

INTEGRAÇÃO DAS FOLKSONOMIAS AOS METADADOS: um perfil de aplicação

Maria Elisabete Catarino*
Ana Alice Rodrigues Pereira Baptista**

RESUMO

Este artigo tem o objetivo de descrever o perfil de aplicação denominado *Social Tagging Application Profile* (STAP) desenvolvido como parte de uma pesquisa de doutoramento em tecnologias e sistemas de informação. O referido perfil de aplicação contém a declaração dos termos de metadados a serem utilizados, de forma complementar ao *Dublin Core* (DC), em repositórios que implementem funcionalidades de *social tagging*. Os termos declarados no STAP representam as novas propriedades identificadas através da análise de etiquetas oriundas de folksonomias. Folksonomia é o resultado da descrição dos recursos da Web pelo próprio usuário. É um conjunto de etiquetas variadas e com carácter geralmente subjetivo. Estas imprimem variedade e riqueza à descrição de recursos Web que até à data era realizada quase exclusivamente por profissionais da informação ou pelos próprios autores. As folksonomias não são, em geral, relacionadas com elementos de metadados estabelecidos e, conseqüentemente, não são inteligíveis por máquinas, nem processáveis em contexto de Web Semântica. Para que tal possa acontecer, é necessário, antes de tudo, que esses elementos de metadados, ou propriedades, sejam conhecidos. Para a identificação das novas propriedades 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. Como resultado foram identificadas e validadas as seguintes novas propriedades: *Action*, *Category*, *Date Tagged*, *Depth*, *Note*, *Rate*, *Share* e *Utility*. O perfil de aplicação apresentado foi proposto para ser utilizado complementarmente ao DC pois entendeu-se que os dois tipos de descrição, tanto o tradicional, como a descrição feita pelos usuários, deveriam coexistir, embora, de forma que não houvesse interferência na descrição tradicional.

Palavras-chave: Folksonomia. Metadados. Dublin Core. Perfil de Aplicação. STAP.

1 INTRODUÇÃO

A descrição dos recursos da Web não tem sido feita apenas pelos gestores dos recursos ou seus autores, mas também pelos usuários. A descrição dos documentos pelo usuário pode ocorrer através da folksonomia, que resulta de uma nova forma de organizar os recursos da Web. Surgiu em meados de 2003 e é o resultado da etiquetagem¹ livre dos recursos da Web pelos seus próprios usuários. Ou seja, a atribuição de etiquetas (em inglês, *tags*) que

¹ Etiquetagem significa atribuir etiquetas aos recursos da Web (CATARINO; BAPTISTA, 2007) e refere-se ao que também é chamado na literatura de *tagging*.

*Departamento de Ciência da Informação- Universidade Estadual de Londrina- Titulação: doutora - e-mail: beteca@uel.br**Departamento de Sistemas de Informação- Universidade do Minho- Titulação: doutora - e-mail: analice@dsi.uminho.pt.

representam tanto a descrição física ou temática do recurso² quanto outros aspectos relativos às funcionalidades e ou relações deste recurso para o seu usuário.

Folksonomia é a tradução do termo *folksonomy* que é um neologismo criado 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 etiquetas³ (*tagging*) a informações ou objectos (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 que etiqueta o recurso da Web.

A folksonomia é algo relativamente recente, mas perfeitamente justificada na organização dos recursos da Web e pressupõe-se que seja adotada pelos repositórios como uma evolução natural da organização dos recursos.

As etiquetas permitem aos usuários representar os documentos conforme sua percepção dos mesmos, ou seja, é uma forma de representar a compreensão particular ou visão que o usuário tem em relação ao recurso (MATHES, 2004; QUINTARELLI, 2005; FEINBERG, 2006).

Pode-se inferir, ao pensar na questão dos elementos de descrição (que são as propriedades dos recursos) e na representação dos recursos sob o ponto de vista do usuário, que a extração destas propriedades nos repositórios seja um contributo à organização e recuperação da informação.

As folksonomias descrevem recursos da Web. Porém, em geral, não estão a ser relacionadas com os metadados porque nem sempre são inteligíveis por máquinas que por vezes não conseguem perceber adequadamente o significado de cada uma das etiquetas. Pressupõe-se que algumas etiquetas podem ser facilmente percebidas pelas máquinas (seja através de comparação com outros valores de registros de metadados ou pela extração automática), como no caso daquelas que estejam relacionadas ao autor, título ou editora, etc. Outras etiquetas serão dificilmente inteligíveis pelo fato de estarem relacionadas com conceitos mais subjetivos (e.g., avaliação da qualidade ou da profundidade do recurso) pelo que é preciso informação adicional.

Para que elas sejam inteligíveis por máquinas e conseqüentemente utilizadas no contexto da Web Semântica, têm que ser alocadas automaticamente a elementos de metadados específicos.

² As etiquetas representam quaisquer formas de descrição seja do ponto de vista físico (as características físicas do recurso, tais como, autor, título, formato, data, etc.), ou do ponto de vista temático (o conteúdo ou temas que o recurso aborda).

³ Nos textos originais em inglês, *Tags*, que segundo Guy e Tonkin (2006) numa simples definição seriam palavras-chave, categorias ou metadados.

Metadados são definidos como “dados sobre dados” ou “dados estruturados sobre dados”. Especificamente no âmbito da Web e das bibliotecas digitais, um dos objetivos da utilização de metadados “é permitir não só descrever documentos eletrônicos e informações em geral, possibilitando sua avaliação de relevância por humanos, mas também permitir que computadores e programas especiais, robôs e agentes de *software*, possam lidar com os metadados associados a documentos e possam então recuperá-los, avaliar sua relevância e manipulá-los com mais eficiência” (CASTRO; SANTOS, 2007).

