

Representação iterativa e folksonomia assistida para repositórios digitais¹

José Eduardo Santarém Segundo*

Silvana Aparecida Borsetti Gregorio Vidotti**

Resumo A recuperação da informação tem sido muito discutida dentro da Ciência da Informação ultimamente. A busca por informação de qualidade e compatível com a necessidade do usuário tornou-se objeto constante de pesquisa. A utilização da Internet como fonte de disseminação do conhecimento indicou novos modelos de armazenamento de informações, como os repositórios digitais, que têm sido utilizados em ambientes acadêmicos e de pesquisa como principal forma de autoarquivar e disseminar informação, porém com uma estrutura de informação que comporta melhor descrição dos recursos e conseqüentemente uma melhor recuperação da informação. Desta forma o objetivo deste trabalho é melhorar o processo de recuperação da informação, apresentando uma proposta de modelo estrutural no contexto da *web* semântica, abordando o uso de recursos da *web* 2.0 e *web* 3.0 em repositórios digitais, que permita recuperação semântica da informação, por meio da construção de uma camada de informação chamada Representação Iterativa. O presente estudo caracteriza-se como uma pesquisa descritiva e analítica, com base em análise documental, dividida em duas partes: a primeira, caracterizada pela observação direta não participativa de ferramentas que implementam repositórios digitais, assim como de repositórios digitais já instanciados, e a segunda, com característica exploratória, em que sugere um modelo inovador para repositórios, com a utilização de estruturas de representação do conhecimento e participação do usuário na construção de um vocabulário próprio de domínio. Através do modelo sugerido e proposto — Representação Iterativa — será possível adequar os repositórios digitais para que utilizem Folksonomia e também vocabulário controlado de domínio, de forma a gerar uma camada de informação iterativa, que possibilite retroalimentação da informação, além de recuperação semântica da informação, através do modelo estrutural desenhado para repositórios. O modelo sugerido resultou na efetivação da tese de que por meio da Representação Iterativa é possível estabelecer um processo de recuperação semântica da informação em repositórios digitais.

Palavras-chave repositórios digitais; representação iterativa; folksonomia; folksonomia assistida; *web* semântica; recuperação da informação; ontologia

¹ Este artigo é derivado da tese *Representação Iterativa: um modelo para repositórios digitais*, defendida por José Eduardo Santarém Segundo no Programa de Pós-Graduação em Ciência da Informação da Universidade do Estado de São Paulo (Unesp) sob orientação da Profa. Dra. Silvana Aparecida Borsetti Gregorio Vidotti.

* Doutor em Ciência da Informação pela Unesp e professor do Departamento de Educação, Informação e Comunicação da Universidade de São Paulo (USP). Endereço postal: USP, Departamento de Educação, Informação e Comunicação, Av. Bandeirantes, 3900, Monte Alegre, Ribeirão Preto, São Paulo, CEP. 14040-901. Tel. (16) 3602-3670 e e-mail santaremsegundo@gmail.com.

** Doutora em Educação pela Unesp e professora do Programa de Pós-Graduação em Ciência da Informação da Unesp. Endereço postal: Unesp, Departamento de Ciência da Informação, Avenida Hygino Muzzi Filho, 737, Campus Universitário, Marília, São Paulo, CEP. 17525-900, Caixa Postal 181. Tel. (14) 3402-1357 e e-mail vidotti@marilia.unesp.br.

Iterative representation and folksonomy assisted for digital repositories

Abstract Information retrieval has been much discussed within Information Science lately. The search for quality information compatible with the users' needs became the object of constant research. Using the Internet as a source of dissemination of knowledge has suggested new models of information storage, such as digital repositories, which have been used in academic research as the main form of autoarchiving and disseminating information, but with an information structure that suggests better descriptions of resources and hence better retrieval. Thus the objective is to improve the process of information retrieval, presenting a proposal for a structural model in the context of the semantic web, addressing the use of web 2.0 and web 3.0 in digital repositories, enabling semantic retrieval of information through building a data layer called Iterative Representation. The present study is characterized as descriptive and analytical, based on document analysis, divided into two parts: the first, characterized by direct observation of non-participatory tools that implement digital repositories, as well as digital repositories already instantiated, and the second with scanning feature, which suggests an innovative model for repositories, with the use of structures of knowledge representation and user participation in building a vocabulary domain. The model suggested and proposed — Iterative Representation — will allow to tailor the digital repositories using Folksonomy and also controlled vocabulary of the field in order to generate a data layer iterative, which allows feedback information, and semantic retrieval of information, through the structural model designed for repositories. The suggested model resulted in the formulation of the thesis that through Iterative Representation it is possible to establish a process of semantic retrieval of information in digital repositories.

Keywords digital repositories; iterative representation; folksonomy; folksonomy assisted; semantic web; information retrieval; ontology

Introdução

É inegável que o mundo tem passado por transformações nos últimos anos, principalmente as ocasionadas pelo uso das tecnologias. Dentro do contexto da Ciência da Informação, a Internet tem atuado diretamente como elemento facilitador no processo de disseminação da informação e do conhecimento, incluindo o conhecimento científico, que deixou de estar disponível apenas nas revistas científicas e livros impressos e passou a utilizar a estrutura tecnológica da Internet para ser disseminado por meio das revistas eletrônicas digitais e dos repositórios digitais.

O cenário atual, baseado no desenvolvimento das tecnologias que englobam as funcionalidades denominadas *web 2.0*, tem intensificado a relação usuário *versus* Internet, visto que esta permite a interatividade entre ambos e estimula o uso constante da rede.

Os conceitos a respeito das funcionalidades da *web 2.0* favorece o uso colaborativo e torna a *web* uma verdadeira plataforma para publicação e consumo de informação. Além das funcionalidades que buscam uma inteligência coletiva e um novo patamar de interação, os conceitos também foram se transformando em realidade no que diz respeito ao uso e a aplicação da *web* semântica.

