

XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2018

GT 02 – ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

PRÁTICAS DE INDEXAÇÃO EM REPOSITÓRIOS DIGITAIS DE ACESSO ABERTO: ANÁLISE DO METADADO ASSUNTO DO REPOSITÓRIO INSTITUCIONAL DA UFRN

Raimunda Fernanda dos Santos- Universidade Federal da Paraíba (UFPB)

Dulce Amélia de Brito Neves – Universidade Federal da Paraíba (UFPB)

*PRACTICES OF INDEXING IN DIGITAL REPOSITORIES OF OPEN ACCESS: ANALYSIS OF THE
METADATE SUBJECT OF THE INSTITUTIONAL REPOSITORY OF UFRN*

Modalidade da Apresentação: Comunicação Oral

Resumo: Estuda aspectos relativos à indexação no contexto dos Repositórios Digitais. Objetiva analisar as práticas de representação temática da informação no Repositório Institucional da Universidade Federal do Rio Grande do Norte (RI/UFRN). Especificamente, visa investigar as implicações da indexação por extração no processo de representação do conteúdo dos documentos nesse contexto; avaliar os descritores utilizados na indexação das produções científicas no metadado “assunto” do Repositório em questão. Utiliza como metodologia a pesquisa bibliográfica pautada em revisão de literatura sobre as práticas de indexação em ambientes digitais e pesquisa exploratória com abordagem qualitativa mediante a análise da indexação realizada no metadado assunto do RI/UFRN. Destaca que os termos atribuídos em linguagem natural viabilizam implicações negativas no processo de representação temática 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 frases, termos com polissemia, presença de descritores abrangentes e com multiplicidade de sentidos, abreviações nos termos, uso de siglas e entre outras. Conclui enfatizando a importância dos gestores dos Repositórios Institucionais de acesso aberto elaborarem uma política de indexação que vise o aperfeiçoamento das práticas já utilizadas nesses ambientes incluindo o auxílio de instrumentos de controle terminológico que possam elevar o nível da qualidade de descrição das produções científicas visando a precisão no processo de busca e recuperação da informação.

Palavras-chave: Representação Temática da Informação. Indexação por extração. Repositórios Institucionais.

Abstract: Studies aspects related to indexing in the context of Digital Repositories. It aims to analyze the practices of thematic representation of information in the Institutional Repository of the Federal University of Rio Grande do Norte (RI / UFRN). Specifically, it aims to investigate the implications of indexing by extraction in the process of representing the content of the documents in that context; to evaluate the descriptors used in indexing the scientific productions in the "subject" metadata of the Repository in question. It uses as a methodology the bibliographic research based on literature review on indexing practices in digital environments and exploratory research with a qualitative approach through the analysis of the indexation performed in the subject metadata of the RI / UFRN. It emphasizes that the terms attributed in natural language allow negative implications in the process of thematic representation of the documents in Digital Repositories. It identifies the presence of some implications through the qualitative analysis of the "subject" metadata of the Digital Repository, among them: use of phrases, terms with polysemy, presence of comprehensive descriptors and multiplicity of meanings, abbreviations in the terms, use of acronyms and among others . It concludes by emphasizing the importance of the managers of the Open Access Institutional Repositories to develop an indexation policy aimed at improving the practices already used in these environments, including the use of terminological control tools that can raise the level of the quality of description of scientific productions aiming at the accuracy in the information retrieval and retrieval process.

Keywords: Thematic Representation of Information. Indexing by extraction. Institutional repositories.

1 INTRODUÇÃO

A Ciência da Informação se constitui em um campo interdisciplinar cuja preocupação central consiste em tornar as informações acessíveis para os usuários que delas necessitam, mediante um conjunto de procedimentos que objetivam facilitar a busca e a recuperação dos itens informacionais, independente do seu formato (análogo ou digital) e da sua tipologia (áudio, texto, imagem, vídeo, etc.). Para tanto, os objetos informacionais devem passar por um conjunto de tarefas de armazenamento, organização e tratamento para que seja possível a sua recuperação, acesso e uso em sistemas de recuperação da informação.

