

Organização da Informação em Repositórios Digitais: Uma Abordagem sobre a Política de Indexação da Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI)

Organization of information in Digital Repositories: an approach on the indexation policy of the Reference Database of Periodical Articles in Information Science (BRAPCI)

Raimunda Fernanda dos Santos (1), Renato Fernandes Corrêa (2)

(1) Universidade Federal da Paraíba, João Pessoa/PB, e-mail: nandaflorania@gmail.com (2) Universidade Federal de Pernambuco, Recife/PE, e-mail: fc_renato@yahoo.com.br

Resumo

Estuda aspectos relativos à representação temática da informação no contexto das bases de dados de artigos de periódicos eletrônicos de acesso aberto que aderem ao modelo Open Archives Initiative e são construídas por uma colheita (harvesting) de metadados via o protocolo Open Archives Initiative for Metadata Harvesting. Objetiva descrever e analisar a política de indexação da Base de Dados Referencial de Periódicos em Ciência da Informação (BRAPCI) cuja ambiência armazena, preserva e divulga parte da memória científica nacional da área de Ciência da Informação. Utiliza como metodologia a pesquisa bibliográfica, documental e exploratória, pautada na revisão, análise e descrição da política de indexação da base de dados em questão. Identifica as seguintes implicações negativas no âmbito da política de indexação da base de dados: a representação da informação sem a interferência do indexador; a ausência de revisão da indexação; a participação do indexador restrita à categorização dos termos e definição de remissivas com descritores livres. Conclui enfatizando a importância dos gestores da base de dados analisada 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 destes documentos visando a precisão no processo de busca e recuperação da informação.

Palavras-chave: Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação; Política de indexação; Representação da Informação.

Abstract

It studies aspects related to thematic representation of information in the context of the databases of articles of open access electronic journals adhering to the Open Archives Initiative model and is constructed by harvesting metadata via the Open Archives Initiative for Metadata Harvesting protocol. The purpose of this paper is to analyze the indexing policy of the Reference Database of Periodicals in Information Science (BRAPCI), whose ambience stores, preserves and disseminates part of the national scientific memory of the Information Science area. It uses as a methodology the bibliographic, documentary and exploratory research, based on the review, analysis and description of the indexation policy of the database in question. It identifies the following negative implications in the context of database indexing policy: the representation of information without the interference of the index-dexter; The absence of indexation review; The participation of the indexer restricted to the categorization of the terms and definition of references with free descriptors. It concludes by emphasizing the importance of the managers of the analyzed database to elaborate an indexation policy aimed at improving the practices already used in this environment including the aid of documentary languages and controlled vocabularies that can raise the level of the quality of description of these documents aiming at the Accuracy in the information retrieval and retrieval process.

Keywords: Base of Reference Database of Periodic Articles on Information Science; Indexing policy; Representation of Information.

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, 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 Bases de Dados 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 dos trabalhos nesses Repositórios, selecionando-se assim as palavras-chave empregadas no resumo do documento para descrever o conteúdo da produção intelectual depositada. Dessa forma, os termos utilizados no momento da indexação de cada documento são decorrentes da coleta automática dos metadados atribuídos pelos autores ou periódicos no ato da publicação dos artigos. Tal atividade se configura como uma indexação por extração de acordo com Lancaster (2004).

Diante do exposto, 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 aplicação no contexto das bases de dados de artigos de periódicos eletrônicos de acesso aberto que aderem ao modelo *Open Archives Initiative* e são construídas por uma colheita automática (harvesting) de metadados dos sistemas de editoração de periódicos eletrônicos via o protocolo *Open Archives Initiative for Metadata Harvesting* (OAI-PMH), como a Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI).

Nessa concepção, a presente pesquisa tem como objetivo geral analisar a política de indexação da BRAPCI. Destarte, face ao objetivo geral, os objetivos específicos da presente pesquisa consistem em: apresentar aspectos relativos às definições e aos elementos constituintes de uma política de indexação; descrever a política de indexação dessa base de dados; analisar aspectos qualitativos referentes ao metadado assunto da base de em questão.

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, nas ferramentas de acesso

aberto como a BRAPCI, ocorrer de forma livre e sem padronização.

Nesse entendimento, a população ou o universo investigado é a BRAPCI. A escolha dessa fonte de informação é relativa à importância que ela representa para a Ciência da Informação no Brasil, haja vista que é por meio dessa base de dados que é armazenada, preservada e divulgada parte da memória científica nacional da área. Torna-se relevante também pesquisar aspectos relativos à representação temática da informação na base de dados supracitada levando em conta que ela consiste em uma ferramenta dinâmica que subsidia alunos, professores e pesquisadores da área de Ciência da Informação a encontrar as produções científicas presentes nos periódicos dessa área, possibilitando também a preservação histórica dos documentos nesse contexto, visando atender à necessidade dos usuários.

As bases de dados científicas se constituem como a metáfora da memória da ciência que se pratica nos dias atuais, objetivando agregar as produções intelectuais dos pesquisadores (SAYÃO, 1997). Sendo assim, a construção dessa memória, bem como a sua recuperação, depende sobretudo da representação da informação e, conseqüentemente, da política de indexação aplicada nesse contexto.