Existem vários padrões de metadados, dentre eles o *Dublin Core* (DC), que é adotado no âmbito dos repositórios. O DC é um conjunto padrão de elementos de metadados para descrição de recursos eletrônicos. Surgiu após alguns *workshops* destinados a discutir as questões relativas à descrição dos recursos da Web, em especial à necessidade de se criar uma nomenclatura padrão de metadados, visando a interoperabilidade dos dados e a recuperação da informação. DC é composto por todos os termos mantidos pelo DCMI, o DCMI *Metadata Terms* (DCMI *Terms*). Fazem parte o conjunto de elementos de metadados DCMES⁴ e outros termos que são propriedades, subpropriedades, classes (incluindo o DCMI *Type Vocabulary*) e esquemas de codificação (*Vocabulary Encoding Scheme* (VES) e *Syntax Encoding Scheme* (SES)) (DCMI USAGE BOARD, 2008b).

Uma das metas do DC é a extensibilidade que permite que sejam definidos elementos de metadados específicos e complementares ao DC de modo a atender as necessidades específicas de diversas aplicações. No entanto, estes elementos complementares devem estar de acordo com as diretrizes do DCMI para garantir a interoperabilidade.

Uma forma de propor aplicações específicas para uso dos metadados DC, adicionando outros elementos, sem necessariamente acrescentar novos termos aos já existentes é o desenvolvimento de perfis de aplicação.

No contexto do DCMI existem diretrizes para o desenvolvimento de Perfis de Aplicação DC que estão especificadas no documento *Guidelines for Dublin Core Application Profiles* (COYLE; BAKER, 2008). Também são importantes os seguintes documentos publicados anteriormente: o *The Singapore Framework for Dublin Core Application Profile* (NILSSON; BAKER; JOHNSTON, 2008) e o *Dublin Core Application Profile Guidelines* (BAKER et al., 2005).

⁴ DCMES está formalmente normalizado nas seguintes normas internacionais: ISO *Standard* 15836-2003, IETF RFC 5013 e ANSI/NISO *Standard* Z39.85-2007, conforme cita o documento DCMI *Metadata Terms* (DCMI USAGE BOARD, 2008a). Ver: <http://www.niso.org/international/SC4/n515.pdf>; <http://www.ietf.org/rfc/rfc5013.txt>; e <http://www.niso.org/standards/resources/Z39-85-2007.pdf>, respectivamente

Um DCAP é uma declaração especificando que termos de metadados uma organização, fornecedor de informação, ou comunidade usa no seu conjunto de metadados. Por definição, um DCAP deve identificar os *namespace* de cada um dos termos de metadados utilizados. Deve, ainda, identificar se estes termos foram definidos em padrões formais, como por exemplo o DC, ou em outros conjuntos de elementos e vocabulários menos formais, ou mesmo se estes termos foram definidos pelo próprio criador do DCAP para uso numa aplicação local.

O objetivo de um DCAP é promover a interoperabilidade no contexto do modelo DC encorajando a harmonização de uso e convergência de semânticas que emergem ao redor deste modelo. Surgem da necessidade de compartilhar aplicações específicas de refinamento e extensões para o DC sem que seja necessário estender o núcleo padrão mantido pela DCMI.

Um DCAP é composto por um conjunto de documentos (figura 1):

- *Functional Requirements* (obrigatório): descreve o que a comunidade pretende realizar com a aplicação (Requisitos Funcionais);
- *Domain Model* (obrigatório): caracteriza os tipos de “coisas” (*things*) descritas pelos metadados e seus relacionamentos (Modelo de Domínio);
- *Description Set Profile (DSP)* (obrigatório): define o conjunto do registro de metadados detalhadamente, a partir do que o DCMI estabelece na linguagem DSP.
- *Usage Guidelines* (opcional): contém as regras de utilização (Diretrizes de Uso);
- *Encoding Syntax Guidelines* (opcional): define a sintaxe que será utilizada para codificar os dados (Diretrizes de sintaxe e Formato de Dados)

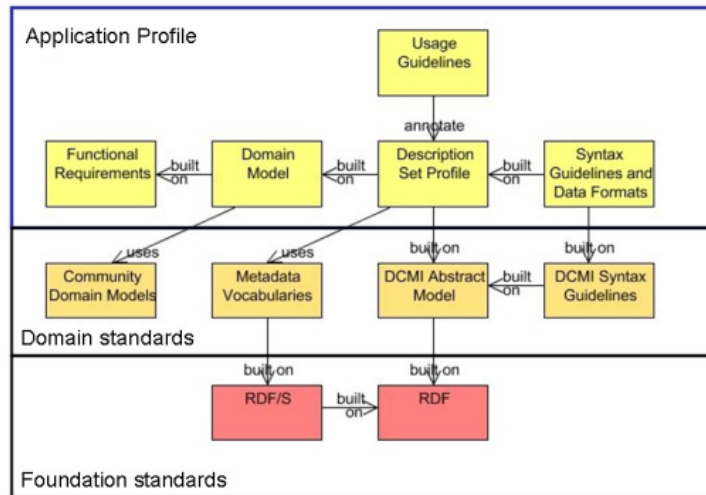


Figura 1: Singapore Framework
 Fonte: The Singapore Framework (NILSSON; BAKER; JOHNSTON, 2008)

Este trabalho apresenta o perfil de aplicação. Foi desenvolvido com o objetivo de declarar os termos de metadados que representam as propriedades que podem conter os valores encontrados nas folksonomias que descrevem os de recursos sob a perspectiva dos seus usuários. Pretende-se que seja uma forma de agregar valor à descrição tradicional realizada nos repositórios.

