

Proposta de criação de um indicador para análise da adequação da coleção do Portal de Periódicos da Capes à produção científica nacional

Establishing an indicator to assess the adequacy of the collection of the CAPES Periodicals Portal for national scientific production

Propuesta para establecer un indicador para evaluar la adecuación de la colección del Portal Capes de revistas electrónicas a la producción científica nacional

Max Cirino de Mattos, mestre e doutorando em Ciência da Informação pela Escola de Ciência da Informação da UFMG. Endereço: Rua José Hemetério de Andrade, 631, apto 401 – Buritis. CEP: 30493-180 – Belo Horizonte, MG. Telefone: (31) 3567-6599/8898-1211. Email: max@cognotec.com.br.

Wesley Rodrigo Fernandes, mestrando em Ciência da Informação pela Escola de Ciência da Informação da UFMG. Endereço: Rua Montese, 51, apto 301 – Itapoã. CEP: 31710-100 – Belo Horizonte, MG. Telefone: (31) 3409-5148. E-mail: wesleyronline@yahoo.com.br.

Beatriz Valadares Cendón, doutora em Biblioteconomia e Ciência da Informação pela Universidade do Texas. Endereço: Escola de Ciência da Informação da UFMG, Caixa Postal 1.606. CEP: 30161-970 – Belo Horizonte, MG. Telefone: (31) 3409-5225. E-mail: cendon@eci.ufmg.br.

Resumo

A partir da análise dos estudos realizados sobre o Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), este trabalho propõe a criação de um indicador

automático de avaliação da adequação do conteúdo do Portal à produção de teses e dissertações cadastradas na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). São descritos os passos básicos para a criação desse indicador e as ponderações sobre a sua viabilidade tecnológica. Ao final, são discutidos possíveis impactos e a importância desse indicador para a avaliação da adequação do Portal de Periódicos da Capes em relação às necessidades informacionais da comunidade científica brasileira. O indicador forneceria informações para a implementação de medidas de adaptabilidade contínua do Portal às necessidades de seus usuários.

Palavras-chave: Portal Periódicos Capes. Avaliação de coleção. BDTD. Bibliometria.

Abstract

Based on the analysis of studies about the CAPES Portal of Electronic Journals, a digital library of scientific journals in Brazil, this paper discusses the creation of an indicator for assessing the adequacy of the content of the CAPES Portal for the production of theses and dissertations indexed at Digital Library of Thesis and Dissertations (BDTD) in Brazil. The article describes the basic steps to create this indicator with some considerations about its technological feasibility. The paper discusses possible impacts and the importance of this indicator for assessing the adequacy of the CAPES Portal for attending the informational needs of researchers in the Brazilian scientific community. The indicator could also provide the information necessary to implement measures for the continuous adaptability of the Portal and other digital libraries to the needs of their users.

Keywords: CAPES Portal of Electronic Journals. Evaluation of Journal Collections. BDTD. Bibliometry.

Resumen

A partir del análisis de los estudios sobre el Portal Capes de revistas electrónicas, una biblioteca digital de revistas científicas en Brasil, este documento propone la creación de un indicador automático para evaluar la adecuación de los contenidos del Portal Capes a la producción de tesis y disertaciones registradas en la Biblioteca Digital de Tesis y Disertaciones Brasileña (BDTD). El artículo describe los pasos básicos para crear este indicador con algunas consideraciones acerca de su viabilidad tecnológica. Además, el documento analiza los impactos potenciales y la importancia de este indicador para evaluar la idoneidad del Portal Capes con respecto a las necesidades de información de los investigadores en la comunidad científica brasileña. El indicador también podría proporcionar la información necesaria para poner en práctica medidas para la adaptación continua del Portal Capes y de otras bibliotecas digitales a las necesidades de sus usuarios.

Palabras clave: Portal Capes de Revistas Electrónicas. Colección de Evaluación. BDTD. Bibliometría.

1 Introdução

A avaliação de sistemas de informação em geral – e de bibliotecas digitais, em particular, do Portal de Periódicos da Capes – não é um tema simples. Uma das formas de se avaliar o seu sucesso é pela mensuração quantitativa do uso de seus periódicos para a produção científica dos programas de pós-graduação. Essa mensuração pode auxiliar na avaliação da adequação de seu conteúdo à produção de conhecimento científico.

Conforme Lancaster (1996, p. 20):

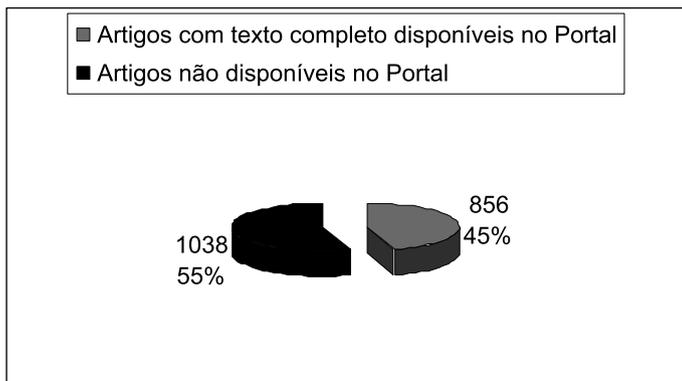
Ao avaliar um acervo, o que se procura de fato é determinar o que a biblioteca deveria possuir e não possui, e o que possui, mas não deveria possuir, tendo em vista fatores de qualidade e adequação da literatura publicada, sua obsolescência, as mudanças de interesses dos usuários, e a necessidade de otimizar o uso de recursos financeiros limitados.

A presença na coleção dos itens buscados influencia o sucesso do serviço de uma biblioteca na satisfação dos interesses de seus usuários. Portanto, um índice importante na avaliação da adequação de coleções é a probabilidade de posse, que pode ser definida como “a probabilidade de se encontrar um item procurado pelo usuário no acervo da biblioteca” (LANCASTER, 1996, p. 17).

Essa mensuração da adequação de uma coleção a partir da probabilidade de posse pode ser realizada por meio de diversas técnicas da bibliometria, que

é o estudo dos aspectos quantitativos da produção, disseminação e uso da informação registrada. Usada pela primeira vez por Pritchard em 1969, a bibliometria desenvolve padrões e modelos matemáticos para medir esses processos, usando seus resultados para elaborar previsões e apoiar tomadas de decisão (TAGUE-SUTCLIFFE, 1992, p. 1).