Quanto ao objetivo a ser alcançado para este estudo, foi empreendida uma pesquisa do tipo exploratória. Quanto aos meios, os procedimentos metodológicos foram empreendidos a partir de uma pesquisa bibliográfica e documental, baseada na análise qualitativa de documentos que relatam aspectos relativos à história e à política de indexação dessa base de dados, os quais foram fornecidos pela sua idealizadora (Leilah Santiago Bufrem). Foram utilizadas também na concepção do trabalho fontes impressas como livros, e em meio digital como: relatórios e projetos, artigos de periódicos científicos eletrônicos, Teses e Dissertações sobre a temática em questão na área de Ciência da Informação encontrados por meio de bases de dados, mecanismos de busca, portais, repositórios institucionais e bibliotecas digitais na Web.

Mediante tais considerações, acredita-se que a presente pesquisa é relevante uma vez que irá contribuir com a produção científica sobre a temática em questão na literatura nacional da Ciência da Informação, bem como trazer à discussão aspectos relativos à política de indexação e representação temática da informação em Bases de Dados científicas.

2 As práticas de indexação e as suas interfaces com as Bases de Dados

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 [1] 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.

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. Além disso, tais termos podem ser homógrafos -descritores que coincidem na sua forma, porém representam conceitos diferentes; apresentar erros ortográficos; conter expressões regionais (que são empregadas como tags de acordo com o contexto social/regional em que o usuário está inserido); termos 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, 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 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 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.

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 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, tesouro) 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 tesouros e as ontologias [2] 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. Entretanto, se faz necessário atentar-se para as práticas de indexação realizadas no âmbito dos Repositórios Digitais como a Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI). Sob esse vies, a seguir são descritos aspectos concernentes a origem e ao histórico da Base de Dados em questão.

De acordo com Lancaster (2004), um dos principais fatores que influenciam no desempenho de um sistema de recuperação da informação e que estão diretamente ligados à representação temática da informação é a política de indexação. Sendo assim, verifica-se a importância de se estabelecer uma política de indexação para assegurar a uniformidade da prática da indexação mediante o estabelecimento de alguns critérios ou princípios que se constituirão como diretrizes norteadoras para a orientação da atividade do indexador - aspecto discutido a seguir.

2.1 Política de Indexação

A indexação e, conseqüentemente, o estabelecimento de critérios e princípios que nortearão essa atividade, são elementos que devem ser levados em consideração no planejamento das bibliotecas e dos sistemas de recuperação de informação. Em linhas gerais, Maximiano (2000) enfatiza que as políticas ou diretrizes objetivam nortear as ações dos indivíduos e/ou grupos em longo período no que concerne aos cenários que apresentam atividades repetitivas ou permanentes por meio da padronização e delimitação dos comportamentos em um determinado contexto. Para tanto, Rubi (2009, p.83) afirma que a política de indexação "deve ser compreendida como uma decisão administrativa que reflita os objetivos da biblioteca, identificando condutas teóricas e práticas das equipes envolvidas no tratamento da informação [...]". Nesse entendimento, observa-se que apesar dessa autora discutir acerca do conceito de política de indexação voltado para as bibliotecas, tais perspectivas podem ser direcionadas a qualquer sistema de recuperação da informação ou outros contextos que tenham a atividade da indexação como um dos seus elementos constituintes.

Correspondentemente, de acordo com Carneiro (1985, p.221) uma política de indexação "deve servir como um guia para tomada de decisões, deve levar em conta os seguintes fatores: características e objetivos da organização, determinantes do tipo de serviço a ser oferecido; identificação dos usuários, para atendimento de suas necessidades de informação e recursos humanos, materiais e financeiros, que delimitam o funcionamento de um sistema de recuperação de informações".

Portanto, visando representar um pequeno passo na direção da maturidade e consolidação do conceito de política de indexação - levando-se em conta que os conceitos relacionados à política de indexação na literatura estão diretamente relacionados especificamente à sua função - a política de indexação pode ser definida neste trabalho como: diretriz que norteia as ações dos profissionais envolvidos na atividade de representação temática dos objetos informacionais existentes em um determinado contexto (biblioteca, sistema de recuperação da informação etc.) com base nos seus objetivos, características, perfil da comunidade de usuários e recursos (humanos, materiais e financeiros), objetivando melhorar o processo de representação e, conseqüentemente, busca e recuperação de documentos no contexto de sua aplicação.

Segundo Lancaster (2004), se o sistema de recuperação de informação tiver uma política de indexação bem descrita, levando em conta o perfil da comunidade de usuários, haverá maiores probabilidades de efetividade em relação à realização dos seus objetivos. Para esse autor, dentre os elementos constituintes dessa política existem "os níveis de exaustividade e especificidade, capacidade de revocação e precisão do sistema, estratégia de busca, tempo de resposta do sistema, forma de saída e avaliação do sistema" (LANCASTER, 2004, p.31).