A seguir a operacionalização do projeto de pesquisa como um todo.

2 Procedimentos Metodológicos

O projeto de pesquisa teve como objetivo a identificação de novas propriedades, com base nas folksonomias, complementares ao conjunto de elementos de metadados DC para a descrição de recursos. Optou-se por uma abordagem de pesquisa qualitativa.

Qualitativa pois o fenômeno estudado não foi tratado de forma quantitativa ou não utilizou procedimentos estatísticos para análise do problema. A pesquisa foi desenvolvida numa perspectiva construtivista, a partir de dados extraídos de documentos diversos mais as inferências em torno das relações destes dados sob a perspectiva do usuário e dos metadados já existentes para a descrição de recursos.

Segundo Myers (1997) os métodos de pesquisa qualitativos foram desenvolvidos nas ciências sociais para estudar os fenômenos sociais e culturais. Na definição de Creswell (2007) um método de pesquisa qualitativa é aquele no qual “O investigador sempre faz alegações de conhecimento com base principalmente ou em perspectivas construtivistas (ou seja, significados múltiplos das experiências individuais, significados social e historicamente construídos, com o objetivo de desenvolver uma teoria ou padrão) ou em perspectivas

reivindicatórias participativas (ou seja, políticas, orientadas para a questão; ou colaborativas, orientadas para mudanças) ou em ambas. Ela também usa estratégias de pesquisa como narrativas, fenomenologias, etnografias, estudos baseados em teorias ou estudos de teoria embasada na realidade. O pesquisador coleta dados emergentes abertos com o objetivo principal de desenvolver temas a partir dos dados”.

Na perspectiva desta abordagem, visando 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. A especificação foi feita com base no conjunto de metadados DC, estabelecido no vocabulário *DCMI Terms*, no modelo *DCMI Abstract Model (DCAM)*⁵ e nas directrizes para *Dublin Core Application Profile (DCAP)*.
- 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.

⁵ Recomendação DCMI publicada em 2007. Trata-se de um documento que especifica os componentes e constructos usados no DC bem como a natureza destes, e como podem ser combinados para criar estruturas de informação. Fornece um modelo independente de codificações específicas de sintaxe. Este modelo também permite uma melhor compreensão dos tipos de descrições que podem ser codificadas e, conseqüentemente, o desenvolvimento de mapas mais adequados para tradução de sintaxes (POWEL et al., 2007).

3 OS PRINCIPAIS RESULTADOS

Tendo em consideração que este trabalho irá apresentar o perfil de aplicação desenvolvido como parte da pesquisa de doutorado, os resultados serão apresentados de forma sucinta. Na sequência será apresentado o perfil de aplicação.

Visando a identificação das propriedades que poderiam ser extraídas das folksonomias decidiu-se investigar dados reais, ou seja, analisar recursos que já se encontravam etiquetados pelos seus usuários.

Neste íterim havia um projeto desenvolvido pelo Departamento de Sistemas de Informação da Universidade do Minho em conjunto com outras instituições europeias, americana e australiana, o *Kinds of Tags* (KoT)⁶.

Optou-se, então, pela utilização do conjunto de dados do projeto KoT que já apresentava alguns resultados, embora preliminares. O conjunto de dados continha dados de 50 recursos mais utilizados no Connotea⁷, na ocasião em que foram recolhidos, e que também estivessem etiquetados no Delicious⁸. Eram em sua maioria recursos do tipo texto, mais precisamente artigos. O tema destes recursos era relacionado basicamente à área de Ciência da Informação. Os dados do KoT foram migrados para a base de dados KoT-Onto desenvolvida para proceder a análise das etiquetas.

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, optou-se por registar o total de ocorrências das etiquetas (79.146).

Após a criação da base KoT_Onto passou-se então à análise das etiquetas.

Para a análise 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 na Web, através dos mecanismos de busca, para identificar o significado da etiqueta. Os recursos léxicos mais utilizados foram: WordNet, Infopedia e Webster.

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-

⁶ O projeto KoT tem vindo a ser realizado em parceria com as seguintes instituições: Universidade do Minho (Portugal), *The Libraries of the Claremont Colleges* (EUA), *University of Bologna* (Itália), *InfoPlex Associates* (Reino Unido), *UKOLN* (Reino Unido), *Reed Business Information* (Itália), *Universidad Carlos III* (Espanha), *Sunrise Research* (Austrália), *Université de Haute Bretagne* (França), *Université Libré de Bruxelles -Faculté de Philosophie et Lettres* (Bélgica).

⁷ O Connotea (<http://www.connotea.org>) é um sistema *on-line* de gestão de referências bibliográficas para académicos e investigadores, criado pela *Nature Publishing Group* em 2004.

⁸ O Delicious (<http://delicious.com>) denominado de *Social Bookmarks Manager* pelo seu fundador, Joshua Schachter.

mail, site ou *blog*), este foi utilizado para dirimir dúvidas quanto ao significado da etiqueta. No entanto, mesmo esgotando todas as possibilidades ainda restaram etiquetas cujos significados não foi possível identificar.

As etiquetas analisadas foram agrupadas em suas formas variantes (singular/plural, maiúsculas/minúsculas, idiomas, grafia, siglas e abreviaturas). Este procedimento foi realizado para facilitar posteriormente a identificação das propriedades. Pressupôs-se que o agrupamento das etiquetas facilitaria a compreensão das mesmas e conseqüentemente a identificação das propriedades. Como resultado deste agrupamento, pode-se perceber melhor o significado e agilizar o processo de identificação das propriedades.

