nos termos, uso de siglas e entre outras. Conclui enfatizando a importância dos gestores do Repositório Digital em questão elaborarem uma política de indexação que vise o aperfeiçoamento das práticas já utilizadas nesse ambiente incluindo o auxílio de linguagens documentárias e vocabulários controlados que possam elevar o nível da qualidade de descrição desses documentos visando a precisão no processo de busca e recuperação da informação.

**Palavras-chave:** Representação temática da Informação. Indexação. Metadado assunto. Repositórios Digitais.

## ABSTRACT:

Studies aspects of the thematic representation of information in the context of Digital Repositories. It aims to analyze the representation of information in such environments. Specifically, it aims to analyze the implications of extraction for indexing in the process of representation of the content of the documents in the Monographs Digital Library of the Federal University of Rio Grande do Norte (BDM / UFRN); analyze the descriptors used in the indexing of library course monographs on the subject of metadata BDM / UFRN. Use as methodology the literature guided by literature review of indexing and information representation in digital repositories and exploratory research with qualitative approach by analyzing the indexing performed on metadata subject of BDM / UFRN. It points out that the terms given in natural language, although listed by the authors of the library area, enables negative

implications on the document representation process in the digital environment. Identifies the presence of some implications by qualitative analysis of metadata "subject" of the Digital Repository, among them: employment terms with polysemy, occurrence of misspellings, the presence of comprehensive descriptors and multiplicity of meanings, abbreviations under, use of acronyms and so on. It concludes by emphasizing the importance of the Digital Repository managers in question develop an indexing policy aimed at the improvement of the practices already used in this environment including the aid of documentary languages and controlled vocabularies that can raise the description of quality level of these documents aimed at precision process of search and retrieval of information.

**Keywords:** Thematic representation of information. Indexing. Metadata subject. Digital Repositories.

## 1 INTRODUÇÃO

A explosão da informação e o advento das novas formas de registrá-la culminaram na complexidade do trabalho de identificação e acesso à informação. Nesse entendimento, tornar as informações acessíveis para os usuários que delas necessitam constituiu-se uma das preocupações centrais da Ciência da Informação. Assim sendo, para que os registros documentais sejam recuperados, é necessária a realização de um conjunto de procedimentos que visam facilitar o seu acesso pela atual e futura geração.

Nessa perspectiva, a indexação consiste na descrição do conteúdo de um documento, de forma concisa e condensada, por meio do emprego de termos - também denominados como palavras-chave ou descritores - que exercem a função de pontos de acesso mediante os quais um documento pode ser identificado e recuperado. Destarte, para a realização dessa atividade existem as linguagens documentárias/vocabulários controlados os quais são instrumentos tradicionais de representação da informação que objetivam facilitar a comunicação por meio da padronização de termos para a descrição dos conteúdos dos documentos.

Contudo, diante do crescente aumento na produção de documentos e do desenvolvimento de produtos e serviços que beneficiam a participação dos usuários na coordenação de ações de produção, organização, representação e disseminação

da informação na *Web*<sup>1</sup>, é evidenciada uma nova forma de organização e representação das informações no âmbito dos Repositórios Digitais em que não se adotam regras de indexação envolvendo controle de vocabulário na descrição dos recursos.

Sendo assim, em linhas gerais, observa-se que na atividade de indexação das produções científicas em Repositórios Digitais Institucionais é permitida a utilização dos mesmos termos empregados pelo autor para a apresentação das suas ideias no texto (ou seja, linguagem natural) para a indexação do seu trabalho no Repositório - selecionando-se assim as palavras-chave empregadas no resumo do documento para representá-lo, configurando-se como uma indexação por extração de acordo com Lancaster (2004).

Dessa forma, o uso da indexação por extração e as suas implicações, como demais questões que daí decorrem, oferecem inúmeras possibilidades de discussões, uma delas refere-se à sua utilização no contexto dos Repositórios Digitais como a Biblioteca Digital de Monografias da Universidade Federal do Rio Grande do Norte (BDM/UFRN).

A BDM/UFRN objetiva promover o acesso e uso das monografias produzidas no âmbito dos cursos de Graduação, bem como divulga e preserva parte da memória acadêmica da Universidade Federal do Rio Grande do Norte. Para tanto, a sua política de indexação permite a utilização dos mesmos termos informados pelos autores nas palavras-chave dos resumos de suas monografias para representar o conteúdo dos trabalhos monográficos - configurando-se como uma indexação por extração mediante o emprego de termos livres para representar a informação.

Sob esse viés, a presente pesquisa tem como objetivo geral analisar a representação da informação em Repositórios Digitais. Especificamente, visa analisar as implicações da indexação por extração no processo de representação do conteúdo dos documentos na Biblioteca Digital de Monografias da Universidade Federal do Rio Grande do Norte (BDM/UFRN); avaliar os descritores utilizados na representação do conteúdo das monografias do curso de Biblioteconomia no metadado<sup>2</sup> assunto da Biblioteca Digital em questão visando verificar como os autores desta área do conhecimento estão descrevendo o conteúdo de suas produções

---

<sup>1</sup> A *World Wide Web* (ou simplesmente *Web*) está relacionada à palavra inglesa cujo significado é “rede de alcance mundial”. É um sistema de informações interligado e executado na Internet.

<sup>2</sup> Dado que “descreve informações de um recurso informacional em meio digital ou não digital” (ROSETTO, 2003, p. 9).

intelectuais nos Repositórios Digitais, levando em consideração que os discentes desse curso estudam aspectos relativos à organização, representação e recuperação da informação.

A importância dessa investigação decorre, em âmbito geral, das tendências de representação da informação em Repositórios Digitais e uma das principais justificativas para a realização desta pesquisa é o fato da representação do assunto, nessas ferramentas de acesso aberto como a BDM/UFRN, ocorrer de forma livre e sem padronização.

Nessa concepção, a escolha dessa fonte de informação é relativa à importância que ela representa para a comunidade acadêmica da instituição, haja vista que é por meio dessa ferramenta depositária que é armazenada, preservada e divulgada parte da memória acadêmica da Universidade Federal do Rio Grande do Norte. Outrossim, a seleção desse universo está atrelada à sua política de indexação cuja atividade de indexação orientada foi mencionada anteriormente.

Torna-se relevante também pesquisar aspectos relativos à representação da informação no Repositório Digital supracitado levando em conta que ela consiste em uma ferramenta dinâmica que subsidia alunos, professores e pesquisadores a encontrar as produções acadêmicas construídas no âmbito da instituição, possibilitando também a preservação histórica dos documentos nesse contexto.