Corroborando com essas ideias, são apresentados a seguir os fatores que devem ser levados em conta no processo de elaboração de uma política de indexação, os quais são mencionados por Carneiro (1985, p. 231, grifo nosso):

- 1 Cobertura de assuntos: assuntos cobertos pelo sistema (centrais e periféricos);
- 2 Seleção e aquisição dos documentos-fonte: extensão da cobertura do sistema em áreas de assunto de seu interesse e a qualidade dos documentos, nessas áreas de assunto, incluídos no sistema;
- 3 Processo de indexação:
 - 3.1 Nível de exaustividade: medida de extensão em que todos os assuntos discutidos em um certo documento são reconhecidos durante a indexação e traduzidos na linguagem do sistema;
 - 3.2 Nível de especificidade: nível de abrangência em que o sistema permite especificar os conceitos identificados documento;
 - 3.3 Escolha da linguagem: a linguagem documentária afeta o desempenho de um sistema de recuperação de informação tanto na estratégia de busca (estabelece a precisão com que o técnico de busca pode descrever os interesses do usuário) quanto na indexação (estabelece a precisão com que o indexador pode descrever o assunto do documento). Portanto, a partir de estudos do sistema, deve-se optar entre linguagem livre ou linguagem controlada e linguagem pré-coordenada ou póscoordenada;
 - 3.4 Capacidade de revocação e precisão do sistema: exaustividade, revocação e precisão estão relacionadas. Quanto mais

exaustivamente um sistema indexa seus documentos, maior será a revocação (número de documentos recuperados) na busca e, inversamente proporcional, a precisão será menor; 4 Estratégia de busca: deve-se decidir entre a busca delegada ou não; 5 Tempo de resposta do sistema: [medida relacionada ao desempenho do sistema que diz respeito ao período de tempo gasto entre o instante em que o usuário iniciou uma solicitação e o instante em que o sistema apresentou a sua resposta ao mesmo ou atendeu à sua solicitação]. 6 Forma de saída: é o formato em que os resultados da busca são apresentados. Tem grande influência sobre a tolerância do usuário quanto à precisão dos resultados. Deve-se verificar qual a preferência do usuário quanto à apresentação dos resultados; 7 Avaliação do sistema: determinará até que ponto o sistema satisfaz as necessidades dos usuários.

Sendo assim, para a construção de uma política de indexação deve-se levar em conta o conjunto de documentos (ou objetos informacionais) disponíveis no sistema, bem como o perfil dos seus usuários, fazendo com que ela possua um caráter personalizado. Destarte, além da caracterização do perfil dos usuários, é importante analisar o conjunto de documentos para fins de identificação da(s) área(s) de conhecimento discutida(s) nesses objetos informacionais a serem indexados, bem como as características do acervo em relação à quantidade de materiais existentes. Ademais, de acordo com Rubi (2009) os princípios de indexação devem ser seguidos pelo profissional que irá realizar a tarefa de análise de assunto e de indexação do conteúdo dos documentos. Esse profissional deve ter consciência da importância da política de indexação para a recuperação dos documentos- principalmente no que diz respeito a exaustividade, a especificidade, a capacidade de revocação e de precisão do sistema de recuperação de informação. Portanto, todos esses aspectos devem ser pensados de maneira conjunta, visando determinar se o sistema será mais específico (contribuindo para uma maior precisão na recuperação da informação), ou se ele será exaustivo (ampliando assim a revocação).

No que concerne à exaustividade, Rubi (2009) afirma que ela se relaciona com a quantidade de termos que são elencados para representar o conteúdo de um determinado documento, possibilitando diferentes opções de acesso por meio de termos genéricos e específicos. Enquanto Lancaster (2004, p.27) afirma que "quanto mais termos forem utilizados para indexar um documento mais acessível ele se tornará e, provavelmente, mais vezes será recuperado", Gil Leiva (2008) e Soergel (1994) alertam para a necessidade de verificar também a extensão do significado de cada termo de indexação a fim de observar se tem ou não uma relação intrínseca com o assunto do objeto informacional a ser indexado - tendo em vista que não basta apenas atribuir muitos termos no ato da

indexação, é preciso também que cada um deles tenha pertinência e relevância para representar tematicamente um determinado objeto informacional.

Mediante tais considerações, verifica-se a importância de determinar as diretrizes que o indexador deverá tomar como base para a identificação e a seleção dos termos, impossibilitando assim que ocorram disparidades no ato da indexação no que diz respeito a pertinência e a relevância dos descritores para representar o conteúdo temático dos documentos. Já a especificidade está relacionada ao "nível de abrangência com que a biblioteca e a linguagem documentária permitem especificar os conceitos identificados no documento" (RUBI, 2009, p.85).

Outrossim, Lancaster (2004) enfatiza que a especificidade consiste em uma propriedade muito importante inerente às linguagens de indexação e pode ser obtida por meio de diferentes formas e em diversos vocabulários controlados. Para tanto, observa-se também a importância da adequação de uma linguagem de indexação para o sistema de recuperação de informação em que a política de indexação está sendo construída, uma vez que essa linguagem será utilizada para realizar a representação temática dos documentos visando facilitar a recuperação da informação.

