

FOLKSONOMIAS: CARACTERÍSTICAS DAS ETIQUETAS NA DESCRIÇÃO DE RECURSOS DA WEB

FOLKSONOMÍAS: LAS CARACTERÍSTICAS DE LAS ETIQUETAS PARA DESCRIBIR LOS RECURSOS WEB

Maria Elisabete Catarino - beteca@uel.br

Professora adjunto do Departamento de Ciência da Informação, Universidade Estadual de Londrina. Doutora em Tecnologias e Sistemas de Informação, pelo Departamento de Sistemas de Informação da Escola de Engenharia da Universidade do Minho, Portugal (2009).

Ana Alice Baptista - analice@dsi.uminho.pt

Professora do Departamento de Sistemas de Informação da Escola de Engenharia da Universidade do Minho. Doutora em Tecnologias e Sistemas de Informação. Pesquisadora nas áreas de Bibliotecas Digitais, Repositórios, Metadados, Web e Web Semântica.

Resumo

No contexto da Web 2.0 surge a *folksonomia* que é o resultado da etiquetagem dos recursos da Web, em um ambiente social, pelos próprios usuários visando a sua recuperação. Trata-se de uma indexação livre em linguagem natural onde não são adotadas regras e/ou política de indexação e nem o controle de vocabulários. Conhecer as possibilidades desta nova forma de descrição dos recursos da Web é imprescindível. Este artigo tem o objetivo de descrever as características das etiquetas que compõem as *folksonomias*. Essas características foram observadas em um projeto de pesquisa de doutorado que teve o intuito de identificar elementos de metadados oriundos das *folksonomias* que fossem complementares ao Dublin Core (DC). Para a identificação desses metadados desenvolveu-se os seguintes procedimentos metodológicos: criação da base de dados, análise das etiquetas, identificação de propriedades complementares ao DC, validação da proposta e construção do perfil de aplicação e ontologia. Na análise das etiquetas pode-se observar várias características das etiquetas que serão aqui descritas: alfabeto, idioma e formas variantes, tais como. singular/plural, simples/composta, símbolos, números, siglas, abreviaturas, mnemônicas e mistas. Pondera-se que o conhecimento e disseminação dessas características subsidiará novos estudos e aplicações de forma a potencializar o uso das *folksonomias*.

Palavras-chave

Folksonomias. Características das Etiquetas. Descrição de recursos.

1 CONTEXTUALIZAÇÃO DO TEMA

As Tecnologias de Informação (TIs) têm evoluído e têm se tornado cada vez mais populares principalmente após o surgimento da Web. Desde sua criação a Web tem

adicionado novos serviços e funcionalidades que, cada vez mais, permitem que os seus usuários participem de forma ativa na construção e organização dos conteúdos lá disponíveis.

De fato, é num contexto de alterações sociológicas que surge o conceito de Web 2.0. Este termo, criado por Tim O'Reilly (2005), reforça o conceito da Internet de propiciar que os seus usuários colaborem efetivamente para a disponibilização de serviços virtuais e organização dos conteúdos. Um exemplo clássico desta nova geração é a Wikipedia, uma enciclopédia dinâmica, na qual os próprios usuários disponibilizam e editam a informação.

Tim O'Reilly define a Web 2.0 como a Web a funcionar como uma plataforma. As aplicações Web 2.0 são aquelas que fazem: distribuição de *software* com atualização constante para melhor utilização e reorganização de dados de múltiplas fontes por usuários individuais que, por sua vez, fornecem seus próprios dados e serviços para que sejam reorganizados por outros, assim criando uma “arquitetura da participação” indo além da metáfora da página da Web 1.0¹ para permitir a efetiva colaboração dos usuários (O'REILLY, 2005).

Em 2006, John Markoff, jornalista do *The New York Times*, criou o termo Web 3.0 para se referir à terceira geração da Web que pretende estruturar todo o conteúdo da Web a partir dos conceitos de semântica de redes. Será uma geração da Web baseada em Web Semântica, microformatos², pesquisas em linguagem natural, *data mining*, inteligência artificial, *machine learning* e *recommendation agents*.

Dentre as diversas evoluções que estão ocorrendo, destaca-se o que pode ser considerado como um novo paradigma para a organização dos conteúdos dos recursos digitais na Web: a possibilidade de os próprios usuários participarem na organização desses conteúdos. Neste novo paradigma surgem as Folksonomias. Trata-se de um novo conceito que tem sido utilizado por diversos profissionais e estudiosos da área de informação. No entanto, parece não haver ainda um consenso, quer sobre a utilização

¹ É importante notar que até ao aparecimento do termo Web2.0, nunca se considerou que esta tivesse versões. O autor utiliza o termo Web1.0 para se referir àquilo que ele considera uma versão anterior da Web. No entanto, o termo não existia até ter aparecido o termo Web2.0 da sua autoria.

² Microformatos permitem a expressão semântica dentro de páginas em xhtml a ser apresentada de forma visível para o usuário, permitindo assim um elevado nível de descrição de tipos específicos de recursos (*business cards*, eventos, etc.) (MENDEZ, BRAVO & LOPEZ, 2007).

deste termo, quer sobre o seu significado. Há os que preferiram utilizar outros termos como, por exemplo, classificação social ou *social tagging*.

Considerando que a Folksonomia é algo novo no contexto da Web faz-se necessário desenvolver pesquisas sobre as possibilidades desta nova forma de participação dos usuários da Web. Conhecer as etiquetas nas suas características ajudará a compreender de que forma elas podem contribuir para a organização da informação da Web.

Este artigo tem o objetivo de descrever as características das folksonomias observadas em uma pesquisa desenvolvida no âmbito do doutorado em Tecnologias e Sistemas de Informação (CATARINO, 2009a)³.

