



XI Encontro Nacional de Pesquisa em Ciência da Informação
Inovação e inclusão social: questões contemporâneas da informação
Rio de Janeiro, 25 a 28 de outubro de 2010

GT8 - Informação e Tecnologia
Modalidade de apresentação: Comunicação Oral

ONTOLOGIA STAP: UM VOCABULÁRIO DE TERMOS DE METADADOS

Maria Elisabete Catarino

Universidade de Londrina

Ana Alice Baptista

Universidade do Minho, Portugal

Resumo: Este artigo tem o objetivo de apresentar a Ontologia STAP, desenvolvida como parte de uma pesquisa de doutorado que identificou novos elementos de metadados oriundos das folksonomias como contributo para a descrição de recursos em repositórios. A motivação da pesquisa foi a Web Semântica, que é a base para a estruturação e a interoperabilidade dos recursos na Web. Para tanto o *Word Wide Web Consortium (W3C)* propõe normas e padrões para a criação de repositórios, construção de vocabulários e definição de regras para gestão de dados. Dentre as várias camadas da Arquitetura da Web Semântica a de Ontologias é a que suporta a evolução de vocabulários que podem definir relações entre diferentes conceitos. A ontologia desenvolvida é composta pelo vocabulário dos termos de metadados do Dublin Core (DC) mais os metadados oriundos das folksonomias. DC é um conjunto de metadados para a descrição de recursos da Web e é adotado nos repositórios para o compartilhamento de dados por meio do protocolo OAI-PMH. Folksonomia é o resultado da atribuição livre e pessoal de etiquetas (*tagging*), pelos próprios usuários, às informações ou objetos, visando à sua recuperação. Pressupõe-se que a folksonomia permite uma nova forma de organização de recursos da Web e que, naturalmente, poderá também ser adotada pelos repositórios. As etiquetas podem ser relacionadas com propriedades dos metadados DC e outras propriedades complementares, enriquecendo, assim, a organização dos recursos sem comprometer a interoperabilidade dos seus metadados. Como metodologia foi usada a abordagem sistemática proposta por Guizzardi (2000) compreendendo a Identificação dos Propósitos e Especificação de Requisitos, Captura, Formalização, Integração, Avaliação e Documentação da Ontologia. A partir desta abordagem foi construída uma ontologia genérica com o propósito de compartilhar os termos de metadados por meio de um único esquema RDF, formada por 138 termos DCMI mais os termos do perfil de aplicação STAP.

Palavras Chave: Ontologia STAP. Metadados. Dublin Core. Folksonomia.



1 INTRODUÇÃO

O foco do estudo apresentado neste artigo é a Web Semântica (WS). Entende-se que a WS é a base para a estruturação e a interoperabilidade dos recursos na Web. O conceito WS representa o projeto da *World Wide Web Consortium* (W3C) que “pretende embutir inteligência e contexto nos códigos *Extensible Markup Language* (XML) utilizados para a realização de páginas Web” (*WORD WIDE WEB CONSORTIUM* [W3C], 2001). Nela, a informação é dada com um significado bem definido, permitindo melhor interação entre computadores e as pessoas (BERNERS-LEE; HENDLER; LASSILA, 2001).

Segundo o W3C a WS é a “Web de Dados”, ou seja, o objetivo é disponibilizar tecnologias e regras para que todos os recursos da Web (inclusive os dados armazenados em bases/bancos de dados) estejam disponíveis em formato padrão e possam ser manipulados. Neste sentido é preciso adotar vocabulários e regras para que os dados possam ser manipulados (W3C, 2010).

O formato padrão recomendado é o *Resource Description Framework* (RDF). No entanto, existem outras especificações formais para formatos de intercâmbio de dados (como por exemplo: RDF/XML, N3, *Turtle*, *N-Triples*), e notações tais como *RDF Schema* (RDFS) e a *Web Ontology Language* (OWL), tudo com a intenção de prover uma descrição formal de conceitos, termos e relacionamentos, num específico domínio do conhecimento (W3C, 2001).

Além dos formatos de intercâmbio de dados, também são imprescindíveis os vocabulários. Catarino e Baptista (2008, p.35) afirmam que “conforme se observa na descrição do princípio fundamental da WS, um dos componentes fundamentais são os vocabulários controlados, e dentre as diversas camadas da Arquitetura da Web Semântica, é a camada de ontologias a que suporta a evolução de vocabulários que podem definir relações entre diferentes conceitos”.

Entende-se que o desenvolvimento de ontologias é imprescindível para a WS e, portanto, justifica-se a realização do projeto que culminou na criação de um vocabulário de metadados composto pelos termos do Dublin Core (DC) e outros que foram identificados no âmbito das folksonomias: a Ontologia STAP.

Para contextualizar, apresenta-se a seguir a motivação para o projeto que resultou na ontologia a ser descrita. A motivação foi a percepção de que as etiquetas (*tags*) que os próprios usuários da Web atribuem aos recursos, representam propriedades que podem agregar valor aos metadados descritivos com atributos que vão além da descrição física e temática. Constatou-se que



as novas propriedades de metadados identificadas são inovadoras, no sentido de que não descrevem apenas o recurso em si, mas agregam valores do recurso em relação ao usuário: ação do usuário em relação ao recurso (*Action*); categorização do recurso que vai além da “classificação” temática (*Category*); 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 finalidade de uso do recurso para o usuário (*Utility*) (CATARINO, 2009a). Identificadas as novas propriedades, desenvolveu-se um perfil de aplicação a ser adotado como complementar ao DC que foi formalizado na Ontologia STAP.