Um exemplo das possibilidades da bibliometria pode ser encontrado no trabalho de Fernandes (2009). O autor utilizou uma técnica bibliométrica denominada análise de citações para avaliar a adequação do conteúdo do Portal de Periódicos da Capes em relação à produção de teses e dissertações defendidas no triênio 2005-2007 na Escola de Ciência da Informação (ECI) da Universidade Federal de Minas Gerais (UFMG). Nessa pesquisa, foi avaliada a cobertura do Portal de Periódicos da Capes em relação aos periódicos e aos artigos citados nas teses e dissertações analisadas e em relação aos periódicos do Portal classificados na área de Ciência da Informação. O trabalho revelou, entre outros resultados, que o Portal continha apenas 45% dos artigos citados (Gráfico 1).



Fonte: Fernandes (2009, p. 87).

Gráfico 1. Periódicos citados: quantidade/porcentagem de artigos disponíveis no Portal

Embora o conhecimento desse percentual seja essencial para a avaliação do acervo do Portal, de acordo com Cendón e Ribeiro (2008), são poucos os trabalhos focados no impacto do uso do Portal na pesquisa e na produtividade das instituições usuárias ou na adequação do conteúdo do Portal.

Todos os trabalhos citados por Cendón e Ribeiro (2008) que se preocuparam com a adequação da coleção e o trabalho de Fernandes (2009) tiveram os dados colhidos de forma manual, o que demandou muito esforço e tempo dos pesquisadores. Talvez seja essa a dificuldade encontrada pela Capes para não utilizar ainda a análise de citação como requisito para a escolha de novos títulos que irão compor seu acervo.

Pensando em viabilizar um processo de avaliação contínua e facilitar os estudos sobre o impacto do Portal sobre a pesquisa científica e a adequação da sua coleção às necessidades de seus usuários, propõe-se a criação de um indicador que represente o percentual de referências utilizadas na produção das teses e dissertações da BDTD e que estão disponíveis no Portal de Periódicos da Capes e que tenha a possibilidade de ser calculado automaticamente, como será discutido a seguir. Esse indicador poderá ser definido para cada tese ou dissertação, a partir da divisão do número de referências citadas disponíveis no

Portal pelo número total de referências citadas. A existência de tal indicador permitirá a avaliação geral da disponibilidade no Portal das referências utilizadas nas teses e dissertações ou a segmentação desse percentual por área geográfica, instituição, nível de titulação e área ou subárea de conhecimento em determinado período.

Acreditamos que a criação desse indicador permitirá a expansão dos estudos semelhantes ao de Fernandes (2009) para outras áreas do conhecimento, sendo essa inclusive uma das considerações finais do referido autor.

O presente trabalho busca a identificação de um processo estruturado para a criação desse indicador e a discussão das possibilidades tecnológicas de implantação de cada uma das etapas desse processo para que o indicador esteja disponível de forma automática e contínua.

2 Revisão de literatura

Além do levantamento realizado por Fernandes (2009), foram consultados alguns trabalhos sobre bibliotecas digitais e sobre o Portal de Periódicos da Capes, entre eles: Cendón e Ribeiro (2008), Ortellado (2008), Rosetto (1997, 2008), Castro e Santos (2007), Hommerding (2007), Rodrigues e Oddone (2007), Marcondes et al. (2006), Smeaton e Callan (2005) e Silva, Sá e Furtado (2004).

A partir das considerações apresentadas por alguns autores (CUNHA, 2009; CENDÓN; RIBEIRO, 2008; MARCONDES et al., 2006), observa-se que, apesar do crescimento do uso e da importância das bibliotecas digitais, as pesquisas empíricas sobre sua utilização e seu impacto são restritas a áreas e segmentos específicos de usuários.

Especificamente sobre a abordagem acadêmica do Portal de Periódicos da Capes, podemos destacar o trabalho de Cendón e Ribeiro (2008). As autoras analisaram 40 trabalhos de pesquisa sobre o Portal, apresentando um panorama sobre os focos, resultados e pesquisadores

envolvidos com o tema. Algumas das conclusões apresentadas são discutidas a seguir:

- Existem poucas pesquisas sobre o tema Portal de Periódicos da Capes, e, dos 40 trabalhos de pesquisa analisados pelas autoras, apenas 11 estão focados no impacto do uso do Portal na pesquisa e na produtividade das instituições usuárias ou na adequação do conteúdo do Portal (CENDÓN; RIBEIRO, 2008);
- A maioria dos estudos focaliza uma ou duas instituições, geralmente aquela(s) de origem do(s) pesquisador(es); e
- As áreas de conhecimento pesquisadas são restritas em relação à abrangência do material disponibilizado no Portal.

Cunha (2009, p. 61) apresenta uma justificativa importante para o estudo do Portal:

A escolha do Portal de Periódicos pode ser justificada pelos seguintes aspectos: a) É o único Portal de Periódicos brasileiro disponível na [comunidade] científica que abrange todas as áreas do conhecimento; b) É o mais importante da comunidade científica em nível de produtividade, qualidade e abrangência; c) Está situado no meio eletrônico; d) Carência de estudos sobre seu uso.

Observamos, assim, quatro limitações na abrangência da análise dos estudos sobre o Portal: quantitativa e geográfica, em relação ao universo das instituições usuárias do Portal; e quantitativa e temática, em relação às áreas de conhecimento cobertas pelo Portal. Essas limitações no volume e na abrangência das pesquisas empíricas sobre o Portal devem-se, em parte, às dificuldades de se levantar os dados necessários para os estudos – em cada instituição usuária e para cada área e subárea do conhecimento – de forma não automatizada. Dessa forma, justifica-se a proposta de criação de um indicador que possa ser calculado automaticamente e que permita a análise contínua de um conjunto mais amplo de áreas e subáreas de conhecimento, instituições de ensino superior e regiões geográficas.

Nessa proposta, consideramos a produção científica dos programas de pós-graduação a partir da BDTD, mesmo cientes de que seu conteúdo atualmente não representa toda a produção nacional, pois a disponibilização dos trabalhos ainda é opcional.