Nesse entendimento, esses itens devem ser tratados levando em consideração as operações que contemplam a sua representação descritiva (cuja atividade também é denominada de tratamento descritivo) e a descrição do seu conteúdo informacional (tratamento temático ou representação temática) para fins de acesso e uso.

A indexação se configura como uma atividade integrante do tratamento temático da informação, que visa descrever os conteúdos dos itens informacionais por meio de termos – também denominados de palavras-chave- provenientes de instrumentos de controle terminológico para fins de disponibilização das informações pertinentes às necessidades dos usuários (FUJITA, 2003).

Todavia, mediante o crescente aumento na produção de informações e do desenvolvimento de produtos e serviços que beneficiam a participação ativa dos usuários nas

atividades de produção, organização, representação e disseminação da informação na web¹, 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.

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 de Acesso Aberto como o Repositório Institucional da Universidade Federal do Rio Grande do Norte (RI/UFRN).

O RI/UFRN objetiva promover o acesso e uso a produção intelectual da comunidade universitária (docentes, técnicos e discente de pós-graduação) da Universidade, bem como divulga e preserva parte da memória acadêmica da instituição, uma vez que a sua missão consiste em armazenar, preservar e disponibilizar na internet, textos completos de acesso livre. Esse Repositório se insere internacionalmente no crescente movimento de constituição de Repositórios Digitais de acesso aberto nos moldes da *Open Access Initiative* (Iniciativa de Acesso Aberto). Para tanto, a sua política de indexação permite a utilização dos termos livres informados pelos autores dos trabalhos, cujos descritores podem ser as palavras-chave extraídas dos resumos dos seus trabalhos para representar o conteúdo das produções científicas - configurando-se como uma indexação por extração mediante o emprego de termos livres para representar a informação.

Nesse segmento, esta pesquisa tem como objetivo geral analisar as práticas de indexação no Repositório Institucional da Universidade Federal do Rio Grande do Norte (RI/UFRN). Especificamente, visa investigar as implicações da indexação por extração no processo de representação do conteúdo das produções científicas nesse contexto; avaliar os

¹ 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.

descritores utilizados no metadado² “assunto” do Repositório Institucional em questão visando verificar como os autores (docentes, técnicos e discentes de pós-graduação) estão descrevendo o conteúdo de suas produções intelectuais nos Repositórios Digitais.

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 o Repositório Institucional da 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 da 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 AS PRÁTICAS DE INDEXAÇÃO EM AMBIENTES DIGITAIS

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.

Nessa perspectiva, 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

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

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.

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).

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". Assim, nesse tipo de operação são determinados os cabeçalhos a serem empregados objetivando-se controlar sinônimos, diferenciar homógrafos³ e ligar termos por meio de relações hierárquicas, associativas e de equivalência (ou seja, utiliza-se uma linguagem artificial/linguagem controlada).

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

³ Termos que coincidem na sua forma, porém representam conceitos diferentes.

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.

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 Re却itório 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 arrolle 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 Re却itórios 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.

Assim, depreende-se que com a liberdade da atribuição dos descritores um mesmo termo pode ter diferentes significados para aqueles que atribuíram para fins de representação do conteúdo.

Noruzi (2007) em sua pesquisa menciona algumas implicações negativas dessas práticas no 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 utilizado; **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**: especificidade dos descritores elencados pelo usuário no processo de representar o conteúdo do objeto informacional, podendo não representar o recurso 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 linguagem natural é comum a **presença de descritores homógrafos** - termos que coincidem na sua forma, porém representam conceitos diferentes; **erros ortográficos**; **expressões regionais** (empregadas 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 **termos assimétricos** (termos que não se relacionam entre si) e abreviações.

Outrossim, alguns termos atribuídos são dotados de **ambiguidade 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, instrumentos que visam facilitar a comunicação mediante a padronização de termos para a descrição dos conteúdos. Esses instrumentos podem 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.

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 o Repositório Institucional da UFRN.