O objetivo deste artigo é apresentar a ontologia desenvolvida no âmbito da pesquisa de doutorado. Antecedendo a apresentação da Ontologia propriamente dita, na próxima seção deste artigo serão introduzidos os temas-base e na sequência descrever-se-á a metodologia de desenvolvimento da Ontologia STAP.

2 METADADOS, FOLKSONOMIA E ONTOLOGIA: conceitos

Serão apresentados a seguir os conceitos básicos referentes aos temas que fundamentaram o desenvolvimento da Ontologia STAP: Metadados, Folksonomia e Ontologia.

2.1 Metadados

Metadados são definidos como “dados sobre dados” ou “dados estruturados sobre dados”. Os metadados incluem dados associados com qualquer sistema de informação ou objeto de informação com os seguintes objetivos: descrição, administração, requisitos legais, funcionalidades técnicas, uso e preservação (WOODLEY; CLEMENT; WINN, 2005).

Há uma variedade de esquemas de metadados que são desenvolvidos para propostas específicas. Conforme consta no documento da NISO (2004), os esquemas de metadados geralmente especificam os nomes dos elementos e sua semântica (a definição ou o que significa um elemento). No entanto, opcionalmente, também podem ser definidas normas para os conteúdos dos elementos, i.e., especificar critérios para os valores a serem incluídos num dado elemento de metadados, tais como: critérios para a formulação de conteúdos (e.g., como identificar um título principal), normas para a representação do conteúdo (e.g., uso de maiúsculas, minúsculas, etc.) ou



definição de valores permitidos para os conteúdos (e.g., por meio do uso de vocabulários controlados).

Dentre os vários esquemas ou padrões de metadados identificados na literatura por Grácio (2002), NISO (2004), e Woodley, Clement e Winn (2005), a seguir alguns exemplos de metadados para descrição de recursos:

- DC (*Dublin Core*): Conjunto de elementos de metadados para descrição de recursos eletrônicos. Ver <http://dublincore.org/>
- MARC (*Machine-Readable Cataloging Record*): Formato padrão para a descrição de registros bibliográficos em formato legível por máquinas, conforme norma ISO 2709:1996 “Informação e documentação – formato de intercâmbio de informações”. Ver <http://www.loc.gov/marc/>
- METS (*Metadata Encoding & Transmission Standard*): Um padrão para definição de metadados descritivos, administrativos e estruturais para objetos de bibliotecas digitais. Ver <http://www.loc.gov/standards/mets/>
- MODS (*Metadata Object Description Schema*): Metadados para descrição registros bibliográficos. Esquema em XML compatível com o formato MARC possibilitando o compartilhamento de dados. Ver <http://www.loc.gov/standards/mods/>

O objeto de estudo da pesquisa que originou a Ontologia STAP foi o Dublin Core. Quem conduz as atividades relacionadas com o desenvolvimento e manutenção do DC é a *Dublin Core Metadata Initiative* (DCMI), que visa criar mecanismos que facilitem a recuperação de recursos na Internet, utilizando padrões de metadados. Um *workshop* foi realizado em Dublin, Ohio, em 1995, com o objetivo principal de definir um conjunto mínimo de elementos para descrição dos recursos da Web. Nesta ocasião, mais de 50 pessoas discutiram a utilidade de um conjunto semântico para descrição visando a pesquisa e localização dos recursos da Web de forma simples (*Dublin Core Metadata Initiative* [DCMI], 2004).

DC é composto por todos os termos mantidos pelo DCMI, o *DCMI Metadata Terms* (DCMI *Terms*). Fazem parte o conjunto de elementos de metadados DCMES 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)).

O DCMES é um vocabulário de 15 propriedades básicas, também denominado *Simple DC* que é um conjunto de elementos suficientemente amplo e flexível para ser usado nas mais diversas situações. O quadro 1 apresenta uma breve descrição das propriedades básicas do DCMES (DCMI *Usage Board*, 2008).



Quadro1: Propriedades do *Simple DC*

Propriedades	Definição
Title	Um nome dado para o recurso
Creator	Uma entidade responsável pelo conteúdo do recurso em primeira instância.
Subject	Descreve o tema referente ao conteúdo do recurso.
Description	Descrição do conteúdo do recurso.
Publisher	A entidade responsável por disponibilizar o recurso.
Contributor	Uma entidade responsável por quaisquer contribuições ao conteúdo do recurso.
Date	Uma data associada com um evento no ciclo de vida do recurso.
Type	A natureza ou gênero do conteúdo do recurso.
Format	Descreve a manifestação física ou digital do recurso.
Identifier	Uma referência única e não ambígua para o recurso num dado contexto.
Source	Indica uma referência para o recurso do qual deriva o recurso que está a ser descrito.
Language	Indica qual o idioma do conteúdo intelectual do recurso.
Relation	Indica referências para os recursos relacionados como, por exemplo, versão de um trabalho, tradução de um trabalho ou parte de um trabalho.
Coverage	A extensão ou cobertura do conteúdo do recurso.
Rights	Informações sobre direitos associados ao recurso.

Fonte: DCMI, 2008

Propriedade é uma característica ou atributo usado para a descrição do recurso. As subpropriedades permitem uma descrição mais específica. Como exemplo, a propriedade *Date*: para este caso, existem outras propriedades relacionadas com a propriedade *Date* como subpropriedade que permitem um refinamento maior especificando se é uma data de aceite, *copyright* ou submissão (*dateAccepted*, *dateCopyrighted*, *dateSubmitted*).