De acordo com Rubi (2009), a decisão da política de indexação que envolve aspectos relativos à especificidade também está diretamente relacionada à precisão e revocação. Nessa perspectiva, a precisão e a revocação são medidas de desempenho que visam medir a capacidade de recuperar documentos úteis e evitar documentos inúteis baseando-se na noção de relevância (LANCASTER, 2004). Nesse entendimento, ambas as medidas supracitadas (precisão e revocação) são relativas à probabilidade dos documentos recuperados serem compatíveis com aquilo que os usuários estão buscando em um sistema de informação.

Diante do exposto, verifica-se que todos os aspectos apresentados anteriormente - os quais envolvem a especificidade, a exaustividade, a capacidade de revocação e a precisão - são interligados entre si e devem ser considerados conjuntamente em uma política de indexação (RUBI, 2009).

Ademais, além da política de indexação a ser adotada pelo sistema de recuperação de informação, faz-se necessária também a elaboração de um manual de indexação que contemple todos esses aspectos e determine como eles serão trabalhados por todos da rede de bibliotecas ou do sistema de recuperação, base de dados, etc. Diante disso, a política de indexação deve estar descrita e registrada em manuais de indexação para que possa ser constantemente avaliada e modificada, garantindo assim a possibilidade de avaliações futuras.

O manual de indexação, por sua vez, deverá constar também a linguagem de indexação a ser adotada, as etapas de indexação e os instrumentos auxiliares para uma representação temática eficaz. Além disso, Lancaster (2004, p.12) destaca a importância dos princípios da indexação orientada para os usuários, afirmando a necessidade dos indexadores terem conhecimentos acerca de questões que estão além dos princípios da indexação. Tais profissionais devem também estar completamente a par dos interesses da comunidade atendida, bem como das necessidades de informação dos seus membros e das novas tendências de representação da informação em meio digital.

Como objeto de estudo da presente pesquisa, são estudados e identificados a seguir, essencialmente, aspectos referentes à representação temática da informação no metadado assunto da BRAPCI, mediante a descrição e análise da sua política de indexação.

4 Descrição e análise da Política de Indexação da BRAPCI

Para fins de realização do tratamento temático dos documentos contidos na Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação e objetivando nortear os profissionais que realizam essa atividade, a base de dados contém uma política de indexação que foi criada em 05 de novembro de 2011 e atualizada em 27 de setembro de 2014, cujo document é disponibilizado publicamente. Sendo assim, no documento sobre a política da indexação da BRAPCI são enfatizadas, de maneira sucinta, algumas questões relativas à cobertura do assunto, processo de indexação, vocabulário controlado, agrupamento de termos (remissivas) e suas respectivas estatísticas.

De acordo com Gabriel Junior (2014), pesquisador responsável pela criação e atualização da política de indexação da BRAPCI, essa base de dados opera com dois diferentes módulos para a sua estruturação interna: o módulo de manutenção - que apresenta maior complexidade e é destinado especialmente para gestores e indexadores - e o módulo público que é disposto para os usuários que consultam a BRAPCI.

Mediante as análises realizadas, observou-se que a política de indexação dessa base de dados deixa claro que o indexador deve fazer o uso dos mesmos termos informados pelos autores nas palavras-chave dos artigos para descrever o conteúdo da produção intelectual depositada nesse ambiente. Para tanto, de acordo com a política de indexação, o processo de representação temática da informação disponível na base de dados é realizado de forma pós-coordenada (posteriormente), uma vez que os termos utilizados no momento da indexação de cada documento são decorrentes da coleta automática dos metadados incorporando os mesmos descritores que foram

atribuídos pelos autores ou periódicos no ato da publicação dos artigos.

No que concerne ao processo supracitado, a política de indexação enfatiza que a base BRAPCI realiza a coleta de metadados por meio de harvesting (colheita) nos periódicos eletrônicos da área de Ciência da Informação objetivando detectar os novos trabalhos que foram publicados nesses periódicos para agregá-los em sua ambiência. Para tanto, o sistema apresenta aos indexadores as novas edições coletadas pela base de dados, bem como todos os metadados dos trabalhos que precisam ser indexados/revisados (Figuras 1 e 2, em anexo).

O conteúdo dos campos de metadados é extraído de arquivo XML [3] ou transcrito diretamente do PDF. Mais especificamente, para termos de indexação, não há intervenção ou correção dos termos extraídos, o que foi constatado como um aspecto negativo que pode influenciar nos processos de representação e recuperação das produções científicas presentes na base de dados em questão. Outrossim, observou-se que o sistema gera de forma automática uma lista com todos os termos utilizados pelos autores/revistas e os seus respectivos idiomas em que foram definidos.

De acordo com Gabriel Junior (2014), no módulo de manutenção da BRAPCI é possível realizar a gestão da informação e das atividades a serem realizadas pelos gestores e indexadores do sistema, assim o sistema também apresenta uma síntese da situação atual do processo de indexação em cada periódico (Figura 3, em anexo).