Nesse segmento, levando em consideração os aspectos teóricos enfatizados anteriormente e objetivando revisitar tais questões na prática, serão apresentados a seguir os resultados da análise do metadado “assunto” do Repositório Institucional da Universidade Federal do Rio Grande do Norte, cuja política de indexação adotada foi mencionada anteriormente.

3 PROCEDIMENTOS METODOLÓGICOS

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 práticas de indexação em ambientes digitais e 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.

Na atividade de indexação das produções científicas armazenadas no RI/UFRN são utilizados os mesmos termos empregados pelo autor nas palavras-chave do resumo do seu trabalho (ou seja, linguagem natural) no metadado “assunto” desse Repositório, configurando-se como uma indexação por extração.

Portanto, 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, aspectos qualitativos referentes à representação da informação no metadado “assunto” das produções científicas depositadas nas comunidades do RI/UFRN. Para tanto, a análise qualitativa do metadado em questão foi realizada mediante a verificação das palavras-chave utilizadas na indexação das produções científicas armazenadas no Repositório em questão.

Até o dia 20 de maio de 2018 o Repositório Institucional da UFRN apresentou 26.149 (vinte e seis mil, cento e quarenta e nove) produções científicas depositadas no sistema, dentre elas: artigos publicados em periódicos; trabalhos completos apresentados em eventos; Dissertações e Teses defendidas pelos docentes da UFRN em outras instituições; Dissertações e Teses defendidas e publicadas na UFRN; livros eletrônicos e capítulos de livros.

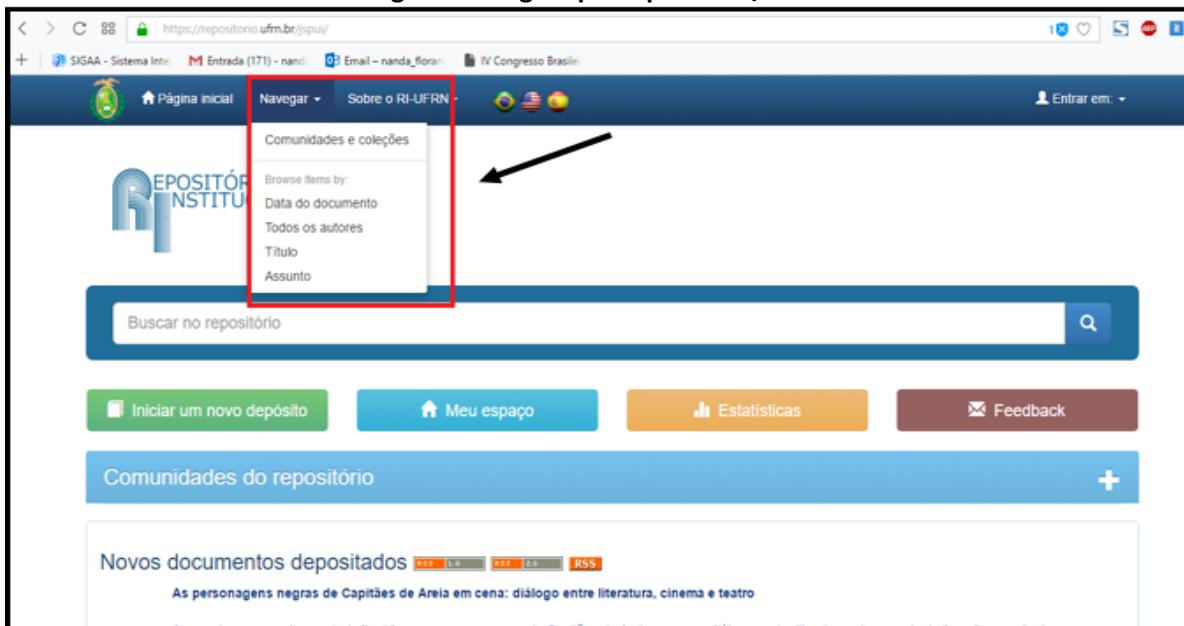
Através de amostra aleatória simples foi selecionado para a análise desta pesquisa 10% do total das produções científicas armazenadas no RI/UFRN, ou seja, 2.614 (dois mil, seiscentos e quatorze) trabalhos dentre os quais: artigos publicados em periódicos; trabalhos completos apresentados em eventos; Dissertações e Teses defendidas pelos docentes da UFRN em outras instituições; Dissertações e Teses defendidas e publicadas na UFRN; livros eletrônicos e capítulos de livros.