## 2 METODOLOGIA

Quanto aos objetivos a serem alcançados para este estudo, foi empreendida uma pesquisa do tipo exploratória. Quanto aos meios, os procedimentos metodológicos foram empreendidos a partir de uma pesquisa bibliográfica, baseada em revisão de literatura sobre as implicações do emprego de termos provenientes de linguagem natural no processo de representação da informação. Foram utilizadas na concepção do trabalho fontes impressas como livros, e em meio digital como artigos de periódicos científicos eletrônicos da área de Ciência da Informação encontradas por meio de bases de dados, mecanismos de busca e bibliotecas digitais na Web.

Além disso, o presente estudo possui caráter qualitativo tendo em vista que foi feita uma investigação pela literatura para seu melhor entendimento, bem como foram analisados, essencialmente, alguns aspectos qualitativos referentes à representação da informação no metadado assunto das monografias do curso de Biblioteconomia

depositadas na comunidade do Centro de Ciências Sociais aplicadas na BDM/UFRN. Para tanto, a análise qualitativa do metadado em questão foi realizada mediante a verificação das palavras-chave empregadas pelos autores das monografias do curso citado anteriormente.

Até o dia 25 de julho de 2016 a comunidade de biblioteconomia na Biblioteca Digital de Monografias da UFRN apresentou 211 monografias depositadas no sistema. Tais trabalhos constituem o *corpus* de documentos analisados na presente pesquisa.

## 2 AS PRÁTICAS DE INDEXAÇÃO E AS SUAS INTERFACES COM OS REPOSITÓRIOS DIGITAIS

Uma das preocupações principais na Ciência da Informação consiste em tornar a informação acessível para aqueles que dela necessitam haja vista que os itens informacionais devem passar por um conjunto de tarefas de armazenamento, organização e representação para que seja possível a sua recuperação em tais sistemas (acesso).

Sob esse viés, Guimarães (2009, p. 1) afirma que no âmbito da organização da informação dois universos se descortinam: "o primeiro, ligado ao acesso aos documentos e o segundo, de natureza mais complexa, voltado para o acesso ao conteúdo informacional, genericamente denominado de tratamento temático da informação" ou representação temática da informação, como é designada na literatura da área de organização da informação.

Nessa perspectiva, verifica-se que o documento deve ser tratado em seu todo levando em consideração tanto as operações que contemplam a descrição física (tratamento descritivo ou representação descritiva), como a descrição do conteúdo temático (tratamento temático ou representação temática) haja vista que a diferenciação entre a forma e o conteúdo "[...] reside na busca do *o que* (materialização) e do *sobre o que* (teor) que convivem no âmbito do documento" (GUIMARÃES, 2009, p. 1, grifo nosso).

Nesse entendimento, a representação temática da informação centra-se nas questões relativas "à análise, descrição e representação do conteúdo dos documentos, bem como suas inevitáveis interfaces com as teorias e sistemas de armazenamento e recuperação da informação" (BARITÉ, 1999, p. 124). A indexação,

por sua vez, consiste em uma atividade integrante da representação temática da informação e que apresenta etapas de caráter integrativo.

Sob esse viés, a indexação consiste em uma atividade integrante do tratamento temático da informação, ou seja, uma "combinação metodológica altamente estratégica entre o tratamento do conteúdo do documento e a sua recuperação por um usuário", demonstrando uma relação direta entre o processo e a finalidade da indexação, visando a disponibilização de informações pertinentes às necessidades dos usuários (FUJITA, 2003, p. 61).

Mediante tais considerações, de acordo com Dias e Naves (2007), a indexação consiste na ação de descrever um recurso em relação ao seu conteúdo por meio de termos (também denominados palavras-chave ou descritores). Nesse entendimento, em um Repositório Digital, os termos são pontos de acesso para os documentos, haja vista que é possível recuperar esses objetos informacionais sob qualquer uma das suas palavras-chave (que foram empregadas no sistema no ato da sua representação) ou até mesmo por meio da combinação delas. Em outras palavras, esses termos são atribuídos pelo indexador (pessoa responsável por realizar a atividade de indexação) e exercem a função de pontos de acesso mediante os quais um documento é identificado e recuperado.

Sendo assim, verifica-se que a indexação, além de ser uma atividade que objetiva a recuperação dos documentos, também estabelece uma ligação no que concerne à comunicação entre o documento e o usuário. Dessa forma, a indexação depreende duas etapas principais: *análise de assunto* (também denominada de análise conceitual, análise documentária, análise de conteúdo) e *tradução*.

Para Dias e Naves (2007), a análise de assunto diz respeito à extração de conceitos que representem o conteúdo de um documento e, em nítido contraste, a etapa da tradução consiste na representação dos documentos em termos provenientes de instrumentos de indexação - os quais são denominados de linguagens de indexação ou linguagens documentárias na vertente francesa. Depois de traduzidos para uma linguagem de indexação, tais termos passam a ser chamados de descritores de assunto, rótulos, cabeçalhos de assunto, termos de indexação ou enunciados.

Apesar dos avanços tecnológicos, observa-se a necessidade da presença de um indexador humano nessa tarefa levando em conta que a indexação consiste em uma atividade intelectual considerada como o limite da máquina, tendo em vista que

a máquina ainda não pode realizar o trabalho humano de abstrair, distinguir, compreender e assimilar o que está presente em um texto. Portanto, todas as fases do processo de indexação sofrem influências de fatores linguísticos, cognitivos e lógicos (simultaneamente), os quais são designados por Dias (2001) como o tripé de sustentação do processo de análise de assunto - gerando assim um caráter interdisciplinar.

Lancaster (2004) realiza a distinção entre *indexação por extração* (também conhecida como indexação derivada) e *indexação por atribuição*. Para esse autor, na *indexação por extração* as palavras ou expressões empregadas pelo autor no documento analisado são selecionadas pelo indexador para representar o seu conteúdo. Em outras palavras, nesse tipo de indexação empregam-se os mesmos termos utilizados pelo autor para a apresentação das suas ideias no texto (ou seja, uma linguagem natural), selecionando-se assim palavras encontradas no título, nas palavras-chave ou na própria estrutura textual do documento. Já a *indexação por atribuição* compreende a atribuição de descritores ao documento utilizando-se uma linguagem originária de uma fonte que não é o próprio documento (LANCASTER, 2004).