3 Materiais e métodos

Trata-se de um estudo teórico, sendo o desenho inicial do processo proposto para a criação do indicador baseado principalmente no trabalho de Fernandes (2009). A Figura 1 resume o processo proposto, constituído de três passos, que serão detalhados a seguir:

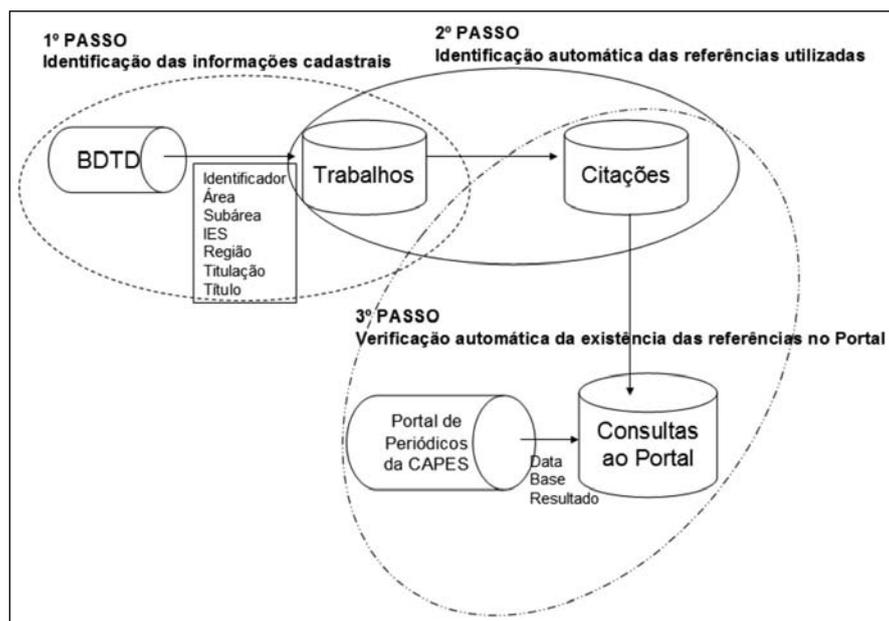


Figura 1. Processo de automação para estudo bibliométrico do Portal Capes

1º passo – Identificação das informações cadastrais das teses e dissertações

O passo inicial é a identificação das informações cadastrais sobre cada trabalho (tese ou dissertação) a ser analisado. A partir das teses e dissertações cadastradas na BDTD, devem ser extraídas as seguintes informações sobre cada uma delas: código de identificação exclusivo de cada trabalho; área e subárea do conhecimento; titulação obtida (doutorado, mestrado acadêmico ou profissional); instituição; região geográfica; etc. As informações que eventualmente não estiverem disponíveis na BDTD deverão ser acrescentadas ao cadastro – é importante que seja mantida a integridade entre essas informações cadastrais e aquelas contidas no Portal de Periódicos da

Capes, como, por exemplo, a mesma classificação de áreas e subáreas de conhecimento.

2º passo – Identificação automática das referências utilizadas

O segundo passo consiste da identificação das referências bibliográficas utilizadas em cada uma das teses e dissertações, armazenando-se essas informações de forma padronizada. Idealmente, propõe-se aqui a criação de uma base de citações da BDTD. Uma possibilidade tecnológica que viabilizaria esse processo é a utilização de programas gerenciadores de citações (como, por exemplo, o EndNoteWeb, o ProCite e o *Reference Manager*) para a produção das teses e dissertações. Essa utilização garantiria, a princípio, a padronização e a identificação mais eficiente (automatizada) das referências utilizadas em cada trabalho. Assim, ao fazer o *upload* do trabalho para a BDTD, as referências seriam automaticamente incluídas na base de citações. Para isso, o arquivo eletrônico da tese ou dissertação seria desenvolvido a partir do uso do gerenciador de citações para todas as referências utilizadas. Essas referências seriam interpretadas no momento do *upload* e inseridas de forma padronizada no cadastro. Outra alternativa seria exportar as referências a partir do programa gerenciador de citações e realizar a leitura desse arquivo para a importação automática das informações sobre as referências.

3º passo – Verificação automática da existência das referências no Portal de Periódicos da Capes

Uma vez levantadas as referências bibliográficas utilizadas nas teses e dissertações, o terceiro e último passo do processo é a consulta contínua ao Portal para verificação da disponibilidade de cada referência, registrando-se a data dessa consulta e se a referência bibliográfica foi encontrada ou não no acervo disponibilizado pelo Portal. Se a referência for encontrada, é interessante também registrar-se em que base de dados ou periódico do Portal ela está contida e se o material está disponível com texto completo ou não. Periodicamente, essas consultas seriam executadas novamente para verificar se houve a inclusão da referência no acervo do Portal. Tecnicamente,

agentes de *software* similares aos robôs, respeitados os protocolos de comunicação previamente estabelecidos, são exemplos de aplicações tecnológicas que permitiriam a consulta contínua às bases disponibilizadas no Portal, registrando os resultados encontrados para cada consulta. Conforme Cendón (2001, p. 41), “os robôs [...] são programas que o computador hospedeiro da ferramenta de busca lança regularmente na Internet, na tentativa de obter dados sobre o maior número possível de documentos para integrá-los, posteriormente, à sua base de dados”.

3.1 A construção do indicador

Após o detalhamento do processo, a construção do indicador é proposta a partir da criação de um banco de dados relacional para o armazenamento das informações coletadas. As informações cadastrais e as referências citadas em cada tese ou dissertação, e ainda a identificação de sua disponibilidade no Portal, serão armazenadas nesse banco de dados relacional. Essas informações permitirão uma análise bibliométrica muito rica (a partir do uso da linguagem *Structured Query Language* (SQL) para a recuperação das informações) sobre o impacto do Portal de Periódicos da Capes na produção científica nacional.

A estruturação de consultas dinâmicas vinculadas a esse banco de dados permitirá a análise contínua do indicador proposto, uma vez que os agentes de *software* – de acordo com o terceiro passo proposto – continuamente incluiriam nessa base as informações cadastrais e referências utilizadas nas novas teses e dissertações disponibilizadas na BDTD.

A seguir, descrevemos o ambiente tecnológico em que estão sendo desenvolvidos os três passos propostos para a implantação desse processo. Até o presente momento, o site www.adaptabilidade.com.br apresenta todas as funcionalidades e os recursos que são descritos a seguir.

3.1.1 A identificação das teses e dissertações

O volume de informações na internet cresce a cada dia, e, para que sua disseminação ocorra de forma mais eficiente, é necessária a criação de padrões de dados, formalmente conhecidos como metadados. De acordo com Souza, Catarino e Santos (1997, p. 94), “metadados são descrições de dados armazenados em banco de dados [ou] ‘dados sobre dados a partir de um dicionário digital de dados’. Esse dicionário de dados normalmente é utilizado para organizar os metadados”. Para Cardoso Junior (2007, f. 7), metadados são “informações estruturadas sobre os recursos presentes em um repositório de dados”. Sousa, Catarino e Santos (1997, p. 97) destacam as seguintes vantagens na adoção desses metadados:

- Estabelecimento de padrões de dados diante da heterogeneidade de informações contidas na rede;
- Facilidade na definição da linguagem de consulta;
- Facilidade e maior precisão na recuperação das informações desejadas; e
- Troca de informações entre aplicações e entre organizações.

Existem vários padrões de metadados para vários tipos de informações; os mais conhecidos e utilizados no campo da Biblioteconomia e Ciência da Informação são o *Machine Readable Card* (MARC) e o *Dublin Core* (DC).