A pesquisa de doutorado teve o intuito de identificar novas propriedades com base nas folksonomias que sejam complementares ao conjunto de metadados DC, para descrição de recursos, em especial para Repositórios Institucionais. Para atingir ao objetivo proposto a pesquisa foi desenvolvida seguindo os seguintes procedimentos metodológicos:

- Criação da Base de Dados – Antecedendo a análise das etiquetas contidas nas folksonomias, procedeu-se o tratamento dos dados e a criação da base de dados;
- Análise das Etiquetas – Nesta fase foram analisadas todas as etiquetas atribuídas pelos usuários a cada um dos recursos componentes do conjunto de dados para que posteriormente fossem identificadas as propriedades a elas relacionadas. Esta análise tomou em consideração os usuários e utilizações do recurso, de modo a que se conseguisse compreender a semântica do maior número possível de etiquetas;
- Identificação de propriedades complementares ao DC – Para as etiquetas às quais não se conseguiu relacionar propriedades presentes no DC, identificaram-se potenciais novas propriedades com elas relacionadas. Em seguida procedeu-se a especificação dessas propriedades com base nas recomendações do *Dublin Core Metadata Initiative* (DCMI);

³ CATARINO, M.E. *Integração das folksonomias aos metadados: identificação de novos elementos de metadados como contributo para a descrição de recursos em repositórios*. 232p. Tese (Doutorado em Tecnologias e Sistemas de Informação). Departamentos de Sistemas de Informação. Escola de Engenharia. Universidade do Minho. Guimarães, Portugal, 2009. Disponível em: <<http://hdl.handle.net/1822/9564>>

- Validação da proposta – A validação foi feita pela comunidade científica, através da apresentação da metodologia e resultados preliminares obtidos, em eventos da área (CATARINO; BAPTISTA, 2008a, 2008b) e pela comunidade DC através de questionários *on-line*;
- Construção de Perfil de Aplicação e criação de uma Ontologia – Declaração das novas propriedades identificadas num perfil de aplicação e criação de uma ontologia com base neste perfil e nas ontologias DC já existentes em RDF (CATARINO, 2009b).

Durante o desenvolvimento dos dois primeiros procedimentos (Criação da Base de Dados e Análise das Etiquetas) foram identificadas algumas características das folksonomias que serão descritas neste artigo.

Considera-se importante divulgar as características observadas durante a realização da pesquisa para futuras investigações e ou desenvolvimento de aplicações práticas.

Antecedendo a descrição das características serão apresentados os conceitos de folksonomia e outros relacionados.

2 FOLKSONOMIA E OUTROS CONCEITOS

O termo Folksonomia é a tradução de *folksonomy*, neologismo criado em 2004 por Thomas Vander Wal, a partir da junção de *folk* (povo, pessoas) com *taxonomy*. Para Wal (2006), Folksonomia é o resultado da atribuição livre e pessoal de etiquetas⁴ (*tagging*) a informações ou objetos (qualquer coisa com URL), visando à sua recuperação. A atribuição de etiquetas é feita num ambiente social (compartilhado e aberto a outros). O ato de etiquetar é do próprio usuário da informação ou objeto.

Optou-se, para fazer referência a *tagging*, utilizar o termo etiquetagem que significa atribuir etiquetas aos recursos da Web. Trata-se de uma indexação livre em linguagem natural, não são adoptadas regras e/ou política de indexação e nem o controle de vocabulários, ou seja, não há efetivamente a tradução dos termos para uma linguagem artificial. Os conteúdos são indexados livremente pelos usuários do recurso, podendo

⁴ Nos textos originais em inglês, *Tags*, que segundo Guy e Tonkin (2006) numa simples definição seriam palavras-chave, categorias ou metadados.

representar assuntos ou quaisquer outros elementos de metadados tais como tipo ou formato (CATARINO; BAPTISTA, 2007).

Outro ponto importante é definir o que são informações ou objetos, que Wal define como “qualquer coisa com um URL”. No âmbito deste artigo, optou-se antes por utilizar o termo Recurso, pois na definição do W3C o termo Recurso é utilizado para se referir a objetos (MILLER, 1998).

Então, em outras palavras, “Folksonomia é o resultado da etiquetagem dos recursos da Web num ambiente social (compartilhado e aberto a outros) pelos próprios usuários visando a sua recuperação” (CATARINO; BAPTISTA, 2007).

Tendo como base as definições acima entende-se que folksonomia é o resultado de um processo. Contudo outros autores dividem-se em dois grupos: 1) os que entendem a folksonomia como o resultado de um processo, ou seja, como um produto; e 2) os que se referem a folksonomia como sendo um sistema, uma metodologia, ou abordagem, ou o próprio processo.

A seguir (quadro 1) são apresentadas as definições de alguns autores, mostrando este interessante viés que ocorre no uso do conceito.

PRODUTO	
Wal (2006)	Folksonomia é o resultado da atribuição livre e pessoal de etiquetas (<i>tagging</i>) a informações ou objetos (qualquer coisa com URL), visando à sua recuperação.
Lund et al. (2005)	Folksonomia se refere a um vocabulário, ou lista de termos, que surge da sobreposição de etiquetas definidas por vários usuários ao marcar as suas hiperligações favoritas, ou seja, seus marcadores para posterior recuperação.
Mathes (2004)	Folksonomia é um conjunto de termos que um grupo de usuários utilizou para etiquetar os conteúdos de recursos digitais da Web.
Trant (2006a, 2006b)	Folksonomia é o resultado de um sistema de classificação socialmente construído, ou, coleção de conceitos expressos num sistema de classificação desenvolvido de forma cooperativa. Folksonomia é um conjunto informal e orgânico de terminologia relacionada.
Sturtz (2006)	Folksonomia é um conjunto de etiquetas – com uma ou mais palavras-chave – que os usuários de um sistema compartilhado de gestão de conteúdos na Web aplicam a recursos individuais a fim de agrupá-los ou classificá-los para posterior recuperação.

PROCESSO	
Russel (2005)	As folksonomias têm propiciado a possibilidade de criar desordenadamente, em texto livre, metadados atribuídos pelos usuários para recursos existentes (livros, imagens, URLs, etc).
Guy e Tonkin (2006)	Folksonomia é um tipo de sistema de classificação distribuída, a folksonomia é normalmente criada por um grupo de indivíduos, tipicamente os usuários do recurso.
Ohmukai, Hamasaki e Takeda (2006)	Trata-se de um sistema que administra etiquetas atribuídas pelos usuários aos recursos por eles indexados, compartilhando-as com outros usuários e também disponibilizando informações de outros recursos disponíveis na Web que foram indexados da mesma forma.
Quintarelli (2005)	Uma nova abordagem emergente para a classificação distribuída de recursos digitais.
Hammond et al. (2005)	Uma classificação não estruturada feita pelos próprios usuários dos recursos digitais.
Valongueiro (2006)	Um novo paradigma de classificação, pois respeita as diferenças culturais e características pessoais de quem utilizou e classificou determinada informação

Quadro 1 - Definições de Folksonomia (produto X processo)

Existem outros termos importantes nesta área que estão relacionados com o conceito de folksonomia. Hammond et al. (2005) consideram os termos “classificação social ” ou “classificação distribuída” mais adequados para representarem o fenômeno desta nova abordagem.

Também Merholz (2004) não considera adequado o uso do termo folksonomia, pois estaria erroneamente relacionado com as taxonomias. Prefere o termo “etnoclassificação”, isto é, classificação popular (*apud* MATHES, 2004).