Sendo assim, são determinados os cabeçalhos a serem empregados objetivando-se controlar sinônimos, diferenciar homógrafos<sup>3</sup> e ligar termos por meio de relações hierárquicas, associativas e de equivalência (ou seja, utiliza-se uma linguagem artificial/linguagem controlada). De acordo com Lancaster (2004, p.19), "a indexação por atribuição envolve o esforço de representar a substância da análise conceitual mediante o emprego de termos extraídos de alguma forma de vocabulário controlado".

Diante do exposto, para esse autor, a linguagem natural configura-se como discurso comum, ou seja, é uma linguagem usualmente empregada na escrita e na fala, diferentemente das linguagens documentárias ou vocabulários controlados que consistem em linguagens construídas de acordo com as necessidades dos usuários. Analogicamente, no âmbito dos Repositórios Digitais, os termos presentes no título, resumo e palavras-chave dos documentos apresentam linguagem natural.

Destarte, Lopes (2002, p.48) apresenta em sua pesquisa as vantagens e desvantagens da linguagem natural, as quais são elencadas no Quadro 1 a seguir.

---

<sup>3</sup> Termos que coincidem na sua forma, porém representam conceitos diferentes.

**Quadro 1- Linguagem natural: vantagens e desvantagens**

VANTAGENS	DESVANTAGENS
Permite o imediato registro da informação em um repositório digital, sem necessidade de consulta a uma linguagem de controle.	Os usuários da informação, no processo de busca, precisam fazer um esforço intelectual maior para identificar os sinônimos, as grafias alternativas, os homônimos etc.
O processo de busca é facilitado com a ausência de treinamentos específicos no uso de uma linguagem de controle.	Haverá alta incidência de respostas negativas ou de relações incorretas entre os termos usados na busca (por ausência de padronização).
Os termos de entrada de dados são extraídos diretamente dos documentos que vão constituir o Repositório Digital.	Os custos de acesso tendem a aumentar com a entrada de termos de busca aleatórios.
Os temas específicos citados nos documentos podem ser encontrados.	Uma estratégia de busca que arrole todos os principais conceitos e seus sinônimos deve ser elaborada para ferramenta (ex.: nomes comerciais de substâncias químicas não ocorrem no <i>Chemical Abstracts</i> ).
Elimina os conflitos de comunicação entre os indexadores e os usuários, pois ambos terão acesso aos mesmos termos.	Perda de confiança do usuário em uma possível resposta negativa.

Fonte: Lopes (2002, p.48).

Nesse sentido, o emprego de termos livres na indexação de documentos no âmbito dos Repositórios Digitais pode incitar algumas implicações negativas no processo de representação da informação, levando em consideração que as informações disponíveis em tais ambientes são descritas com palavras provenientes da linguagem natural sem o auxílio de um instrumento de representação da informação. Para tanto, os usuários representam conteúdos com termos que usam e/ou acreditam que entendem, sendo assim alguns termos podem fazer sentido apenas a quem os elencou e nem sempre para a comunidade ou grupo de pessoas que utilizam o sistema.

Sob esse viés, depreende-se que com a liberdade da atribuição dos descritores um mesmo termo pode ter diferentes significados para aqueles que atribuíram as etiquetas/descriptores para fins de representação do conteúdo.



Nessa concepção, Brandt e Medeiros (2010, p. 120) acrescentam que uma das implicações do emprego de linguagem natural no processo de descrição de recursos em um sistema aberto é a **falta de padronização** haja vista a ausência de regras para o emprego de termos. Destarte, o pensamento das autoras supracitadas conduz à reflexão de que a liberdade para a representação de conteúdos implica na descentralização do processo de representação da informação.

Noruzi (2007) em sua pesquisa menciona algumas implicações negativas dessas práticas processo de representação da informação em ambientes digitais, dentre as quais:

**a) ocorrência de plurais** - ausência de padronização em relação aos termos com flexões de número (singular ou plural): palavras com flexões gramaticais e variações linguísticas de gêneros nos descritores utilizados, por exemplo;

**b) existência de polissemia:** palavra que representa mais de um significado;

**c) emprego de sinonímia:** palavras distintas que têm o mesmo significado;

**d) profundidade/ especificidade da descrição:** refere-se à especificidade dos descritores elencados pelo usuário no processo de representar o conteúdo do objeto informacional em tags. Tais descritores (em sua maioria) podem não representar o objeto em sua totalidade, sendo elencados apenas para interesse particular do usuário que o atribuiu.

Além das implicações supracitadas, Amstel (2007), Gouvêa; Loh (2007) e Catarino; Baptista (2009); Santos (2013) enfatizam que em ambientes que utilizam a representação colaborativa da informação é comum a **presença de descritores homógrafos** - os quais são termos que coincidem na sua forma, porém representam conceitos diferentes; **erros ortográficos**; **expressões regionais** (que são empregadas como *tags* de acordo com o contexto social/regional em que o usuário estar inserido); **etiquetas em diferentes idiomas e com letras maiúsculas e minúsculas** (sem padronização na descrição) e **etiquetas assimétricas** (tags que não se relacionam entre si) e abreviações.

Outrossim, alguns termos atribuídos são dotados de **ambigüidade morfológica**, ou seja, "pertencem a mais de uma categoria gramatical, como proposta, que pode ser substantivo ou adjetivo ou verbo" (BRASCHER, 2002, p.4)

impossibilitando classificá-los pelo fato de ter formas mínimas portadoras de significado concorrentes para uma mesma palavra.

Nessa concepção, para fins de indexação existem as linguagens documentárias/vocabulários controlados os quais são instrumentos que visam facilitar a comunicação mediante a padronização de termos para a descrição de conteúdos. Diante disso, verifica-se que tais instrumentos objetivam controlar as dispersões lexicais existentes nas linguagens naturais orientando também o indexador no que concerne à escolha dos termos para representar o assunto dos documentos, bem como indicar os melhores termos para que ele possa realizar a indexação do texto, atuando também como instrumento de comunicação entre a linguagem dos usuários e a linguagem do sistema.

Lancaster (2004) enfatiza que o tipo de vocabulário controlado (esquema de classificação, cabeçalhos de assuntos, tesauro) não é o aspecto mais importante a influenciar na etapa de tradução da indexação. Muito mais importante é a sua *abrangência* (alcance) e a sua *especificidade*. Porém, de acordo com Café e Sales (2009), no que concerne à representação temática da informação por meio de um controle terminológico, os tesauros e as ontologias<sup>4</sup> são considerados como os principais instrumentos utilizados e estudados no âmbito da Ciência da Informação.