A seguir são apresentados os resultados desta pesquisa mediante a realização da análise do metadado “assunto” do Repositório Institucional da Universidade Federal do Rio Grande do Norte (RI/UFRN).

4 ANÁLISE DO METADADO “ASSUNTO” DO REPOSITÓRIO INSTITUCIONAL DA UFRN (RI/UFRN)

Após a análise do repositório em evidência verificou-se que o processo de busca e recuperação das produções científicas 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 do RI/UFRN



Fonte: <<https://repositorio.ufrn.br/jspui/>>

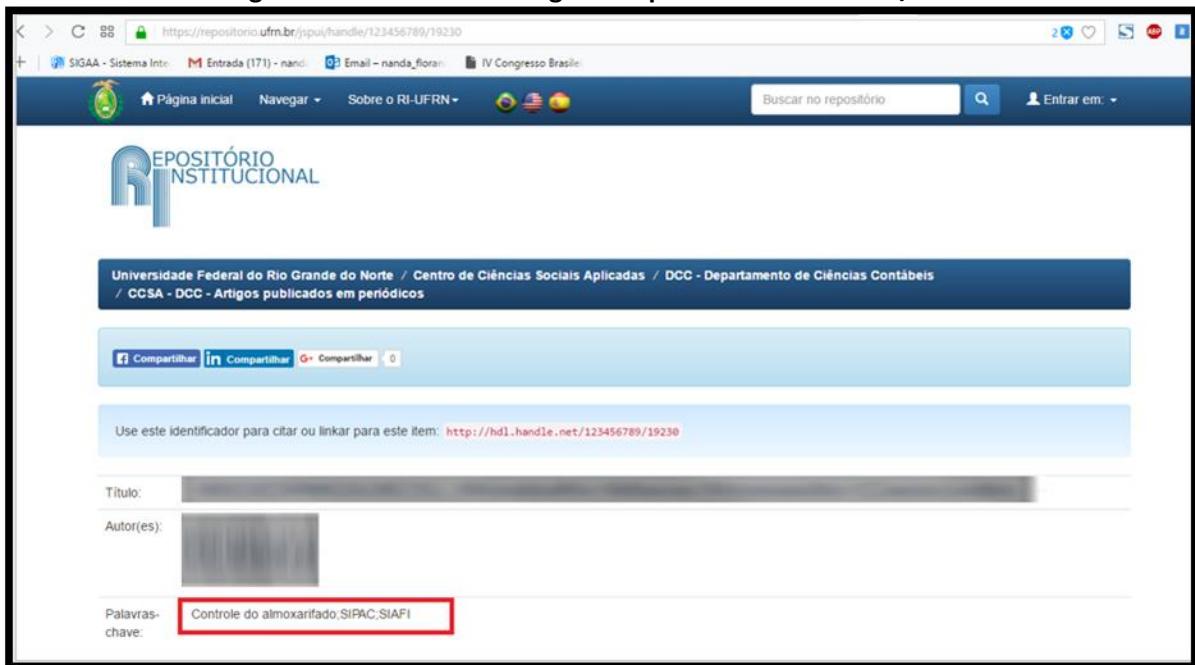
Ao clicar na opção de navegação “Comunidades e coleções” é possível visualizar a lista de comunidades, subcomunidades e coleções presentes no RI/UFRN. Em tais comunidades são armazenados os seguintes documentos: a) artigos publicados em periódicos; b) trabalhos completos apresentados em eventos; c) Dissertações e Teses defendidas pelos docentes da UFRN em outras instituições; d) Dissertações e Teses defendidas e publicadas na UFRN; e) livros eletrônicos; f) capítulos de livros.