Os conceitos da *web* semântica, cunhada por Tim Berners-Lee e homologada pelo W3C, têm sido objeto de estudo das Ciências da Informação e da Computação e despertado interesse da comunidade, de um modo geral. A *web* 3.0, como tem sido chamada a *web* semântica, consiste num conjunto de padrões destinados a fazer com que o material publicado na *web* possa ser recuperado de forma semântica, agrupando informações com o mesmo significado, independente de sua estrutura sintática. Permite associação de termos que são facilmente relacionados na estrutura cerebral do ser humano, porém são de difícil relacionamento em sistemas de informação.

Através de recursos tecnológicos, a *web* 3.0 tem efetivado uma mudança de paradigma em relação ao armazenamento e à recuperação de informações na *web*.

Web 2.0: conceitos e funcionalidades

O termo *web* 2.0 surgiu durante uma conferência promovida pelas empresas de mídia Media-Live e O'Reilly Media, realizada em São Francisco (EUA), em 2004. Nesta conferência discutiu-se a ideia de que a *web* deveria ser mais dinâmica e interativa, de modo que os internautas pudessem colaborar com seus conteúdos.

Neste evento, em palestra de abertura, John Battle e O'Reilly fizeram uma lista preliminar de princípios em que o primeiro era “A *web* como plataforma”. Neste novo modelo, o usuário passa a ser o centro das atenções, tornando-se participante ativo sobre a criação e seleção do conteúdo postado em um determinado site, através de plataformas abertas. O cerne da *web* 2.0 está na intensa participação do usuário e na sua interatividade com os serviços on-line, muito mais voltada para a coletividade do que propriamente para o tecnológico, permitindo a construção da informação de maneira coletiva.

De acordo com O'Reilly (2005),

não há como delimitar fronteiras para a *web* 2.0, pois trata-se de princípios e práticas para que diversos sites sigam. Um dos princípios fundamentais é a *web* como plataforma, ou seja, o usuário poder realizar atividades online que antes só eram possíveis com programas rodando em seu computador. (O'REILLY, 2005, p. 1)

Os exemplos de utilização dos conceitos de *web* 2.0 são: os serviços da Google, como Orkut, Gmail, Blogger, além de outros serviços, como Del.icio.us, um gerenciador de bookmark e o Flickr.

Alguns elementos evidenciaram a consolidação da *web* 2.0 como plataforma de interação, dentre eles: interfaces ricas, blogs, *wikis*, inteligência coletiva, *mashups* e principalmente folksonomia, descrita a seguir.

Folksonomia

Folksonomia é a tradução do termo criado por Thomas Vander Wal, a partir da junção das palavras *folk* (povo) com *taxonomy* (taxonomia). Wal (2006, p.1) define Folksonomia como “resultado de atribuição livre e pessoal de *tags* (etiquetas) a informações ou objetos (recursos na *web*), visando a sua recuperação”.

Entre os recursos da *web 2.0*, a Folksonomia é um dos que mais caracterizam essa condição, de construção coletiva de inteligência informacional. A Folksonomia é considerada elemento fundamental no desenvolvimento desta pesquisa, funcionando como recurso primordial na construção do modelo Representação Iterativa.

O propósito principal da Folksonomia neste contexto é permitir que usuários comuns criem *labels/tags* que possam descrever ou apontar para o conteúdo que estão inserindo durante o autoarquivamento de documentos digitais em repositórios digitais, de modo que os recursos possam ser recuperados posteriormente pelo próprio usuário ou ainda por outros usuários que procurem informações no ambiente digital.

Entende-se nesta pesquisa que a Folksonomia caracteriza-se como uma forma de inserir e relacionar recursos por meio da descrição dos mesmos pelas palavras-chave, de forma aberta, que tem como principal objetivo facilitar o processo de gerenciamento e recuperação das informações em ambientes digitais.

As etiquetas (*tags*) podem ser definidas ainda como palavras-chave, categorias ou metadados, e podem ser classificados como qualquer palavra que define uma relação entre o recurso on-line e um conceito na mente do usuário (GUY; TONKIN, 2006).

O fato de a Folksonomia promover a participação do usuário de forma livre permite que a criação das *tags* receba o nome de vocabulário descontrolado, em uma alusão aos vocabulários controlados, que são um recurso disponível para alinhar indexação de informação dentro de um conjunto de palavras fixas que representam um determinado domínio de informação.

Aquino (2007) faz essa abordagem:

Poderíamos dizer que a folksonomia é um tipo de vocabulário descontrolado. Isso não quer dizer que o esquema seja uma desordem total [...] Na verdade, trata-se de um mecanismo de representação, organização e recuperação de informações que não é feito por especialistas anônimos, o que muitas vezes pode limitar a busca por não trazer determinadas palavras-chave, mas sim um modo onde os próprios indivíduos que buscam informação na rede ficam livres para representá-la, organizá-la e recuperá-la, realizando estas ações com base no senso comum. (AQUINO, 2007, p. 10)

A Folksonomia mudou o paradigma em relação à recuperação da informação em ambientes *web*, tanto que é comum ver sites apresentando buscas baseadas em palavras-chave que foram inseridas pelo próprio usuário dentro do ambiente. Portanto trata-se de um recurso rico, que

contribui de forma acentuada para o fortalecimento e solidificação da Internet como plataforma para construção de informação coletiva.

Ontologia

Estudos baseados em ontologias têm surgido constantemente nas pesquisas relacionadas à Ciência da Informação, permeando várias disciplinas e áreas dentro da ciência.

Várias são as definições encontradas e que podem ser aplicadas ao termo.

Para Guarino (1998, p.7), ontologia é “uma maneira de se conceituar de forma explícita e formal os conceitos e restrições relacionados a um domínio de interesse”. Numa visão mais tecnológica, o termo refere-se a um artefato de engenharia que, em uma visão simplista, pode ser descrito como uma hierarquia de conceitos relacionados entre si através de uma classificação de parentesco (hipernímia e hipônimo), também chamada de taxonomia.