Há outros autores que preferem o termo *social bookmarking*, dando ênfase, desta forma, ao aspecto colaborativo destas ferramentas com a palavra “social”, conforme afirma Elaine Peterson (2006) em artigo publicado na *D-Lib Magazine*.

Contexto	Termos	Definições
Etiquetagem	<i>Tagging</i>	Tipo de ferramentas dá poder sem precedentes para os usuários que podem moldar as informações com as quais eles interagem (WINGET, 2006).
	<i>Tagging Systems</i>	Sistemas que habilitam usuários para acrescentar palavras-chave nos recursos digitais da Web sem o uso de vocabulários controlados (MARLOW et al., 2006)

	<i>Social Tagging</i>	Refere-se à prática de publicamente etiquetar ou categorizar recursos num ambiente compartilhado (TRANT, 2006b); ou um tipo de indexação aberta que se manifesta na Web (TENNIS, 2006).
	<i>Social Tagging Systems</i>	Permitem que os usuários compartilhem suas etiquetas de recursos particulares, além de que cada etiqueta serve como uma hiperligação para recursos adicionais que foram indexados por outros (MARLOW et al., 2006).
	<i>Collaborative Tagging Systems</i>	São sistemas colaborativos de etiquetagem que permitem, aos usuários, indexar as suas hiperligações, fotografias, referências e outros recursos digitais com palavras-chave ou etiquetas (VOSS, 2006); ou então processo pelo qual os usuários adicionam metadados em forma de palavras-chave ou etiquetas para compartilhar conteúdos (GOLDER; HUBERMAN, 2006a).
Classificação	<i>Social Classification</i>	Sinônimo de folksonomia que, para o autor, são metadados criados pelos próprios usuários da informação (SPITERI, 2006). Uma nova abordagem que está desafiando os esquemas tradicionais de classificação e de indexação baseados em vocabulários controlados (LIN et al., 2006). Processo pelo qual uma comunidade de usuários categoriza seus recursos naquela comunidade para o seu próprio uso (BOGERS; THOONE; BOSCH, 2006)
Bookmarking	<i>Bookmarking</i>	Um dos métodos mais populares para armazenar informação relevante da Web para acessá-la novamente e reutilizá-la (SPITERI, 2006).
	<i>Social Bookmarking</i>	Ferramentas que possibilitam que os usuários marquem suas páginas e atribuam etiquetas para representar seus temas de interesse (CAMPBELL, 2006).
	<i>Social Bookmarking Manager</i>	Denominação dada ao Delicious pelo seu criador. Golder e Huberman (2006b) definem este serviço como um sistema colaborativo para indexar os <i>bookmarks</i> da Web.
Ontologia	<i>Social Ontologies</i>	Mote (2006) considera que o termo folksonomia representa <i>social ontologies</i> , ou seja, ontologias construídas de forma colaborativa, e significa uma classificação consensual gerada pelos usuários dos recursos digitais.
Taxonomia	Taxonomia Dinâmica	Joseph et al. (2006) afirmam que folksonomia é uma taxonomia dinâmica que representa as categorias que usuários individuais empregam para organizar seus espaços de informação

Quadro 2 - Termos relativos a indexação de recursos da Web.

Todos os termos analisados referem-se a etiquetagem de recursos da Web. No entanto, dão ênfase a aspectos diferentes. Um grupo de termos reporta-se diretamente à ação propriamente dita de atribuir etiquetas aos recursos da Web: “Etiquetagem” e “Classificação”. Outro grupo de termos relaciona-se diretamente aos Marcadores: *Bookmarking*. Há dois outros termos pouco utilizados que são os de Ontologias Sociais e Taxonomia Dinâmica (ver quadro 2).

Há diversos serviços que dispõem de folksonomias e que permitem a etiquetagem dos recursos da Web (quadro 3).

Sites	Recursos	URL
CiteULike	Hiperligações de documentos acadêmicos: artigos, <i>papers</i> , teses etc	http://www.citeulike.org
Clipmarks	Clips / notícias	http://clipmarks.com
Connotea	Referências / informações bibliográficas	http://www.connotea.org
Delicious	Coleção de hiperligações favoritas	http://delicious.com
Flicker	Fotos	http://www.flickr.com
Furl	Coleção de hiperligações favoritas	http://www.furl.net
Last.fm	Música	http://www.last.fm
LiveJournal	Weblogs	http://www.livejournal.com
Odeo	Música e vídeo	http://www.odeo.com
Simpy	Websites e blogs	http://www.simpy.com
Social Marker	Sites Social bookmarking	http://socialmarker.com
Spurl.net	Coleção de hiperligações favoritas	http://www.spurl.net
Technorati	Weblog	http://www.technorati.com
Yahoo's My web 2.0	Hiperligações favoritas / <i>bookmarks</i>	http://myweb2.search.yahoo.com
YouTube	Vídeos	http://www.youtube.com

Quadro 3 – Sites que adotam a Folksonomia.

No entanto, é importante destacar que estes não são os únicos recursos. O *Social Marker* (ver <http://socialmarker.com/>), uma ferramenta que permite a inclusão de um URI e sua etiquetagem em vários *sites* de *social bookmarking* ao mesmo tempo, tinha relacionado 160 serviços até Dezembro de 2008.

3 VANTAGENS E DESVANTAGENS NO USO DAS FOLKSONOMIAS

Com base na literatura, destacam-se algumas características que são apontadas como vantagens e desvantagens no uso de Folksonomias.

A vantagem que tem maior destaque é o cunho colaborativo/social das folksonomias. Os usuários compartilham com outros as suas etiquetas, que podem ser ou não adotadas na classificação de um mesmo recurso por outros.

Outra vantagem é a possibilidade de formar, automaticamente, comunidades em torno de assuntos de interesse. Ao utilizar serviços de folksonomia, o usuário tem acesso aos outros usuários que têm os mesmos interesses identificados através das etiquetas.

Uma outra vantagem apontada é que não há uma regra preestabelecida de controle dos vocabulários. Os usuários dos recursos expressam, ao etiquetar estes

conteúdos, a sua percepção em relação àquela informação. Há uma liberdade de expressão que possibilita abarcar todas as formas de ver um mesmo conteúdo, respeitando as diferenças culturais, interpretativas, etc. Sabe-se que a leitura (textual, imagética, etc) é diferente de indivíduo para indivíduo, pois depende de vários fatores, dentre eles os antecedentes intelectual e cultural de quem lê. E no caso das folksonomias, estas diferenças são respeitadas já que não há regras para expressão das etiquetas ao etiquetar um determinado conteúdo.