No que concerne ao emprego de vocabulário controlado, observou-se que a política de indexação dessa base de dados não define o instrumento que será utilizado como auxílio para a realização da indexação, deixando claro apenas que são utilizados mecanismos para o processamento automático de descritores e que a BRAPCI pode incorporar diferentes tesouros.

Sendo assim, de acordo com Gabriel Junior (2010, p.139) dentre algumas ferramentas disponíveis na BRAPCI pode-se mencionar: "coleta de dados por meio do OAI-PMH; exportação de dados para o módulo público; geração das nuvens de tags; criação de ontologias e tesouros e processamento de linguagem de marcação empregada na análise de citação" (Figura 4, em anexo).

Sem especificar a periodicidade de avaliação, é ressaltada na política de indexação da BRAPCI a importância da análise periódica dos termos coletados para fins de agrupamento e definição dos descritores autorizados para descreverem a temática dos documentos disponíveis em sua ambiência, uma vez que o termo é considerado como uma representação provisória de um tema ou assunto podendo ser atualizado ou substituído por outro de forma dinâmica.

Ainda de acordo com a política de indexação da BRAPCI, os indexadores desse ambiente utilizam termos compostos (lexicais) para a representação temática dos documentos, os quais podem ser descritos por meio das categorias de tipo “assunto”, “data”, “geográfico”, “autoridade” e “NC”, as quais são apresentadas no Quadro 1 a seguir.

Quadro 1. *Tipologias dos Descritores*

<i>Tipo</i>	<i>Descrição</i>	<i>Exemplos</i>
Assunto	Identifica um assunto ou tema, podendo ser simples ou composto (polilexical)	Indexação; Catalogação; Indexação Pós-coordenada.
Data	Delimita um período ou uma data específica.	1500-1808 2010 Maio de 2015
Geográfico	Representa uma localização geográfica, como nome de país, estados, cidade.	Alemanha; Paris; Rio de Janeiro; Rio Claro
Autoridade	Nome de uma instituição, organização, evento, pessoa, individual ou coletiva (grupos).	ANCIB; 1ª reunião sobre catálogo coletivo regional de publicações seriadas para a América latina; Ana Maria Athayde Polke
NC	Não Definido, quando um termo novo é identificado, entra nesta categoria.	

Fonte: Gabriel Junior (2014, p.5).

Observou-se ainda que a política de indexação da BRAPCI destaca a análise individual dos termos que foram coletados (conforme a sua data de incorporação) e categorização dos mesmos nas tipologias elencadas no Quadro 1. Nessa perspectiva, os termos novos (não catalogados) são marcados na categoria "NC", ou seja, como "Não Catalogado".

Em relação ao agrupamento de termos, observou-se que a política de indexação da base de dados enfatiza a relevância do emprego de remissivas para identificar a relação dos descritores que têm o mesmo significado ou conceito, mas que têm variação sintática. Diante disso, para fins de indexação, a política de indexação orienta a entrada de todos os descritores, considerando que todos os termos atribuídos a um documento têm relevância e pesos iguais.

Além disso, verificou-se que nessa base de dados não é realizada a substituição dos termos originais elencados pelo autor nas palavra-chave do seu artigo por um termo oriundo de vocabulário controlado, objetivando garantir a ideia e o contexto original do autor. Sendo

assim, para fins de correção no que concerne aos descritores, a base de dados aprimora o sistema de busca para o usuário procurando outras variações das palavras, mediante o uso do sistema de pós-coordenação (ignorando as stop words [4] como prenomes, preposições, artigos definidos e indefinidos presentes no título, resumo e palavras-chave de cada documento).

Ademais, em um sistema pós-coordenado os termos pesquisados pelos usuários podem ser combinados entre si, de qualquer maneira, no momento da realização da busca conforme a sua necessidade, permitindo a flexibilidade na recuperação da informação.

Mediante tais considerações, analisar e descrever a política de indexação da BRAPCI torna-se imprescindível, uma vez que a recuperação das produções científicas disponíveis nessa base de dados para o acesso pela geração hodierna e futura depende, dentre outras questões, da indexação realizada no âmbito desse repositório digital.

5 Considerações finais

Em linhas gerais, a presente pesquisa apontou aspectos relativos à representação da informação no metadado assunto da Base de Dados Referencial de Periódicos em Ciência da Informação (BRAPCI) cuja ambiência armazena, preserva e divulga parte da memória científica nacional da área de Ciência da Informação. Nesse segmento, o presente estudo descreve e analisa a política de indexação da base de dados supracitada, cujo documento é importante – uma vez que é irrisória a quantidade de bases de dados que possuem política de indexação registrada oficialmente (e disponibilizada). Portanto, a política de indexação da BRAPCI não serve somente para designar as práticas de representação da informação realizadas nessa ambiência, como também serve de auxílio nas decisões administrativas desse repositório digital.