A definição de Jacob (2003) aproxima-se muito do conceito de ontologia que mais se aplica à Ciência da Informação quando no contexto da recuperação semântica de informações:

Ontologias são categorias de coisas que existem ou podem existir em um determinado domínio particular, produzindo um catálogo onde existem as relações entre os tipos e até os subtipos do domínio, provendo um entendimento comum e compartilhado do conhecimento de um domínio que pode ser comunicado entre pessoas e programas de aplicação. (JACOB, 2003, p.19)

Ontologias fornecem o conhecimento estruturado e uma infra-estrutura para integrar bases de conhecimentos, independentes da implementação. Constituem uma ferramenta poderosa para suportar a especificação e a implementação de sistemas computacionais de qualquer complexidade.

Gruber (1993, p.2) define ontologias como uma “especificação explícita de uma conceituação”. Uma conceituação pode ser representada como um conjunto de objetos, restrições, relacionamentos e entidades que se assumem necessárias em alguma área de aplicação.

Apesar de serem aplicadas em diversas áreas dentro da Ciência da Informação, as ontologias têm um papel especialmente importante para a *web* semântica. De acordo com Berners-Lee, Hendler e Lassila (2001), para o funcionamento da *web* semântica, computadores devem ter acesso a coleções estruturadas de informação e conjuntos de regras que possam usar para conduzir raciocínio automático, sendo esse o principal desafio da área.

Ressalte-se porém que, apesar dos diferentes vocabulários e vertentes, praticamente todas as definições citam a construção de uma estrutura de relação entre conceitos dentro de um domínio. A abordagem que se faz em relação à Ontologia é que essa estrutura de informação está inserida dentro de um contexto de Estruturas de Representação do Conhecimento.

Nesta pesquisa a definição de uma ontologia é primordial para que se consolide a idéia de que é possível estruturar um ambiente digital por meio de um conjunto regulado de termos. Dessa forma a ontologia servirá como elemento que caracterizará o depósito de documentos em repositórios digitais utilizando-se de palavras-chave pré-definidas.

Repositórios digitais de informação científica

A dificuldade e a necessidade de acesso ao material já produzido, juntamente com a introdução da tecnologia digital, estabeleceram uma nova ordem na edição e publicação da comunicação científica.

O surgimento das publicações científicas em meio eletrônico e a aproximação e interação da comunidade científica, pela *web*, em novas escalas de tempo e espaço, além da criação de um novo conceito de publicação — o *Open Access Initiative* (OAI) — que tem como premissa promover o acesso livre e irrestrito à literatura científica e acadêmica, têm efetivamente alterado a forma e a maneira de explorar o material científico produzido.

O OAI estabeleceu novos critérios em relação à maneira com que as instituições e os pesquisadores lidam com o material produzido em seu âmbito, porém o estabelecimento desta filosofia está amparado por estruturas tecnológicas que permitem a publicação e conseqüente disseminação da informação. Essas estruturas tecnológicas são encabeçadas, principalmente, pelas ferramentas para criação de repositórios digitais e revistas eletrônicas.

Repositórios são conjuntos de documentos coletados, organizados e disponibilizados eletronicamente. No contexto específico dos repositórios, os documentos adquirem novas configurações e são denominados objetos digitais ou estrutura de dados digitalmente codificados, composta pelo conteúdo de informação, metadados e identificador (BEKAERT; VAN DE SOMPEL, 2006).

Nos últimos anos, os repositórios institucionais têm sido alvo de grande atenção por parte de universidades, reassumindo o controle acadêmico sobre a publicação, aumentando a competição e reduzindo o monopólio das revistas científicas das editoras comerciais.

As grandes universidades brasileiras, em especial as públicas, que contam com programas de pós-graduação, cumprindo solicitação da Capes, já têm ou procuram iniciativas que buscam publicar pelo menos as dissertações de mestrado e as teses de doutorado de maneira eletrônica, tornando de conhecimento público os trabalhos desenvolvidos.

Para a implementação técnica de repositórios, são vários os programas de *software* disponíveis, tanto em iniciativas de *software* livre, *open source* e até alguns que sugerem a aquisição de uma licença de uso. Entre os principais, atualmente encontram-se: *Dspace*, *GNU E-prints*, *OPUS*, *Open Repository*, *DiVA* e *Fedora*.

Independente da ferramenta adotada, é notório que grande parte do *software* ainda não oferece relacionamento de termos e busca semântica em seu conteúdo. Dessa forma esta pesquisa utiliza os repositórios digitais como objeto principal de estudo na construção de um modelo que possa efetivar a construção e aplicação dos conceitos da *web* semântica neste ambiente, efetivando o que chamamos de Representação Iterativa.

Representação iterativa

A Representação Iterativa caracteriza-se como um conjunto de técnicas formando um modelo de estrutura para adaptação dos repositórios digitais para que estes possam efetivar o relacionamento de termos e a recuperação semântica da informação.

Dentro deste contexto a Representação Iterativa deve transformar um repositório digital científico em uma ferramenta apta a descrever, armazenar e recuperar informação, permitindo a recuperação semântica e a construção coletiva de uma estrutura relacional semântica de informações por meio de Folksonomia Assistida; e as técnicas utilizadas no desenvolvimento da estrutura sugerida (SANTAREM SEGUNDO, 2010).

Importante ressaltar que, diferente de um simples registro de banco de dados ou então de livre armazenamento de um documento, o processo de autoarquivamento de objetos digitais em um repositório digital científico é um pouco mais complexo e exige dedicação do usuário, que deverá descrever a informação de maneira coesa ao autoarquivar seu objeto digital.

A atividade de inserir informações em um repositório digital compreende o processo de inicialmente descrever o conjunto de informações que representa os metadados do objeto a ser inserido e, na sequência, realizar o envio do arquivo principal e também dos arquivos complementares, se houver, para que todo o conjunto de informações seja armazenado no repositório.

Representação Iterativa, estruturando o modelo

Observa-se que a pesquisa sugere a construção de um modelo estrutural para repositórios digitais científicos, de forma que esses ambientes possam agregar funcionalidades que atuem no sentido de garantir ao usuário uma melhor interface de comunicação com o sistema e ainda evoluir no processo de recuperação da informação, possibilitando a apresentação de resultados baseados em relação semântica, baseada em associação de conteúdos, e não apenas em comparação sintática, como é realizado atualmente.

