

# Rede de *tags* para recuperação da informação no contexto da Representação Iterativa

*Network tags for information retrieval in the context of Iterative Representation*

**José Eduardo Santarem Segundo**

Doutor em Ciência da Informação pela Faculdade de Filosofia e Ciências da UNESP.  
Professor do Curso de Graduação em Ciências da Informação e da Documentação da FFCLRP/USP  
E-mail: [santarem@usp.br](mailto:santarem@usp.br)

**Silvana Aparecida Borsetti Gregorio Vidotti**

Doutora em Educação pela Faculdade de Filosofia e Ciências da UNESP.  
Professora do Programa de Pós-Graduação em Ciência da Informação da UNESP  
E-mail: [vidotti@marilia.unesp.br](mailto:vidotti@marilia.unesp.br)

## Resumo

A recuperação da informação é um tema recorrentemente pesquisado dentro da Ciência da Informação. Estudos dessa natureza visam melhorar os resultados em buscas tanto na Web como em vários outros ambientes digitais informacionais. Neste contexto, a Representação Iterativa, modelo sugerido para repositórios digitais, aparece como um diferencial que altera o paradigma do auto-arquivamento de objetos digitais, criando um conceito de relacionamento entre os termos que ligam o pensamento do usuário ao material depositado no ambiente digital. As ligações efetivadas pela Representação Iterativa com auxílio da Folksonomia Assistida geram uma estrutura em formato de redes que conecta, de forma vertical e horizontal, os objetos depositados, apoiando-se em algum tipo de estrutura de representação do conhecimento das áreas de especialidades e, portanto, criando uma rede de informações baseada no conhecimento dos usuários. A rede de informações criada, chamada de rede de *tags* é dinâmica e efetiva um modelo diferente de recuperação da informação e estudo dos repositórios digitais informacionais.

**Palavras-chave** Repositórios Digitais; Representação Iterativa; Folksonomia; Folksonomia Assistida; Rede de *Tags*

## Abstract

Information retrieval is a recurrent subject in search of information science. This kind of study aim to improve results in both searches on the Web and in various other digital information environment. In this context, the Iterative Representation model suggested for digital repositories, appears as a differential that changes the paradigm of self-archiving of digital objects, creating a concept of relationship between terms that link the user thought the material deposited in the digital environment. The links effect by the Iterative Representation aided Assisted Folksonomy generate a shaped structure that connects networks, vertically and horizontally, the objects deposited, relying on some kind of structure for representing knowledge of specialty areas and therefore, creating an information network based on knowledge of users. The network of information created, called the network of tags is dynamic and effective a different model of information retrieval and study of digital information repositories.

**Keywords** Digital Repositories; Iterative Representation; Folksonomy; Folksonomy Assisted; Semantic Web; Network Tags.

## **Introdução\***

O modelo “Representação Iterativa” foi concebido para ser aplicado em repositórios digitais informacionais. O principal objetivo deste trabalho é apresentar esse modelo, formalizando a Representação Iterativa na interação do usuário com o ambiente digital permitindo uma condução que amparada pela Folksonomia Assistida cria uma estrutura em forma de rede que organiza e associa ideias e termos, assim como é feito no cérebro humano.

A estrutura em rede, pensada e disposta neste trabalho é baseada no conceito de grafos, que é um modelo matemático muito usado nas mais variadas formas de resolução de problemas, sendo apresentado na forma de um diagrama composto por pontos e linhas que conectam esses pontos. Aos pontos é dado o nome de vértice e as linhas são conhecidas como edges ou arestas.

A Representação Iterativa destaca um de seus modelos na recuperação da informação, o modelo em rede, que hoje é considerado a base para a construção de redes sociais e tem sido utilizado para vários estudos dentro da Ciência da Informação, tais como: relações entre pesquisadores, relação entre temas de pesquisa, inteligência coletiva, desenvolvimento colaborativo. As análises baseadas nos modelos em rede tem sido objeto inclusive para descobrir relações e associar trabalhos entre pesquisadores, grupos de pesquisadores e produção dentro das instituições tornando-se um recurso importante para efetivação do modelo colaborativo na construção do conhecimento em rede.

## **Web 2.0: conceitos e funcionalidades e a Folksonomia**

A estrutura de rede formada pela Representação Iterativa nasce instintivamente através do uso da Web 2.0 e, principalmente, de um de seus recursos, a Folksonomia, para a construção do conhecimento de forma colaborativa.

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

---

\* 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 Unesp de Marília, sob orientação da Profa. Dra. Silvana Ap. Borsetti Gregório Vidotti.

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.

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, p.1),

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.

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.

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 e 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 dos documentos em repositórios digitais, de modo que os recursos possam ser recuperados posteriormente pelo próprio usuário ou, ainda, por outros usuários que procurem informações nesse tipo de ambiente digital.

Entende-se 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 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, p.10) 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.

A Folksonomia mudou o paradigma em relação à recuperação da informação em ambientes Web. 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.

### **Representação Iterativa**

A Representação Iterativa caracteriza-se como um conjunto de técnicas que, juntas, adaptam um ambiente digital, do tipo repositório digital, 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 (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 houverem, para que todo o conjunto de informações seja armazenado no repositório.