Através do interesse em compreender e visualizar aspectos práticos relativos às implicações da indexação por extração no processo de representação temática da informação nesse ambiente, considerou-se importante realizar uma análise qualitativa no metadado “assunto” do RI/UFRN por meio da análise das palavras-chave empregadas pelos autores das produções científicas incluídas no Repositório.

Nesse segmento, essa tarefa possibilitou pesquisar também como os autores estão representando os conteúdos de suas produções intelectuais através do uso de termos provenientes da linguagem natural.

É importante ressaltar que nas figuras abaixo foram resguardados dados como os nomes dos autores e os respectivos orientadores das produções científicas visando uma análise ética da representação dos conteúdos nesse ambiente.

Figura 3 – Ocorrência de siglas em palavras-chave no RI/UFRN



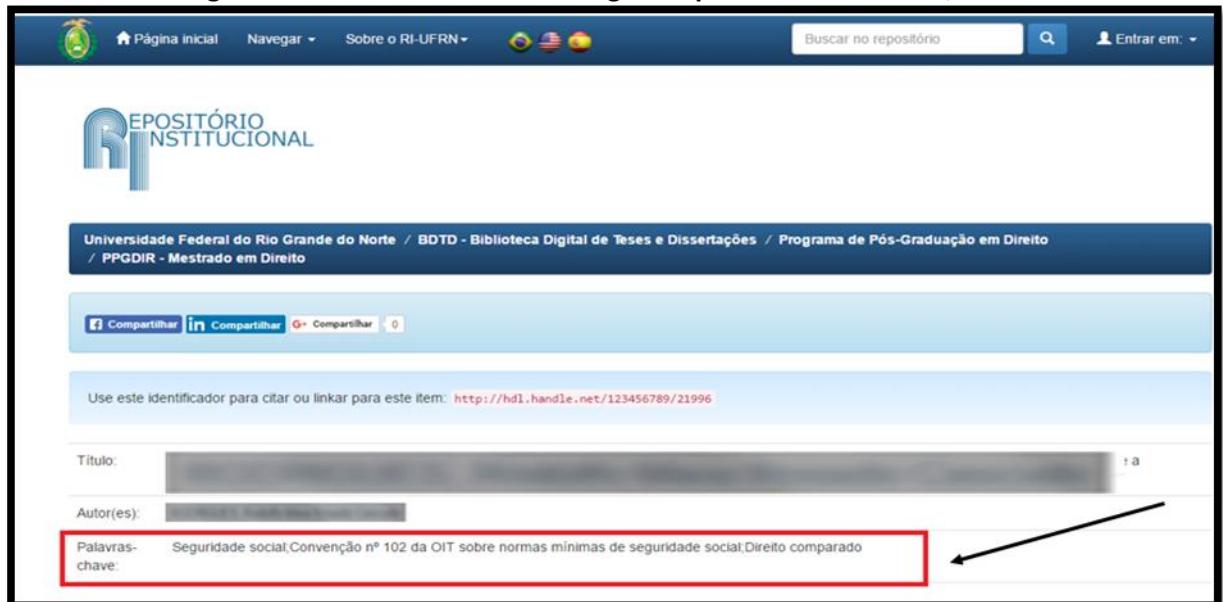
Fonte: < <https://repositorio.ufrn.br/jspui/community-list>>.

Os metadados são considerados elementos que visam proporcionar o desenvolvimento de sistemas de recuperação de informação, permitindo a interoperabilidade entre eles e o acesso mais amplo às informações disponibilizadas em diversos ambientes. Nesse entendimento, os metadados auxiliam na descrição, identificação, localização e acesso dos documentos. O metadado “assunto”, em especial, é destinado à representação do conteúdo dos recursos digitais possibilitando a inclusão de palavras-chave (termos, descritores) para fins de indexação e recuperação da informação.