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. Como por exemplo: a *Key-tag* Digital Libraries possuía as seguintes formas variantes: singular/plural (digital libraries e digital library); idiomas (biblioteca digital e digital library); sigla (dl) e grafia das etiquetas (digital libraries, digital_libraries e digitallibraries).

Como resultado as 5098 etiquetas foram agrupadas em 3224 *Key-tags*. No entanto, considerando que uma *Key-tag* pode ser atribuída a vários recursos e, em alguns casos relacionar-se com mais do que uma propriedade, o total de ocorrência de *key-tags* foi de 7466. Deste total, 4519, ou seja, 60,5% foram identificadas como sendo propriedades do DC. Para 1974 (26,4%) *Key-tags* foram propostas novas propriedades a serem validadas e 973 (13%) não tiveram nenhuma propriedade relacionada.

As novas propriedades identificadas foram: *Rate* com 26,4% (521), *Utility* 26,1% (515), *Action* 15,9% (314), *Category* 9,7% (192), *Note* 9,7% (191), *Depth* 4,1% (81), *User Name* 4,0% (78), *Share* 1,6% (32), *Self Reference* 1,4% (27) e *Date Tagged* 1,2% (23).

A descrição de cada uma destas propriedades será apresentada no quadro 2, na descrição do perfil de aplicação.

4 PERFIL DE APLICAÇÃO

Considerando-se que as propriedades identificadas como necessárias para a descrição de recursos com base nas folksonomias possam vir a ser aplicadas em vários contextos, entende-se que existem várias opções para a apresentação do perfil de aplicação, o que vai depender das necessidades de cada aplicação.

A seguir apresenta-se uma opção considerada aplicável para a descrição de recursos etiquetados. Esta opção representa um perfil de aplicação onde serão declarados os termos

complementares ao DC. A aplicação que adoptar este perfil fará uso deste em conjunto com metadados do DC ou outros utilizados na descrição tradicional.

Contudo, deve-se ressaltar que o perfil de aplicação que será apresentado no âmbito deste trabalho não deve ser considerado como opção única para a aplicação das novas propriedades identificadas. Outros perfis de aplicação poderão compor outras formas de utilização das propriedades.

O *Social Tagging Application Profile* (STAP)⁹ foi criado para declarar termos de metadados que são propriedades complementares às já existentes no DC para a descrição de recursos de repositórios institucionais que implementem funcionalidades de *social tagging* ou importem etiquetas de outros sistemas. Portanto, foi proposto para ser utilizado pelos repositórios institucionais que possuam uma folksonomia resultante das etiquetas atribuídas pelos próprios usuários dos recursos. A intenção é acrescentar valor à descrição tradicional permitindo que os próprios usuários registem os valores relativos às propriedades que descrevem o recurso. Pressupõem-se que desta forma serão ampliadas as possibilidades de organização e recuperação da informação de forma diferenciada (CATARINO, 2009).

O STAP foi criado conforme estabelecido pelas diretrizes do DCMI.

Em primeiro lugar definiram-se os requisitos funcionais e o domínio do STAP com base nos resultados da pesquisa. Estes itens foram delineados de forma a suportar a declaração dos termos de metadados para a descrição de recursos de repositórios institucionais que adotem folksonomias.

A seguir apresentam-se: os requisitos funcionais e domínio; a definição dos termos e o *Description Set Profile*.

4.1 Requisitos funcionais e domínio.

O STAP pretende documentar os elementos de metadados que servirão para a descrição dos recursos sob a perspectiva do usuário. Portanto a descrição do recurso etiquetado, ou *Resource Tagged*, como será denominado neste perfil de aplicação.

Considerou-se que para este tipo de descrição serão necessárias as seguintes propriedades:

- a) Propriedades baseadas nos termos já existentes no DC, e
- b) Novas propriedades identificadas.

⁹ Ver <http://odisseia.dsi.uminho.pt/STAP>

No primeiro conjunto de propriedades foram escolhidas as seguintes propriedades originárias do DC: *Audience, Subject, Type*.

A escolha de algumas propriedades DC em detrimento de outras será justificada de seguida. Em primeiro lugar, considerou-se que na maioria das propriedades DC (tais como: *creator, title, data, publisher, identifier*), os valores são objetivos. Significa que os valores relacionados a este tipo de propriedades não seriam subjetivos e independentemente de serem atribuídos pelos profissionais gestores dos repositórios ou pelos usuários, não sofreriam alterações a ponto de agregar valor à descrição. E portanto, mesmo que atribuídos numa forma diferente pelo usuário, não acrescentariam valor à descrição, ou até mesmo poderia distorcê-la.

Sugere-se que as propriedades DC que poderiam se relacionar as etiquetas constantes nas folksonomias e que acrescentariam valores diferenciados seriam portanto: *audience, subject* e *type*.

Audience poderá expressar para quem o recurso é útil na visão do usuário. Como exemplo a etiqueta For Nurses, atribuída a um recurso, pode significar que o usuário que atribuiu esta etiqueta entende que ele é útil para esta classe profissional.

Subject é a propriedade que representa a maioria das etiquetas atribuídas pelos usuários. Os valores relativos a esta propriedade complementarão a indexação temática dos recursos, pois permitirão que o próprio usuário identifique termos que representam o tópico ou tema abordado, seja ele o assunto principal ou secundário. Este tipo de indexação é importante neste contexto, pois cada usuário tem uma percepção única em relação ao recurso.