Classe é definida como um grupo que contém membros que possuem atributos, comportamentos, relacionamentos ou semântica em comum; um tipo de categoria (POWELL et al., 2007). O DCMI *Terms* contém 22 termos que representam classes (ver <http://dublincore.org/documents/dcmi-terms/#H6>).

Os Esquemas de Codificação indicam esquemas já existentes que podem auxiliar na descrição de um dado elemento. Estes esquemas incluem os vocabulários controlados, ou *Vocabulary Encoding Schemes* (VES) e as regras de formatação das notações, ou *Syntax Encoding Schemes* (SES). Como exemplo dos VES para a propriedade *Subject*, há a indicação de esquemas que podem ser utilizados como vocabulários controlados, como é o caso do sistema de classificação *Dewey Decimal Classification* (DDC) e a lista de cabeçalhos de assunto da Biblioteca do Congresso Norte-Americano, *Library of Congress Classification* (LCC). Os SES são regras que indicam como formatar o valor a ser relacionado com uma propriedade ou subpropriedade. Como exemplo, podemos citar o *W3C Date and Time Formats Specification* (W3CDTF) que indica como valores que representam datas ou período de tempo devem ser formatados.



Considerando que a abordagem desta investigação tem foco na questão de metadados, a seguir, abordar-se-á o tema interoperabilidade dos metadados. Como já referido, existem diferentes padrões de metadados que são desenvolvidos para atender a necessidades específicas de diferentes grupos, conforme pode se verificar nos exemplos apresentados acima. Ocorre que um recurso pode ser descrito com diferentes padrões de metadados.

Um caso de interoperabilidade dos metadados é a linguagem RDF. Desenvolvida pelo W3C é um modelo de dados para descrição de recursos da Web (NISO, 2004). Neste modelo, cada recurso (*Resource*) é identificado por um URI específico. Aos recursos são associadas propriedades (atributos ou características) que são identificadas por *property-types* e para cada uma das *property-types* existe um valor correspondente (*values*). Os valores podem ser atômicos (textos, números etc) ou podem ser também outros recursos. Ao conjunto dessas propriedades dá-se o nome de descrição (*Description*) (MILLER, 1998)

2.2 Folksonomias

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. Neste contexto de alterações sociológicas é 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. 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.

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.

Folksonomia é a tradução do termo *folksonomy* que é um 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 etiquetas (*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 que etiqueta o recurso da Web.

Portanto, Folksonomia é o resultado da etiquetagem dos recursos da Web num ambiente social (compartilhado e aberto a outros) pelos próprios usuários da informação visando a sua recuperação. Destacam-se três fatores essenciais: 1) é resultado de uma indexação livre do próprio usuário do recurso; 2) objetiva a recuperação *a posteriori* da informação e 3) é desenvolvida num ambiente aberto que possibilita o compartilhamento e, até, em alguns casos, a sua construção conjunta. O Delicious, por exemplo, favorece a construção conjunta das etiquetas. Aqui, quando um usuário seleciona um URL para *bookmark*, é-lhe logo fornecido um conjunto de etiquetas possíveis já criadas por outros usuários (CATARINO; BAPTISTA, 2007).

Pressupõe-se que a folksonomia permite uma nova forma de organização de recursos da Web e que, naturalmente, poderá também ser adotada pelos Repositórios Institucionais para que seus usuários tenham uma forma de organizar os recursos conforme suas necessidades. Além de servir como uma forma de organização individual, julga-se que as etiquetas atribuídas pelos usuários possam ser aproveitadas pelos gestores dos Repositórios para enriquecer a informação relativa aos recursos neles depositados. As etiquetas podem ser relacionadas com propriedades do DC e outras propriedades complementares, enriquecendo, assim, a organização dos recursos sem comprometer a interoperabilidade dos seus metadados.

2.3 Ontologia

A origem da palavra Ontologia remonta ao grego *Onto* (ser) + *Logos* (razão/palavra) e atualmente existem dois sentidos diferentes para a palavra. Ontologia como campo de estudo da Filosofia e Ontologia como uma tecnologia para cientistas da computação e da informação. Neste artigo abordamos o segundo senso.

Uma Ontologia para as Ciências da Computação e Informação é a especificação de uma conceitualização que é um conjunto de ideias, conceitos, relações ou outras abstrações que compõem o domínio de um modelo ou discurso. Uma Ontologia define um vocabulário representacional para a conceitualização, e especifica restrições no uso deste vocabulário de forma que os fatos sobre um determinado domínio podem ser compartilhados, comunicados e debatidos (GRUBER et al., 2007).



Segundo Heflin (2004) a palavra ontologia tem sido utilizada para descrever artefatos com diferentes níveis de estrutura, desde simples taxonomias (como a hierarquia do Yahoo), esquemas de metadados (tais como Dublin Core) até teorias lógicas. Portanto, deve-se considerar que existem diferentes tipos e finalidades para as ontologias que se refletem na metodologia de construção, ferramentas e linguagens.

Para proceder à construção de ontologias, existem ferramentas que geralmente fornecem interface gráfica aos usuários, não sendo, desta forma, necessário o conhecimento de linguagens específicas. Dentre as ferramentas disponíveis, destacam-se a *Protégé*, *WebODE* e *OntoEdit*, ferramentas desenvolvidas no âmbito de universidades. No entanto existem outros produtos desenvolvidos por empresas privadas, tais como *Altova Semantic Works*®.

