

Metadados, *Web* Semântica, Categorização Automática: combinando esforços humanos e computacionais para a descoberta e uso dos recursos da *web*

Rafael Port da Rocha

RESUMO

Pesquisadores em Ciência da Computação e Ciência da Informação encontram-se frente ao desafio de prover a usuários da *web* facilidades para descobrir e usar os seus recursos. Este artigo apresenta metadados, *Web* Semântica e categorização automática como técnicas em que habilidades humanas, como a indexação e a descrição de recursos, são aplicadas de uma forma que máquinas possam estabelecer julgamentos, para melhor ajudar humanos na descoberta e no uso dos recursos da *web*.

PALAVRAS-CHAVE: Internet. Busca de Informações. Metadados. *Web* Semântica.

1 INTRODUÇÃO

Hoje vivemos em um mundo em que uma quantidade significativa de informação é produzida em meio digital e disponibilizada através da Internet. Este cenário proporciona um acesso fácil e barato a uma grande quantidade de informação, visto que são superados aspectos como localização geográfica e suporte físico.

A super oferta de informação proporcionada por este cenário, entretanto, traz dificuldades às pessoas em encontrar aquelas informações que lhes são relevantes.

Esta dificuldade deve-se principalmente à pouca organização da informação da *web*, que impede a construção de estratégias e mecanismos de busca eficientes

Para amenizar o problema da sobrecarga de informação, as comunidades de Ciência da Informação e da Ciência da Computação têm convergido em diversos aspectos, no sentido de encontrar mecanismos para organizar as informações da *web*, como classificadores automáticos, diretórios abertos, metadados, ontologias e *Web Semântica*. Este artigo trata desses aspectos e estende o problema da busca da informação para a descoberta, localização e uso de recursos da *web*, que podem ser tanto documentos contendo informações, quanto *sites*, sistemas, computadores, páginas pessoais etc.

A seção 2 apresenta como atualmente é realizada a busca de informações na Internet. A seção 3 mostra avanços na indexação de páginas da Internet, destacando diretórios abertos e a categorização automática. A seção 4 apresenta como metadados são usados para descrever propriedades dos recursos da *web*, a fim de possibilitar não somente a busca e a localização destes recursos, mas também suas utilizações. A seção 5 apresenta a *Web Semântica*, em que técnicas de Inteligência Artificial e metadados são combinados para permitir que máquinas auxiliem pessoas a operar a *web*. Conclusões são apresentadas na seção 6, destacando a convergência entre esforços humanos e computacionais.

2 A BUSCA DE INFORMAÇÕES NA INTERNET

Atualmente, a forma usual de buscar informações da *web* é feita através dos motores de busca, como Google e Yahoo. Um motor de busca possui robôs, que visitam freqüentemente as páginas da *web* e as entregam a um indexador de palavras. Este indexador atualiza um índice, chamado de arquivo invertido, que é formado por três conjuntos: o conjunto das palavras indexadas, o conjunto dos endereços das páginas visitadas pelo robô e um conjunto composto por relacionamentos binários entre uma palavra do conjunto de palavras e um endereço do conjunto de endereços, em que cada relacionamento indica que a página do referido endereço contém a referida palavra.

Consultas formuladas por usuários são avaliadas através da utilização do arquivo invertido. Essas consultas possuem resultados não muito precisos,

devido ao fato destes mecanismos não serem capazes de identificar o significado das palavras do documento. Nesses mecanismos, a palavra “Holanda” pode ser tanto um país, como o sobrenome do compositor Chico Buarque. Esses motores de busca também não permitem a busca através de propriedades atribuídas ao documento, como a busca de um documento cujo autor foi “Holanda” ou o documento que trata do assunto “Holanda”.

Para prover mecanismos mais eficientes de busca da Internet, são necessários recursos mais poderosos que a indexação de palavras. São necessários mecanismos que indexam documentos através de conceitos e que catalogam estes documentos através da descrição de propriedades destes, como seu autor, seu título etc.