Como desvantagem, parece haver um consenso de que o maior problema é justamente a falta de um controle do vocabulário, que é resultado da característica de liberdade na classificação dos conteúdos. Então, a característica de ausência de controle de vocabulários apresenta vantagens e desvantagens de acordo com o ponto de vista.

A liberdade de atribuição de etiquetas faz com que haja pouca precisão na recuperação da informação, pois um mesmo termo pode ter significados diversos para os vários usuários que atribuíram as etiquetas. Para Feinberg (2006) termos comuns como “*java*” e “*design*” são atribuídos para centenas de milhares de recursos discrepantes, tornando quase impossível uma consulta produtiva sem refinamento adicional.

Guy e Tonkin (2006) destacam que a maior falha das folksonomias está no fato de os termos utilizados para etiquetar os conteúdos serem imprecisos. São os usuários que atribuem as palavras-chave e, portanto, são frequentemente ambíguas, muito personalizadas e inexatas. Por enquanto, há pouco ou nenhum controle de sinônimos ou homônimos, também não são impostas regras de indexação: são utilizados termos no singular ou plural, simples ou compostos, palavras sem sentido que não têm significado, exceto para um grupo específico de usuários. Tudo isso pode resultar num conjunto confuso de termos que poderá interferir no resultado da recuperação da informação. A falta de controle de vocabulário, ou seja, a não utilização de instrumentos de terminologia tais como listas de cabeçalhos de assunto ou tesouros, e de regras gerais para a aplicação das palavras-chave, causam problemas que poderão afetar a recuperação da informação, ou não - é necessária mais investigação sobre este assunto.

É importante destacar que não é objetivo deste trabalho concluir sobre as questões relativas às vantagens e desvantagens das folksonomias. Aqui pretende-se, apenas, dar a conhecer as diferentes perspectivas. Assume-se aqui que a mais significativa vantagem das folksonomias está justamente no fato de os usuários de recursos poderem expressar

livremente, através de etiquetas para organização de seus recursos, o seu entendimento em relação à representação descritiva, seja física, temática ou quaisquer outras notações.

4 ANÁLISE DAS ETIQUETAS: PROCEDIMENTOS

Definiu-se por desenvolver uma pesquisa com base em dados reais, ou seja, analisar etiquetas de recursos que já se encontravam etiquetados pelos seus usuários.

Neste ínterim havia um projeto desenvolvido pelo Departamento de Sistemas de Informação da Universidade do Minho (Portugal) em conjunto com outras instituições europeias, americana e australiana, o *Kinds of Tags* (KoT).

Optou-se, então, pela utilização do conjunto de dados do projeto KoT que já apresentava alguns resultados, embora preliminares. O conjunto de dados foi formado pelos recursos mais utilizados no Connotea⁵, na ocasião em que foram recolhidos, e que também estivessem etiquetados no Delicious⁶. Eram em sua maioria recursos do tipo texto, mais precisamente artigos. O tema destes recursos era relacionado basicamente à área de Ciência da Informação.

No cômputo geral, os dados representavam 50 recursos, etiquetados por 15.381 usuários, com 5.098 etiquetas atribuídas. Considerando que uma etiqueta podia ser atribuída a vários recursos e por vários usuários, o total de ocorrências das etiquetas foi de 79.146.

Para a análise das etiquetas foi necessário o uso de recursos léxicos que auxiliaram na identificação dos significados e tradução dos termos. Em alguns casos também houve a necessidade de fazer pesquisas nos mecanismos de busca da Web para identificar o significado da etiqueta. Os recursos léxicos mais utilizados foram: WordNet, Infopedia e Webster.

WordNet é uma base de dados lexical de inglês. É composta por substantivos, verbos, adjetivos e advérbios agrupados em conjuntos de sinônimos cognitivos (*synsets*) que individualmente expressam conceitos distintos. Foi um instrumento de suma importância pois a visualização dos diversos conjuntos de sinônimos permitiu que a

⁵ <http://www.connotea.org>

⁶ <http://delicious.com>

análise de etiquetas de significados dúbios fosse feita com respaldo léxico (WORDNET, 2008).

A Infopedia é uma base de conteúdos de referência disponibilizada num *site* que abrange todas as áreas do conhecimento e inclui um amplo conjunto de matérias de caráter enciclopédico, linguístico e gráfico (dicionários, enciclopédias, atlas e recursos gráficos) (INFOPEDIA, 2008).

O Webster é um dicionário *on-line* multilíngue, composto por aproximadamente 30 milhões de entradas oriundas de 400 línguas modernas e dez ancestrais. Este recurso léxico permitiu uma tradução mais eficiente das etiquetas que estavam escritas em idiomas que não o inglês (WEBSTER's, 2008).

Em algumas situações nenhum dos recursos léxicos citados nem mesmo mecanismos de busca da Web foram eficazes para a tradução e/ou identificação do significado das etiquetas. Para estes casos, quando havia a indicação de algum tipo de contato do usuário (*e-mail*, *site* ou *blog*), este foi utilizado para dirimir dúvidas quanto ao significado da etiqueta. Foram enviados 89 *e-mails*, dos quais 49 foram respondidos. As informações recolhidas foram bastante elucidativas. No entanto, mesmo esgotando todas as possibilidades ainda restaram etiquetas cujos significados não foi possível identificar.

Para a análise, as etiquetas foram agrupadas em suas formas variantes (singular/plural, maiúsculas/minúsculas, idiomas, grafia, siglas e abreviaturas). Este procedimento foi realizado para facilitar a compreensão das mesmas.

A cada grupo de etiquetas deu-se o nome de *Key-tag*. As *Key-tags* podem ser definidas como etiquetas chave que reúnem as várias formas de uma mesma etiqueta.

No quadro 4 apresentam-se alguns exemplos de *Key-tags*. A coluna *key-tag* mostra o termo escolhido para representar o agrupamento das formas variantes de uma mesma etiqueta representados na coluna Etiqueta.

ey-tag	Etiqueta
Article	_article article articles artikel article:sw

Cataloguing	cataloging cataloguing classification-cataloguing katalogisering katalogisierung kataloogimine z-libcataloging
Digital Libraries	biblioteca digital biblioteques digitals digital libraries digital library digital_libraries digital_library digitallibraries digital-libraries digitallibrary digital lybraries dl

Quadro 4 – Exemplos de *key-tags*

Pode-se perceber Variações na Escrita (_article e article; cataloging e cataloguing; digital library, digital_library, digitallibrary e dl); Variações de Forma Singular/Plural (article e articles; digital library e digital libraries); Variações de Idiomas (article (inglês) e artikel (alemão); Cataloguing (inglês), Katalogisierung (alemão) e Kataloogimine (estoniano); Biblioteca digital (português), biblioteques digitals (catalão) e Digital Library (inglês)).