Tais instrumentos possibilitam a realização da representação temática intimamente relacionada com a ideia do autor do documento a ser indexado; a ampliação dos métodos de busca, bem como o controle dos termos que apresentam polissemia e sinonímia; a proximidade entre a linguagem natural (do usuário) e a linguagem adotada pelo sistema de recuperação de informação.

A partir de tais considerações, observa-se que o tratamento temático da informação, mediante o uso dos processos e instrumentos desenvolvidos em sistemas tradicionais, consiste em uma tarefa indispensável para a recuperação da informação em ambientes digitais. Sob esse viés, se faz necessário atentar-se para as práticas de indexação realizadas no âmbito dos Repositórios Digitais institucionais como a Biblioteca Digital de Monografias da Universidade Federal do Rio Grande do Norte (BDM/UFRN).

---

<sup>4</sup> Ontologias são instrumentos de indexação que apresentam uma estrutura de termos e as relações entre eles na perspectiva do sujeito e da linguagem de um determinado domínio. Apresentam-se como alternativas viáveis para organizar o conhecimento em ambiente Web na forma de uma teia de relações permitindo a ligação entre os conceitos.

A BDM/UFRN objetiva promover o acesso e uso das monografias produzidas no âmbito dos cursos de graduação da instituição, bem como divulga e preserva parte da memória acadêmica da Universidade Federal do Rio Grande do Norte. Para tanto, a sua política de indexação permite a utilização dos mesmos termos informados pelos autores nas palavras-chave dos resumos de suas monografias para representar o conteúdo dos trabalhos monográficos nos metadados do sistema- configurando-se como uma indexação por extração de acordo com Lancaster (2004).

Nesse segmento, levando em consideração os aspectos teóricos enfatizados anteriormente e objetivando revisar tais questões na prática, será apresentada a seguir os resultados da análise do metadado assunto da comunidade do curso de Biblioteconomia na Biblioteca Digital de Monografias da Universidade Federal do Rio Grande do Norte, cuja política de indexação adotada foi mencionada anteriormente.

### **3 ANÁLISE DO METADADO ASSUNTO DA BIBLIOTECA DIGITAL DE MONOGRAFIAS DA UFRN**

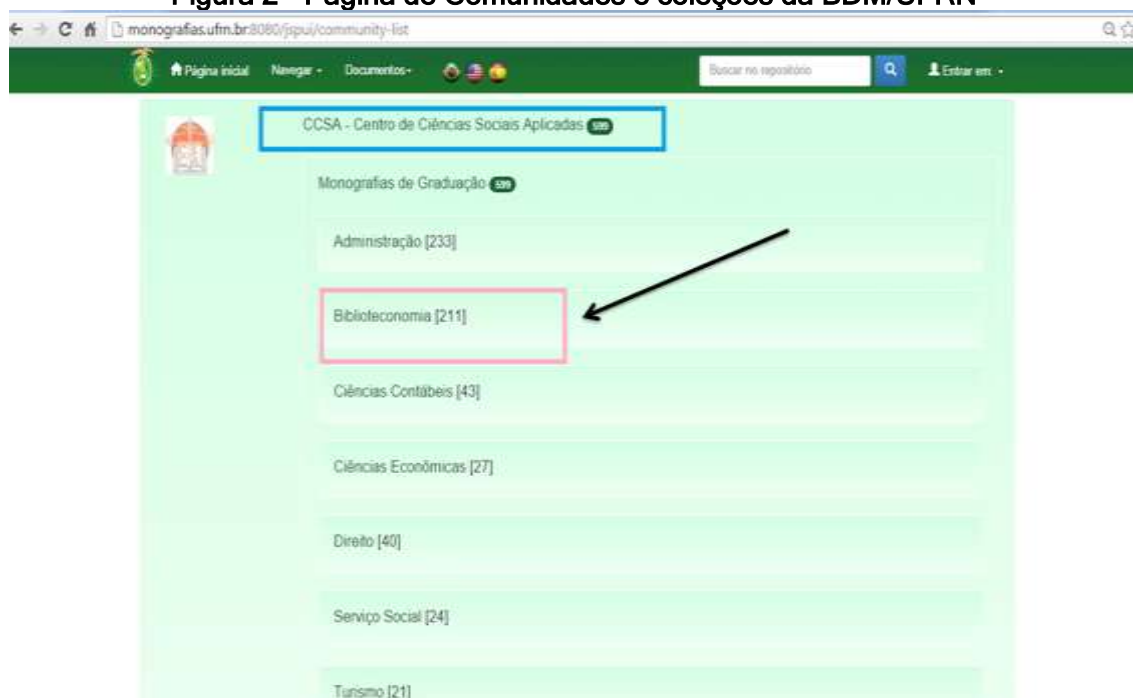
Após a análise do Repositório em evidência verificou-se que o processo de busca e recuperação das monografias neste ambiente ocorre de maneira clara e simples, haja vista que são disponibilizados alguns mecanismos de acesso em sua página principal, tais como: busca por comunidades e coleções, data do documento, nome dos autores, título e assunto - conforme pode ser visualizado na Figura 1.

Figura 1 - Página principal da BDM/UFRN



Fonte: < <http://monografias.ufrn.br:8080/jspui/>>

Ao clicar na opção de navegação “Comunidades e coleções” é possível visualizar a lista de comunidades, subcomunidades e coleções presentes na BDM/UFRN. Destarte, dentre elas é encontrada a comunidade do Centro de Ciências Sociais Aplicadas (CCSA) que agrega monografias dos seguintes cursos: Administração, Biblioteconomia, Ciências Contábeis, Ciências Econômicas, Direito, Serviço Social e Turismo - totalizando 599 (quinhentos e noventa e nove) monografias presentes neste grupo conforme pode ser visualizado na Figura 2 a seguir.