Conforme é possível observar na Figura 3, o(a) autor(a) do documento empregou três termos para representar o conteúdo da produção intelectual depositada no âmbito do RI/UFRN, cujos descritores foram: “Controle do almoxarifado”; “SIPAC” e “SIAFI”. Verificou-se, portanto, o emprego de abreviações de palavras (por meio das siglas “SIPAC” e SIAFI), as quais dizem respeito às seguintes expressões - respectivamente: Sistema Integrado de Patrimônio, Administração e Contratos; e Sistema Integrado de Administração Financeira. Destarte, a aplicação de siglas e abreviações em descritores podem implicar no processo de recuperação dos documentos no âmbito dos Repositórios Institucionais, uma vez que elas são originalmente empregadas para reduzir palavras/termos – sem que os mesmos sejam descritos na íntegra.

Mediante as análises realizadas, foi identificado o emprego de frase e de sigla no metadado “assunto” do RI/UFRN, conforme pode ser visualizado na Figura 4.

Figura 4 – Ocorrência de frases e sigla em palavras-chave no RI/UFRN



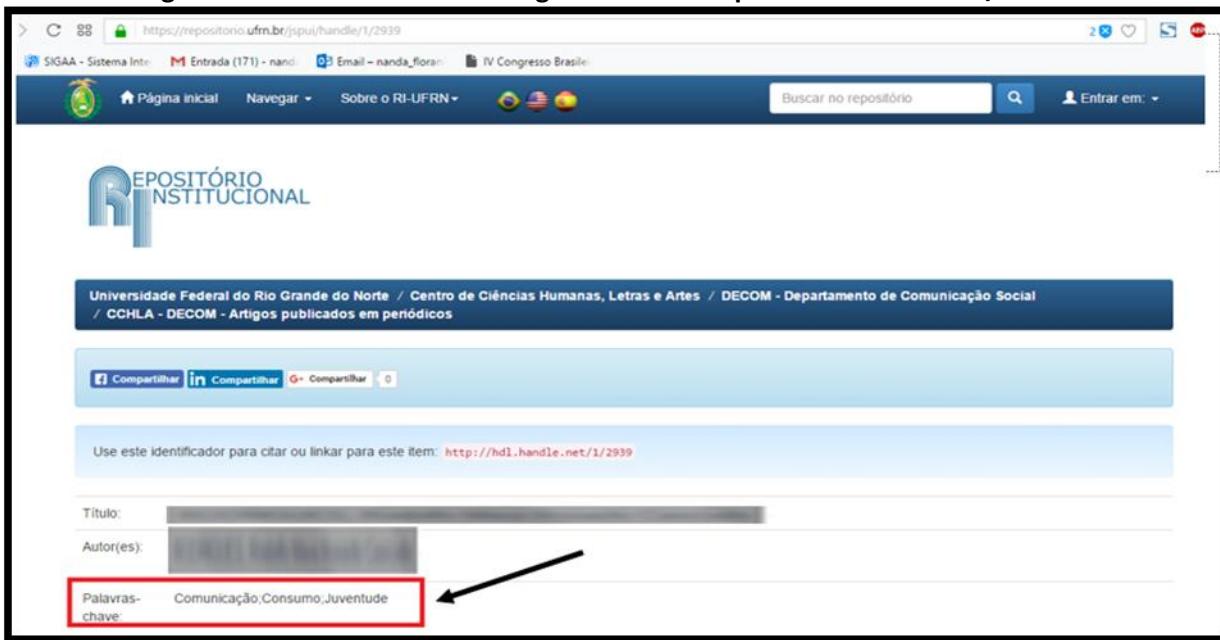
Fonte: <<https://repositorio.ufrn.br/jspui/community-list>>.

Como é perceptível, o(a) autor(a) do trabalho apresentado na Figura 4 elencou a frase “Convenção nº 102 da OIT⁴ sobre normas mínimas de seguridade social” para representar a sua produção intelectual, bem como os termos “Seguridade Social” e “Direito Comparado”. Nesse segmento, verifica-se que o emprego de frases e sigla no metadado “assunto” dos Repositórios Digitais pode implicar em problemas de revocação e precisão, as quais são medidas relativas à probabilidade dos documentos recuperados serem compatíveis com aquilo que os usuários estão buscando em um sistema de informação. Dessa forma, a qualidade das palavras-chave atribuídas, bem como a terminologia adotada são fundamentais para o sucesso de qualquer sistema, independentemente de sua forma e estrutura.