Outra forma variante observada durante o agrupamento das etiquetas em *Key-Tags* foi a forma simples ou composta das etiquetas: Etiqueta simples agrupada a uma *key-tag* simples (etiqueta Library = *key-tag* Library); Etiqueta simples agrupada a uma *key-tag* composta (etiqueta finalpaper = *key-tag* Final Paper); Duas ou mais etiquetas simples agrupadas a uma *key-tag* composta (etiqueta Emanuele + etiqueta Quintarelli = *key-tag* Emanuele Quintarelli); Uma etiqueta composta agrupada a uma *key-tag* composta (etiqueta digital repositories = *key-tag* Digital Repositories); Uma etiqueta composta desmembrada em duas ou mais *key-tags* simples (etiqueta classification:folksonomy = *key-tag* Classification e *key-tag* Folksonomy).

A definição pela composição de *Key-tags* simples ou compostas dependia exclusivamente da análise das etiquetas em relação ao recurso ao qual foi atribuída.

Durante o processo de análise e agrupamento pode-se observar as características das etiquetas que serão descritas a seguir.

5 CARACTERÍSTICAS DAS FOLKSONOMIAS

As etiquetas são grafadas indistintamente em diferentes alfabetos e idiomas, na sua forma simples ou composta, singular ou plural, além de divergentes grafias.

O conjunto de etiquetas analisadas é representativo do que ocorre nos serviços de *Social Bookmarking* e portanto havia etiquetas em diversos alfabetos (latino, grego/Ελληνική, cirílico/Кирилица, chinês/中國, japonês/日本語, etc). No entanto, observa-se que a maioria delas está no alfabeto latino.

Relativamente ao idioma havia etiquetas em várias línguas e através dos dicionários utilizados foi possível traduzir 425 etiquetas identificadas em outros idiomas que não o inglês, correspondendo este valor a 8,3% das 5098 etiquetas analisadas. Contudo, é importante destacar que nem todas as etiquetas puderam ser traduzidas porque os seus significados não foram identificados.

Foram traduzidas etiquetas de vários outros idiomas, dentre os quais, os mais constantes: Português (PT), Francês (FR), Alemão (DE), Espanhol (ES) e Catalão (CA). Para a identificação dos idiomas foi utilizado o conjunto das Normas ISO 639-1 e 2⁷.

É importante observar que em alguns casos as etiquetas foram identificadas como *Multiple Languages* (MUL) pois na tradução as mesmas correspondiam a vários idiomas, como por exemplo a etiqueta Artikel palavra que no dicionário Webster está relacionada nos idiomas: DA, DE, NL, SV, traduzido para o inglês (EN) *Article*. O quadro 5 mostra a quantificação dos vários idiomas identificados.

Sigla ISO 639	Idioma	Nº etiquetas	Sigla ISO 639	Idioma	Nº etiquetas
CA	Catalan	43	HU	Hungarian	9
CS	Czech	3	IT	Italian	16
DA	Danish	3	MUL	Multiple Languages	57
DE	German	51	NL	Dutch	16
ES	Spanish	47	NO	Norwegian	9
ET	Estonian	2	PL	Polish	2
EU	Basque	1	PT	Portuguese	77
FI	Finnish	9	RO	Romanian	4
FR	French	68	SV	Swedish	8
HR	Croatian	1	TR	Turkish	1

Quadro 5 - Idiomas identificados.

⁷ Disponíveis em http://en.wikipedia.org/wiki/List_of_ISO_639-1codes e http://en.wikipedia.org/wiki/List_of_ISO_639-2codes

Além do alfabeto e do idioma, outras características foram percebidas. Havia etiquetas compostas por sinais e símbolos, números, siglas, abreviaturas, formas mnemônicas e mistas, assim como etiquetas com erros de grafia.

Foram identificadas 17 etiquetas compostas apenas por sinais gráficos, sinais de pontuação ou símbolos (asterisco, traço, ponto, vírgula, arroba, entre outros) que não compunham palavras (quadro 6).

-	--	'	!!!	????+
+	++	+++++	*	**
***	****	*****	,	...
???????	@@			

Quadro 6 – Etiquetas compostas apenas por sinais e símbolos.

Este tipo de etiqueta apresenta um alto grau de subjetividade. Por exemplo, em relação à etiqueta *********, pode-se “ler” ou interpretar como sendo “cinco estrelas”. No entanto, qual o significado de “cinco estrelas” para o usuário? Pode-se inferir que sirva para classificar o recurso quanto à sua qualidade. Mas esta inferência é difícil de ser comprovada.

Relativamente às etiquetas numéricas, havia um total de 65 etiquetas compostas somente por números (ex: 2005, 11072006, 4163). Os números dificultam a identificação dos significados, exceto no caso em que estivessem relacionados à data, seja de publicação ou de etiquetagem, ou que estivessem explicitamente relacionados ao recurso (título, assunto, editor, etc).

Algumas etiquetas eram compostas por siglas: **bmj**, **bmo**, **bmjtech**, **dl**, **cmc**, **cnpq**, abreviaturas: **biblio**, **ref**, **tech** e outras apareciam em forma mnemônicas: **2bread**, **2do**, **2print**, **4work**. Siglas e formas mnemônicas também tornam a interpretação difícil. Nem sempre era possível perceber o significado destas etiquetas sem que houvesse necessidade de pesquisas complementares aos mecanismos de buscas da Web.

Foram identificadas etiquetas compostas por mais do que um termo porém os mesmos foram grafados unidos, como por exemplo: **boeingreadinglist** ou **collaborativefiltering**. Geralmente, nestes casos, o significado era de fácil interpretação.

As etiquetas mistas, ou seja, etiquetas que eram compostas por palavras mais sinais gráficos, de pontuação, símbolos ou números, foram analisadas como palavras. Nestes casos buscou-se a identificação e/ou tradução dos termos. Como exemplos de etiquetas mistas: !favs, !school, *essay, @article, [D-Lib], 005-lagoze, 2.0library, 2006august, etc.

Havia ainda um conjunto de etiquetas com erros de grafia, tais como: buisness, clasiffication, defintion e folksnomy.

Excetuando-se a questão dos idiomas, as demais formas de grafia das etiquetas representam características que dificultam a identificação dos seus significados. Porém, geralmente, foi possível interpretar o significado destas etiquetas através da análise do conjunto de etiquetas do usuário em relação ao recurso sob análise, ou mesmo, à análise do conjunto de etiquetas daquele usuário em toda a sua coleção de *bookmarks*.