Conforme descreve Malucelli (2005), o *Protégé* (ver <http://protege.stanford.edu>) é uma ferramenta desenvolvida pela *Stanford University / Stanford Medical Informatics University* (Califórnia, EUA) que tem as seguintes características: código aberto, *standalone*, arquitetura extensível, editor de ontologia, mais *plugins* com funcionalidades e importa/exporta de/para *Flogic*, *Jess*, *Ontology Interchange Language* (Oil), XML e *Prolog*.

A ferramenta *WebODE* (<http://delicias.dia.fi.upm.es/webODE/>) foi desenvolvida pelo Laboratório de Inteligência Artificial da Universidade Técnica de Madrid e tem como características: aplicação Web; arquitetura extensível; importa/exporta de/para XML, RDF, OIL, DARPA *Agent Markup Language* (DAML) + OIL, CARIN, *Flogic*, *Jess* e *Prolog*; armazenamento das ontologias em bases de dados relacionais; serviços de documentação, avaliação e fusão de ontologias (Malucelli, 2005).

OntoEdit (ver <http://ontoserver.aifb.uni-karlsruhe.de/ontoedit>) é uma ferramenta desenvolvida pela *Institut für Angewandte Informatik und Formale Beschreibungsverfahren* (AIFB) da Universidade de *Karlsruhe*. Malucelli (2005) destaca algumas das suas características: arquitetura extensível baseada em *plugins*, importa/exporta de/para *Flogic*, XML, RDF, DAML + OIL. Está disponível em duas versões: *OntoEdit Free* e *OntoEdit Professional*.

Altova Semantic Works ® (ver http://www.altova.com/products/semanticworks/semantic_web_rdf_owl_editor.html) é um editor RDF e OWL para Web Semântica. Uma aplicação desenvolvida pela empresa Altova, sediada em Massachusetts, e fundada em 1992. Projeta graficamente instâncias de documentos, vocabulários e ontologias em RDF, RDFS ou OWL, com saída em formatos RDF/XML ou *N-Triples*.



Diversas linguagens para a construção e compartilhamento de ontologias têm vindo a ser desenvolvidas. E com o advento da Web as linguagens passaram a ser concebidas para explorar as características da Web. Tais linguagens são denominadas *Web-Based Ontology Languages* ou *Ontology Markup Languages*. A seguir algumas linguagens são descritas.

RDFS (*RDF Schema*), desenvolvido pelo W3C, tem por objetivo a representação de conhecimento por meio da ideia de redes semânticas. É uma linguagem que permite a representação de conceitos, taxonomias de conceitos e relações binárias (LASSILA; SWICK, 1999 apud ALMEIDA; BAX, 2003).

OIL é a linguagem precursora da DAML+OIL e base para uma linguagem para a WS. Combina primitivas de modelação das linguagens baseadas em *frames* com a semântica formal e serviços de inferência da lógica descritiva. Pode representar classificação e taxonomias de conceitos (FENSEL et al., 2001 apud ALMEIDA; BAX, 2003).

DAML + OIL é uma linguagem de marcação semântica para a Web que apresenta extensões a linguagens como o DAML, RDF e RDFS, por meio de primitivas de modelação baseadas em linguagens lógicas (HORROCKS et al., 2001 apud ALMEIDA; BAX, 2003). Antecedeu a OWL.

OWL criada pelo grupo de trabalho *Web-Ontology* da W3C que teve o objetivo de construir uma nova linguagem de marcação de ontologias para a WS. Ela é baseada nas linguagens OIL e DALM-OIL e é uma recomendação da W3C. Três “sublinguagens” são derivadas: OWL *Lite*, OWL *DL* e OWL *Full* (MCGUINNESS; HARMELEN, 2004).

2.4.A Ontologia do Dublin Core

Conforme dito anteriormente, a palavra ontologia tem vindo a ser usada para representar vários artefatos, dentre eles Heflin (2004) refere-se aos esquemas de metadados, citando inclusivamente como exemplo o DC. No âmbito desta investigação, considera-se que a Ontologia DC é a conceitualização explícita dos termos do vocabulário DCMI (*DCMI Terms*) que contém as propriedades, subpropriedades, classes e seus relacionamentos. É tipificada por Sicília (2005) como um tipo de *upper ontology* que são ontologias de alto nível, também conhecidas como ontologias genéricas. As ontologias genéricas descrevem conceitos gerais e são independentes de um domínio específico, ou seja, podem ser aplicadas a vários domínios (e.g. *Cyc Upper Ontology*, ver <http://www.cyc.com/cyc/opencyc/overview>).



McComb (2006) sugere um novo conceito: *minimalist upper ontology*, ou seja, ontologias de alto nível com um número reduzido de conceitos. Cita como exemplo deste tipo de ontologia o DC e o *Friend of a Friend* (FOAF).

Conforme se pode verificar numa pesquisa efetuada no SWOOGLE¹ (*Semantic Web Search*), existem várias ontologias que estão relacionadas com o DC (ver quadro 2).