Type permitirá que o usuário registre o tipo do recurso, conforme a sua concepção de natureza ou gênero. Como exemplo de *key-tags* que se relacionaram à propriedade *Type* porém não constam no DCMI *Type Vocabulary*: e-zine, e-book

O segundo conjunto é formado pelas novas propriedades identificadas e que foram consideradas validadas: *Action, Category, Date Tagged, Depth, Note, Rate, Share* e *Utility*.

As novas propriedades são diferenciadas pois extrapolam a descrição do recurso em si. Estas propriedades agregam valores que representam a relação entre o recurso e o usuário: 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*).

Propor-se-á ainda uma propriedade mais abrangente, e que poderá ser utilizada para relacionar os valores das folksonomias que não estejam relacionados com propriedades mais específicas do perfil de aplicação. Esta propriedade será identificada pelo termo *Tag*.

O STAP deve ser complementar à descrição recursos dos repositórios institucionais, ou seja, os metadados propostos neste perfil de aplicação coexistirão com o conjunto de metadados da descrição tradicional - esta seguirá os procedimentos normais já adoptados pelo repositório podendo, nestes casos, ser utilizados os Esquemas de Codificação indicados no DCMI *Metadata Terms* ou quaisquer outros que os gestores do repositório considerarem adequados.

Propõem-se que para os valores originados das folksonomias:

- Não será adoptado nenhum esquema de codificação (*Syntax Endoding Scheme* ou *Vocabulary Encoding Schemes*) sugerido no DCMI *Metadata Terms* ou outros, pois pretende-se respeitar a descrição feita pelo usuário na sua forma original;
- Não haverá restrições de máxima ou mínima ocorrência porque os usuários atribuem um número de etiquetas que varia do um ao infinito; e
- Nenhuma das propriedades será obrigatória.

4.2 Definição dos Termos de Metadados e *Description Set Profile*

Para a descrição dos *Resource Tagged* serão utilizados tanto os termos novos (*Action, Category, Date Tagged, Depth, Notes, Rate, Share e Utility*) quanto os originários do DC. Estes últimos referentes às propriedades (*Audience, Subject e Type*) cujos valores, pela sua característica de subjetividade, poderão acrescentar valores diferenciados aos atributos dos recursos. Para os termos originários do DC optou-se por declarar novos termos ao invés de utilizar os do DCMI-*Terms*. Os termos serão pós-fixados com “*Tag*”: *Audience Tag, Subject Tag e Type Tag*. Desta forma pretende-se distinguir os valores atribuídos de forma tradicional daqueles atribuídos pelos usuários.

Tradicionalmente os valores que são adicionados aos metadados por profissionais, ou pelos próprios depositantes, seguem as determinações estabelecidas pelos gestores do repositório, inclusive no que concerne a adopção de esquemas de codificação. Os valores oriundos das folksonomias serão sempre Literais, Opcionais e Repetitivos, respeitando a opção do usuário.

4.3 Termos para a descrição do *Resource Tagged*

Nesta secção cada um dos termos relativos às propriedades que compõem o STAP será descrito em tabelas. Cada termo foi definido com os seguintes atributos: *Term URI*, *Name*, *Label*, *Defined By*, *Definition*, *Comments*, *Type of Term*, *Refines*, *Type of Value*, *Min Occurrence* e *Max Occurrence* (quadro 1). Estes atributos foram considerados suficientes para a posterior criação do conjunto de descrição do registro de metadado conforme a *Description Set Profile*.

Term URI	<i>Uniform Resource Identifier</i> (URI) que identifica exclusivamente o elemento
Name	Identificador atribuído para o elemento; único dentro do DCMI <i>Namespace</i>
Label	Etiqueta para leitura humana atribuída ao elemento;
Defined By	Um <i>namespace</i> que aponta para o documento no qual o termo encontra-se definido.
Definition	A declaração que claramente representa o conceito e a natureza essencial do elemento;
Comments	Informações adicionais sobre o termo ou sobre a sua aplicação;
Type of Term	Tipo de Termo, conforme define o DCMI <i>Abstract Model</i> [DCAM]
Refines	uma propriedade da qual o termo descrito é uma subpropriedade
Type of Value	O tipo do valor permitido: <i>Literal</i> (o valor é exactamente uma <i>string</i>) ou <i>Non-Literal</i> (uma URI que remete ao valor ou um literal que representa o valor)
Min Occurrence	Número mínimo de vezes que um valor poderá ocorrer numa descrição
Max Occurrence	Número máximo de vezes que um valor poderá ocorrer numa descrição

Quadro 1: Atributos dos termos STAP

Todos os termos possuem valores idênticos para os seguintes atributos: *Defined By* = <<http://odisseia.dsi.uminho.pt/STAP/terms/action>>; *Type of Term* = *Property*; *Type of Value* = *Literal*; *Min. Occurrence* = 0 e *Max. Occurrence* = *Infinity*. Na sequência (quadro 2) serão apresentados os demais atributos de cada um dos termos.