3 A INDEXAÇÃO NA INTERNET

Há séculos, especialistas em Ciência da Informação vêm desenvolvendo técnicas de indexação, para organizar documentos de empresas, acervos de museus e de bibliotecas. Através destas técnicas, especialistas, denominados indexadores, associam termos a documentos, que indicam os assuntos destes documentos. Estes termos não são simples palavras. Eles são nomes dados a conceitos estabelecidos por uma comunidade e fazem parte de um vocabulário controlado desta comunidade, construído geralmente na forma de tesouro.

Quando tratamos da *web*, entretanto, existe uma grande discussão no que diz respeito à utilização/adaptação dos mecanismos tradicionais de indexação. Para Khan (2002) a indexação manual é mais precisa, mas perde em uniformidade, já que diferentes classificadores possuem diferentes perspectivas e abordagens sobre o assunto. Segundo Souza (2000), a internet difere-se das bibliotecas tradicionais em dois aspectos fundamentais: a informação e os usuários. Esses aspectos interferem na utilização de técnicas tradicionais na indexação de páginas da *web*. Para Souza (2000)¹,

[. . .] um conjunto de documentos em uma biblioteca tradicional é selecionado em função dos objetivos próprios de cada sistema, em função das áreas de conhecimento em que atua e a caracterização do usuário a que serve, e muitas vezes, por tipo de documento. Em princípio, a Internet não seleciona nenhum tipo de documento, abrange todas as áreas do conhecimento e caracteriza como sendo seus usuários todas as pessoas que a acessam, independente de sua caracterização de tipo de usuário.

¹ Documento eletrônico.

Os documentos de uma biblioteca normalmente possuem informações qualificadas e completas, visto que geralmente passam por revisões por pares, bancas, comissões de avaliação, corpo editoriais etc. A qualidade desse material requer um processo bastante apurado de indexação, baseada em conceitos. Esta indexação pressupõe que os documentos serão buscados por usuários familiarizados com estes conceitos, e que poderão ter o apoio de bibliotecários especializados. Já no caso da Internet, muitos documentos contêm informações superficiais, redundantes e dúbias. Isso requer um processo de filtragem para eliminar aquelas informações irrelevantes. Além disso, a maioria dos usuários da Internet não tem a priori uma conceitualização muito clara do que quer obter em suas buscas. Isso leva a um questionamento da eficiência de buscas a informações indexadas por conceitos.

Outro problema enfrentado na indexação tradicional de páginas da *web* diz respeito à grande quantidade de páginas a ser indexada, que exigiria um esforço humano muito grande. Como propostas de soluções para esse problema, temos os diretórios abertos e a indexação automática.

Os diretórios são mecanismos de busca que utilizam uma estratégia para organizar as informações da *web* em que especialistas em Ciência da Informação estabelecem categorias, e indexadores indexam manualmente páginas da *web* a estas categorias. Por exigirem um processo manual de indexação, os diretórios abrangem um universo mais restrito da *web*, quando comparados com os motores de busca baseados em palavras. Por isso, com o objetivo de atingir uma maior quantidade de páginas indexadas, surgiram os diretórios abertos. *Open Directory* (www.dmoz.org) é um projeto de diretório aberto que, em vez de possuir um grupo seletivo de indexadores, é formado por uma comunidade de milhares de indexadores, denominados editores. Qualquer pessoa pode ser um editor desse diretório. A desvantagem dessa alternativa é que estes editores não são especialistas em indexação.

Técnicas de categorização automática, isto é, que indexam páginas a conceitos sem a participação humana, também têm sido desenvolvidas. Segundo Adams², a categorização automática é importante, pois o trabalho humano não possui escala suficiente para indexar uma *web*, que possui uma quantidade estimada em 2001 de um bilhão de páginas, com um incremento diário de aproximadamente 1,5 milhão de páginas. Dentre as técnicas de categorização automática, observa-se aquelas que utilizam estatísticas para encontrar padrões em artigos já indexados e utilizam esses padrões para indexar novos documentos. Utilizando essa técnica, Moens e Dumortier (1998) obtiveram sucesso para esquemas de classificação simples, não hierárquicos, em que mais

² Documento eletrônico.

de 70% de todos os documentos de uma base de dados que pertencem a uma determinada categoria são indexados a esta categoria pela ferramenta (medida chamada de *recall*) e em que 60 % dos documentos associados pela ferramenta a uma categoria são realmente desta categoria (medida chamada de *precision*). Um exemplo de indexador automático é o projeto *Scorpion* (SCHAFER, 1997), que, a partir de uma base de amostra que contém páginas indexadas ao esquema *Decimal Classification System*, classifica outras páginas da *web*.