Quadro 2: Ontologias Dublin Core

URL	Descrição
http://purl.org/dc/elements/1.1/	Esquema RDF para <i>namespaces</i> do <i>Dublin Core Element Set</i> v1.1. Fornece URIs para os elementos DCMES usando linguagem RDF <i>schema</i> para dar suporte às aplicações em RDF.
http://purl.org/dc/terms/	O <i>Dublin Core Terms namespaces</i> fornece URIs para o <i>Dublin Core Element Set qualifier Vocabulary</i> . Os termos do vocabulário são declarados usando linguagem RDF <i>schema</i> para dar suporte às aplicações em RDF.
http://purl.org/dc/dcmitype/	O <i>Dublin Core Types namespaces</i> fornece URIs para os verbetes do <i>DCMI Type Vocabulary</i> . Os verbetes são declarados usando linguagem RDF <i>schema</i> para dar suporte às aplicações RDF
http://dublincore.org/2000/03/13-dces.fr	Descrição em RDF dos elementos de metadados Dublin Core no idioma francês.
http://dublincore.org/2000/03/13-marcel	Declaração em RDF dos termos do MARC (LOC) relacionados ao Dublin Core.
http://dublincore.org/2000/03/13-dcagent	O <i>Dublin Core Agent Core Vocabulary</i> em RDF.
http://daml.umbc.edu/ontologies/webofbelief/1.3/dc.owl	Ontologia DC em OWL DL
http://orlando.drc.com/semanticweb/daml/ontology/DC/dces-ont	Versão oficial da ontologia do DCMES em DAML.
http://pike.kw.nl/files/documents/pietzwart/dev/DublinCore3g.owl	Ontologia do DC em OWL
http://tdwg.napier.ac.uk/ontology/resources/dublincore/terms.owl	<i>Dublin Core Terms namespaces</i> fornece URIs para o <i>Dublin Core Element Set Qualifier Vocabulary</i> . Os termos do vocabulário são declarados usando a linguagem OWL <i>Lite</i> para dar suporte às aplicações OWL e RDF.
http://tdwg.napier.ac.uk/ontology/resources/dublincore/elements.owl	<i>Dublin Core Element Set</i> v1.1 <i>namespaces</i> fornece URIs para o DCMES v1.1. Os verbetes do vocabulário são declarados em OWL <i>Lite</i> .
http://www.cores-eu.net/registry/schema/dcterms.xml	RDF <i>schema</i> que contém a descrição dos elementos DC, elementos de refinamento e esquemas de codificação adicionadas informações requeridas pelo MEG <i>Registry</i> .
http://www.ukoln.ac.uk/metadata/education/regproj/es/dcterms.xml	Esquema que contém descrições dos elementos DC, elementos de refinamento e esquemas de codificação acrescido de informações requeridas pelo MEG <i>Registry</i> .
http://www.ukoln.ac.uk/metadata/education/regproj/es/dc.xml	Esquema que contém descrições dos elementos DC acrescido de informações requeridas pelo MEG <i>Registry</i> .
http://www.ukoln.ac.uk/projects/iemsr/dcap/dc11/2005-03-09	RDF <i>Data</i> que descreve os metadados DCMES v1.1. As descrições são feitas usando o RDF <i>Schema</i> e o IEMSR RDF <i>Vocabulary</i> .

¹ Swoogle é resultado de um projeto da *University of Maryland, Baltimore County* (UMBC), que é um mecanismo de busca que tem o objetivo de analisar e indexar conhecimento codificado em RDF e OWL em documentos da WS publicados na Web

Existem, ainda, outros projetos de Ontologias que foram desenvolvidos a partir do DC ou integrados ao DC. Alguns projetos foram identificados na literatura: *MarcOnt* (KRUK; SYNAK; ZIMMERMANN, 2005), *Integrating Dublin Core Metadata for Cultural Heritage Collections Using Ontologies* (KAKALI et al., 2007), *On the Use of Existing Upper Ontologies as a Metadata Integration Mechanism* (SICILIA, 2005), *HealthCyberMap's Dublin Core Ontology* (BOULOS; ROUDSARI; CARSON, 2002) e *Dublin Core Ontology for SHOE* (LUKE, 2000).

3 METODOLOGIA

Algumas metodologias para construção de ontologias são descritas na literatura, tais como: *Metodologia Inicial de Uschold*; METHONTOLOGY e TOVE. Contudo, na opinião de Guizzardi (2000), estas não apresentam um processo suficientemente estruturado a ponto de suportar a construção de ontologias como uma verdadeira disciplina de engenharia.

Guizzardi (2000) apresenta uma abordagem sistemática (figura 1) para construção de ontologias também descrita por Falbo, Guizzardi e Duarte (2002).

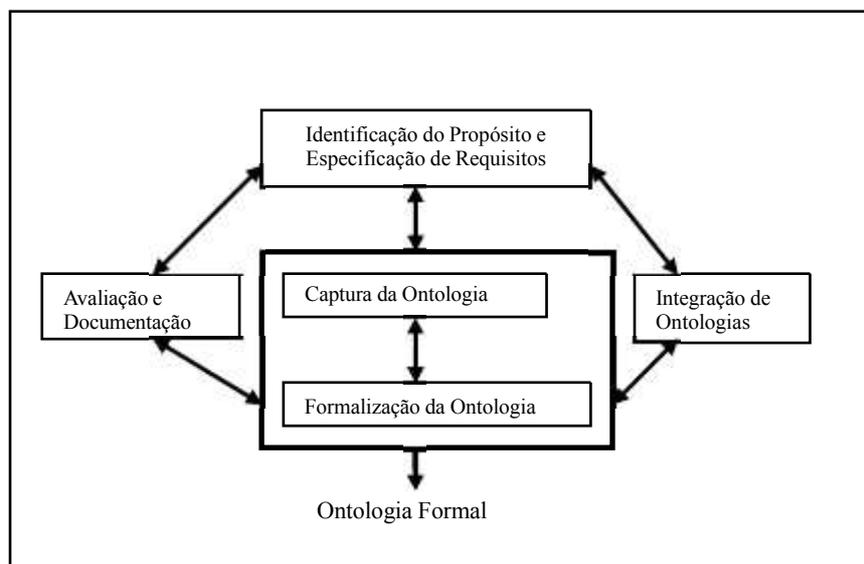


Figura 1: Processo de Desenvolvimento de Ontologias²

Esta abordagem serviu de base para a construção da Ontologia STAP.