O modelo — Representação Iterativa — parte do princípio de que o usuário deverá ter uma interface diferente para inserção de dados no repositório digital. A princípio, a única alteração em relação à interface padrão de descrição do recurso será no momento de informar as palavras-chave, visto que estes campos deverão vir com uma informação de que, além de configurar como palavras-chave, os dados descritos ali serão também utilizados como tags (SANTAREM SEGUNDO, 2010).

O fato de caracterizar o uso de tags já cria neste ambiente a ideia de que o ambiente tratará as palavras-chave como parte da concepção de Folksonomia, e, portanto, deverá implementar recursos que permitam a recuperação da informação em novos formatos, como uma nuvem de *tags*, por exemplo.

No momento em que o usuário iniciar o processo de descrição da *tag* deverá ocorrer uma intervenção do sistema, de forma que se caracterize um processo que denominamos Folksonomia Assistida.

Folksonomia Assistida, enriquecendo a descrição do recurso

Folksonomia Assistida é um processo de apoio ao usuário, no momento de definir os termos mais adequados para as *tags* que referenciarão seu trabalho depositado em um repositório digital. O processo é composto por duas partes principais (SANTAREM SEGUNDO, 2010).

A primeira parte implica que, para a implementação da Folksonomia Assistida, deverá ser alterada a interface de comunicação do usuário com o repositório, ou então desenvolvida uma nova interface, para a inserção de informações no campo palavra-chave, utilizado como referência para a inserção de conteúdo para as *tags*.

Nesse primeiro passo, deve-se apresentar ao usuário, no momento da digitação da *tag*, de forma sistemática, um conjunto de informações já previamente inseridas no sistema, como uma sugestão de *tags*. A busca de informação para fazer a sugestão é baseada em busca sintática.

Essa maneira de inserção de dados já é realizada no site Delicious, e tem como característica a apresentação de sugestões conforme o usuário vai digitando o termo a ser registrado como *tag*. Tecnicamente, essa funcionalidade da *web 2.0*, de interação com o usuário de forma rápida e sem recarregamento da página, são as já citadas interfaces ricas. O recurso de “sugestão”, utilizado pelo Delicious e também na concepção da Folksonomia Assistida, foi inicialmente apresentado pelo Google em sua ferramenta de busca, porém neste contexto tem sido adaptado para facilitar o processo de descrição do recurso pelo usuário.

O conjunto de informações que deverá ser apresentado ao usuário no momento que este estiver digitando será baseado nas *tags* já inseridas no sistema e também nos termos que fazem parte de uma estrutura de representação do conhecimento das áreas de especialidades, que deverá estar associada ao repositório como parte do modelo estrutural proposto.

Assim que o usuário descrever as *tags*, aceitando ou não as sugestões, o sistema receberá a informação e dará início a um segundo passo para a concepção da Folksonomia Assistida.

No segundo passo, o repositório deverá receber os termos enumerados pelo usuário e proceder à pesquisa de relacionamento da informação dada pelo usuário em relação ao conjunto de informações internas que a ferramenta dispõe.

O processo de relacionamento em questão é justamente uma busca de relações dentro de uma estrutura de representação do conhecimento das áreas de especialidades, visto que esta pode ser caracterizada por um tesouro ou ainda por uma ontologia, que são instrumentos que permitem uma busca hierárquica horizontal, mas, principalmente, uma busca hierárquica vertical de relacionamento de termos.

Neste modelo, sugere-se o uso de uma estrutura de representação do conhecimento das áreas de especialidades, em qualquer um de seus instrumentos, porém nesta pesquisa utilizar-se-á uma ontologia, descrita em *Web Ontology Language (OWL)*, principalmente por ser uma linguagem que vem sendo aprimorada constantemente, e conta com indicação de uso pelo W3C.

A busca por termos relacionados em uma ontologia escrita com a linguagem OWL deve ser realizada com a linguagem Sparql, que tem como princípio justamente recuperar informações relacionadas em uma linguagem para descrição de ontologias.

Esse segundo passo da Folksonomia Assistida, além de recuperar termos relacionados de forma semântica em uma estrutura de representação do conhecimento das áreas de especialidades, deverá também buscar informações no conjunto de *tags* já inseridas no sistema, principalmente em seus relacionamentos horizontais. A busca por termos na estrutura de representação do conhecimento deverá acontecer em níveis pré-estabelecidos pelo administrador do ambiente, e a busca por relacionamentos horizontais no conjunto de *tags* já descritas também poderá ser mediada pelo administrador, que deverá informar a quantidade de termos oferecidos para cada termo digitado pelo usuário. Esses conceitos poderão ser previamente parametrizados e adaptados conforme o repositório for sendo ampliado com novos depósitos.

A seguir, após essa busca interna por relacionamentos em relação ao termo descrito pelo usuário, o sistema apresentará novamente ao usuário um conjunto de termos que poderão ser aceitos de forma total ou parcial, ou ainda descartados pelo usuário, como sugestão final de *tags* para o recurso a ser inserido. Em todo esse processo, cabe ao usuário decidir as *tags* que melhor representem seu recurso digital dentro do domínio do repositório digital científico em que está sendo realizado o depósito.

A utilização de termos de uma estrutura de representação do conhecimento e também de *tags* já inseridas no sistema não tem o objetivo de engessar a criatividade do usuário, nem tampouco de descaracterizar o termo Folksonomia, pois o sistema permite claramente que o usuário decida livremente os termos que deverão ser utilizados como *tags*. A Folksonomia Assistida tem como principal característica oferecer ao usuário um conjunto de termos que já estão sendo empregados no sistema, de forma que ele possa usar a base de conhecimento do próprio repositório para qualificar a descrição de seu recurso.

Santarem Segundo (2010, p. 276) afirma que “a Folksonomia Assistida prima pela consistência das *tags*, de forma que o usuário do sistema evite abreviações, plurais/singulares ou ainda palavras que possam dificultar a recuperação da informação, posteriormente”.

O processo de gravação das informações é efetivado quando o usuário definitivamente escolhe os termos que gostaria de usar como *tags* e grava as informações.