4 O USO DE METADADOS NA DESCRIÇÃO DE INFORMAÇÕES DA INTERNET

Para proporcionar a descoberta e a localização de itens do seu acervo, bibliotecas e arquivos produzem informações (catálogos) que descrevem propriedades desses itens, como seu autor, seu título, sua data de publicação, seu assunto etc. Quando se trata da *web*, uma atividade semelhante é realizada através de metadados.

O conceito de metadados é bastante simples: metadados são dados sobre dados. Quando se trata do mundo digital, chama-se de recurso o objeto descrito por metadados, pois este pode ser tanto um simples dado, quanto um documento, uma página da *web*, ou até mesmo uma pessoa, uma coleção, um sistema, um equipamento ou uma organização. Na *web*, o conceito recurso significa qualquer objeto que pode ser atingido através de uma URI (*Uniform Resource Identifier* - www.w3.org/Addressing), como qualquer recurso que é acessado via seu endereço eletrônico. Isso inclui documentos, páginas pessoais, *sites* ou sistemas. A descrição de uma pessoa ou organização é feita através da descrição da página dessa pessoa ou organização.

Metadados descrevem os recursos da *web* com a finalidade de facilitar a sua descoberta, localização e utilização. Motores de busca, ao utilizarem estes metadados, proporcionam consultas bem mais precisas, envolvendo não somente palavras, mas propriedades descritas, como o autor do recurso, o formato do recurso, a data do recurso etc.

No mundo digital, metadados não se restringem àqueles usados para descrever recursos para permitir sua descoberta e localização. Por exemplo, metadados também descrevem os formatos dos recursos, a fim de permitir que computadores identifiquem quais aplicativos podem ser usados para manipular estes recursos. Aplicativos, ao analisarem metadados, verificam se o formato do recurso é compatível, as condições de uso do recurso, restrições de acesso etc.

Segundo Milstead e Feldman (1999), em computadores, metadados servem de representante de um recurso e para caracterizar o trabalho original, para que usuários entendam seu significado, propósito, origem e condições de uso. Metadados são vistos como dados que descrevem propriedades de um recurso para diversos propósitos, como o contexto em que recurso se insere, sua qualidade, suas condições de uso, sua identificação, suas estratégias de preservação etc. Essa diversidade de tipos de metadados dá suporte à realização de inúmeras funções, como funções de descoberta do recurso, de localização do recurso, de avaliação do recurso, de uso do recurso etc.

Segundo Milstead e Feldman (1999), embora o conceito seja bastante novo, bibliotecários estão há séculos produzindo e padronizando metadados, à medida que extraem de documentos (que seriam os dados) informações de indexação e catalogação (que seriam os metadados), para oferecer aos usuários caminhos, para que estes possam buscar os documentos de que necessitarem. Na Biblioteconomia, o padrão MARC (www.loc.gov/marc) é um exemplo de esquema de metadados. Estes metadados indicam propriedades do documento, como seu autor, sua data de publicação, seu título, seu assunto, e têm como finalidade permitir a descoberta e a localização destes documentos.

Os metadados podem estar inseridos nos próprios documentos que descrevem. Em páginas HTML é possível inserir metadados através de marcações (chamadas de meta *tags*), que não visíveis quando a página é exibida no navegador. Esses metadados podem ser lidos por motores de busca ou por aplicativos que os usam.