² Fonte: adaptado de Falbo, Guizzardi e Duarte (2002)



A Identificação dos Propósitos e Especificação de Requisitos é a etapa inicial do processo. Nesta etapa, definem-se os propósitos e finalidades de uso da ontologia a ser construída, ou seja, a sua competência. Também é útil identificar os usuários potenciais e o contexto que motiva a construção da ontologia. Desta etapa resulta o documento de especificação da ontologia.

A Captura da Ontologia é considerada a etapa mais importante do desenvolvimento de uma ontologia. Tem o objetivo de capturar a conceitualização com base nas competências da ontologia. Nesta etapa devem ser identificados os conceitos e relações relevantes que formam a base da ontologia.

Na etapa de Formalização da Ontologia, a conceitualização capturada é representada explicitamente numa linguagem formal. A linguagem deve ser capaz de representar de forma precisa e não ambígua os elementos que modelam as entidades do domínio. Deve ter a capacidade de escrever axiomas formais que restrinjam a interpretação da estrutura formada por estas entidades (FALBO; GUIZZARDI; DUARTE, 2002).

Existem várias ontologias já desenvolvidas, e as suas conceitualizações previamente estabelecidas podem ser aproveitadas por meio da Integração de Ontologias (GUIZZARDI, 2000). Para Guizzardi (2000) a ontologia deve ser avaliada com o intuito de verificar se satisfaz os requisitos da especificação. A Avaliação de Ontologia faz parte de um processo iterativo, portanto é uma atividade que deve ocorrer em paralelo às etapas de captura e formalização.

A fase de Documentação consiste em documentar todas as etapas, incluindo “propósitos, requisitos e cenários de motivação, as descrições textuais da conceitualização, a ontologia formal e os critérios de projeto adotados”. Esta etapa ocorre, assim como a avaliação, paralelamente a todas as outras atividades (GUIZZARDI, 2000).

4 ONTOLOGIA STAP

A seguir será apresentada a Ontologia STAP³ que foi desenvolvida para a descrição dos termos de metadados para a descrição de recursos etiquetados com base no DC e nos termos oriundos das folksonomias.

Trata-se do vocabulário composto pelo conjunto de termos de metadados, com as declarações de propriedades, subpropriedades, classes e seus relacionamentos e restrições. No

³ A estrutura completa da ontologia encontra-se detalhada na tese de doutorado e no site do projeto STAP (ver CATARINO, 2009a, 2009b).



entanto, não se trata de uma ontologia de domínio e sim uma ontologia genérica (ou de alto nível/*upper ontology*), pois será útil para vários domínios.

A ontologia foi construída em RDF e contém todos os termos para descrição dos recursos etiquetados num repositório. Pretende-se, com esta ontologia, compor uma especificação de todos os termos (propriedades, classes e esquemas de codificação), necessários para descrição de recursos etiquetados em repositórios, para que sejam compartilhados em aplicações RDF. A ontologia aqui proposta é uma extensão de algumas ontologias já existentes acrescidas de alguns novos conceitos, que correspondem às novas propriedades identificadas no projeto de doutorado.

4.1 Identificação de Propósito e Especificação de Requisitos

Esta é a atividade inicial do processo de construção de Ontologia e corresponde ao planejamento da ontologia. Nesta fase serão especificados os propósitos e finalidades de uso da ontologia a ser construída, ou seja, a sua competência, bem como os usuários potenciais e o contexto que motiva a construção da ontologia (GUIZZARDI, 2000).

Esta ontologia é do tipo genérica ou de alto nível pois os conceitos nela especificados podem ser aplicados de forma geral a quaisquer áreas de domínio.

Especificação da Ontologia:

Domínio: Ontologia genérica

Propósito: compartilhar os termos de metadados pra descrição de recursos etiquetados por meio de um único esquema RDF.

Nível de formalidade: semi-formal

Escopo: 138 Termos, sendo: 97 termos DCMI *Metadata Terms*, 12 termos do DCMI *Type Vocabulary*, 02 termos DCMI *Abstract Model*, 15 termos do *Dublin Core Metadata Elements Set* e 12 termos do perfil de aplicação STAP.

4.2 Captura da Ontologia

Esta atividade tem por objetivo capturar a conceitualização que irá compor a ontologia com base nos seus propósitos, identificando os conceitos e as suas relações. A ontologia é composta por todos os termos do DCMI *Terms* (compilados todos os esquemas RDF da DCMI) mais os termos declarados no STAP, conforme determinado no “escopo”:

- <http://purl.org/dc/terms/> (DCMI *Metadata Terms*), esquema RDF para a coleção de todas as propriedades, classes e esquemas de codificação exceto as propriedades do



Dublin Core Metadata Element Set version 1.1 [DCMES], as classes do *DCMI Type Vocabulary [DCMI-TYPE]* e os termos usados no *DCMI Abstract Model*;

- <http://purl.org/dc/dcmitype/> (*DCMI Type Vocabulary*), esquema RDF para a coleção de classes no *DCMI Type Vocabulary [DCMI-TYPE]*;
- <http://purl.org/dc/dcam/> (*DCMI terms used in the DCMI Abstract Model*), esquema RDF para a coleção de termos usados no *DCMI Abstract Model*;
- <http://purl.org/dc/elements/1.1/> (*The Dublin Core Metadata Element Set, Version 1.1*), esquema RDF com os 15 elementos originais do DC.
- <http://odisseia.dsi.uminho.pt/STAP/> (*STAP Terms*). Vocabulário que contém os termos do *Social Tagging Application Profile*.