O MARC foi o primeiro padrão de formato de dados bibliográficos criado para possibilitar o intercâmbio de dados em uma catalogação automatizada. Ele foi criado pela *Library of Congress* dos Estados Unidos na década de 60. Atualmente, o formato MARC possui aproximadamente 800 campos de códigos associados, o que torna impraticável implementá-lo em sua totalidade em muitos repositórios informacionais. São exemplos desses códigos:

- Código 100: autor;
- Código 245: título;
- Código 500: notas; e
- Código 650: assunto.

Sendo o padrão MARC muito complexo, outros padrões mais simples têm sido desenvolvidos. Um bom exemplo é o DC, criado na cidade de Dublin, em 1995, com o propósito de organizar as informações nas páginas da web, bem como estabelecer padrões de classificação e catalogação dessas informações na internet (SOUSA; CATARINO; SANTOS, 1997).

O DC possui apenas 15 elementos básicos de dados usados para descrever uma variedade de fontes digitais e facilitar assim a recuperação de documentos eletrônicos. Esses elementos estão descritos na Tabela 1, logo a seguir.

Tabela 1. Elementos básicos do DC

Elemento	Descrição
Título	O nome dado ao documento eletrônico pelo autor ou editor
Autor	Pessoas ou organizações responsáveis pelo conteúdo intelectual do objeto
Assunto	Representa o assunto do documento eletrônico
Descrição	É uma descrição do conteúdo, podendo ser um resumo ou uma descrição com texto livre
Editor	Entidade responsável por disponibilizar o recurso criado
Outros colaboradores	Outras pessoas, entidades ou outros serviços que contribuíram para a realização da obra
Data	Data de quando o documento foi disponibilizado
Tipo de recurso	A natureza ou o gênero do conteúdo do recurso, tais como: home page, novela, poema, arquivo de dados, etc.
Formato	É a manifestação física do documento eletrônico e pode incluir um tipo de mídia ou dimensões do recurso
Identificação	Referência única que identifica o documento (URL, DOI, etc.)
Fonte	Referência do documento que deu origem ao documento eletrônico
Idioma	O idioma do conteúdo intelectual do recurso
Relação	Referência para outros documentos eletrônicos ou impressos relacionados
Cobertura	Localização no espaço, tempo ou na jurisdição do documento
Direito autoral	Informações sobre os direitos de uso, copyright

Fonte: Souza; Catarino e Santos (1997, adaptado).

A importância e relevância desses dois metadados citados podem ser comprovadas na pesquisa realizada por Knutson, Palmer e Twidale (2003, p. 2) (ver Gráfico 2), que analisaram os padrões de metadados utilizados entre 227 instituições que possuíam algum tipo de acervo digital.

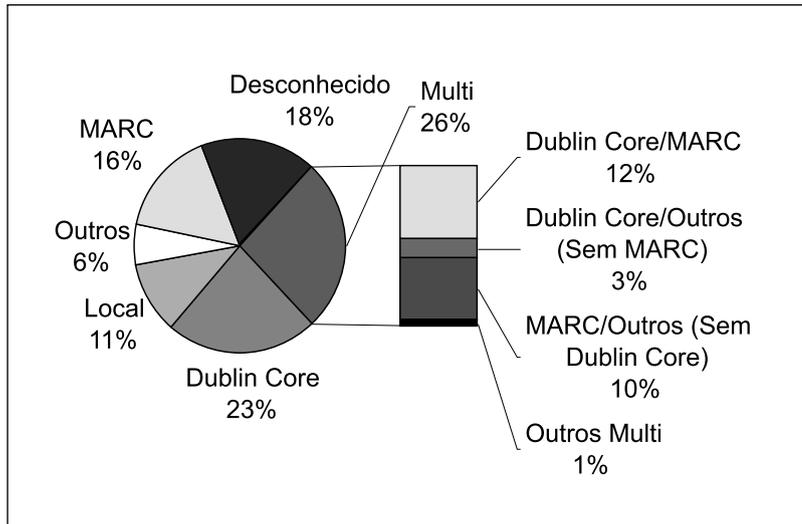


Gráfico 2. Percentagem dos metadados utilizados nas instituições analisadas

A partir do conceito exposto até aqui, para viabilizar o primeiro passo proposto no processo foi criada uma tela inicial que permite a inclusão de metadados para a tese ou dissertação enviada. O usuário deve informar a região geográfica, instituição de ensino, titulação relacionada ao trabalho, área e subárea de conhecimento – sendo essas últimas padronizadas de acordo com a classificação do Portal. Sugere-se a criação de um acesso direto para consulta às áreas e subáreas de conhecimento do Portal, preservando assim a integridade dessas informações quando ocorrerem eventuais mudanças na classificação do Portal.

Além dessas informações, também devem ser cadastrados o ano da defesa, o título do trabalho, o autor, o orientador, até quatro palavras-chave e o resumo do trabalho – como mostra a Figura 2 a seguir.

1º PASSO – Identificação das informações cadastrais das teses e dissertações

Região:	<----- Escolha ----->	IES:		Resumo:
Área:	<----- Escolha ----->	Subárea:		
Título:	<----- Escolha ----->	Ano da defesa:		
Título:				
Autor:				
Orientador:				
Palavras-chave:				

Figura 2. Identificação das informações cadastrais das teses e dissertações

Essa tela de cadastro foi criada a partir do uso da linguagem PHP, e os dados informados são armazenados em um banco de dados MySQL. Após a finalização dos dados cadastrais, o usuário deve informar o arquivo que contém a tese ou dissertação, com as citações geradas a partir do gerenciador de citações EndNote, e realizar o *upload* das informações. O programa armazenará as informações no banco de dados relacional MySQL, finalizando assim o primeiro passo do processo.

Os arquivos contendo os testes realizados até o momento foram gerados no formato *Rich Text Format* (RTF), e as citações organizadas a partir do uso do gerenciador de citações EndNoteWeb. Para os exemplos gerados, foi utilizado o editor de textos Microsoft Word 2003 e o sistema operacional Windows 7 Home Premium.

3.1.2 A identificação automática das referências utilizadas

Ao realizar-se o *upload* do arquivo da tese ou dissertação, desde que as citações tenham sido geradas a partir do EndNoteWeb, as referências são identificadas por outro programa já desenvolvido em PHP. Logo abaixo das informações cadastrais, aparecem os títulos de todos os arquivos de trabalho já cadastrados no sistema em ordem de classificação descendente, de acordo com o momento do cadastro – dessa forma, o arquivo mais recentemente cadastrado aparecerá no início da listagem.