Os metadados também podem ser armazenados em repositórios. Nesses repositórios, os metadados são produzidos através de anotadores, que são sistemas que guiam usuários na descrição dos recursos. Esse tipo de solução tem sido usada em organizações para permitir que estas descrevam seus recursos. Essas organizações consideram, como recursos, páginas da *web* que identificam seus departamentos, seus especialistas, seus sistemas informatizados disponibilizados, bibliotecas digitais, repositórios de documentos digitais etc. Através desses metadados, estas organizações possuem um maior entendimento de suas próprias potencialidades. Um exemplo disso é o projeto *Australian Government Locator Service* (CUNNINGHAM, 1998), em que os recursos do governo australiano são descritos através de metadados e ferramentas de localização de recursos são desenvolvidas para auxiliar os cidadãos a identificar os recursos do governo que poderão dar atendimento às suas necessidades.

Na *web*, comunidades especializadas em diversas áreas do conhecimento têm definido esquemas próprios de metadados para padronizar a descrição dos seus recursos. Essa tendência fica clara na conferência denominada

International Conference on Dublin Core and Metadata Applications de 2001 (INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2001), em que podemos observar várias iniciativas em criar esquemas de metadados para áreas como ambiental, governos, educação, biblioteconomia, agricultura, matemática etc

O surgimento de uma grande diversidade de esquemas de metadados, entretanto, pode levar a uma sobreposição de esquemas, com esquemas sendo definidos para as mesmas finalidades. Esforços para evitar este surgimento caótico dos esquemas de metadados são os Registros de Metadados, os Perfis de Aplicação e a iniciativa Dublin Core.

4.1 REGISTROS DE METADADOS E PERFIS DE APLICAÇÃO

Os Registros de Metadados (HEERY; WAGNER, 2002) são repositórios em que os vários esquemas de metadados são disponibilizados, possibilitando a divulgação destes aos interessados em usá-los, evitando o surgimento de novos esquemas de metadados para mesmas finalidades e proporcionando uma universalização dos esquemas registrados.

Normalmente, um esquema de metadados é desenvolvido por uma comunidade de especialistas de um domínio do conhecimento, para descrever recursos deste domínio. Entretanto, muitas vezes temos a necessidade de descrever recursos que abrangem vários domínios de conhecimento. Neste caso, teríamos que combinar e adaptar elementos de vários esquemas de metadados. A especificação de como esquemas de metadados são combinados e adaptados para descrever um determinado conjunto de recursos chama-se Perfil e Aplicação (HEERY; PATEL, 2000). Um Perfil de Aplicação indica os esquemas adotados, quais os elementos destes esquemas serão usados e como estes elementos serão usados na descrição. O projeto Schema (<http://www.Schemaforum.org>) é um Registro de Metadados em são registrados esquemas de metadados e Perfis de Aplicação.

4.2 DUBLIN CORE:

UMA LÍNGUA FRANCA PARA METADADOS

A iniciativa Dublin Core (<http://dublincore.org>) tem como objetivo oferecer uma coleção básica de elementos de metadados (esquema básico de metadados) para serem utilizados por qualquer comunidade, independente de sua área de domínio. O esquema Dublin Core foi o resultado de uma oficina realizada em 1995, patrocinada pela comunidade da Ciência da In-

formação e por especialistas em padrões para a *web*, cuja tarefa principal concentrou-se em identificar e definir um conjunto de elementos de metadado, denominado Dublin Core, para descrever recursos da *web*, a fim de proporcionar a descoberta destes recursos (WEIBEL, 1995).

A definição destes elementos obedeceu a critérios como serem simples, fáceis de ser criados, e aplicáveis a vários domínios. Os elementos de Dublin Core estão organizados em três grupos: Conteúdo, propriedade intelectual e instância. Os elementos que descrevem o conteúdo do recurso são: Título, assunto, descrição, linguagem, fonte, recursos relacionados (relação) e abrangência (espacial ou temporal). Os elementos de propriedade intelectual são: o criador (responsável intelectual pela criação do recurso), publicador (quem tornou o recurso público), contribuidor e direitos autorais. Os elementos de instância são data, tipo (ex.: página da *web*, artigo, livro), formato (ex.: pdf, world, mp3) e identificador (ex.: URI, ISSN).