Ao decretar definitivamente o conjunto de dados que descrevem o objeto digital, o sistema receberá e armazenará no banco de dados o conjunto de informações que o usuário escolheu para descrever o objeto digital.

O processo denominado Folksonomia Assistida efetiva a necessidade de fazer com que as *tags* tenham um grau maior de significado em relação ao objeto depositado, principalmente dentro do contexto em que está sendo utilizada.

Guy e Tonkin (2006, p. 1) afirmam que,

Começamos por olhar para a questão das "tags malfeitas", um problema para o qual os críticos da Folksonomia fazem questão de aludir, e perguntar à comunidade que pesquisa sobre Folksonomia se há maneiras de compensar esses problemas [...] Provavelmente, a grande falha dos sistemas de folksonomia atuais, é que os termos de marcação utilizados nesses sistemas são

imprecisos. Os usuários dos sistemas que utilizam Folksonomia inserem livremente as tags, o que significa que as tags são muitas vezes ambíguas, excessivamente personalizadas e inexatas.

O uso da Folksonomia Assistida busca justamente melhorar a eficiência do uso de tags, permitindo ao usuário uma descrição livre para os objetos digitais a que deposita, porém de forma que possa se amparar no próprio conhecimento já disponível no ambiente em que está utilizando.

Iteratividade, a retroalimentação da informação

A implementação do processo de Folksonomia Assistida será a base para a consolidação da Representação Iterativa, que deverá ser retroalimentada, sempre baseada no contexto de uma estrutura de representação do conhecimento, através de uma ontologia, taxonomia ou de um tesouro, que consiste em definir os limites de um domínio do conhecimento.

Santarem Segundo (2010) afirma que

É possível visualizar a Representação Iterativa de forma conceitual. Dada uma visão geral, o modelo é iniciado no usuário, através da extração de informações de um documento, e amparados por estruturas de representação do conhecimento, além de informações já inseridas no sistema por outros usuários, que fazem a descrição do objeto digital para efetivar um depósito em um repositório digital científico. As informações cadastradas são utilizadas para amparar o depósito de outros usuários, além de possibilitar a um usuário administrador que, sob observação do conjunto de informações depositadas, faça alterações na estrutura de representação do conhecimento utilizada. (SANTAREM SEGUNDO, 2010, p. 187)

Essa visão geral é detalhada na figura 1 que apresenta os passos para que realmente aconteça o uso completo da Representação Iterativa.

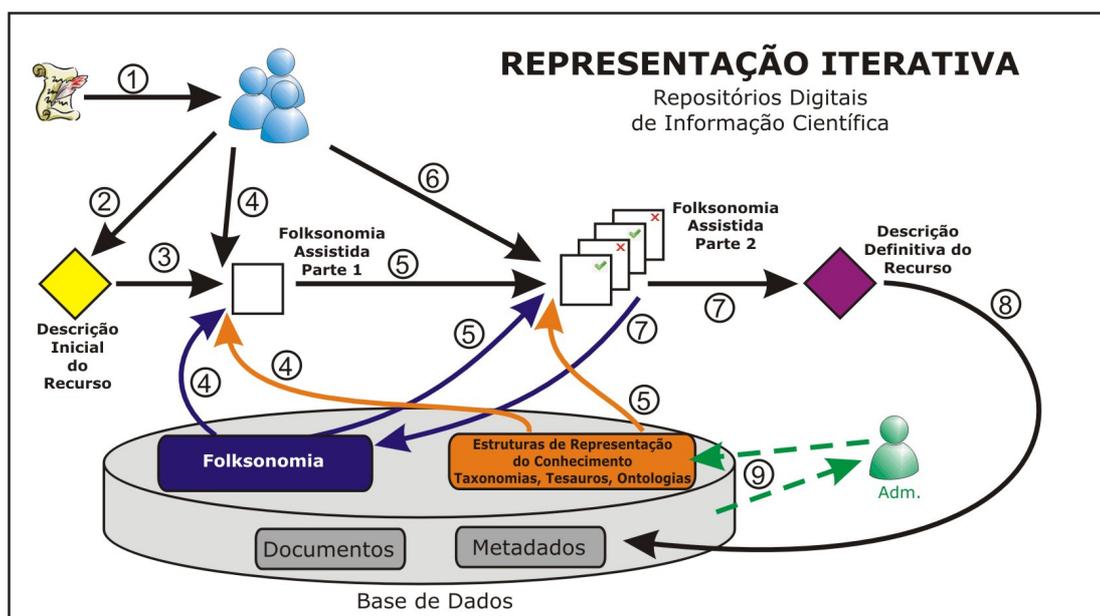


Figura 1: Representação Iterativa – Visão Detalhada

Fonte: Santarem Segundo (2010)

A construção do modelo nomeado Representação Iterativa, apresentado de forma detalhada na figura 1, deverá ser construído conforme os seguintes passos:

1. Os usuários fazem uma leitura e verificação do documento a ser depositado e extraem os metadados necessários que descrevam o máximo possível o objeto, para que seja realizado o depósito.
2. O usuário através de formulário disponível no ambiente inicia o processo de descrição do recurso. Esse passo é chamado de descrição inicial do recurso porque é neste momento que o usuário deverá inserir todos os metadados relativos ao objeto, com exceção da *tag* assunto.
3. Com as informações dos metadados já alimentadas, o sistema encaminha o usuário para fazer a descrição da *tag* assunto, que é a informação que representará de forma mais significativa o recurso dentro da Representação Iterativa.
4. Esse passo representa o início da Folksonomia Assistida. Nesse momento, o sistema deverá colaborar na descrição da *tag*, utilizando uma estrutura de sugestão, semelhante ao da pesquisa do Google, sendo que as informações sugeridas serão os próprios termos já inseridos anteriormente por usuários (Folksonomia – representação livre), além dos termos que fazem parte da estrutura de representação do conhecimento (taxonomias, ontologias ou tesouros) que estará associada ao repositório.
5. Nesse passo acontece o segundo momento da Folksonomia Assistida. Após a descrição da *tag* assunto, o ambiente reconhece essas informações e busca relacionamentos e associações dentro do instrumento de estrutura de representação do

conhecimento utilizado, agrega termos, e em seguida faz o mesmo dentro do conjunto de *tags* já definidas por outros usuários (representação livre), busca associações e, na sequência, também relaciona termos. Esse conjunto de termos que foram selecionados são devolvidos para o usuário.