De outro modo disposto, verificou-se o emprego de termos generalistas, como é apresentado na Figura 5 a seguir.

⁴ Sigla destinada a abreviação do termo “Organização Internacional do Trabalho” - (OIT).

Figura 5 – Ocorrência de termos generalistas em palavras-chave no RI/UFRN

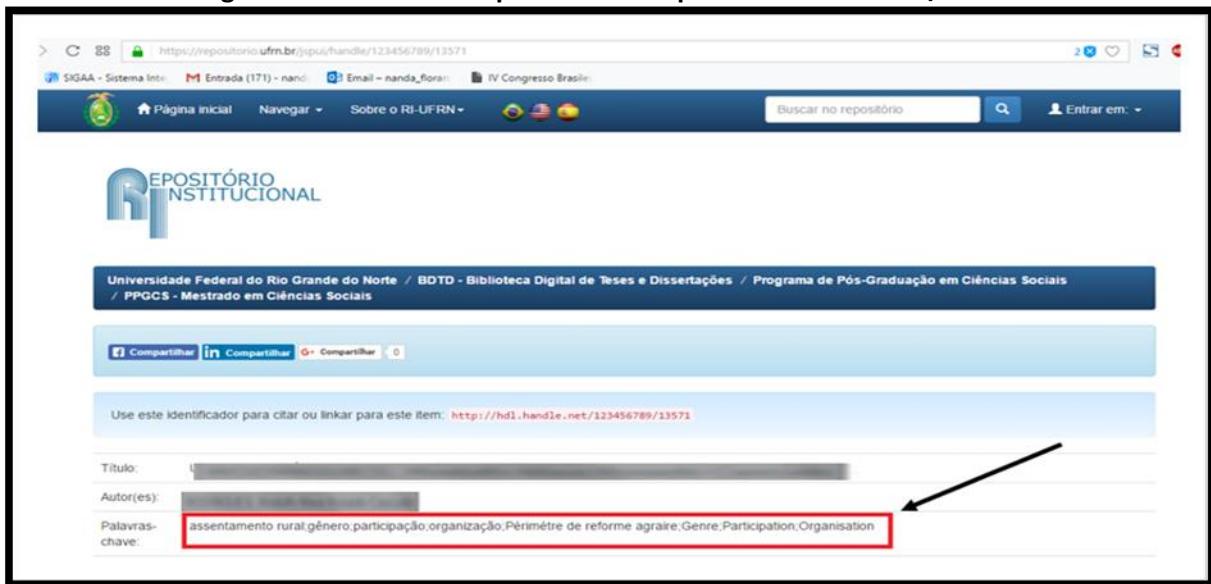


Fonte: <<https://repositorio.ufrn.br/jspui/community-list>>.

Nesse entendimento, outro elemento vinculado à 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. Na Figura 5 verificou-se a ocorrência desse fenômeno nos termos “Comunicação”, “Consumo” “e”, “Juventude”. Assim, percebe-se que estes termos poderiam estar inseridos em diversos contextos sob enfoques e filtros diversos, podendo resultar em implicações negativas no processo de busca no Repositório Institucional.

Mediante as análises realizadas, constatou-se também a ocorrência de polissemia em palavras-chave no RI/UFRN:

Figura 6– Ocorrência de polissemia em palavras-chave no RI/UFRN



Fonte: <<https://repositorio.ufrn.br/jspui/community-list>>.

Como é perceptível na Figura 6, o(a) autor(a) da produção científica apresentada elencou 8 (oito) palavras-chave/descritores para representarem a sua produção intelectual, dentre elas palavras em língua estrangeira e o descritor “Organização” que consiste em uma polissemia – tendo em vista que esse termo está vinculado a, pelos menos, dois significados - são eles:

- **Organização** → Ordem = Ordenação/Arrumação
Associação = Entidade/Instituição

Nessa concepção, infere-se que a representação do conteúdo de um documento por uma palavra polissêmica (como apresentado na Figura 6) e a ausência de controle desse fenômeno pode resultar em problemas na sua recuperação. Assim, ao fazer uma busca com o termo “Organização”, o Repositório Digital pode recuperar para o usuário trabalhos sobre entidade/instituição, como também outras produções científicas que abordem aspectos relativos à ordenação/arrumação de algo.