Esse esquema foi feito tanto para ser usado por alguém que construiu uma *home page* e gostaria de descrevê-la de forma genérica, quanto por especialistas em um domínio de conhecimento, que necessitam descrever recursos desse domínio com mais detalhes. Para contemplar descrições feitas por não especialistas, Dublin Core tem como objetivo ser simples, pequeno (só contém 15 elementos) e genérico não impondo restrições e dificuldades a quem quiser usá-lo. Entretanto, para contemplar a descrição de recursos de comunidades com domínios de conhecimento específicos, que requerem descrições mais aprofundadas, o conceito qualificador foi estabelecido (HABING; COLE; MITCHO, 2001).

Em Dublin Core Qualifiers, os recursos de um domínio específico são descritos pelos mesmos elementos de Dublin Core. Estes elementos, entretanto, são qualificados através de regras próprias desse domínio. Por exemplo, o elemento "criador", de Dublin Core, é usado para descrever o responsável pela produção intelectual de um recurso. Em um domínio como biblioteconomia, este elemento pode ter um significado mais específico na descrição de um livro ou artigo, isto é, o elemento criador poderia ser visto como o autor do livro. Neste mesmo domínio, o elemento assunto poderia ter qualificado por uma regra que estabelece que seus valores devam ser controlados pelo vocabulário *Decimal Classification System*.

Atualmente, muitos perfis de aplicação que adotam Dublin Core como base têm sido desenvolvidos para áreas como saúde, governo e bibliotecas. Esses perfis estendem Dublin Core através de qualificadores e da definição de elementos próprios.

5 WEB SEMÂNTICA: CONVERGÊNCIA ENTRE METADADOS E ONTOLOGIAS

Na *web*, as informações não possuem estruturas previamente definidas, suas estruturas podem ser irregulares, incompletas, parciais e podem mudar com frequência. A pouca estruturação das informações da *web* traz dificuldades na sua operação por máquinas, isto é, por agentes computacionais. Por isso, diz-se que a *web* atual é facilmente operada por humanos, mas dificilmente operada por máquinas. A falta de estruturação da *web* dificulta a construção de agentes, que auxiliam as pessoas na descoberta e manipulação de recursos.

A busca de soluções para este problema está na proposta de Berners-Lee, o criador da *web*, para uma nova geração da *web*, denominada *Web Semântica*. Na *Web Semântica*, metadados são usados para descrever o significado dos recursos da *web*, e agentes inteligentes, desenvolvidos através de técnicas de Inteligência Artificial, usam essas descrições para auxiliar usuários da *web* na localização e manipulação desses recursos.

Segundo Berners-Lee (1998), a *web* desenvolveu linguagens excelentes para expressar a informação que se destina a ser utilizada pelo homem, porém foi falha quando da manipulação desta pelas máquinas. Por isso, a *Web Semântica* surge como uma proposta de trazer à rede global uma estrutura e significado, que permitem a sua evolução de uma rede de documentos para uma rede de dados na qual toda a informação tem um significado bem definido, podendo ser interpretada e processada por humanos e computadores (BERNERS-LEE; HENDLER; LASSILA, 2001).

A *Web Semântica* não exige alterações nem impõe padrões para a *web* atual. As informações da *web* continuam como estão, pois é impossível estipular estruturas padrões para representá-las e exigir que estas sejam definidas nestas estruturas. Entretanto, a *Web Semântica* estipula uma arquitetura em que metadados semânticos são usados para descrever o significado das estruturas da *web* atual. Segundo Berners-Lee (1998), no contexto da *Web Semântica*, metadados significam informações compreensíveis por máquinas sobre recursos da *web* ou outros objetos. A *Web Semântica* chama de recurso qualquer elemento da *web* identificado através de uma URI.

Para *Web Semântica*, esquemas de metadados formam uma boa base para proporcionar um entendimento semântico comum. Entretanto, para permitir que máquinas utilizem metadados para auxiliar humanos no uso dos recursos descritos, técnicas de Inteligência Artificial são necessárias. Nesse sentido a *Web Semântica* estende o conceito de metadados, à medida que incor-

pora a estas técnicas de raciocínio e inferência. Na Inteligência Artificial, os instrumentos para descrição semântica são chamados de ontologias.