Os termos do STAP estão disponíveis em <http://odisseia.dsi.uminho.pt/STAP/> e contém a coleção de propriedades declaradas especificamente para a descrição de recursos etiquetados.

4.3 Formalização da Ontologia

Nesta etapa, a conceitualização capturada é representada explicitamente numa linguagem formal. A linguagem deve ser capaz de representar, de forma precisa e não ambígua, os elementos que modelam as entidades do domínio. Deve ter a capacidade de escrever axiomas formais que restrinjam a interpretação da estrutura formada por estas entidades (FALBO; GUIZZARDI; DUARTE, 2002).

Para este projeto, optou-se por utilizar a linguagem *RDF Schema (RDF schema language)* utilizada também pela DCMI (ver <http://dublincore.org/schemas/rdfs/>). A definição de classes, subclasses e seus relacionamentos, além da descrição de outros atributos, seguiu o que é determinado no *DCMI Metadata Terms* e nos esquemas já existentes do DC. O esquema RDF completo pode ser consultado no site do projeto⁴, bem como uma figura com a representação gráfica de parte da ontologia contendo os termos do STAP (CATARINO, 2009b).

Para a formalização da ontologia optou-se pelo uso do *Altova Semantic Works®*, por se tratar de uma aplicação que projeta graficamente instâncias de documentos, vocabulários e ontologias em RDF, RDFS ou OWL, com saída em formatos RDF/XML ou *N-Triples*. Existem outros editores de Ontologias já referidos neste trabalho, contudo decidiu-se pelo uso deste por se tratar de uma aplicação amigável.

⁴ <http://odisseia.dsi.uminho.pt/STAP/Ontologia/ontologia.htm>



A Documentação, conforme define Guizzardi (2000), consiste em documentar todas as etapas, incluindo “propósitos, requisitos e cenários de motivação, as descrições textuais da conceitualização, a ontologia formal e os critérios de projeto adotados”. Toda a informação contida nesta seção compõe a documentação. Complementarmente, também, a descrição dos termos do STAP e o DCMI *Terms* (<http://dublincore.org/documents/dcmi-terms/>). O conjunto de toda esta documentação encontra-se registrado no sítio do projeto: <http://odisseia.dsi.uminho.pt/STAP/ontologia>.

CONSIDERAÇÕES FINAIS

A Ontologia STAP foi desenvolvida como trabalho complementar ao projeto de pesquisa de doutorado em Tecnologias e Sistemas de Informação, desenvolvido numa linha de investigação da Escola de Engenharia da Universidade do Minho, Portugal.

O trabalho teve como foco a WS, em específico a questão dos metadados DC somados às propriedades oriundas das folksonomias, como uma forma de agregar os valores atribuídos pelos usuários aos recursos da Web.

Trata-se de uma ontologia composta pelo vocabulário do conjunto de termos de metadados, com as declarações de propriedades, subpropriedades, seus relacionamentos e restrições. No entanto, não se trata de uma ontologia de domínio e sim uma ontologia genérica (ou de alto nível/*upper ontology*) pois a conceitualização que ela contém será útil para vários domínios.

A ontologia foi construída em RDF e contém todos os termos para descrição dos recursos etiquetados num repositório. Pretende-se, com esta ontologia, compor uma especificação de todas as propriedades necessárias para descrição de recursos etiquetados em repositórios, para que sejam compartilhados em aplicações RDF. A ontologia aqui proposta é uma extensão de algumas ontologias já existentes acrescida de alguns novos conceitos, que correspondem às novas propriedades identificadas na investigação.

Abstract: This paper aims to present the STAP Ontology, developed as part of a doctoral research that has identified new elements of metadata from folksonomies as a contribution to the description of repository resources. The motivation of the research was the Semantic Web, which is the basis for structuring and interoperability of resources on the Web. The World Wide Web Consortium (W3C) proposes standards for the creation of repositories, building of vocabularies and formulation



of rules for data management. In the Semantic Web Architecture, the layer Ontology is related with the construct of vocabularies that can define relationships between different concepts. The ontology developed is composed of the vocabulary of metadata terms of the Dublin Core (DC) and the folksonomy terms. DC is one set of metadata for describing Web resources and is adopted in the repositories for sharing data via the OAI-PMH. Folksonomy is the result of personal free tagging of information and objects (anything with an URL) for one's own retrieval. Folksonomy allows a new form of organization of Web resources and may be adopted by repositories. Tags may be related to properties of DC metadata and other additional properties, enriching thus the organization of resources without compromising the interoperability of metadata. The Methodology used was the systematic approach proposed by Guizzardi (2000): Identification of the Purposes and Requirements Specification, Capture, Formalization, Integration, Evaluation and Documentation Ontology. From this approach we constructed an ontology which aims at sharing metadata terms through a single RDF schema, consisting of 138 DCMI terms and the terms of the STAP.

Key Words: STAP Ontology. Metadata. Dublin Core. Folksonomy.

REFERÊNCIAS

ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília, v.32, n.3, p.7-20, set./dez. 2003. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/17/12>>. Acesso em: 10 de abr. 2005.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. (2001). The Semantic Web. **Scientific American**, may 2001. Disponível em:<http://www.sciam.com/print_version.cfm?article?articleID=00048144-10D2-1C70-84A9809EC588EF21>. Acesso em: 10 de abr. 2005.