De outro modo disposto, diante dos aspectos analisados, verificou-se as seguintes implicações negativas no âmbito da política de indexação da BRAPCI: a representação da informação sem a interferência do indexador (haja vista que o indexador da base de dados utiliza apenas os mesmos termos informados pelos autores nas palavras-chave dos artigos para descrever o conteúdo da produção intelectual depositada neste ambiente, sem realizar interferência ou correção dos descritores); a ausência de revisão da indexação (não há um processo de intervenção ou correção dos termos extraídos das palavras-chave empregadas pelos autores dos trabalhos); a participação do indexador restrita à categorização dos termos e definição de remissivas com descritores livres.

Excluir os termos provenientes da coleta automática das palavras-chave (empregadas na indexação das produções científicas da base de dados) para adicionar outra forma de representação temática pode significar um risco para o Repositório Digital. 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" da BRAPCI. Contudo, se faz necessário que o usuário administrador (ou indexador da base de dados) 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.

Se bem construído e atualizado com base nas garantias literária, estrutural e de uso um instrumento de controle terminológico a ser criado especificamente para a BRAPCI (tesauro, ontologia ou taxonomia) pode aumentar a confiança do usuário autor e do indexador da base de dados na identificação de conceitos relacionados para a indexação e recuperação dos documentos nessa ambiência, uma vez que poderá evitar a ocorrência de termos polissêmicos, descritores com homógrafos, etiquetas assimétricas e com ambiguidade morfológica no âmbito do sistema. Ademais, a presença de um instrumento de representação da informação disponível para o usuário no momento da busca dos trabalhos no sistema configura-se como uma estratégia útil, principalmente quando ele não souber o termo correto a ser utilizado ou quando a busca por palavras-chave não apresentar resultados satisfatórios para ele. Essa estratégia possibilita agrupar produções científicas similares com os mesmos assuntos expressados por descritores diferentes ou discriminar registros que estejam descritos com termos homógrafos no âmbito da base de dados.

Mediante as análises realizadas, observou-se também a necessidade de elaboração de um manual de indexação que contemple todos esses aspectos e determine como eles serão trabalhados por todos os tipos de usuários (administradores, indexadores, etc.) da base de dados. Nesse segmento, a política de indexação deve estar descrita e registrada em manuais de indexação para que possa ser constantemente avaliada e modificada, garantindo assim a possibilidade de avaliações futuras. O manual de indexação, por sua vez, deverá constar também a linguagem de indexação a ser adotada, as etapas de indexação e os instrumentos auxiliares para uma representação temática eficaz, levando em

consideração aspectos como exaustividade, especificidade, qualidade dos termos e capacidade de revocação e precisão do sistema.

Outrossim, além das sugestões elencadas anteriormente, as quais podem ser incluídas para fins de atualização da política de indexação da BRAPCI, o trabalho de Santos (2016) contribui essencialmente para a construção da memória coletiva dos pesquisadores da área de Ciência da Informação, bem como vislumbra sugestões a serem aplicadas para a melhoria das práticas de indexação já realizadas no âmbito da BRAPCI.

Portanto, verifica-se a importância dos gestores da base de dados analisada atualizarem a sua política de indexação levando em consideração o aperfeiçoamento das práticas já utilizadas neste ambiente incluindo os elementos supracitados e o auxílio de linguagens documentárias e vocabulários controlados que podem elevar o nível da qualidade de descrição destes documentos visando a precisão no processo de busca e recuperação da informação.

É importante ressaltar que as propostas de adequações elencadas neste trabalho também podem ser implementadas ou aperfeiçoadas em outras bases de dados de artigos de periódicos eletrônicos de acesso aberto ou repositórios digitais científicos que adotam políticas de indexação semelhantes à da BRAPCI como a base PERI-ECI e repositórios institucionais, por exemplo. Nesse entendimento, tais ideias podem servir de referência para a construção e desenvolvimento de um módulo de indexação para esses sistemas.

Como estudos futuros a serem realizados, será criado e proposto um modelo colaborativo de indexação para a Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação visando destacar contribuições e apresentar melhorias que contemplem tanto o usuário autor do documento, como o usuário pesquisador e o indexador com vistas a melhorar o processo de representação e, conseqüentemente, de busca e de recuperação da informação nessa base de dados.

Notas

- [1] Homógrafos são termos que coincidem na sua forma, porém representam conceitos diferentes.
- [2] 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. Apresenta-se como alternativas viáveis para organizar uma teia de relações permitindo a ligação entre os conceitos.
- [3] Do inglês *eXtensible Markup Language*, é uma linguagem de marcação recomendada pela W3C para a criação de documentos com dados organizados hierarquicamente, tais como textos, banco de dados (TECMUNDO, 2017).