Na inteligência artificial, ontologia é definida como uma especificação explícita e formal de uma conceitualização compartilhada (GUARINO, 1998). Esta conceitualização é definida como uma visão abstrata e sistemática do mundo que se quer representar por algum propósito. A especificação desta conceitualização é explícita, pois os tipos dos conceitos e as restrições de seus usos são definidos explicitamente; ela é formal, pois deve ser compreendida por máquinas; e ela é compartilhada, pois captura o conhecimento consensual, não restrito a um indivíduo, mas aceitável a um grupo (FENSEL et alii, 2001).

6 CONCLUSÃO

Frente ao problema de como tornar utilizável a super-oferta de informação disponível na *web*, este artigo analisou avanços obtidos pela comunidade científica no que diz respeito à combinação de técnicas tradicionais da Ciência da Informação, como indexação e catalogação, com técnicas da Ciência da Computação, como categorização automática, metadados e ontologias.

Observa-se que a indexação da *web* por humanos tem como principais problemas a quantidade de informação, a qualidade da informação e o fato de que se deve levar em consideração que a *web* abrange todas as áreas do conhecimento e é acessada por qualquer tipo de usuário. Como soluções observam-se avanços na categorização automática e na universalização da indexação humana, através dos Diretórios Abertos. Observa-se que estes esforços partem de uma combinação entre esforços humanos e computacionais. A categorização automática torna-se possível à medida que humanos indexam previamente uma coleção significativa de documento, para que então máquinas possam extrair padrões para indexar novos documentos. Os diretórios abertos podem ser usados como fontes para esses indexadores.

A catalogação tradicional, através da utilização de esquemas como MARC, também não se aplica na *web*, pois esta abrange outros tipos de recursos, além daqueles tradicionalmente catalogados em bibliotecas, como, por exemplo, páginas de pessoas, páginas de organizações, páginas que são sistemas informatizados, páginas que representam produtos, lojas virtuais etc. Além disso, a *web* envolve comunidades diferentes daquelas que tradicionalmente usam bibliotecas, como consumidores, que necessitam adquirir um produto, ou cidadãos, que necessitam algum serviço de um governo. Estas comunidades são formadas por pessoas que não são especialistas em catalogação. Por isso, para catalogar recursos da *web*, a comunidade científica tem apresentado soluções como a linguagem Dublin Core.

Essas soluções partem do pressuposto de que comunidades da *web* possam definir seus próprios esquemas de metadados. Entretanto, para evitar a sobreposição de esquemas, os registros de esquemas apresentam-se como componentes de disseminação de esquemas já definidos, e Dublin Core apresenta-se como língua mãe para metadados, que pode ser estendida, através da criação de Perfis de Aplicação, para contemplar necessidades de comunidades específicas.

A utilização da máquina para auxiliar pessoas a operar informações da *web* está condicionada às estruturas e à forma com que estas informações estão organizadas. A qualidade das buscas também é incrementada, à medida que metadados oferecem descrições precisas sobre os recursos. A fim de incrementar o poder de expressão em mecanismos de descrição, especialistas de Inteligência Artificial trabalham como ontologias, que estendem metadados com recursos como regras de inferência e axiomas. Ontologias são os recursos usados na nova versão da *web*, a *Web Semântica*, para permitir que máquinas entendam o significado dos recursos da *web*.

Metadata, Semantic Web, Automatic Categorization: joining human and computer efforts for the discovery and use of web resources

ABSTRACT

Computer and Information Science researchers are facing the challenge of providing web users with facilities for discovering and using web resources. This article deals with this subject by presenting metadata, Semantic Web and Automatic Categorization as techniques in which human abilities, such as indexing and describing resources, are applied in such a way that machines can make judgments to better help humans to discover and use web resources.

KEY WORDS: Internet. Information Retrieval. Metadata. Semantic Web.

REFERÊNCIAS

ADAMS, C. **Word Wranglers-Automatic Classification Tools Transform Enterprise Documents from “bags of words” into Knowledge Resources.** Disponível em: < www.intelligentkm.com/feature/010101/feat1.shtml <<http://www.intelligentkm.com/feature/010101/feat1.shtml>>. Acesso em: 10 dez. 2002.

BERNERS-LEE, T.; HENDLER, J.; LASSILA O. The Semantic Web. **Scientific American**, NewYork, May 2001. Disponível em: <www.sciam.com>. Acesso em: 15 jul. 2004.