6. O usuário volta a atuar novamente assim que recebe o conjunto de termos do ambiente. Neste momento ele deve completar o processo de Folksonomia Assistida escolhendo de forma definitiva os termos que serão utilizados na *tag* assunto. Essa decisão implica estabelecer relacionamentos entre as *tags*, e portanto, criar a relação semântica de termos que irá caracterizar a recuperação semântica posterior. Portanto, esse momento é muito importante para a consolidação da Representação Iterativa, porque estabelece os termos e relacionamentos que caracterizam o recurso.
7. Esse passo apenas apresenta a confirmação da descrição completa do recurso, visto que o usuário já descreveu inicialmente os metadados e em seguida, com auxílio da Folksonomia Assistida, escolheu os termos que compõe a *tag* assunto. É nesse momento que a Folksonomia (representação livre) será alimentada efetivamente com o novo conjunto de termos e relacionamentos que o usuário efetivou e dessa forma reorganizada, atualizando o peso dos termos e relacionamentos de acordo com os novos elementos que foram inseridos. Cada vez que esse passo é efetivado em um novo depósito acontece um enriquecimento e fortalecimento do conjunto de termos e relações existentes, e as informações que foram inseridas passam a ficar disponíveis para serem utilizadas por novos usuários em novos depósitos.
8. Nesse passo o conjunto completo de metadados assim como os objetos digitais são armazenados na base de dados.
9. A cada período de tempo, o processo deverá ser avaliado por um administrador de sistema que poderá também retroalimentar a estrutura de representação do conhecimento das áreas de especialidades, dando uma nova visão a respeito dos limites estabelecidos ao domínio do conhecimento. Esse processo cria uma nova perspectiva na Ciência da Informação, que é a avaliação e reconstrução da estrutura de representação do conhecimento, baseado na construção da informação, por usuários de um ambiente digital.

É importante ressaltar que o administrador deve ser um profissional ou equipe multidisciplinar responsável pela catalogação do ambiente informacional e pela manutenção das estruturas de representação do conhecimento (bibliotecário, arquivista e/ou cientista da informação).

O processo de iteratividade é estabelecido de forma que fica a cargo de um usuário administrador a retroalimentação da estrutura de representação do conhecimento, e, como função sistemática e automática dos usuários, as retroalimentações da Folksonomia.

O processo de iteratividade resulta na reconstrução do conhecimento, de forma coletiva e moderada, permitindo o enriquecimento e amadurecimento da estrutura de representação do conhecimento para o domínio em que o repositório digital científico está inserido.

A arquitetura proposta neste trabalho parte do princípio da iteratividade, que é o processo em que ocorre a realimentação constante do sistema em busca da melhor qualidade do conjunto de informações.

O princípio da iteratividade está dentro do contexto de desenvolvimento de software, do qual foi realizada uma adaptação para a construção deste modelo estrutural para repositórios digitais científicos. É importante ressaltar que o estudo de processos e metodologias para melhorar o desenvolvimento de software é constante dentro da área de Ciência da Computação.

O conceito de desenvolvimento iterativo é bastante utilizado na Engenharia de Software, disciplina da Ciência da Computação, e faz parte de alguns processos de desenvolvimento de software já sedimentados e muito utilizados como RUP (Rational Unified Process), Programação Extrema (XP) e Scrum.

Segundo Larman (2007),

O ciclo de vida iterativo é baseado em refinamentos e incrementos sucessivos de um sistema por meio de múltiplas iterações, com realimentação (feedback) e adaptação cíclicas como principais propulsores para convergir para um sistema adequado. O sistema cresce incrementalmente ao longo do tempo, iteração por iteração, razão pela qual esta abordagem também é conhecida como desenvolvimento iterativo e incremental. Como a realimentação e adaptação fazem as especificações e o projeto evoluir, esse sistema é conhecido como desenvolvimento iterativo e evolutivo. (LARMAN, 2007, p. 47)

Larman (2007) afirma que o processo iterativo é também evolutivo, assim como acontece no modelo proposto nesta pesquisa, que propõe a evolução das representações de informação.

Essa evolução pode ocorrer através da Folksonomia Assistida, ou seja, do processo repetitivo de inserção de conteúdos para tags, assim como da evolução e adaptação da estrutura de representação do conhecimento utilizada, por intermédio de um administrador.

O processo de desenvolvimento iterativo, do qual a Representação Iterativa é baseada, deve obedecer limites temporais.

Larman (2007, p. 50) alerta que: “A maioria dos métodos iterativos recomenda que a duração de uma iteração seja entre duas e seis semanas”.

A Representação Iterativa tem um contexto diferente, porque não trata de desenvolvimento de software, mas sim da construção do *corpus* de informação de um domínio, através de uma inteligência coletiva, porém o princípio da temporalidade também pode ser abordado e utilizado.

O processo de construção da inteligência coletiva pela Folksonomia Assistida não deve ser temporal, ele deve ser feito dinamicamente, sugerindo que o usuário possa ter acesso ao conjunto de informações a qualquer momento, ou seja, assim que uma *tag* é inserida no sistema, ela passa a ficar disponível para ser utilizada como sugestão a outros usuários.

O acesso do administrador do sistema para fazer ajustes ao modelo que está sendo construído pode ter sim uma temporalidade definida, corroborando com a ideia de iteratividade. O modelo Representação Iterativa não define um intervalo exato de temporalidade de intervenção do administrador do sistema, porém cada ambiente deve estabelecer seu próprio intervalo de temporalidade de acordo com o a quantidade de acessos e o volume de informações dentro do repositório (SANTAREM SEGUNDO, 2010).

Cada iteração gera um novo conjunto de informações, relacionamentos e também uma forma diferente de conhecimento. Dentro do contexto evolutivo do modelo, é possível que a interatividade entre os usuários e o sistema gere uma camada de informações cada vez mais rica, principalmente porque permite *feedback* ao usuário, assim como a possibilidade da informação já armazenada serve como base para que a próxima seja inserida.