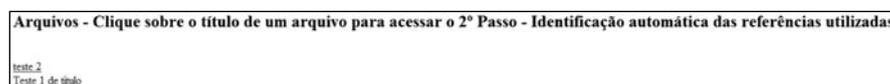


Figura 3. Acesso à identificação automática das referências utilizadas

Ao clicar sobre o *link* de qualquer arquivo, é apresentada a tela com as informações resultantes do segundo passo do processo – a extração automática dos metadados das citações identificadas no arquivo RTF informado.

A versão atual do sistema permite a identificação dos metadados do trabalho selecionado (título, autor, orientador, ano da defesa, palavras-chave, área e subárea de conhecimento, região geográfica e instituição de ensino), de acordo com o exemplo a seguir.

PROPOSTA DE CRIAÇÃO DE UM INDICADOR PARA ANÁLISE DA ADEQUAÇÃO DA COLEÇÃO DO PORTAL DE PERIÓDICOS DA CAPES À PRODUÇÃO CIENTÍFICA NACIONAL	
2º PASSO – Identificação automática das referências utilizadas	
Título	Teste 1 de título
Autor	autor
Orientador	orientador
Ano da defesa	2010
Palavras-chave	p
Área de conhecimento	Ciências sociais aplicadas - Ciência da informação
Região geográfica	Sudeste - Universidade Federal de Minas Gerais
Resumo	Resumo do primeiro teste.

Figura 4. Identificação dos metadados do trabalho selecionado

Além desses metadados, também são mostradas as referências citadas, identificadas automaticamente para o trabalho. A figura a seguir ilustra as citações e os recursos disponibilizados pelo sistema.

Ocorrências	Referência
10	Journal Article - Transinformação <i>Cendom 2009 Uso dos Periféricos do Portal CAPES pelos Programas de Pós-graduação da Faculdade de Odontologia de Araraquara Unesp, no período de 2000 a 2005 (Cendon, Silvieri-Araújo, 2009) http://revistas.pue-campinas.edu.br/transfinfo/viewissue.php?id=22 Transpormação 133-149 2 Cendom, Beatriz Valadares Silvieri-araújo, Gustavo Leves, Maria Helena Matsumoto Komasti Ribeiro, Nêlsa Ameno Moreira, Lucilia Vilarino Montagnoli, Marley Cristina Cgjusoli Berbert, Fêlbio Luiz Camargo Villela 1276039860 17 2009 31 1246039860 21</i>
1	Thesis - Centro de Ciências Sociais Aplicadas <i>Costa. 2008 USABILIDADE DO PORTAL DE PERIÓDICOS DA CAPES (Da, 2008) http://dci2.ccsa.ufpb.br:8080/jspui/hndle/123456789/99 Centro de Ciências Sociais Aplicadas 231 Costa. Luciana Ferreira da 1276039860 João Pessoa 32 2005 35 Universidade Federal da Paraíba 1276039860 Mestrado</i>

1	<p>Book Section - Asist 2002: Proceedings of the 65th Assist Annual Meeting, Vol 39, 2002</p> <p><i>Major 2002 Integrating traditional user-basead search and retrieval with information technologies for metadata sharing in a global environmental change information system (Major e Olsen, 2002)</i> :/000180277800104_0044-7870_1-57387-167-2 <i>Asist 2002: Proceedings of the 65th Asist Annual Meeting, Vol 36, 2002 545-546</i> Major, G. R. Olsen, L. M. English 1276008094 Medford 5 SSAI, Lanham, MD 20706 USA. NASA, GSFC, Greenbelt, MD 20770 USA. Jajor, GR, SSAI, 10210 Greenbelt Rd, Suite 400, Lanham, MD 20706 USA. 2002 11 Information Today Inc 1276008094 <i>ISI:000180277800104</i> Toms, E. G. 39 <i>Proceedings of the Asist Annual Meeting</i></p>
1	<p>Journal Article - Investigacion Bibliotecologia</p> <p><i>Rojas 2008 Libracy and information science within the context of the social sciences and humanities. Epistemology, methodology and interdisciplinarity (Rojas, 2008) Jan-Apr library science spistemology science methodology inerdisciplinarity</i> :/000258880800004_0187-358X <i>Article Investigacion Bibliotecologica 65-76 44</i> Rojas, M. A. R. Spanish 1276007992 17 Univ Nacl Autonoma Mexico, Ctr Univ Invest Bibliotecol, Mexico City, DF, Mexico. Rojas, MAR, Univ Nacl Autonoma Mexico, Ctr Univ Invest Bibliotecol, Mexico City, DF, Mexico. marr@servidor.unam.mx 2008 8 <i>Investig. Bibliotecol.</i> 1276007992 <i>ISI:000258880800004_22</i></p>
2	<p>Journal Article - Information Research-na International Electronic Journal</p> <p><i>Suominen 2007 The problem of 'userism', and how to overcome it in library theory (Suominen, 2007) Oct</i> :/000207111500033_1368-1613 <i>Article Information Research-an International Electronic Journal 10 4</i> Suominen, V. English 12760008000 17 Univ Oulu, Dept Finnish Informat Studies & Logoped, Oulu, Finland Suominen, V, Univ Oulu, Dept Finnish Informat Studies & Logoped, Oulu, Finland. 20047 9 <i>Inf. Res.</i> 1276008000 <i>ISI:000207111500033_colis33_12</i></p>

Figura 5. Ocorrências das referências identificadas automaticamente

A primeira coluna informa o número de ocorrências da referência no trabalho – quantas vezes ela foi citada ao longo do texto. A segunda coluna contém informações sobre a referência, sendo a primeira a sua classificação (identificada a partir da tag *type-ref* exemplificada adiante) – no exemplo acima, são apresentadas referências dos tipos *Journal Article*, *Thesise Book Section*. É importante observar que, caso exista o endereço de acesso na internet para a referência, essa classificação estará sinalizada como um *hiperlink*, gerado automaticamente pelo sistema, e permitirá o acesso direto à referência citada. No exemplo anterior, apenas a citação identificada como *Book Section* não permite esse acesso. Também é identificado o local de publicação da referência.

Finalmente, todas as informações disponibilizadas pelo EndNoteWeb são apresentadas, sem a intenção de adequação a nenhuma norma específica nesse momento. A formatação das referências de acordo com o padrão ABNT pode ser estudada para a implantação em futuras versões do sistema.

3.1.3 A verificação automática da existência das referências no Portal de Periódicos da Capes

Esse passo ainda está em fase de desenvolvimento. Pretendemos, em breve, aprimorar o sistema para que realize consultas automáticas a algumas bases de citações, utilizando o protocolo Z39.50, descrito adiante. Para a visualização do indicador, descrito na próxima seção, a versão atual do sistema marca aleatoriamente algumas referências identificadas no segundo passo, simulando a realização da consulta às bases do Portal. Assim, o indicador apresentado no site www.adaptabilidade.com.br é apenas um índice fictício para demonstrar a viabilidade técnica da presente proposta.