Action	
Term URI	http://odisseia.dsi.uminho.pt/STAP/terms/action
Name	<i>Action</i>
Label	<i>Action</i>
Definition	Uma ação que o usuário pretende fazer ou sugere fazer em relação ao recurso.
Comments	<i>Action</i> pode ser usada para descrever a ação do usuário em relação ao recurso. Exemplos: <i>_toread</i> ; <i>a_lire</i> ; <i>to print</i> .
Refines	http://odisseia.dsi.uminho.pt/STAP/terms/tag
Category	
Term URI	http://odisseia.dsi.uminho.pt/STAP/terms/category
Name	<i>Category</i>
Label	<i>Category</i>
Definition	Categoria de um grupo de recursos.
Comments	<i>Category</i> pode ser usada para classificar um conjunto de recursos, conforme classificações diferentes do tema ou assunto, uma vez que para isso a propriedade <i>Subject</i> deve ser utilizada. Exemplos:

	Faq (conjunto de recursos que contém respostas as <i>Frequently Asked Questions</i>)
Refines	http://odisseia.dsi.uminho.pt/STAP/terms/tag
Date Tagged	
Term URI	http://odisseia.dsi.uminho.pt/STAP/terms/dateTagged
Name	<i>dateTagged</i>
Label	<i>Date Tagged</i>
Definition	Data ou período em que o recurso foi etiquetado.
Comments	<i>Date Tagged</i> pode ser usado para representar a data ou período que ocorreu a etiquetagem do recurso. Para este perfil de aplicação não será recomendado o uso de nenhum esquema de codificação, para que seja respeitada a forma de inserção da etiqueta pelo usuário do recurso.
Refines	http://odisseia.dsi.uminho.pt/STAP/terms/Tag
Refines	http://purl.org/dc/elements/1.1/date
Refines	http://purl.org/dc/terms/date
Depth	
Term URI	http://odisseia.dsi.uminho.pt/STAP/terms/depth
Name	<i>depth</i>
Label	<i>Depth</i>
Definition	Grau de profundidade intelectual do recurso.
Comments	<i>Depth</i> pode ser usada para representar o grau de profundidade intelectual do recurso na estimativa do usuário.
Refines	http://odisseia.dsi.uminho.pt/STAP/terms/tag
Note	
Term URI	http://odisseia.dsi.uminho.pt/STAP/terms/note
Name	<i>note</i>
Label	<i>Note</i>
Definition	Uma nota ou anotação referente ao recurso.
Comments	<i>Note</i> pode ser utilizada para expressar um comentário ou observação com o objetivo de fazer lembrar algo, ou registar uma observação, comentário ou explicação relativo ao recurso etiquetado.
Refines	http://odisseia.dsi.uminho.pt/STAP/terms/tag
Rate	
Term URI	http://odisseia.dsi.uminho.pt/STAP/terms/rate
Name	<i>rate</i>
Label	<i>Rate</i>
Definition	A qualidade do recurso etiquetado.
Comments	<i>Rate</i> pode ser usada para expressar a avaliação qualitativa do usuário em relação ao recurso etiquetado.
Refines	http://odisseia.dsi.uminho.pt/STAP/terms/tag
Share	
Term URI	http://odisseia.dsi.uminho.pt/STAP/terms/share
Name	<i>Share</i>
Label	<i>Share</i>
Definition	Uma entidade com a qual o recurso etiquetado será compartilhado.
Comments	<i>Share</i> pode ser usado para indicar uma entidade. A entidade pode ser expressa com o nome de uma pessoa, organização ou serviço com quem o usuário quer compartilhar o recurso.
Refines	http://odisseia.dsi.uminho.pt/STAP/terms/tag
Refines	http://purl.org/dc/terms/audience
Tag	
Term URI	http://odisseia.dsi.uminho.pt/STAP/terms/tag
Name	<i>Tag</i>
Label	<i>Tag</i>
Definition	Etiqueta atribuída pelo usuário do recurso.
Comments	As etiquetas são grafadas pelo usuário sem controle de vocabulário e podem servir para a descrição física ou temática, qualificação, categorização, bem como para anotar ações do usuário em relação ao recurso e quaisquer outros apontamentos.
Utility	
Term URI	http://odisseia.dsi.uminho.pt/STAP/terms/utility
Name	<i>Utility</i>
Label	<i>Utility</i>
Definition	Representa a finalidade de uso do recurso para o usuário.

Comments	<i>Utility</i> pode ser usado para agrupar os recursos de acordo com a utilidade do mesmo para o usuário.
Refines	http://odisseia.dsi.uminho.pt/STAP/terms/tag
Audience Tag	
Term URI	http://odisseia.dsi.uminho.pt/STAP/terms/audienceTag
Name	audienceTag
Label	Audience Tag
Definition	Etiqueta que representa uma classe de entidade para quem o recurso se destina ou é útil
Refines	http://odisseia.dsi.uminho.pt/STAP/terms/tag
Refines	http://purl.org/dc/terms/audience
Subject Tag	
Term URI	http://odisseia.dsi.uminho.pt/STAP/terms/subjectTag
Name	subjectTag
Label	Subject Tag
Definition	Etiqueta que representa o tópico, ou tema, do recurso
Refines	http://odisseia.dsi.uminho.pt/STAP/terms/tag
Refines	http://purl.org/dc/elements/1.1/subject
Refines	http://purl.org/dc/terms/subject
Type Tag	
Term URI	http://odisseia.dsi.uminho.pt/STAP/terms/typeTag
Name	typeTag
Label	Type Tag
Definition	Etiqueta que representa a natureza ou género do recurso
Refines	http://odisseia.dsi.uminho.pt/STAP/terms/tag
Refines	http://purl.org/dc/elements/1.1/type
Refines	http://purl.org/dc/terms/type

Quadro 2: Descrição dos Termos

Na próxima seção apresenta-se o conjunto de descrição para o *Resource Tagged* conforme o *Description Set Profile*.

4.4 Description Set Profile

O perfil de aplicação será descrito conforme a *Description Set Profile* (DSP) publicado pelo DCMI (NILSSON, 2008). A DSP é uma linguagem para descrever restrições num conjunto de descrições (*description set*). Uma DSP contém um *Description Template* para cada “coisa” a ser descrita - bem como suas relações - e que está especificada no modelo de domínio. O STAP terá apenas um *Description Template*, o *Resource Tagged*. Um *Description Template* é composto pelos diversos *Statement Templates*, que contém todas as restrições às propriedades (*max. e min. Occurrence, type of value, etc*).