Dessa forma, a Representação Iterativa oferece aos repositórios um novo formato de organização da informação, de modo que passe a existir uma relação entre os trabalhos autoarquivados, não apenas pela simples sintaxe das palavras-chave e nem tampouco pela comunidade e coleção de que fazem parte.

A estrutura funcional deste modelo parte do princípio da agregação de valores ao repositório, de forma que ocorra uma contextualização do material digital inserido, criando relações que possam sustentar uma recuperação semântica de informações.

Recuperação da informação na representação iterativa

A estrutura da Representação Iterativa permite criar novos sistemas de recuperação da informação dentro dos repositórios. Os novos modelos não devem substituir o anterior, mas sim agregar mais opções de pesquisa e interação do usuário com o ambiente.

Um dos formatos propostos é o de rede, que tem sido muito abordado no conceito de colaboração científica, principalmente nos relacionamentos entre coautorias e citações, porém a mesma ideia utilizada neste conceito aplica-se às redes de *tags*, que podem agregar a informação a respeito dos autores e criar o conceito de autores que tenham o mesmo perfil de depósito dentro de um repositório. O modelo em formato de rede aproxima termos que estão relacionados criando uma estrutura de informação que tem apresentação visual agradável e de entendimento intuitivo.

Portanto, se a ideia de coautoria das redes colaborativas torna possível uma grande quantidade de estudos, as redes construídas por meio da estrutura da Representação Iterativa poderão gerar um conjunto grande de informações a respeito do conteúdo dos objetos depositados nos repositórios digitais informacionais.

Para a criação de uma rede de informações que permita ao usuário navegar pelos termos, os dados registrados deverão formar uma matriz de adjacência que possibilite a construção do grafo, que é a estrutura matemática e computacional escolhida para representar as redes.

O primeiro passo para a construção da rede de *tags* é gerar uma matriz de adjacência, que dá sustentação à criação do grafo/rede. A matriz de adjacência é construída de forma que as linhas e colunas da matriz sejam representadas pelas *tags* e o cruzamento indica a quantidade de relacionamentos existentes entre as *tags*. Baseado na matriz de adjacências é possível construir o grafo de *tags*.

A apresentação do grafo possibilita algumas variações, e neste trabalho sugere-se que os vértices tenham tamanhos diferentes, de forma proporcional, de acordo com a quantidade de incidências dos termos armazenados no repositório.

Portanto pode-se definir uma quantidade de níveis de apresentação dos termos (círculos), sendo que cada nível terá um tamanho diferente, ficando os termos mais populares com o maior diâmetro e os menos populares com menor diâmetro, conforme Figura 2.

As arestas que ligam os vértices representam a quantidade de ligações existentes entre cada uma das *tags*. A quantidade de relacionamentos existentes entre as *tags* é representada visualmente pela largura das arestas.

No plano de visualização da rede pelo usuário, é inviável que seja apresentada toda a rede de *tags*, portanto a Representação Iterativa sugere a apresentação de termos que estejam a uma distância (*d*) de dois ou três termos do termo que é apresentado como termo (nó) principal da rede de *tags*, porém, à medida que o usuário vai navegando na rede, o nó principal passa a ser trocado e então mudam a profundidade e a largura, para que novos vértices do grafo passem a fazer parte da visualização. A distância (*d*) entre termos é a quantidade de nós que se devem passar para se chegar de um termo a outro.

Quando o usuário proceder com dois cliques em um nó da rede, então deverá ser executado o procedimento de recuperação e apresentação dos resultados. A definição do nó principal da rede deve se dar por meio do termo que é mais citado no repositório, iniciando a rede sempre por esse termo.

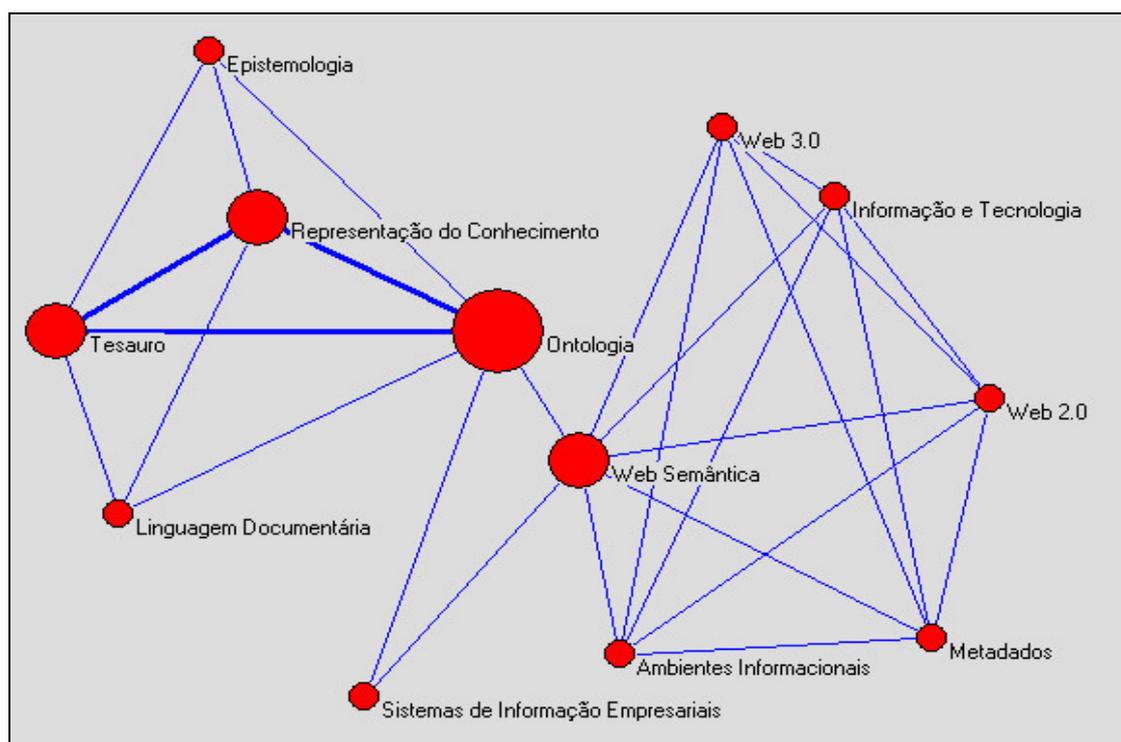


Figura 2: Rede de tags de um conjunto de quatro artigos utilizados como exemplo.