Para viabilizarmos a consulta às bases do Portal, para a identificação da existência das referências citadas no Portal e a disponibilização automática do indicador, será necessária a utilização de certos protocolos. O protocolo mundialmente aceito para a recuperação de informação é o Z39.50 e a sua primeira versão foi lançada em 1988 pela *The National Information Standards Organization* (NISO). O protocolo Z39.50 pode ser definido como

Um protocolo de comunicação entre computadores desenhado para permitir pesquisa e recuperação de informação – documentos com textos completos, dados bibliográficos, imagens, multimeios – em redes de computadores distribuídos. Baseado em arquitetura cliente/servidor e operando sobre a rede Internet, o protocolo permite um número crescente de aplicações. E como esse ambiente é muito dinâmico, no qual o protocolo é aplicado, é preciso que a norma seja constantemente analisada e atualizada para proporcionar as mudanças de que os criadores, provedores e usuários de informação necessitam (MOEN, 1995 apud ROSETTO, 1997).

Esse protocolo permite ao usuário acessar qualquer base de dados do mundo que o utiliza com apenas um clique; ele até mesmo possibilita uma união virtual de vários catálogos de diferentes

bibliotecas. É o Z39.50 que viabiliza a mais nova ferramenta de busca do Portal, a busca integrada, que é a busca simultânea em todas as bases do Portal que trabalham com esse protocolo.

Outro protocolo de comunicação também existente é o *Open Archives Initiative Protocol for Metadata Harvesting* (OAI PMH) que, diferentemente do Z39.50, não permite uma busca on-line, ou seja, uma busca imediata ao documento. O OAI PMH é um protocolo que permite “a exposição de metadados dos trabalhos nele armazenados para coleta automática de dados (harvesting) e reuso por provedores de serviços de informação, que como eles podem criar diferentes serviços de valor agregado” (MARCONDES; SAYÃO, 2002, p. 46).

Nesse ponto, é importante diferenciar provedor de dados e provedor de serviços. Provedores de dados são gestores de arquivos com, no mínimo, as seguintes facilidades: mecanismos de autossucessão de trabalhos e artigos, sistema de armazenamento de documentos eletrônicos e a exposição dos metadados do arquivo para permitir a colheita por terceiros (KURAMOTO, 2006). Esses terceiros seriam os provedores de serviço que são “provedores que coletam metadados de um ou mais provedores de dados e com estes metadados prestam serviços de valor agregado” (MARCONDES; SAYÃO, 2002, p. 48).

Para exemplificar os dois conceitos discutidos anteriormente, podemos citar a BDTD de cada Instituição de Ensino Superior (IES) como um provedor de dados e a BDTD do IBICT como um provedor de serviços, pois esta não tem o arquivo armazenado, realizando apenas a coleta dos metadados dos documentos das IES. Analogamente, o Portal de Periódicos da Capes pode ser considerado um provedor de serviços, de acordo com a definição citada, pois não possui o documento eletrônico, mas aponta para esses documentos nas bases de dados ou nas editoras.

3.1.4 Breves considerações sobre o ambiente de testes e a extração automática das citações

Os programas usados no site www.adaptabilidade.com.br foram desenvolvidos com a linguagem PHP, e os dados, gravados no banco de dados relacional MySQL. Os arquivos gerados para os testes foram gravados a partir do editor de textos Word 2003 e do sistema operacional Windows 7 Home Premium, com o módulo do EndNoteWeb usado para gerar as citações.

Um dos arquivos gerados para os testes continha o seguinte trecho: “De acordo com alguns autores (Cendon, Silvieri-Araújo et al., 2009)”. Esse trecho contém uma citação que foi gerada a partir do uso do EndNoteWeb e, internamente no arquivo, ela é gravada como uma série de tags, conforme a figura a seguir.

```
<EndNote><Cite>
  <Author>Cendon</Author>
  <Year>2009</Year>
  <IDText>Uso dos Periódicos do Portal CAPES pelos Programas de P&F3s-gradua&e7e3o da Faculdade de Odontologia de Araraquara UNESP, no per&e7edodo de 2000 a 2005</IDText>
  <DisplayText>(Cendon, Silvieri-Ara&e7fajo <style face=&quot;italic&quot;> et al.</style>, 2009)</DisplayText>
  <record>
    <urls><related-urls>
      <url>http://revistas.puc-campinas.edu.br/transinfo/viewissue.php?id=22</url>
    </related-urls></urls>
    <titles><title>Uso dos Periódicos do Portal CAPES pelos Programas de P&F3s-gradua&e7e3o da Faculdade de Odontologia de Araraquara UNESP, no per&e7edodo de 2000 a 2005</title></titles>
    <titles><secondary-title>Transinforma&e7e3o</secondary-title></titles>
    <pages>133-149</pages><number>2</number>
    <contributors><authors>
      <author>Cendon, Beatriz Valadares</author>
      <author>Silvieri-ara&e7fajo, Gustavo</author>
      <author>Leves, Maria Helena Matsumoto Komasti</author>
      <author>Ribeiro, N&e7e9dia Ameno</author>
      <author>Moreira, Lucilia Vilarino</author>
      <author>Montagnoli, Marley cristina Cgiusoli</author>
      <author>Berbert, F&e7elbio Luis Camargo Villela</author>
    </authors></contributors>
    <added-date format=&quot;utc&quot;>1276039860</added-date>
    <ref-type name=&quot;Journal Article&quot;>17</ref-type>
    <dates><year>2009</year></dates>
    <rec-number>34</rec-number><last-updated-date format=&quot;utc&quot;>1276039860</last-updated-date>
    <volume>21</volume>
  </record>
</Cite> </EndNote>
```

Figura 6. Tags destacadas de uma citação gerada pelo EndNoteWeb

Inicialmente, o programa PHP identifica cada trecho do arquivo contido entre as tags <Cite><EndNote> e </Cite></EndNote>. Cada uma delas é considerada uma citação. Para cada uma dessas citações, o programa identifica as seguintes tags, gravando seu conteúdo no banco de dados:

- <url> e </url>: grava o conteúdo do *link* para o arquivo da referência na internet, se existir;

- <title> e </title>: grava o conteúdo do título da referência;
- <secondary-title> e </secondary-title>: grava o conteúdo do local da publicação; e
- <ref-type> e </ref-type>: grava o conteúdo do tipo de referência.

A prioridade na implantação do protótipo foi mostrar a viabilidade técnica da construção do indicador e alguns problemas iniciais, como, por exemplo, a acentuação das referências, que foi deixada para tratamento em uma segunda etapa.

3.1.5 Resultados alcançados

O site www.adaptabilidade.com.br contém um indicador fictício gerado automaticamente a partir das referências identificadas pelo sistema, a partir dos arquivos de teste cadastrados na primeira fase.