BERNERS-LEE, Tim. Semantic Web Road Map. **World-Wide Web Consortium (W3C)**. 1998. Disponível em: <<http://www.w3.org/DesignIssues/Semantic.html>>. Acesso em: 15 jul. 2004.

CUNNINGHAM, A. **Australian Government Locator Service**: enabling seamless online access to government. 1998. Disponível em: <http://www.naa.gov.au/recordkeeping/gov_online/agls/Metadata_paper22sept98.html> . Acesso em: 28 jan. 2002.

FENSEL, D. et alii. OIL: An Ontology Infrastructure for Semantic Web. **IEEE Intelligent Systems**, Los Alamitos, v.16, n.2, p. 38-45, Mar./Apr. 2001.

GUARINO, N. Formal Ontology and Information Systems. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS, 1998, Trento. **Proceedings** . . . Amsterdam: IOS Press, 1998. P.3-15

HABING, T.; COLE, T.; MISCHO, W. Qualified Dublin Core using RDF for Sci-Tech Journal Articles. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2001, Tokyo. **Proceedings** . . . Tokyo: National Institute of Informatics, 2001.

HEERY, R.; PATEL, M. Application Profiles: mixing and matching metadata schemas. **Ariadne**, Kassel, v. 25, Sept. 2000. Disponível em: <www.adriane.ac.uk/issue25/app-profiles>. Acesso em: 15 jul. 2004.

HEERY, R.; WAGNER, H. A Metadata Registry for the Semantic Web. **D-Lib Magazine**, Reston, Va., v.2, n.5, May 2002. Disponível em: <www.dlib.org>. Acesso em: 15 dez. 2002.

HENDER, J. Agents and the Semantic Web. **IEEE Intelligent Systems**, Los Alamitos, v.16, n.2, p.30-37, Mar./Apr. 2001.

INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2001, Tokyo. **Proceedings** . . . Tokyo: National Institute of Informatics, 2001.

KHAN, T. **Taxonomies and Classification. LIS385T** - Information architecture and design fall 2002, GSLIS, UT, Austin, 2002. Disponível em: <www.gslis.utexas.edu/~l385tdt/assignment.html>. Acesso em: 10 out. 2002.

MILSTEAD, J.; FELDMAN, S. Metadata: cataloging by any other name. **On Line Magazine**, New Haven, v.23, n.1, 1999. Disponível em: <<http://www.onlineinc.com/onlinemag/OL1999/milstead1.html>>. Acesso em: 24 jul. 2001.

MOENS, M.-F.; DUMORTIER, J. Automatic Categorization of Magazine Articles: the creation of highlight abstract. In: ANUAL INTERNATIONAL CONFERENCE IN INFORMATION RETRIEVAL, 21., 1998, Meulborne. **Proceedings** . . . [Meulborne], 1998.

NATIONAL ARCHIVES OF AUSTRALIA. **The Australian Government Locator Service (AGLS): Manual for Users, Version 1.2.** Camberra, 2000.

SHAFFER, K. Scorpion Helps Catalog the Web. **Bulletin of the American Society of Information Science**, Silver Springs, v. 24, n.1, Oct./Nov. 1997. Disponível em: <orc.rsch.oclc.org/b-asis.html>. Acesso em: 05 out. 2002.

SOUZA, R. A Classificação como Interface da Internet. **DatagramaZero - Revista de Ciência da Informação**, Rio de Janeiro, v.2, n.2, abr., 2000. Disponível em: <www.dgz.org.br>. Acesso em: 15 abr. 2003

WEIBEL, S. Metadata: The Foundations of Resource Description. **D-Lib Magazine**, Reston, July 1995. Disponível em: <www.dlib.org>. Acesso em: 15 dez. 2002.

Rafael Port da Rocha
Doutor em Ciência da Computação pelo Programa de Pós-Graduação em Ciência da Computação da UFRGS.
Professor Adjunto do Departamento de Ciências da Informação da Faculdade de Biblioteconomia e Comunicação da UFRGS.
E-mail: rafael.rocha@ufrgs.br