BOULOS, M.N.K.; ROUDSARAI, A.V. ; CARSON, E.R.. Towards a Semantic Medical Web: HealthCyberMap's tool for building na RDF metadata base of health information resources based on the Qualified Dublin Core Metadata Set. **Med. Sci. Monit**, v.8, n.7, 2002. Disponível em:<http://www.medscimonit.com/pub/vol_8/no_7/2615.pdf>. Acesso em: 04 de set. 2007.

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. **DataGramZero – 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.

_____. Web Semântica e a qualidade no intercâmbio da informação. In: TOMAÉL, M. I. (Org.). **Fontes de Informação na Internet**. Londrina : EDUEL, 2008. Cap. 2.



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.

DUBLIN CORE METADATA INITIATIVE (2004). **History of the Dublin Core Metadata Initiative**. DCMI, 2004. Disponível em: <<http://dublincore.org/about/history/>>. Acesso em: 02 de Set. 2007.

FALBO, R. A.; GUIZZARDI, G.; DUARTE, K. C. (2002). An Ontological Approach to Comain Engineering. In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND KNOWLEDGE ENGINEERING, Ischia, Italy, 2002. **Proceedings ...** Disponível em: <<http://www.inf.ufes.br/~falbo/download/pub/Seke2002.pdf>>. Acesso em: 17 de set. 2007.

GRACIO, J. C. A.. **Metadados para descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade**. Dissertação (Mestrado em Ciência da Informação, não publicada, Universidade Estadual Paulista. Campus de Marília. Disponível em: <http://www.biblioteca.unesp.br/bibliotecadigital/document/get.php/3680/gracio_jca_dr_mar.pdf>. Acesso em: 02 de set. 2007.

GRUBER, T., et al.. **Ontology Framework Draft Statement**.. Ontolog, 2007. Disponível em: <http://ontolog.cim3.net/file/work/OntologySummit2007/Framework-1c_ontolog%20proposal20070419-2.doc>. Acesso em: 14 de ago.2007.

GUIZZARDI, G.. **Uma abordagem metodológica de desenvolvimento para e com reuso baseada em Ontologias formais de domínio**. Dissertação (Mestrado em Informática). Departamento de Informática, Universidade Federal do Espírito Santo., 2000. Disponível em: <<http://www.loa-cnr.it/Guizzardi/MSc.htm>>. Acesso em: 09 de ago. 2007.

HEFLIN, J. (ed.). OWL Web Ontology Language: use cases and requirements. **W3C Recommendations**, 10 fev. 2004. Disponível em: <<http://www.w3.org/TR/webont-req/#onto-def.>>. Acesso em: 20 de dez. 2008.

KAKALI, C. et al. . Integrating Dublin Core metadata for cultural heritage collections using ontologies. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATION, Singapore, 2007. **Proceedings...** DCMI, 2007. Disponível em: <<http://www.dcmipubs.org/ojs/index.php/pubs/article/view/16/11/>>. Acesso em: 19 de set. 2007.

KRUK, S. R.; SYNAK, M.; ZIMMERMANN, K.. *MarcOnt – Integration Ontology for Bibliographic Description Formats*. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATION, Madri, 2005. **Proceedings...** DCMI, 2005. Disponível em: <<http://www.marcont.org/marcont/pdf/DC2005skmskz.pdf>>. Acesso em: 19 de set. 2007.

LUKE, S. **Dublin Core Ontology**. 2000. Disponível em: <<http://www.cs.umd.edu/projects/plus/SHOE/onts/dublin.html>>. Acesso em: 25 de ago. 2007.

MALUCELLI, A. **Ontologias**. 2005. Disponível em: <<http://paginas.fe.up.pt/~eol/TNE/APONT/Ontologia.pdf>>. Acesso em: 15 de fev. 2008.



- McCOMB, D.. A minimalist upper ontology. In: _____ **Semantic Technology**. 2006. Disponível em: <http://semantic-conference.blogspot.com/semtech06/2006/02/a_minimalist_up.html>. Acesso em: 24 de mar. 2009.
- MCGUINNESS, D. L.; HAMELEN, F. (eds.) (2004). OWL Web Ontology Language Overview. **W3C Recommendation**, 10 fev. 2004. Disponível em: <<http://www.w3.org/TR/owl-features/>>. Acesso em: 15 de fev. 2008.
- 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.
- NATIONAL INFORMATION STANDARDS ORGANIZATION. **Understanding Metadata**. Bethesda: NISO Press, 2004. Disponível em: <<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>>. Acesso em: 05 de set. 2007.
- 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.
- 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.
- SICILIA, Miguel-Angel (2005). On the use of existing upper ontologies as a metadata integration mechanism. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATION, Madri, 2005. **Proceedings...** DCMI, 2005. Disponível em: <<http://www.nl.go.kr/dcpapers/pdf/2005/paper23.pdf>>. Acesso em: 19 de set. 2007.
- WAL, T. V. **Folksonomy definition and wikipedia**. Disponível em: <<http://www.vanderwal.net/random/entrysel.php?blog=1750>>. Acesso em: 22 nov. 2006.
- WOODLEY, M. S.; CLEMENT, G.; WINN, P.. **DCMI Glossary**. DCMI, 2005. Disponível em: <<http://dublincore.org/documents/usageguide/glossary.shtml>>. Acesso em: 25 de ago. 2007.
- WORD WIDE WEB CONSORTIUM [W3C] (2001). **W3C Semantic Web Activity**. Disponível em: <<http://www.w3c.org/2001/SW/>>. Acesso em: 12 fev. 2008.