O acesso a esse indicador pode ser realizado na tela inicial da primeira fase, logo abaixo do botão de *upload* do arquivo RTF, conforme a figura a seguir.

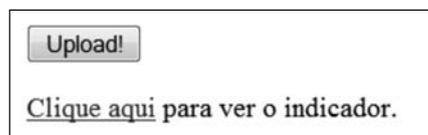


Figura 7. Acesso ao indicador de avaliação bibliométrica

O *hiperlink* anterior direciona o usuário para a tela em que o indicador é apresentado. Essa versão inicial do sistema, que tem por objetivo apenas demonstrar a viabilidade técnica da proposta, permite o filtro apenas pela região geográfica, sendo possível visualizar o indicador de três formas:

- Indicador geral: mostra o percentual de cada tipo de referência que está marcada como existente no Portal. Lembramos que, até o presente momento, esse é apenas um indicador fictício, gerado aleatoriamente pelo sistema, pois o terceiro e último passo do processo ainda está em fase de desenvolvimento. Para visualizar o valor geral do indicador, basta selecionar a opção “Não agrupar” no filtro da região geográfica, conforme a figura a seguir.

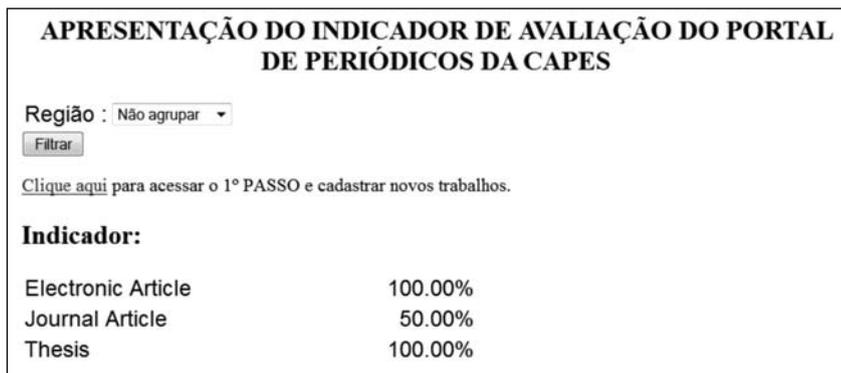


Figura 8. Indicador geral

- Outra forma possível de visualização do indicador é por região geográfica. Para essa forma, basta selecionar a opção “TODAS” no combo de regiões geográficas. Será apresentado o indicador para cada tipo de referência por região geográfica, como mostra a figura a seguir.

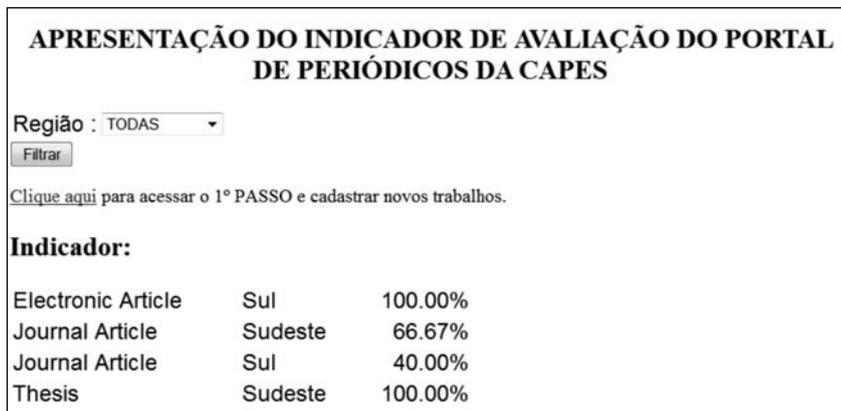


Figura 9. Indicador por região geográfica geral

- A terceira forma de visualização do indicador é para uma região geográfica específica, que pode ser escolhida no combo de regiões. No exemplo a seguir, são apresentados os resultados para a região Sudeste.

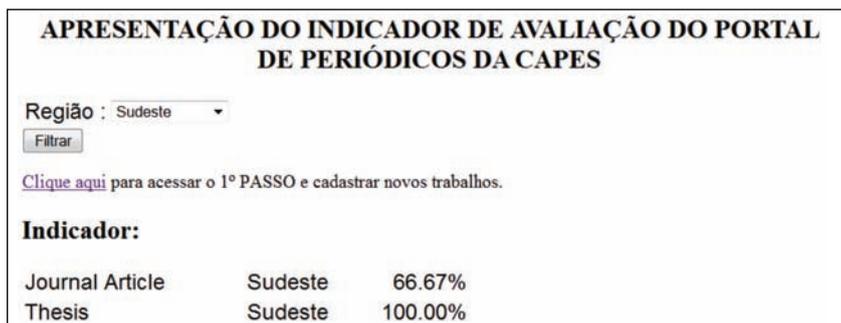


Figura 10. Indicador por região geográfica específica

Até o presente momento, somente está disponível o filtro por região geográfica, mas em breve pretendemos disponibilizar mais filtros – e a combinação entre filtros – por área e subárea de conhecimento, instituição de ensino, titulação, ano de defesa e outros. Nossa intenção foi apenas apresentar, de forma prática, a viabilidade técnica para a construção do indicador para análise da adequação da coleção do Portal de Periódicos da Capes à produção científica nacional.

4 Resultados esperados

Uma vez implantado esse processo, a análise bibliométrica poderia ser realizada a partir das informações cadastrais (região geográfica, área e subárea de conhecimento, nível de titulação), além de permitir relatórios agregados com a existência ou não das referências citadas nas teses e dissertações da BDTD – em texto completo ou não – no Portal.

O indicador poderá ser utilizado pela Capes como um importante instrumento de avaliação para subsidiar as decisões de expansão de investimento, segmentando as informações por região, área e subárea de conhecimento e nível de titulação. Poderá também auxiliar na identificação do impacto de cada base disponibilizada no Portal para a produção científica nacional, isto é, como cada base está sendo usada para a produção de teses e dissertações nas diversas áreas de conhecimento.

A partir da implantação desse processo, espera-se contribuir para a discussão sobre a adequação e gestão das coleções das bibliotecas digitais, em especial do Portal de Periódicos da Capes. Ao acompanhar a evolução do indicador proposto, será possível interferir no gerenciamento do conteúdo disponibilizado na biblioteca digital, buscando uma melhor adaptação de seu conteúdo às necessidades da comunidade científica nacional.

Uma vez levantadas as informações propostas neste trabalho, será possível uma análise mais abrangente – em termos geográficos e de área do conhecimento – do impacto do conteúdo do Portal sobre a produção de teses e dissertações, considerando-se a utilidade de seu conteúdo para o desenvolvimento desses trabalhos.