Ainda pode-se verificar que havia etiquetas simples, ou seja, formadas por um único termo e as compostas, formadas por dois ou mais termos. No entanto nem sempre foi possível manter esta formação para proceder a análise destas etiquetas e consequentemente a criação das *key-tags*.

Na ocorrência de etiquetas simples, ou seja, aquelas com um único termo, foi observada uma peculiaridade: a forma de inserção das etiquetas nos dois serviços de *social bookmarking*.

No Delicious, por exemplo, quando o usuário insere as etiquetas, o único separador é o espaço. Tudo o que for digitado separado por espaços será considerado como etiquetas distintas. Se for inserido o termo composto Digital Library contendo apenas o espaço como separador, o sistema considerará duas etiquetas: Digital e Library. Para que possa ser inserido no sistema como uma etiqueta composta é necessário utilizar recursos tais como *underline*, traço, dois pontos, etc, como por exemplo Digital_Library, Digital-library, Digital:Library, Digital.Library. Tonkin (2006) verificou que no Delicious os seguintes separadores são os mais utilizados: Traço (39%), *Underline* (25%), Barra (14%), Ponto (14%) e outros (8%).

Num outro exemplo, o Connotea, as etiquetas também são separadas por espaço ou por vírgula, no entanto, há sugestão de que as etiquetas compostas sejam digitadas entre aspas. Por exemplo, se o utilizador inserir Information Science sem colocar as

palavras entre aspas, serão consideradas duas etiquetas simples, mas se forem digitadas entre aspas “Information Science” o sistema gerará uma única etiqueta composta.

Portanto, a forma de inserção das etiquetas interfere na indexação das etiquetas pelo sistema, separando palavras que em alguns casos, pressupõe-se, os usuários gostariam que fossem processadas como etiquetas compostas.

Um exemplo é de um utilizador do Delicious que ao atribuir etiquetas ao recurso “*The Semantic Web*” de autoria de Tim Berners-Lee inseriu as seguintes etiquetas: the, semantic, web, article, by, tim, berners-lee, sem usar os recursos de união das palavras (_ ; - etc). O sistema gerou, neste caso, sete etiquetas simples, no entanto, fica claro que estas etiquetas podiam ser pós-coordenadas como etiquetas compostas, tais como Semantic Web e Tim Berners-Lee.

Este exemplo demonstra que as etiquetas simples podem, dependendo da análise realizada, ser pós-coordenadas com outras etiquetas simples do mesmo utilizador. Contudo, cabe esclarecer que a pós-coordenação dos termos simples só pode ser feita em relação a um recurso em particular e jamais de uma forma global, o que poderia descaracterizar a análise.

Para exemplificar esta situação tomemos o caso do recurso intitulado *As We May Think*. Foram atribuídas as seguintes etiquetas simples: As, may, think, we. Para o recurso em questão fica claro que estas etiquetas representam, se pós-coordenadas, o título do recurso. No entanto, para outros recursos a relação destas etiquetas com o Título não seria verdadeira. Como por exemplo a etiqueta May que para outros recursos poderia também estar relacionada a Data.

6 CONSIDERAÇÕES FINAIS

Nas folksonomias, as etiquetas são atribuídas de forma livre, sem o controle de vocabulário. Esta liberdade pode ser considerada tanto uma vantagem quanto uma desvantagem.

A vantagem da folksonomia está na liberdade de expressão que possibilita a descrição dos recursos da Web conforme a visão dos seus próprios usuários, o que agrega valor à descrição e, pressupõe-se, amplia as possibilidades de recuperação. As etiquetas dos usuários permitem tanto a descrição física ou temática do recurso quanto

outros aspectos relativos às funcionalidades e ou relações deste recurso para o seu utilizador. Na pesquisa que foi a base deste artigo, foram identificadas propriedades (ou elementos de metadados) que demonstram as possibilidades de descrição que são inovadoras no âmbito das folksonomias: ação do usuário em relação ao recurso (*Action*); categorização do recurso que vai além da “classificação” temática (*Category*); a data em que o recurso foi etiquetado (*Date Tagged*); avaliação do recurso sobre o ponto de vista do usuário quanto ao nível intelectual ou à qualidade (*Depth e Rate*); apontamentos que registram observações, comentários ou explicações próprias de quem utilizou o recurso (*Note*); compartilhamento do recurso numa rede social (*Share*) e a finalidade de uso do recurso para o usuário (*Utility*) (CATARINO, 2009a).

E a desvantagem está também focada na falta de controle de vocabulário. A liberdade de atribuição de etiquetas faz com que haja pouca precisão na recuperação da informação, pois um mesmo termo pode ter significados diversos para os vários utilizadores que as atribuíram.

As folksonomias, com suas vantagens e desvantagens, estão definitivamente inseridas na Web. Atualmente é bastante comum encontrar sites que permitem a atribuição de etiquetas aos recursos. Compreender as características destas etiquetas poderá maximizar as possibilidades de uso das folksonomias em várias aplicações.

Neste estudo pode-se averiguar que não existem apenas etiquetas em alfabeto latino. Encontram-se etiquetas em japonês, chinês, grego, etc. o que justifica desenvolver formas de transliteração automática entre os vários alfabetos.

Verificou-se que as etiquetas em inglês predominam. Porém existem etiquetas em vários outros idiomas que devem ser consideradas.

Pode-se observar que a forma de inserção das etiquetas nos diferentes serviços de *social bookmarking* define se as mesmas serão processadas na forma simples ou composta. Na análise das etiquetas fica claro que alguns usuários não percebem essas variações. Isto faz com que, muitas vezes, as etiquetas que deveriam ser compostas apareçam separadas no índice. Esta característica pode afetar a recuperação.

As outras características observadas estão relacionadas à forma de escrita. Existem etiquetas compostas por sinais e símbolos (ex.: *****), números (ex.:2005), siglas

(ex.:Cnpq), abreviaturas (ex.:biblio), formas mnemónicas (ex.:4work) e mistas (ex.: 2.0library).

A forma de escrita das etiquetas, muitas vezes, é muito subjetiva, e faz sentido apenas para o usuário que a atribuiu. Esta subjetividade é um fator que dificulta o desenvolvimento de aplicações que pretendam aproveitar a descrição do usuário incorporando-a aos metadados descritivos de repositórios digitais.

Para concluir, cabe ressaltar que o estudo foi realizado num conjunto de recursos específico: textos, etiquetados no Connotea e Delicious e na área de ciência da informação. Neste sentido outros estudos devem ser realizados para ampliar a análise para outros tipos de recursos, áreas do conhecimento e serviços que implementam folksonomia.

Espera-se que o conhecimento das características das etiquetas das folksonomias possa colaborar para que outros estudos e desenvolvimento de aplicações tenham êxito.