O DSP:

```

DescriptionSet: STAP
  Description template: ResourceTagged
    minimum = 1; maximum = 1
  Statement template: tag
    minimum = 0; maximum = unlimited
    Property: http://odisseia.dsi.uminho.pt/STAP/terms/tag
    Type of Value = "literal"
  Statement template: action
    minimum = 0; maximum = unlimited
    Property: http://odisseia.dsi.uminho.pt/STAP/terms/action
    SubPropertOf: http://odisseia.dsi.uminho.pt/STAP/terms/tag

```

Type of Value = "literal"
Statement template: category
minimum = 0; maximum = unlimited
Property: <http://odisseia.dsi.uminho.pt/STAP/terms/category>
SubPropertOf: <http://odisseia.dsi.uminho.pt/STAP/terms/tag>
Type of Value = "literal"

Statement template: dateTagged
minimum = 0; maximum = unlimited
Property: <http://odisseia.dsi.uminho.pt/STAP/terms/dateTagged>
SubPropertOf: <http://odisseia.dsi.uminho.pt/STAP/terms/tag>
SubPropertOf: <http://purl.org/dc/elements/1.1/date>
SubPropertOf: <http://purl.org/dc/terms/date>
Type of Value = "literal"

Statement template: depth
minimum = 0; maximum = unlimited
Property: <http://odisseia.dsi.uminho.pt/STAP/terms/depth>
SubPropertOf: <http://odisseia.dsi.uminho.pt/STAP/terms/tag>
Type of Value = "literal"

Statement template: note
minimum = 0; maximum = unlimited
Property: <http://odisseia.dsi.uminho.pt/STAP/terms/note>
SubPropertOf: <http://odisseia.dsi.uminho.pt/STAP/terms/tag>
Type of Value = "literal"

Statement template: rate
minimum = 0; maximum = unlimited
Property: <http://odisseia.dsi.uminho.pt/STAP/terms/rate>
SubPropertOf: <http://odisseia.dsi.uminho.pt/STAP/terms/tag>
Type of Value = "literal"

Statement template: share
minimum = 0; maximum = unlimited
Property: <http://odisseia.dsi.uminho.pt/STAP/terms/share>
SubPropertOf: <http://odisseia.dsi.uminho.pt/STAP/terms/tag>
SubPropertOf: <http://purl.org/dc/terms/audience>
Type of Value = "literal"

Statement template: audienceTag
minimum = 0; maximum = unlimited
Property: <http://odisseia.dsi.uminho.pt/STAP/terms/audienceTag>
SubPropertOf: <http://odisseia.dsi.uminho.pt/STAP/terms/tag>
SubPropertOf: <http://purl.org/dc/terms/audience>
Type of Value = "literal"

Statement template: subjectTag
minimum = 0; maximum = unlimited
Property: <http://odisseia.dsi.uminho.pt/STAP/terms/subjectTag>
SubPropertOf: <http://odisseia.dsi.uminho.pt/STAP/terms/tag>
SubPropertOf: <http://purl.org/dc/elements/1.1/subject>
SubPropertOf: <http://purl.org/dc/terms/subject>
Type of Value = "literal"

Statement template: typeTag
minimum = 0; maximum = unlimited
Property: <http://odisseia.dsi.uminho.pt/STAP/terms/typeTag>
SubPropertOf: <http://odisseia.dsi.uminho.pt/STAP/terms/tag>
SubPropertOf: <http://purl.org/dc/elements/1.1/type>
SubPropertOf: <http://purl.org/dc/terms/type>
Type of Value = "literal"

Pretende-se que o STAP seja futuramente validado, através de prova de conceito, numa aplicação prática visando a implementação de metadados específicos para a descrição de recursos etiquetados.

5 CONSIDERAÇÕES FINAIS

O perfil de aplicação apresentado nesta pesquisa, o STAP, foi proposto para ser utilizado complementarmente ao DC. Entendeu-se que os dois tipos de descrição, tanto o tradicional, como a descrição feita pelos usuários, deveriam coexistir, embora, de forma que não houvesse interferência na descrição tradicional. A intenção é a de garantir que a descrição

feita pelos profissionais não se misture àquela feita pelos usuários por se considerar que elas têm abordagens completamente distintas.

Outra definição para o perfil de aplicação foi de que não seria adequado aproveitar as etiquetas advindas das folksonomias cujos valores seriam objetivos e que não agregariam valores à descrição tradicional. Como exemplo desta situação pode-se citar o caso da propriedade *Creator* já existente no DC e que não seria diferente se registrado pelo usuário.

Neste sentido para compor o perfil de aplicação foram eleitas apenas algumas propriedades originárias do DC consideradas como passíveis de agregar valor à descrição tradicional. Estas propriedades originárias do DC foram declaradas como novos termos: *Audience Tag*, *Subject Tag* e *Type Tag*. Portanto o STAP contém apenas novas propriedades, tanto as propriedades originárias do DC quanto as identificadas originalmente nesta pesquisa.

No entanto, não há impedimentos para que ambas as abordagens de descrição, tanto a tradicional quanto a baseada na folksonomias, sejam misturadas, o que vai depender da decisão dos gestores de repositórios ou sistemas que adotarem as folksonomias como complemento aos metadados.