**Figura 2 - Página de Comunidades e coleções da BDM/UFRN**

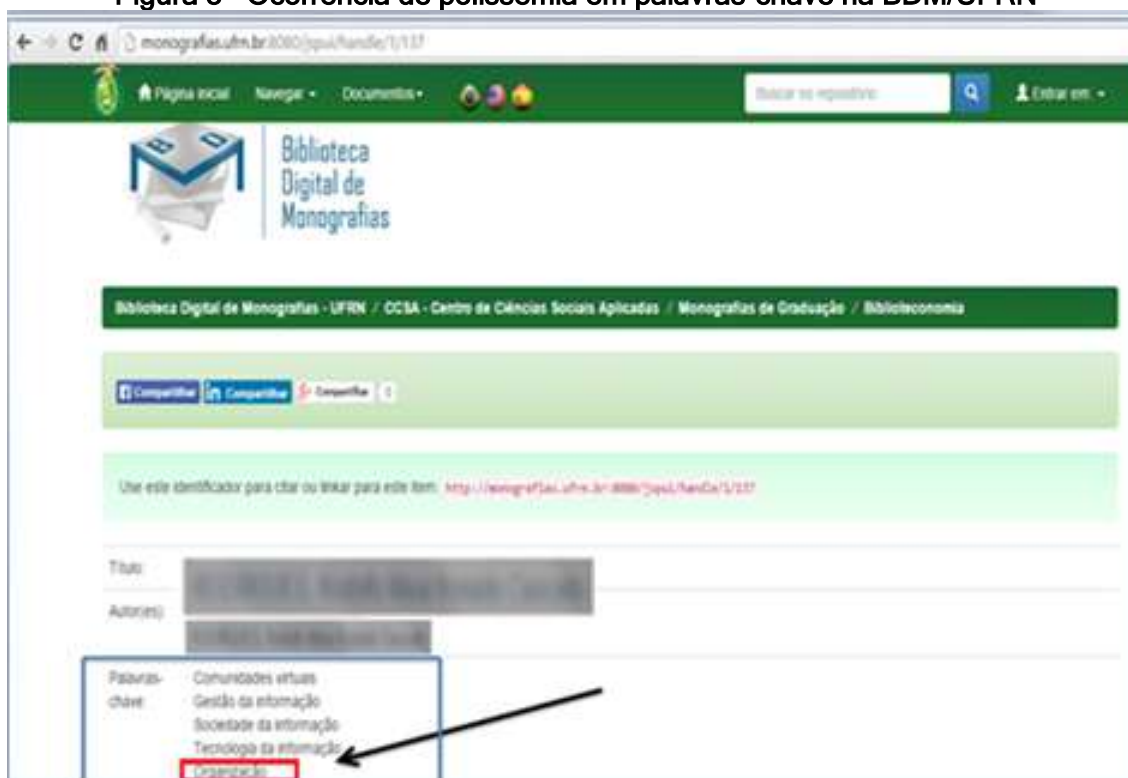
**Fonte:** < <http://monografias.ufrn.br:8080/jspui/community-list>>.

A partir do interesse em compreender e visualizar aspectos práticos relativos às implicações da indexação por extração no processo de representação da informação considerou-se importante realizar uma análise qualitativa da indexação no metadado “assunto” da BMD/UFRN por meio da análise das palavras-chave empregadas pelos autores das monografias do curso de Biblioteconomia (cuja comunidade agrega 211 monografias inseridas no sistema).

Nesse segmento, essa tarefa possibilitou pesquisar também como os autores do curso de Biblioteconomia (que possuem conhecimentos relativos à organização, representação e recuperação da informação) estão representando os conteúdos de suas produções intelectuais através da livre representação da informação (emprego de termos em linguagem natural) no metadado assunto do repositório.

É importante ressaltar que nas figuras a seguir foram resguardados dados como os nomes dos autores, títulos dos trabalhos e os respectivos orientadores das monografias visando uma análise ética da representação de conteúdos nesse contexto.

**Figura 3 - Ocorrência de polissemia em palavras-chave na BDM/UFRN**



Fonte: <<http://monografias.ufrn.br:8080/jspui/handle/1/137>>

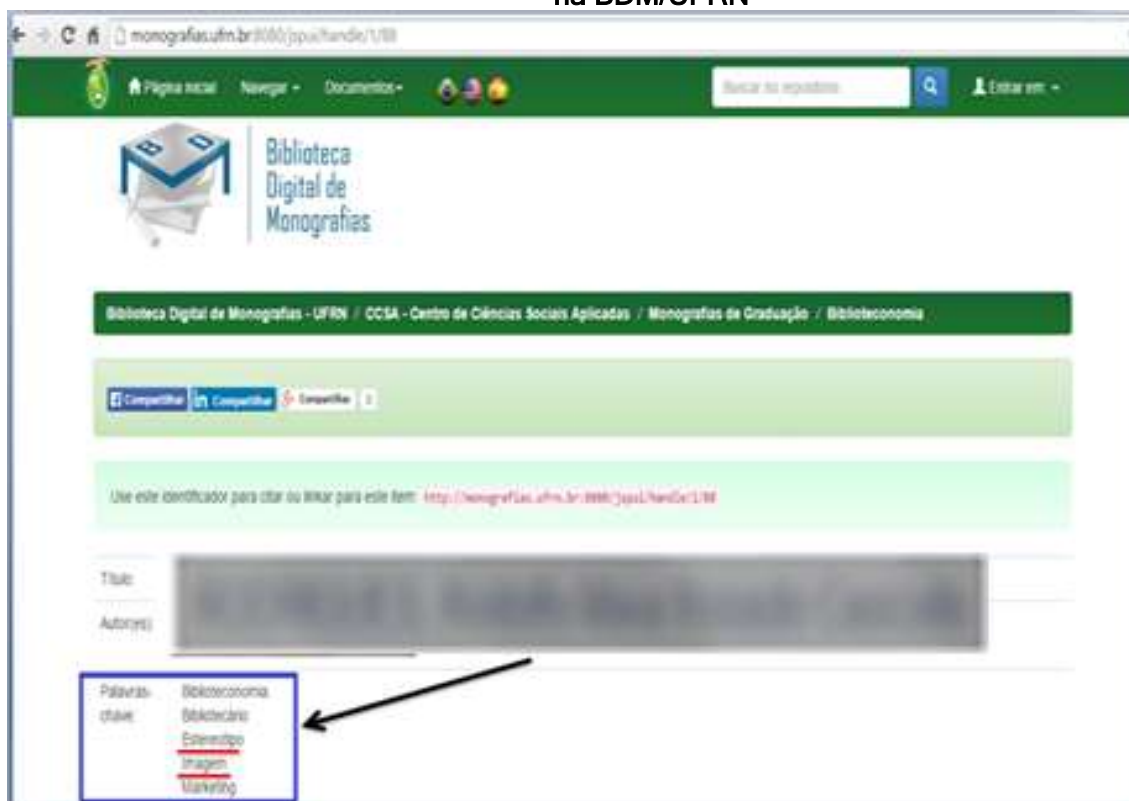
Conforme é observado na Figura 3, o autor do trabalho monográfico apresentado anteriormente elencou 5 (cinco) palavras-chave/descriptores para representarem a sua produção intelectual. Dentre os termos foi empregado o descritor “Organização” considerado como um termo polissêmico, haja vista que essa palavra está vinculada a (pelo menos) dois significados, são eles: Organização no sentido de Ordem (Ordenação/Arrumação) e no sentido relativo a uma Associação (Entidade/Instituição).