REFERÊNCIAS

BOGERS, T.; THOONE, W.; BOSCH, A. Expertise classification: collaborative classification vs. automatic extraction. In: SIG/CR CLASSIFICATION RESEARCH WORKSHOP, 17., 2006, Austin (USA). *Papers...* Austin, 2006. Disponível em: <<http://www.slais.ubc.ca/users/sigcr/sigcr-06bogers.pdf>>. Acesso em: 06 nov., 2006.

CAMPBELL, D. G. A phenomenological framework for the relationship between the semantic web and user-centered tagging systems. In: SIG/CR CLASSIFICATION RESEARCH WORKSHOP, 17, 2006, Austin (USA). *Papers...* Austin, 2006. Disponível em: <<http://www.slais.ubc.ca/users/sigcr/sigcr-06campbell.pdf>>. Acesso em: 06 nov. 2006.

CATARINO, M. E. *Integração das folksonomias aos metadados: identificação de novos elementos de metadados como contributo para a descrição de recursos em repositórios*. 232p. Tese (Doutorado em Tecnologias e Sistemas de Informação). Departamento de Sistemas de Informação. Escola de Engenharia. Universidade do Minho. Guimarães, Portugal, 2009a.

_____. *Social Tagging Application Profile (STAP)*. Disponível em: <<http://odisseia.dsi.uminho.pt/STAP>>. Acesso em: 14 abr. 2009b.

CATARINO, M. E. ; BAPTISTA, A. A. Folksonomia: um novo conceito para a organização dos recursos digitais na Web. *DataGramaZero – Revista de Ciência da Informação*, Rio de Janeiro, v.8, n.3, jun. 2007. Disponível em: <http://www.dgz.org.br/jun07/Art_04.htm>. Acesso em: 19 jun. 2007.

_____; _____. Relating Folksonomy with Dublin Core. In: DC-2008 INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, Berlin, 2008a. *Proceedings...* Berlin, 2008a. Disponível em: <<http://repositorium.sdum.uminho.pt/handle/1822/8167>>. Acesso em: 5 nov. 2008.

_____; _____. Social Tagging and Dublin Core: A Preliminary Proposal for an Application Profile for DC Social Tagging. In: INTERNATIONAL CONFERENCE ON ELECTRONIC PUBLISHING, 12., Toronto, 2008b. *Proceedings...* Toronto, 2008. Disponível em: <http://elpub.scix.net/data/works/att/100_elpub2008.content.pdf>. Acesso em: 5 nov. 2008.

FEINBERG, M. An examination of authority in social classification systems. In: SIG/CR CLASSIFICATION RESEARCH WORKSHOP, 17, 2006, Austin (USA). *Papers...* Austin, 2006. Disponível em: <http://www.slais.ubc.ca/users/sigcr/sigcr-06_feinberg.pdf>. Acesso em: 06 nov. 2006.

GOLDER, S. A.; HUBERMAN, B. A. *The structure of collaborative tagging systems*. Disponível em: <<http://arxiv.org/abs/cs.DL/0508082>>. Acesso em: 14 nov. 2006a.

_____; _____. Usage patterns of collaborative tagging systems. *Journal of Information Science*, v.32, n.2, p.198-208, 2006b.

GUY, M.; TONKIN, E. Folksonomies: tidying up tags? *D-Lib Magazine*, v.12, n.1, 2006. Disponível em: <<http://www.dlib.org/dlib/january06/guy/01guy.html>>. Acesso em: 12 dez. 2006.

HAMMOND, T. et al. Social Bookmarking Tools (I): a general review. *D-Lib Magazine*, v.11, n.4, 2005. Disponível em: <<http://www.dlib.org/dlib/april05/hammond/04hammond.html>>. Acesso em: 14 nov. 2006.

INFOPEdia ENCICLOPÉDIAS E DICIONÁRIOS. Porto: Porto ed., 2008. Disponível em: <<http://www.infopedia.pt>>. Acesso em: 07 fev. 2008.

JOSEPH, S. et al. Searching emergent vocabularies: exploring methods to reduce cognitive load during web navigation and resource contribution. In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 39., 2006, Hawaii. *Proceedings...* 2006. Disponível em: <<http://ieeexplore.ieee.org/iels/10548/33367/01579603.pdf>>. Acesso em: 14 nov. 2006.

LIN, X. et al. Exploring characteristics of social classification. *Papers of SIG/CR Classification Research Workshop*, 17., Austin (USA), 2006. Disponível em: <<http://dlist.sir.arizona.edu/1790/lin.pdf>>. Acesso em: 06 nov. 2006.

LUND, B. et al. Social Bookmarking Tools (II): a case study: Connotea. *D-Lib Magazine*, v.11, n.4. Disponível em: <<http://www.dlib.org/dlib/april05/lund/04lund.html>>. Acesso em: 14 de nov. 2006.

MARKOFF, J. Entrepreneurs see a Web guided by common sense. *New York Times, NY Business*, Nov., 12, 2006. Disponível em: <http://www.nytimes.com/2006/11/12/business/12web.html?pagewanted=1&_r=1>. Acesso em: 18 dez. 2008.

MARLOW, et al. Position paper, tagging, taxonomy, Flickr, Article, ToRead. *Papers of WWW2006 International World Wide Web Conference*, 15., Edinburgo Scotland, 2006. Disponível em: <<http://www.danah.org/www2006.pdf>>. Acesso em: 14 nov. 2006.

MATHES, A. *Folksonomies - cooperative classification and communication through shared metadata*. Computer Mediated Communication – LIS590CMC, Urbana : University of Illinois, 2004. Disponível em: <<http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>>. Acesso em: 25 out. 2006.

MÉNDEZ RODRÍGUES, E.; BRAVO, A.; LOPEZ, L.M. Microformatos: Web 2.0 para el Dublin Core. *El profesional de la información*, v. 16, n. 2, 2007, p.107-113.

MILLER, E. An Introduction to the Resource Description Framework. *D-Lib Magazine*, v.4, n.5, 1998. Disponível em: <<http://www.dlib.org/dlib/may98/miller/05miller.html>>. Acesso em: 23 jan. 2007.

MOTE, N. *The new school of ontologies*. Disponível em: <<http://www.isi.edu/~mote/papers/Folksonomy.pdf>>. Acesso em: 26 out. 2006.

O'REILLY, T. *What Is Web 2.0?: design patterns and business models for the next generation of software*. Disponível em: <<http://oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>>. Acesso em: 06 nov., 2006.