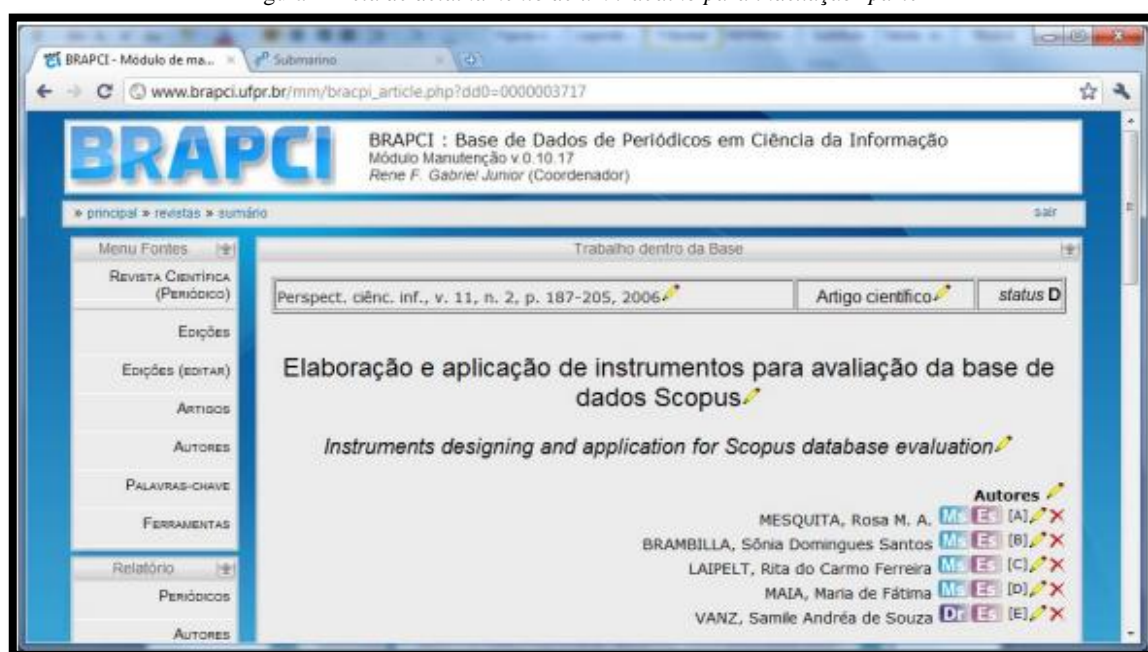
- [4] *Stop Words* (ou palavras vazias, em português) diz respeito às palavras que são consideradas como irrelevantes para o conjunto de resultado a ser exibido em um sistema de recuperação da informação.

Referências

- Referências devem ser elaboradas de acordo com a NBR6023 (2002) – Informação e documentação – Referências – Elaboração e devem seguir o alinhamento aqui proposto.
- 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.
- BASE DE DADOS REFERENCIAL DE ARTIGOS DE PERIÓDICOS EM CIÊNCIA DA INFORMAÇÃO. Disponível em: < <http://www.brapci.ufpr.br/brapci/>> . Acesso em: 13 maio. 2017.
- BRANDT, M.; MEDEIROS, M. B. B. Folksonomia: esquema de representação do conhecimento? Transinformação, Campinas, v. 22, n. 2, maio/ago. 2010.
- BRASCHER, M. A ambiguidade na recuperação da informação. DataGramaZero - Revista de Ciência da Informação, Rio de Janeiro, v.3, n.1, fev. 2002.
- CAFÉ, L.; SALES, R. de. Tesaurus e ontologias sob o olhar da teoria comunicativa da terminologia. In: CONGRESS ISKO-SPAIN, 9., 2009, Valência. Anais... Valência, 2009.
- CARNEIRO, M. V. Diretrizes para uma política de indexação. R. Esc. Bibliotecon. UFMG, Belo Horizonte, v.14, n.2, p.221-241, set. 1985.
- DIAS, E. W. Contexto digital e tratamento da informação. DataGramaZero, Rio de Janeiro, v. 2, n. 5, out. 2001. Disponível em: < http://www.dgz.org.br/out01/Art_01.htm>. Acesso em: 20 maio 2017.
- 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: .Acesso em: 26 set. 2014.
- GABRIEL JUNIOR, R. F. Política de Indexação: projeto BRAPCI. Curitiba, set. 2014.
- GABRIEL JUNIOR, R. F. Socialização dos saberes: metodologia para desenvolvimento de um repertório temático em Ciência da Informação. 189f. Dissertação (Mestrado em Ciência, Gestão e Tecnologia da Informação) - Universidade Federal do Paraná, Curitiba, 2010.
- GIL LEIVA, I. Manual de indización: Teoría y práctica. Gijón: Trea, 2008. p.385-400.
- 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.
- MAXIMIANO, A. C. A. Introdução à administração. 5.ed. São Paulo: Atlas, 2000.
- 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.
- RUBI, M. P. Os princípios da política de indexação na análise de assunto para catalogação: especificidade, exaustividade, revocação e precisão na perspectiva dos catalogadores e usuários. In.: FUJITA, M.S.L., (org.). A indexação de livros: a percepção dos catalogadores e usuários de bibliotecas universitárias. Um estudo de observação do contexto sociocognitivo com protocolos verbais. São Paulo: Editora UNESP; São Paulo: Cultura acadêmica, 2009, p.81-93.
- SAYÃO, F. Bases de dados: a metáfora da memória científica. Ci. Inf., Brasília, v.26, n.3, set./dez. 1997.
- 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.
- SOERGEL, D. Indexing and retrieval performance: the logical evidence. Journal of the American Society for Information Science, Washington, v. 45, n. 8, p. 589-599, 1994.
- TECMUNDO. O que é XML? 2017. Disponível em: < <https://www.tecmundo.com.br/programacao/1762-o-que-e-xml-.htm>>. Acesso em: 20 maio 2017.