Nessa concepção, a representação do conteúdo de um documento por uma palavra polissêmica (como apresentado na Figura 3) e a ausência de controle desse fenômeno pode resultar em problemas na sua recuperação. Em consonância de evidências, quando o usuário do repositório digital em questão realizar uma busca com o termo “Organização”, essa ferramenta pode recuperar documentos relativos a entidade/instituição, como também outros trabalhos monográficos que abordem aspectos relativos à ordenação/arrumação de algo.

De outro modo disposto, além da ocorrência de palavras polissêmicas como implicações provenientes da utilização da indexação por extração, constatou-se também a presença de erros ortográficos ao analisar o metadado “assunto” das

monografias presentes na BDM/UFRN. Tais implicações negativas são destacadas na Figura 4 a seguir.

**Figura 4 - Ocorrência de polissemia e erro ortográfico em palavras-chave na BDM/UFRN**



Fonte: <<http://monografias.ufrn.br:8080/jspui/handle/1/88>>.

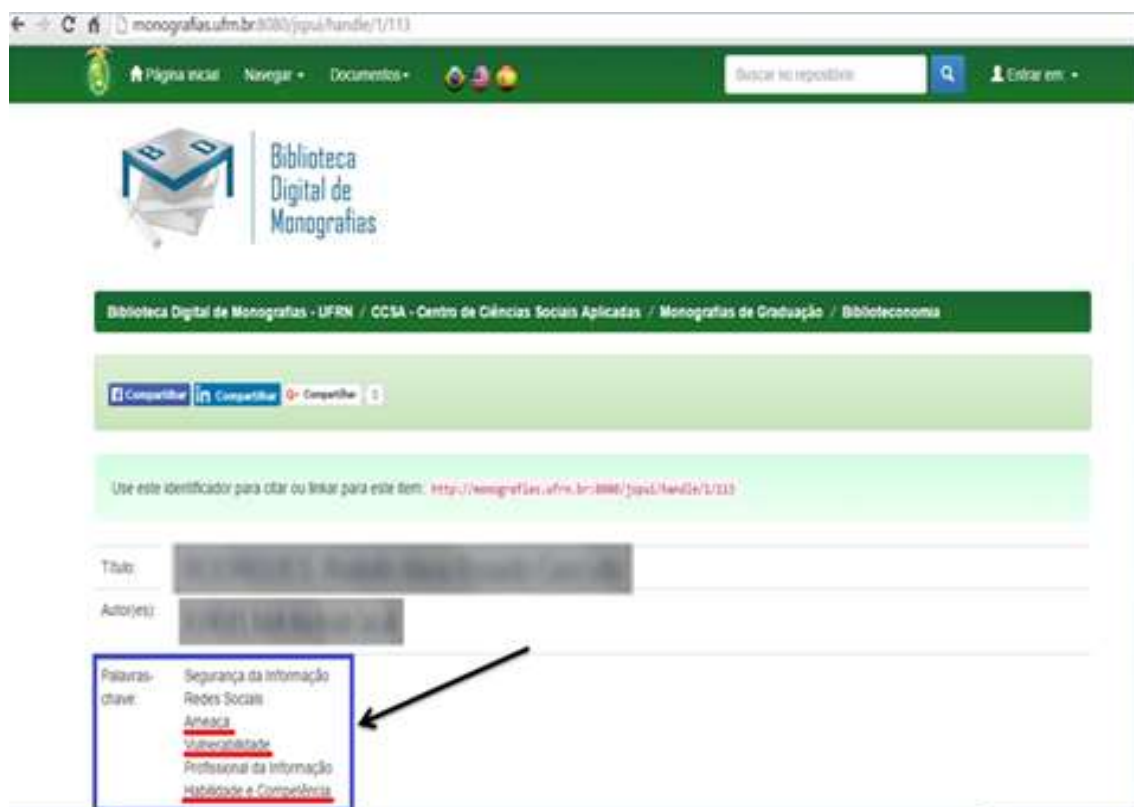
Como pode ser visualizado, foram aplicados 5 (cinco) descritores na descrição do conteúdo da monografia supramencionada, dentre eles os termos “Estereotipo” e “Imagem”. Verificou-se que o termo “Estereotipo” possui erro ortográfico, podendo também implicar negativamente no processo de recuperação da informação no Repositório Digital em questão. Os acentos gráficos são necessários para auxiliar a compreensão de um termo, sendo assim a grafia correta desse termo seria “Estereótipo” (com acentuação gráfica).

De outro modo disposto, o descritor “Imagem” consiste em uma palavra generalista e polissêmica tendo em vista que está vinculado a mais de um significado, podendo assim interferir também no processo de recuperação da informação. Nesse sentido, a palavra em evidência apresenta alguns significados, dentre os quais: “Imagem” no sentido de Representação de uma coisa ou pessoa

(figura/gravura/ilustração) e “imagem” no sentido relativo a uma ideia (abstração/recordação/visão).

Ademais, outro aspecto vinculado ao emprego de linguagem natural (de acordo com o que foi abordado nos aspectos teóricos deste estudo) é o emprego de termos generalistas ou com multiplicidade de sentidos. No Repositório Digital analisado verificou-se a ocorrência frequência desse fenômeno na indexação de trabalhos monográficos. A figura 5, por exemplo, apresenta a indexação de um trabalho com mediante a aplicação dos termos “Ameaça”, “Vulnerabilidade”, “Habilidade e Competência” para a sua descrição. Assim, percebe-se que esses termos podem estar inseridos em diversos contextos sob enfoques e filtros diversos.