OHMUKAI, I.; HAMASAKI, M.; TAKEDA, H. A Proposal of Community-based Folksonomy with RDF Metadata. *ISWC*, 5., 2006. Disponível em: <<http://www-kasm.nii.ac.jp/papers/takeda/05/ohmukai05iswc.pdf>>. Acesso em: 26 out. 2006.

PETERSON, E. Beneath the Metadata: some philosophical problems with Folksonomy. *D-Lib Magazine*, v.12, n.11, 2006. Disponível em: <<http://www.dlib.org/dlib/november06/peterson/11peterson.html>>. Acesso em: 12 dez. 2006.

QUINTARELLI, E. Folksonomies: power to the people. *Papers of Incontro ISKO Italia - UNIMIB*, Milão, 2005. Disponível em: <<http://www.iskoi.org/doc/folksonomies.htm>>. Acesso em: 23 out. 2006.

RUSSELL, T. *Contextual authority tagging: cognitive authority through folksonomy*. Disponível em: <<http://www.terrellrussell.com/projects/contextualauthoritytagging/conanthtag200505.pdf>>. Acesso em: 26 out. 2006.

SPITERI, L. *Controlled Vocabulary and Folksonomies*. Disponível em: <<http://www.collectionscanada.ca/obj/014005/f2/014005-05-209-e-e.pdf>>. Acesso em: 16 out. 2006.

STURTZ, D. N. Communal categorization: the folksonomy. Disponível em: <<http://www.davidsturtz.com/drexel/622/communal-categorization-the-folksonomy.html>>. Acesso em: 07 dez. 2006.

TENNIS, J. Social tagging and the next steps for indexing. *Papers of SIG/CR Classification Research Workshop*, 17., Austin (USA), 2006. Disponível em: <<http://www.slais.ubc.ca/users/sigcr/sigcr-06tennis.pdf>>. Acesso em: 06 nov. 2006.

TONKIN, E. Searching the Long Tail: hidden structure in social tagging. *Papers of SIG/CR Classification Research Workshop*, 4., Austin, Texas, 2006. Disponível em: <<http://hdl.handle.net/1822/679>>. Acesso em: 10 abr. 2006.

TRANT, J. Exploring the potential for social tagging and folksonomy in art museums: proff of concept. *New Review of Hypermedia and Multimedia*, v.12, n.1, p. 63-81, 2006a.

_____. Social Classification and folksonomy in art museums: early data from the steve.museum tagger prototype. *Papers of SIG/CR Classification Research Workshop*, 17., Austin (USA), 2006. Disponível em: <<http://www.slais.ubc.ca/users/sigcr/sigcr-06trant.pdf>>. Acesso em: 06 nov. 2006b.

VALONGUEIRO, A. *Sobre folksonomia, tags e afins*. Disponível em: <<http://valongueiro.blogspot.com/2006-10-01-archive.html>>. Acesso em: 02 nov. 2006.

VOSS, J. Collaborative thesaurus tagging the Wikipedia way. *Wikimetrics*, v.1, n.1. Disponível em: <<http://arxiv.org/trackback/cs/0604036>>. Acesso em: 15 nov. 2006.

WAL, T. V. *Folksonomy definition and wikipedia*. Disponível em: <<http://www.vanderwal.net/random/entrysel.php?blog=1750>>. Acesso em: 22 nov. 2006.

WEBSTER'S ONLINE DICTIONARY: with multilingual Thesaurus Translation. Disponível em: <<http://www.websters-online-dictionary.com>>. Acesso em: 05 jan. 2008.

WINGET, M. User-defined classification on the onlis photo sharing site flickr ... or How I Leraned to stop worrying and love the million typing monkeys. *Papers of SIG/CR Classification Research Workshop*, 17., Austin (USA), 2006. Disponível em: <<http://www.slais.ubc.ca/users/sigcr/sigcr-06winget.pdf>>. Acesso em: 06 nov. 2006.

WORDNET. *A lexical database for the english language*. Princeton University, Cognitive Science Laboratory. Disponível em: <<http://wordnet.princeton.edu/>>. Acesso em: 07 de fev. 2008.

Title

Folksonomies: characteristics of tags for Web resources' description

Abstract

The Folksonomy emerges in the context of Web 2.0 which comes as a result of Web resources' tagging, in a social environment made by users who aim to recovery these tags. This free index in natural language does not adopt rules or indexing policies or even the vocabulary control. Due to the fact that knowing the possibilities of this new form of description of Web resources is essential,

this paper aims to describe the characteristics of the tags that compose the folksonomies. These characteristics were observed in a doctoral research project in order to identify metadata elements originating from the folksonomies that were complementary to Dublin Core (DC) model. For the identification of new properties the following methodological procedures were adopted: creation of the database; dataset tag analysis; identification of Dublin Core (DC) complementary properties; validation of the proposal and application profile and ontology development. During the tags' analysis it was possible to observe several features, such as: alphabet, language and variant forms such as singular/plural, simple/compound, symbols, numbers, acronyms, abbreviations, mnemonics and mixed. It ponders that the knowledge and dissemination of those characteristics will give basis to new studies and applications, increasing the use of folksonomies.

Key Words

Folksonomies. Tags' Characteristics. Resource Description.

Título

Folksonomías: Características de las etiquetas em la descripción de recursos Web

Resumen

En el contexto de la Web 2.0 surge la folksonomía que es el resultado de la marcación de los recursos de la Web, en un ambiente social, por los propios usuarios visando su recuperación. Se trata de una indexación libre en lenguaje natural donde no son adoptadas reglas y/o política de indexación y ni el control de vocabularios. Conocer las posibilidades de esta nueva forma de descripción de los recursos de la Web es imprescindible. Este artículo tiene el objetivo de describir las características de las marcaciones que componen las folksonomías. Esas características fueron observadas en un proyecto de investigación de doctorado que tuvo el objetivo de identificar elementos de metadatos oriundos de las folksonomías que fueran complementarios a Dublin Core (DC). Para la identificación de esos metadatos se desarrolló los siguientes procedimientos metodológicos: creación de la base de datos, análisis de las marcaciones, identificación de propiedades complementarias al DC, validación de la propuesta y construcción del perfil de aplicación y ontología. En el análisis de las marcaciones puede observarse varias características que se describen a continuación: alfabeto, idioma y formas variantes, tales como singular/plural, simple/compuesta, símbolos, números, siglas, abreviaturas, mnemónicos y mixto. Se pondera que el conocimiento y la diseminación de esas características va a subvencionar los nuevos estudios y aplicaciones de forma a potenciar el uso de las folksonomías.

Palabras Clave

Folksonomías. Características de las Etiquetas. Descripción de los Recursos. Características de las marcaciones.

Recebido em: 15.09.2009

Aceito em: 01.10.2009