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 ambientes diversificados como os Repositórios Digitais esboçando assim novos campos de competência profissional.

Sendo assim, no momento da representação do conteúdo do documento, se o(a) autor(a) deixar de aplicar um termo mais específico para representar o assunto, empregando

um termo generalista ao conteúdo temático, provavelmente ocorrerão falhas na recuperação da informação.

A aplicação de descritores genéricos na representação da informação é uma questão que precisa ser observada/analisada, uma vez que estes 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.

Por essa razão, a ausência de controle dos fenômenos elencados anteriormente pode resultar em problemas na recuperação das produções científicas em Repositórios Digitais de Acesso Aberto. Dessa forma, as palavras-chave atribuídas, bem como a terminologia adotada são fundamentais para o sucesso de qualquer sistema, independentemente de sua forma e estrutura.

6 CONSIDERAÇÕES FINAIS

Mediante tais considerações, a partir dos estudos realizados, observou-se nesta pesquisa que os termos atribuídos em linguagem natural viabilizam implicações no processo de representação e recuperação dos documentos em ambiente digital. Destarte, na análise qualitativa do metadado “assunto” do Repositório Institucional da Universidade Federal do Rio Grande do Norte verificou-se a presença de algumas implicações, dentre elas: emprego de frases, 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.

Excluir os termos provenientes da coleta automática das palavras-chave empregadas na indexação das produções científicas do Repositório Institucional para adicionar outra forma de representação temática pode significar um risco para esse sistema. Dessa forma, observa-se a importância de manter os mesmos descritores informados pelos autores nas palavras-chave dos resumos dos seus documentos no metadado "assunto" do Repositório. Contudo, se faz necessário que o usuário administrador (ou indexador do sistema) analise e corrija eventuais erros ortográficos existentes nesses termos, bem como adicione remissivas para eles (com descritores semelhantes àqueles já sugeridos pelos usuários ou explicitamente identificados como equivalentes disponíveis no instrumento de controle terminológico a ser associado ao sistema). Além disso, com essas adequações propostas o sistema pode permitir a

correção automática de erros ortográficos dos descritores coletados ou atribuídos para a indexação de uma produção científica.

Mediante tais considerações, infere-se a importância dos gestores do RI/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 incluindo o auxílio de linguagens documentárias e vocabulários controlados que podem elevar o nível da qualidade de descrição do conteúdo desses documentos visando, consequentemente, 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 RI/UFRN (bem como os administradores de repositórios digitais que adotam políticas semelhantes) repensem o processo de indexação dos documentos nessas ferramentas de acesso aberto.

Tais ações podem ser voltadas para a preocupação na construção de modelos colaborativos de indexação social, os quais são descritos por Santos (2016)⁵ 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. Para a autora citada anteriormente, a alternativa do emprego de um modelo colaborativo de indexação 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/descritores

⁵ 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.

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.

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. Sendo assim, tais reflexões se constituem como objetivos de estudo dos trabalhos atuais e futuros da pesquisadora.

Em linhas gerais, 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, V. F. **Folksonomia**: vocabulário descontrolado na arquitetura da informação ou samba do criolo doido. 2007. Disponível em:
http://www.guilhermo.com/ai_biblioteca/referencia.asp?referencia=317. Acesso em: 20 maio 2018.

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

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

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.; 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>. Acesso em: 26 jun. 2018.

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

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>. Acesso em: 11 jun. 2018.

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 jul. 2018.

REPOSITÓRIO INSTITUCIONAL DA UFRN. 2017. Disponível em: <<https://repositorio.ufrn.br/jspui/>>. Acesso em: 15 maio 2018.

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, R. F. **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.

SANTOS, R. F.; CORRÊA, R. F. Análise e síntese dos diversos usos do termo "Folksonomia" no âmbito da ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., 2015. **Anais...** João Pessoa: Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação, 2015.