**Figura 5 - Ocorrência de termos generalistas em palavras-chave da BDM/UFRN**



**Fonte:** < <http://monografias.ufrn.br:8080/jspui/handle/1/113>>.

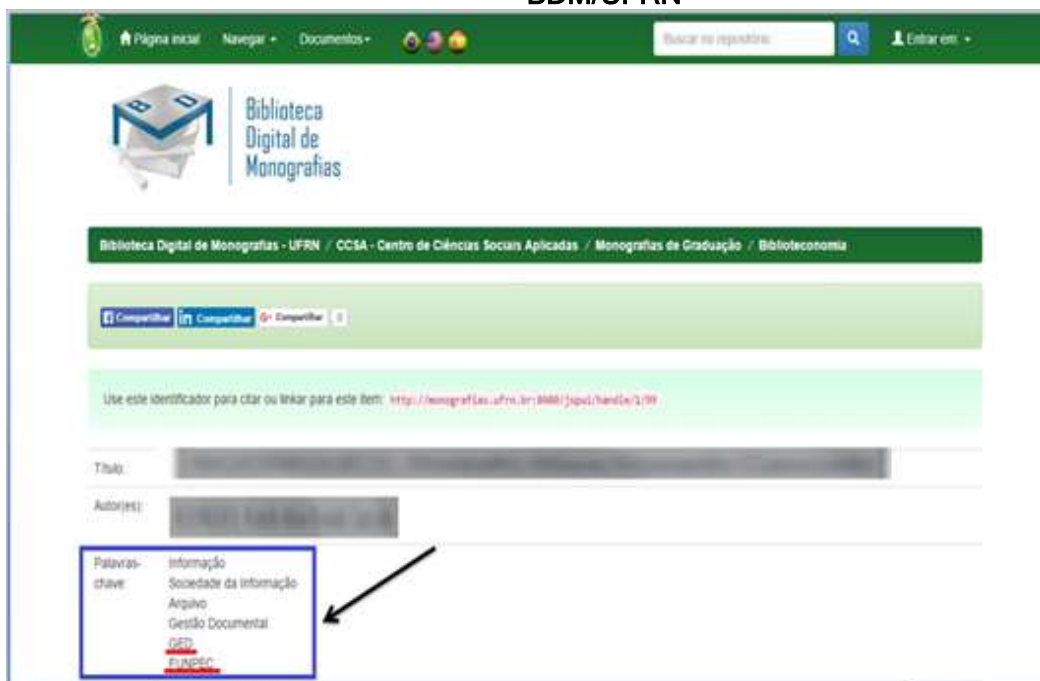
Sendo assim, observa-se a importância da aplicação de termos específicos, precisos e coerentes com o conteúdo do documento indexado, pois termos generalistas contribuem para a ocorrência de falhas e equívocos na recuperação da informação, possibilitando que o repositório digital retorne documentos que não condizem com as necessidades dos seus usuários. Além disso, a aplicação de



descritores genéricos na representação da informação é uma questão que precisa ser observada/analísada, uma vez que esses termos também servirão de base para a comunidade da área de conhecimento em questão (que por sua vez possui terminologia específica) para fins de recuperação da informação e, possivelmente, auxílio para o desenvolvimento de outras pesquisas. Destarte, essa consideração também é direcionada para a aplicação de siglas e abreviações em descritores tendo em vista que essa atividade implica negativamente no processo de recuperação dos documentos.

A Figura 6 a seguir certifica a ocorrência de abreviações de palavras (por meio das siglas “GED” e FUNPEC”) no metadado assunto para a representação do conteúdo de uma monografia do curso de Biblioteconomia presente na BDM/UFRN:

**Figura 6 - Emprego de siglas e abreviações em palavras-chave na BDM/UFRN**



**Fonte:** < <http://monografias.ufrn.br:8080/jspui/handle/1/99>>.

Por essa razão, a ausência de controle dos fenômenos elencados anteriormente pode resultar em problemas de revocação e precisão (medidas de desempenho mencionadas anteriormente) em Repositórios Digitais. As palavras-chave atribuídas, bem como a terminologia adotada são fundamentais para o sucesso de qualquer sistema, independentemente de sua forma e estrutura. Nessa perspectiva, o desenvolvimento das tecnologias exige do profissional da informação além de conhecimentos específicos na sua área, outros conhecimentos e habilidades

relativas ao tratamento da informação em suportes e ambiências diversificadas como os Repositórios Digitais esboçando assim novos campos de competência profissional.

## 5 CONSIDERAÇÕES FINAIS

Mediante tais considerações, a partir dos estudos realizados, observou-se nesta pesquisa que os termos atribuídos em linguagem natural - ainda que elencados por autores da área de biblioteconomia (concluintes do curso) que estudam em sua graduação aspectos relativos à classificação, armazenamento, disseminação, recuperação e acesso às informações - viabiliza implicações no processo de representação e recuperação dos documentos em ambiente digital.

Destarte, na análise qualitativa do metadado “assunto” da Biblioteca Digital de Monografias da Universidade Federal do Rio Grande do Norte na coleção do curso de Biblioteconomia verificou-se a presença de algumas implicações, dentre elas: emprego de termos com polissemia, ocorrência de erros ortográficos, presença de descritores abrangentes e com multiplicidade de sentidos, abreviações nos termos, uso de siglas e entre outras.

Diante da análise supracitada, infere-se a importância dos gestores da BDM/UFRN (e dos demais Repositórios Digitais Institucionais que adotam práticas semelhantes no que concerne à indexação dos seus documentos) elaborarem uma política de indexação que vise o aperfeiçoamento das práticas já utilizadas neste ambiente (utilização das palavras-chave empregadas pelos autores no resumo de suas monografias no metadado “assunto” do repositório) incluindo o auxílio de linguagens documentárias e vocabulários controlados que podem elevar o nível da qualidade de descrição do conteúdo desses documentos visando, conseqüentemente, a precisão no processo de busca e recuperação da informação neste ambiente.

Nessa perspectiva, a política de indexação deve ser considerada como um elemento de grande importância não somente para designar as práticas de representação da informação, como também para servir de auxílio nas decisões administrativas dos Repositórios Digitais Institucionais. Desse modo, diante das implicações elencadas anteriormente, torna-se relevante que os administradores do Repositório Digital em questão (bem como os administradores de repositórios

digitais que adotam políticas semelhantes) repensem o processo de indexação dos documentos na ferramenta.

Tais ações podem ser voltadas para a preocupação na construção de modelos colaborativos de indexação social, os quais são descritos por Santos (2016)<sup>5</sup> como modelos que delineiam atividades colaborativas de representação de conteúdo a serem executadas pelos usuários - como estratégias de indexação dos conteúdos dos recursos em um sistema - e servem de referência para o desenvolvimento de sistemas colaborativos de indexação, podendo ser adaptados ou aperfeiçoados a diversos contextos. Esses modelos delineiam a interação e comunicação entre os seus usuários na realização de atividades que compõem o processo de indexação, que podem ocorrer em diferentes cardinalidades: de um para um; um para muitos; muitos para um; ou de muitos para muitos.