5 Considerações finais

O indicador proposto – que, além de uma visão geral, poderá também ser segregado por área e subárea de conhecimento, região geográfica, instituição e nível da titulação (doutorado, mestrado acadêmico ou profissional) – é de grande importância, na medida em que pode subsidiar algumas decisões de investimento e gerenciamento do conteúdo do Portal de Periódicos da Capes.

Esse indicador também permitirá o desenvolvimento de novos estudos focados no impacto do uso do Portal na pesquisa e na produtividade das instituições usuárias ou na adequação do conteúdo do Portal – estudos atualmente escassos.

Sugerimos a criação de alguma forma de incentivo para que as instituições sejam estimuladas a divulgar esses trabalhos, como, por exemplo, uma maior pontuação na avaliação dos programas de pós-graduação – no conceito atribuído pela Capes – para aquelas instituições que obtiverem elevado índice de disponibilização de suas teses e dissertações na BDTD, com o uso do gerenciador de citações para a composição automática do indicador proposto.

O processo apresentado neste trabalho pode ser generalizado para outras bibliotecas digitais, expandindo a análise do impacto do Portal para outros tipos de publicação que não somente teses e dissertações. A título de exemplo, podemos considerar a análise da revista *Perspectivas em Ciência da Informação*, da Escola de Ciência da Informação da UFMG. Poderia ser buscada a implantação desse processo, exigindo-se a entrega dos artigos aceitos para publicação em um formato gerado por programa de gerenciamento de citações. A partir da entrega e identificação desses artigos (primeiro passo), as citações seriam cadastradas automaticamente na base (segundo passo) e a consulta constante ao Portal (terceiro passo) permitiria uma análise de seu impacto na produção da revista.

Para finalizar, acreditamos que o indicador proposto pode contribuir para a discussão do conceito de adaptabilidade de um sistema de informações, ou seja, a sua capacidade de adaptar-se continuamente às necessidades de seus usuários – no caso específico, a partir de um melhor gerenciamento do conteúdo ofertado.

Recebido 31/07/2010

Aprovado 01/08/2011

6 Referências bibliográficas

CARDOSO JUNIOR, M. J. M. **Clio-i**: interoperabilidade entre repositórios digitais utilizando o protocolo OAI-PMH. 2007. 137 f. Dissertação (Mestrado em Ciência da Computação) – Centro de Informática, Universidade Federal de Pernambuco, Recife, 2007.

CASTRO, F. F.; SANTOS, P. L. V. A. C. Os metadados como instrumentos tecnológicos na padronização e potencialização dos recursos informacionais no âmbito das bibliotecas digitais na era da internet. **Informação & Sociedade: Estudos**, João Pessoa, v. 17, n. 2, p. 13-19, maio/ago. 2007.

CENDÓN, B. V. Ferramentas de busca na web. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 39-49, jan./abr. 2001. Disponível em: <<http://>

revista.ibict.br/index.php/ciinf/article/view/222/197>. Acesso em: 03 jul. 2010.

CENDÓN, B. V.; RIBEIRO, N. A. Análise da literatura acadêmica sobre o Portal Periódico CAPES. **Informação & Sociedade: Estudos**, João Pessoa, v. 18, n. 2, p. 157-178, maio/ago. 2008.

CUNHA, A. Á. L. **Uso de bibliotecas digitais de periódicos: um estudo comparativo no Portal de Periódicos CAPES entre áreas de conhecimento**. 2009. 207 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

FERNANDES, W. R. **Ciência da informação no portal de periódicos Capes: estudo bibliométrico para avaliação da coleção e da interdisciplinaridade da área**. 2009. 121 f. Monografia (Especialização em Gestão Estratégica da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

HOMMERDING, N. M. S. **Em busca da avaliação de bibliotecas digitais: caminhos e descaminhos**. 2007. 217 p. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação e Arte, Universidade de São Paulo, São Paulo, 2007.

KNUTSON, E.; PALMER, C.; TWIDALE, M. Tracking metadata use for digital collections. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2003, Seattle, Washington. **Electronic proceedings...** Disponível em: <http://www.siderean.com/dc2003/706_Poster49-color.pdf>. Acesso em: 29 abr. 2010.

KURAMOTO, H. Informação científica: proposta de um novo modelo para o Brasil. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 91-102, maio/ago. 2006.

LANCASTER, F. W. **Avaliação de serviços de bibliotecas**. Brasília: Brinquet de Lemos Livros, 1996.

MARCONDES, C. H. et al. (Orgs.). **Bibliotecas digitais: saberes e práticas**. 2. ed. Salvador: EdUfba; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2006.

MARCONDES, C. H.; SAYÃO, L. F. Documentos digitais e novas formas de cooperação entre sistemas de informação em C&T. **Ciência da Informação**, Brasília, v. 31, n. 3, p. 42-54, set./dez. 2002.

ORTELLADO, P. As políticas nacionais de acesso à informação científica. **Liinc em Revista**, Rio de Janeiro, v. 4, n. 2, p. 186-195, set. 2008.

RODRIGUES, K. O.; ODDONE, N. Aferindo os descritores temáticos do Portal da CAPES para a coleção de periódicos da Ciência da Informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. **Anais eletrônicos...** Salvador: Ufba, 2006. 1 CD-ROM.

ROSETTO, M. Uso do protocolo Z39.50 para recuperação de informação em redes eletrônicas. **Ciência da Informação**, Brasília, v. 26, n. 2, maio/ago.1997. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651997000200004&lng=en&nrm=iso>. Acesso em: 29 abr. 2010.

_____. Bibliotecas Digitais: cenário e perspectivas. **Revista Brasileira de Biblioteconomia e Documentação, Nova Série**, São Paulo, v. 4, n. 1, p. 101-130, jan./jun. 2008.

SILVA, N. C.; SÁ, N. O.; FURTADO, S. R. S. Bibliotecas digitais: do conceito às práticas. In: SIMPOSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 2004, Campinas. Disponível em: <<http://libdigi.unicamp.br/document/?code=8304>>. Acesso em: 08 jun. 2010.

SMEATON, A. F.; CALLAN, J. Personalisation and recommender systems in digital libraries. **International Journal of Digital Libraries**, v. 5, n. 4, p. 299-308, ago. 2005.

SOUZA, T. B.; CATARINO, M. E.; SANTOS, P. C. Metadados: catalogando dados na internet. **Transinformação**, Campinas, v. 9, n. 2, p. 93-105, maio/ago. 1997.

TAGUE-SUTCLIFFE, J. An introduction to informetrics. **Information Processing & Management**, v. 28, n. 1, p. 1-3, 1992.