Como sugestão de trabalho futuro o desenvolvimento de um projeto multidisciplinar que desenvolva uma aplicação para a recolha automática de etiquetas atribuídas pelos usuários em repositórios institucionais e que as relacione com as propriedades do STAP. Tal projeto deverá envolver áreas diversas, tais como Ciência da Informação, Linguística, Sistemas de Informação, Ciência da Computação, Inteligência Artificial, Processamento da Linguagem Natural e *DataMining*.

Como trabalhos derivado, especificamente em relação ao STAP, pode-se citar o uso do *namespace* STAP/Tag no projeto “*Social Tagging Importer add-on*” que está desenvolvendo uma aplicação para o DSpace.

THE INTEGRATION OF FOLKSONOMIES IN METADATA: an application profile

ABSTRACT

This article aims to describe the Social Tagging Application Profile (STAP) developed as part of a research of Ph.D. in technology and information systems. The application profile contains a statement of metadata terms to be used in a complementary way of Dublin Core (DC) in repositories that implement functionality of social tagging. The STAP's terms describes the new properties identified through the analysis of labels from Folksonomy. Folksonomy is the result of the description of Web resources made by their own users. It is a wide set of tags normally with a subjective character. These apply richness and variety to the Web resources'

descriptions which until today was mainly expressed either by their authors or by information professionals. Folksonomies, are not generally related with metadata elements and, consequently, they are not machine-readable or processed in a Semantic Web context. To make this possible, the first step is to identify these metadata elements, or properties. For the identification of new properties are developed the following methodological procedures: creation of the database; dataset tag analysis; identification of Dublin Core (DC) complementary properties; validation of the proposal and application profile and ontology development. The following new properties were identified and validated: Action, Category, Date Tagged, Depth, Note, Rate, Share and Utility. The application profile presented was proposed to be used in addition to DC. It was thought that the two types of description, both the traditional and the description given by users, should coexist, though, so there was no interference in the traditional description.

Keywords: Folksonomy. Metadata. Dublin Core. Application Profile. STAP

REFERÊNCIAS

BAKER, T. et al. **Dublin Core Application Profile Guidelines**. Dublin Core Metadata Initiative, 2005. Disponível em: <<http://dublincore.org/usage/documents/2005/09/03/profile-guidelines/>>. Acesso em: 01 Jun. 2008.

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 web semântica. **Informação & Sociedade: Estudos**, João Pessoa, v.17, n.2, p. 12-21, Maio/Ago. 2007.

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

CATARINO, M. E.; BAPTISTA, A. A. Relating Folksonomy with Dublin Core. In: DC2008 - INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, Berlin, 2008. **Proceedings...** Dublin Core Metadata Initiative, 2008a. Disponível em: <<http://repositorium.sdum.uminho.pt/handle/1822/8167>>. Acesso em: 05 Nov. 2008.

CATARINO, M. E.; BAPTISTA, A. A. Social Tagging and Dublin Core: a Preliminary Proposal for an Application Profile for DC Social Tagging. In: ELPUB2008 - INTERNATIONAL CONFERENCE ON ELECTRONIC PUBLISHING, 12., Toronto, 2008b. **Proceedings...** University of Toronto, 2008. Disponível em: <http://elpub.scix.net/data/works/att/100_elpub2008.content.pdf>. Acesso em: 05 Nov. 2008.

CATARINO, M.E. Social Tagging Application Profile (STAP). Disponível em: <<http://odisseia.dsi.uminho.pt/STAP>>. Acesso em 14 Abr. 2009.

COYLE, K.; BAKER, T. **Guidelines for Dublin Core Application Profile**: working draft. Dublin Core Metadata Initiative, 2008. Disponível em: <<http://dublincore.org/documents/2008/11/03/profile-guidelines/>>. Acesso em: 06 Nov. 2008.

CRESWELL, J. W. **Projeto de Pesquisa, Método Qualitativo, Quantitativo e Misto**. São Paulo : Artmed, 2007.

DCMI USAGE BOARD. **DCMI Metadata Terms**. Dublin Core Metadata Initiative, 2008a. Disponível em: <<http://dublincore.org/documents/2008/01/14/dcmi-terms/>>. Acesso em: 15 Mar. 2008.

DCMI USAGE BOARD. **DCMI Type Vocabulary**. Dublin Core Metadata Initiative, 2008b. Disponível em: <<http://dublincore.org/documents/2008/01/14/dcmi-type-vocabulary/>>. Acesso em: 15 Mar. 2008.

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

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

MYERS, M. D. Qualitative research in Information Systems. **MIS Quarterly**, v.21, n.2, p.241-242, 1997. Disponível em: <<http://www.qual.auckland.ac.nz>>. Acesso em: 22 Out. 2007.

NILSSON, M. **Description Set Profile: a constraints language for Dublin Core Application Profile**. Dublin Core Metadata Initiative, 2008. Disponível em: <<http://dublincore.org/documents/2008/03/31/dc-dsp/>>. Acesso em: 06 Nov. 2008.

NILSSON, M.; BAKER, T.; JOHNSTON, P.. **The Singapore Framework for Dublin Core Application Profile**. DCMI, 2008. Disponível em: <<http://dublincore.org/documents/2008/01/14/singapore-framework/>>. Acesso em 06 Nov. 2008.

POWELL, A. et al.. **DCMI Abstract Model**. Dublin Core Metadata Initiative, 2007. Disponível em: <<http://dublincore.org/documents/abstract-model/#sect-2>>. Acesso em: 03 Out. 2007.

QUINTARELLI, E. (2005). Folksonomies: power to the people. INCONTRO ISKO ITALIA - UNIMIB, Milão, 2005. **Papers...** Milão, 2005. Disponível em: <<http://www.iskoi.org/doc/folksonomies.htm>>. Acesso em: 23 Out. 2006.

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