Para a autora citada anteriormente, a alternativa do emprego de um modelo colaborativo de indexação social é controlar o nível de liberdade do usuário autor do documento ao atribuir termos na indexação de conteúdos em Repositórios Digitais para que o processo de iteratividade possa resultar na reconstrução do conhecimento de forma coletiva e moderada - permitindo uma nova estrutura de organização e representação de informação mais consistente para o domínio em que o sistema está inserido. Outrossim, essas propostas permitem que os termos/descriptores tenham uma relação maior no que concerne ao conteúdo do objeto depositado, principalmente dentro do contexto em que elas estão sendo empregadas.

Nessa perspectiva, verifica-se que a percepção da atuação dos usuários enquanto mentes tradutoras e idealizadoras de novos arranjos e categorizações informacionais, desafia os profissionais da informação a repensarem a construção de sistemas e metodologias frente à representação da informação em ambiente digital, sobretudo no âmbito dos Repositórios Digitais que se constituem como ferramentas depositárias e disseminadoras de informação.

---

<sup>5</sup> Mais informações acerca dos modelos colaborativos de indexação social ler a Dissertação Referenciada a seguir: SANTOS, Raimunda Fernanda dos. **Modelos colaborativos de indexação social e a sua aplicabilidade na Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI)**. 184f. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal de Pernambuco, Recife, 2016.

Nesse entendimento, necessita-se que novos processos, metodologias e instrumentos venham a ser desenvolvidos no âmbito dos Repositórios Digitais - seja aplicando técnicas para a ampliação ou aperfeiçoamento dos termos fornecidos pelos autores no metadado assunto dessas ferramentas, seja adicionando os termos oriundos das linguagens documentárias. Tais reflexões se constituem como objetivos de estudo dos trabalhos atuais e futuros da pesquisadora.

Portanto, acredita-se que o presente trabalho é relevante no contexto da Ciência da Informação uma vez que aborda aspectos representação da informação no âmbito dos Repositórios Digitais e a relaciona com os desafios do profissional da informação frente às novas possibilidades de representação da informação, tendo em vista que é possível aperfeiçoar as técnicas já utilizadas, além de adaptá-las às necessidades da sociedade atual.

## REFERÊNCIAS

AMSTEL, Van Frederick. **Folksonomia**: vocabulário descontrolado na arquitetura da informação ou samba do crioulo doido. 2007. Disponível em: <[http://www.guilhermo.com/ai\\_biblioteca/referencia.asp?referencia=317](http://www.guilhermo.com/ai_biblioteca/referencia.asp?referencia=317)> . Acesso em: 20 maio 2015.

BARITÉ, M. **Formación de recursos humanos en el área de información en el Mercosur**: compatibilización curricular y competencias del profesional de la información en el Mercosur. Santiago, Chile: Universidad Tecnológica Metropolitana, 1999, p.121-128.

BIBLIOTECA DIGITAL DE MONOGRAFIAS DA UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE. Disponível em: < <http://monografias.ufrn.br:8080/jspui/>>. Acesso em 10 jul. 2016.

BOCCATO, V. R. C. O Contexto sociocognitivo do indexador no processo de representação temática da informação. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v.17, n. esp., 2012.

BRANDT, Mariana Baptista ; MEDEIROS, Marisa Brascher Basílio. Folksonomia: esquema de representação do conhecimento? **TransInformação**, Campinas, v.22, n.2, p.111-121, maio/ago., 2010. Disponível em: < <http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/489>>. Acesso em: 13 out. 2013.

BRASCHER, M. A ambiguidade na recuperação da informação. **DataGramZero - Revista de Ciência da Informação**, Rio de Janeiro, v.3, n.1, fev. 2002.

CAFÉ, L.; SALES, R. de. Tesouros e ontologias sob o olhar da teoria comunicativa da terminologia. In: CONGRESS ISKO-SPAIN, 9., 2009, Valência. **Anais...** Valência, 2009.

CATARINO, M.; BAPTISTA, A. Folksonomias: características das etiquetas na descrição de recursos da web. **Inf. Inf.**, Londrina, v.14, n. esp., p.46-67, 2009.

DIAS, E. W. Contexto digital e tratamento da informação. **DataGramZero**, Rio de Janeiro, v. 2, n. 5, out. 2001. Disponível em: <[http://www.dgz.org.br/out01/Art\\_01.htm](http://www.dgz.org.br/out01/Art_01.htm)>. Acesso em: 25 set. 2014.

DIAS, E. W.; NAVES, M. M. L. **Análise de assunto: teoria e prática**. Brasília: Thesaurus, 2007.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia & Ciência da Informação**, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003. Disponível em: <[http://server01.bc.unicamp.br/revbib/artigos/art\\_5.pdf](http://server01.bc.unicamp.br/revbib/artigos/art_5.pdf) >. Acesso em: 26 jun. 2016.

GOUVÊA, C.; LOH, S. Folksonomias: identificações de padrões na seleção de tags para descrever conteúdos. **Revista Eletrônica de Sistemas de Informação**, [S.l.], v. 11, n.2, 2007.

GUIMARÃES, J. A. C. **Abordagens teóricas de Tratamento Temática da Informação (TTI): catalogação de assunto, indexação e análise documental**. Ibersid, 2009, p.105-117.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2.ed. Brasília: Briquet de Lemos, 2004.

LOPES, I. L. Uso das linguagens controlada e natural em base de dados: revisão de literatura. **Ci. Inf.**, Brasília, v.31, n.1, p.41-52, jan./abr. 2002. Disponível em: <[http://eprints.rclis.org/14449/1/linguagens\\_controladas.pdf](http://eprints.rclis.org/14449/1/linguagens_controladas.pdf)>. Acesso em: 11 set. 2014.

NORUZI, A. Folksonomies: Why do we need controlled vocabulary? **Webology**, v.4, n.2, 2007. Disponível em: <<http://www.webology.org/2007/v4n2/editorial12.html>>. Acesso em: 20 ago. 2013.

ROSETTO, M. Metadados e recuperação da informação: padrões para bibliotecas digitais. In: CIBERNÉTICA: SIMPÓSIO INTERNACIONAL DE PROPRIEDADE INTELECTUAL, INFORMAÇÃO E ÉTICA, 2., Florianópolis. **Anais...** Florianópolis: [s.n.], 2003.

SANTOS, H. P. Etiquetagem e Folksonomia: o usuário e sua motivação para organizar e compartilhar informação na web 2.0. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.1, n.2, p.91-104, abr./jun. 2013.

SANTOS, Raimunda Fernanda dos. **Modelos colaborativos de indexação social e a sua aplicabilidade na Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI)**. 184f. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal de Pernambuco, Recife, 2016.