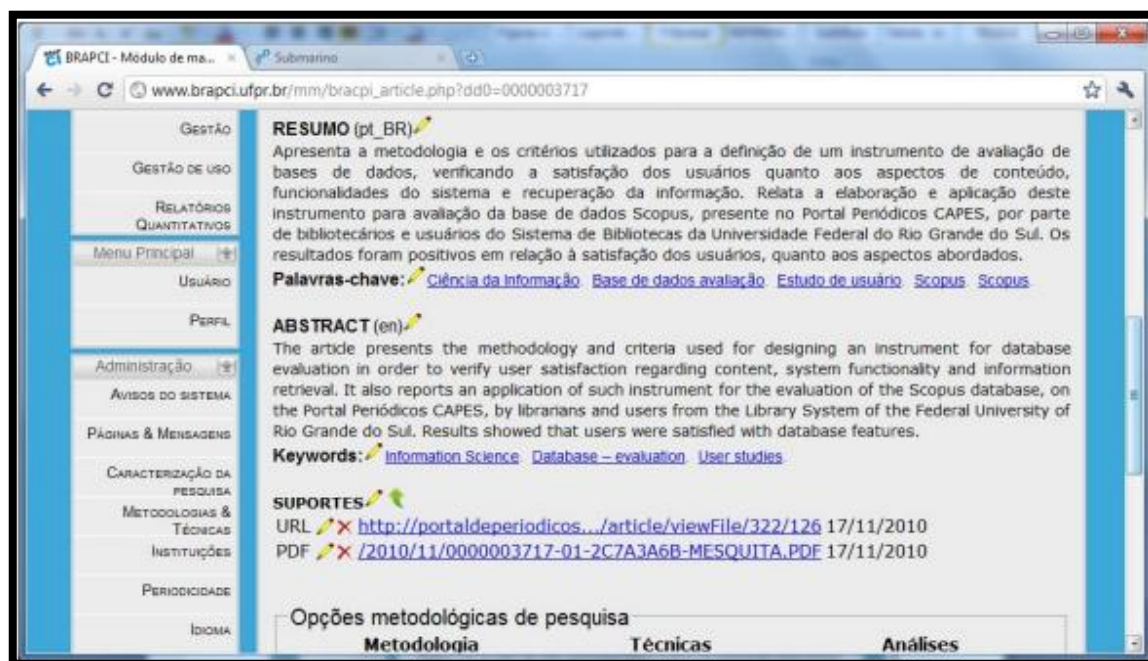
Anexo A – Telas de Detalhamento da indexação realizada na BRAPCI

Figura 1- Tela de detalhamento de um trabalho para indexação- parte I



Fonte: Gabriel Junior (2010, p.134)

Figura 2 - Tela de detalhamento de um trabalho para indexação- parte II



Fonte: Gabriel Junior (2010, p.134).

Figura 3- Resumo do processo de indexação

Periódicos		indexação	100% revisão	revisado	total	atualizado
AUTORES	Título					
PALAVRAS-CHAVE	Arquivistica.net	0	0	70	70	09/09/2010
FERRAMENTAS	Arquivo & Administração	51 58.0%	7 8.0%	30	88	09/09/2010
Relatório	Biblionline	0	0	108	108	21/10/2010
PERIÓDICOS	BIBLOS - Revista do Instituto de Ciências Humanas e da Informação	230 57.4%	8 2.0%	163	401	19/10/2010
AUTORES	Brazilian Journal of Information Science	0	0	32	32	30/10/2010
LOG DE ACESSO	Cadernos de Biblioteconomia	84 100.0%	0	0	84	19/10/2010
PENDÊNCIAS	Ciência da Informação	0	0	1201	1201	19/10/2010
Gestão	Comunicação & Informação	89 43.4%	43 21.0%	73	205	19/10/2010
Gestão de uso	DataGramaZero	0	11 3.4%	313	324	19/10/2010
RELATÓRIOS QUANTITATIVOS	Em Questão: Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS	19 8.8%	8 3.7%	189	216	20/12/2010
Menu Principal	Encontro Nacional de Pesquisa em Ciência da Informação	0	4 100.0%	0	4	19/10/2010
	Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação	17 4.8%	13 3.6%	327	357	19/10/2010
	Estudos Avançados em Biblioteconomia e Ciência da Informação	25 92.6%	2 7.4%	0	27	19/10/2010

Fonte: Gabriel Junior (2010, p.137).

Figura 4- Ferramentas do sistema



Fonte: Gabriel Junior (2010, p.137).