Fonte: Santarem Segundo (2010)

A rede de *tags*, assim como outros modelos que podem ser sugeridos, apresentam novos conceitos de recuperação da informação, baseados na Representação Iterativa. Assim, funda-se o trabalho com a completude de um modelo que pode mudar a estrutura funcional dos repositórios digitais, de forma a permitir que estes sejam ambientes mais ricos e aptos a construir a recuperação semântica de informações.

Conclusões

A construção desta pesquisa nasceu da necessidade de melhorar a recuperação da informação em repositórios digitais informacionais. Para que isso fosse possível, foi proposto um modelo novo nomeado Representação Iterativa para repositórios digitais.

O primeiro passo para iniciar a construção do modelo proposto foi verificar que os repositórios institucionais são ambientes que necessitam de melhorias, tanto do ponto de vista das funcionalidades oferecidas aos usuários, quanto do ponto de vista técnico para a recuperação da informação.

Foi possível verificar que as ferramentas disponíveis para implementação de repositórios foram construídas sobre um modelo que oferece condições de implantação dos recursos da *web* semântica. Alguns pontos são fundamentais, como a utilização do formato de metadados Dublin Core.

Verificou-se ainda que a Folksonomia é um processo importantíssimo para ser aplicado ao contexto dos repositórios, visto que permite a construção de inteligência coletiva e oferece subsídios para que haja uma busca por termos relacionados, porém se for efetivamente utilizado de forma totalmente livre pode gerar termos sem relacionamentos futuros, ou ainda inexatos e inconsistentes dentro da Representação Iterativa.

Concluiu-se que a necessidade de um novo conceito de Folksonomia, a Folksonomia Assistida, proposta neste trabalho, vem ao encontro da necessidade de auxiliar o usuário na descrição da *tag* assunto do recurso a ser depositado, em relação ao domínio do conhecimento do qual o repositório faz parte. A Folksonomia Assistida é um processo que pode definitivamente elevar o nível de qualidade de descrição do recurso, mantendo a criatividade do usuário na inserção da *tag*, mas também oferecendo a ele elementos que possam relacionar seu recurso a outros já depositados anteriormente ou/e ainda a uma estrutura de representação do conhecimento.

O modelo de Representação Iterativa, principal proposta deste trabalho, é de fundamental importância no papel de qualificar e melhorar a estrutura de representação do conhecimento das áreas de especialidades. Do ponto de vista da evolução, uma estrutura de representação do conhecimento pode encontrar subsídios na utilização das *tags* propostas no sistema para melhor adequar-se ao domínio e aos pesquisadores que utilizam o repositório. É notório que participação do usuário no modelo Representação Iterativa é fundamental, visto que o perfil do usuário pode condicionar o bom funcionamento da Representação Iterativa.

Verificou-se também que desenvolver modelos gráficos para amparar a recuperação da informação pode facilitar e auxiliar os usuários no processo de recuperação da informação em ambientes que se-utilizam de Folksonomia, como a Representação Iterativa.

Embora tenha sido concebido dessa forma, o modelo não é restrito a repositórios digitais, que foi o foco da pesquisa. A Representação Iterativa e a Folksonomia Assistida podem ser aplicadas em outros tipos de ambientes digitais que ofereçam ao usuário a possibilidade de descrever suas próprias *tags* e trabalhem com uma estrutura de representação do conhecimento das áreas de especialidades.

Artigo recebido em 14/02/2011 e aprovado em 18/02/2011.

Referências

AQUINO, M. C. Hipertexto 2.0, folksonomia e memória coletiva: um estudo das tags na organização da web. *E-Compós*, v. 9, 2007. Disponível em: < <http://www.compos.org.br/seer/index.php/e-compos/article/view/165/166>>. Acesso em: 03 nov. 2009.

BEKAERT, J.; VAN DE SOMPEL, H. *Augmenting interoperability across scholarly repositories: report*, 2006. Disponível em: <<http://msc.mellon.org/Meetings/Interop/FinalReport>>. Acesso em: 14 fev. 2009.

BERNERS-LEE T.; LASSILA, O.; HENDLER, J. The semantic web. *Scientific American*, v. 5, May 2001. Disponível em: <http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>. Acesso em: 03 mar. 2009.

GRUBER, T. R. *Toward principles for the design of ontologies used for knowledge sharing*. Padova: Stanford University, 1992. Disponível em: <http://ksl.stanford.edu/KSL_Abstracts/KSL-93-04.html>. Acesso em: 15 fev. 2009.

GUARINO, N. Formal ontology and information systems. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS - FOIS'98, 1998, Trento. *Proceedings...* Amsterdam: IOS Press, 1998. p. 3-15.

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: 13 fev. 2009.

JACOB, E. K. Ontologies and the semantic web. *Bulletin for the American Society for Information Science and Technology*, v. 29, n. 4, p. 19-22, Apr./Mayo 2003.

LARMAN, C. *Utilizando UML e padrões: uma introdução à análise e ao projeto orientados a objetos e ao desenvolvimento iterativo*. 3. ed. São Paulo: Bookman, 2007.

O'REILLY, T. *What is web 2.0: design patterns and business models for the next generation of software*. 2005. Disponível em:

<<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html#mememap>>. Acesso em: fev. 2009.

SEGUNDO, J. E. S. *Representação iterativa: um modelo para repositórios digitais*. 2010. 224 f. Tese (Doutorado em Ciência da Informação)– Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

WAL, T. V. *Folksonomy definition and wikipedia*. Disponível em: <<http://www.vanderwal.net/random/entrysel.php?blog=1750>>. Acesso em: 02 jun. 2009.