O armazenamento de informações estruturadas guarda no banco de dados do repositório as informações pertinentes aos metadados que foram descritos pelo usuário, assim como as informações complementares a respeito da comunidade e coleção de que o objeto faz parte. Informações a respeito dos arquivos binários, como tamanho, tipo de arquivo e nome, também são armazenadas no banco de dados.

Neste trabalho a ferramenta Dspace foi utilizada como exemplo de repositório digital, portanto, os exemplos apresentados são baseados neste *software*. Ressalta-se, porém, que todas as outras principais ferramentas *opensource* (*software* de código livre e aberto) que implementam repositórios digitais tem estruturas de armazenamento semelhantes.

Para conceber o armazenamento interno das informações relativas aos objetos que estão sendo depositados no repositório, em sua grande parte por pesquisadores, o banco de dados define um conjunto de tabelas que deve armazenar desde a informação do próprio usuário que está fazendo o depósito, incluindo data, até o conjunto de informações que compõe os metadados do recurso a ser depositado. No Dspace, as tabelas físicas responsáveis por armazenar as informações do objeto digital depositado são:

- *item*: responsável por armazenar as informações sobre o usuário que fez o depósito, e definir um número único para o objeto, além da data em que foi realizada a última alteração no objeto;
- *collection2item*: armazena a informação referente a qual coleção pertence o recurso que está sendo inserido (estabelece o relacionamento);
- *metadatatype*: armazena as informações dos metadados do recurso que está sendo inserido. Essa tabela faz uma ligação direta com outras duas tabelas: *metadataschemaregistry* e *metadatafieldregistry*.

Além das tabelas de banco de dados citadas, o Dspace ainda mantém outras tabelas que, relacionadas, tem como principal objetivo armazenar todas as informações e detalhes necessários para que o objeto digital auto-arquivado pelo usuário possa ser armazenado e posteriormente recuperados pelos usuários do ambiente.

A Representação Iterativa sugere a construção de um modelo estrutural para repositórios digitais, 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, através da associação de conteúdos, e não apenas em comparação sintática, como é realizado atualmente.

O modelo — Representação Iterativa — parte do princípio de que no momento do auto-arquivamento de seus objetos digitais o usuário deverá dispor de uma interface diferenciada em relação à apresentada pela maioria das ferramentas que implementam repositórios digitais. Ao usuário é importante identificar que as palavras-chave utilizadas serão também empregadas como etiquetas (*tags*).

O fato de caracterizar o uso de *tags* já cria 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*, uma rede de *tags* ou outros formatos similares.

A descrição da *tag* é um dos momentos importantes na descrição do recurso dentro da Representação Iterativa, visto que no momento em que o usuário iniciar o processo deverá ocorrer uma intervenção do sistema, de forma que se caracterize um processo que se denomina 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.

O funcionamento da Folksonomia Assistida depende de uma interface diferente que deverá ser implementada ao repositório digital, visto que este recurso não está disponível nas ferramentas atualmente utilizadas para implementação de repositórios.

A implementação deve permitir que o usuário no momento de digitar as palavras-chave que irão compor a descrição do recurso seja amparado por uma lista de palavras, como o procedimento efetivado pelo Google em sua ferramenta de busca ao apresentar uma lista de sugestões ao usuário.

O conjunto de informações que deverá ser apresentado ao usuário no momento que este estiver digitando será baseado nas *tags* anteriormente inseridas no sistema e também nos termos que fazem parte de uma estrutura de representação do conhecimento das áreas de especialidades, ou seja, uma ontologia, tesouro ou algum tipo de recurso dessa natureza que deverá estar associado ao repositório como parte do modelo estrutural proposto.

Esse procedimento indica que Folksonomia Assistida prima pela consistência das *tags*, de forma que o usuário, ao inserir informações no ambiente evite abreviações, plurais/singulares ou ainda palavras que possam dificultar a recuperação da informação, posteriormente.

Assim que o usuário descrever as *tags*, aceitando ou não as sugestões listadas, 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, a utilização de ontologias através da linguagem OWL permite agregar recursos e facilitar o processo de

recuperação da informação, principalmente por ser uma linguagem que vem sendo aprimorada constantemente, e conta com indicação de uso pelo W3C.

Esse segundo passo da Folksonomia Assistida, além de recuperar termos relacionados 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 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 (objeto digital) 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 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 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.

O processo de gravação das informações é efetivado quando o usuário definitivamente escolhe e define 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, vem ao encontro à 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 dos objetos digitais que deposita, de forma a amparar-se no próprio conhecimento já disponível no ambiente em que está utilizando. Esse processo enaltece a construção colaborativa da inteligência coletiva.

### **Reestruturando o banco de dados**

A Representação Iterativa tem como princípio armazenar as *tags* definidas pelo usuário. Por conseguinte, deve-se realizar uma alteração na estrutura de banco de dados que as ferramentas utilizam, criando um conjunto de tabelas que possa estabelecer o armazenamento e relacionamento dessas informações.

É justamente essa reestruturação do banco de dados que vai permitir que as informações arquivadas no repositório formem uma estrutura reticular dos objetos digitais depositados.

Nesse modelo se estabelece que uma nova *tag*, sempre que for inserida no repositório, deverá ser cadastrada em uma tabela. No entanto, se a *tag* já existir no banco de dados, cadastrada em depósito anterior, apenas será atribuído um incremento em relação à quantidade de vezes que a *tag* já fora utilizada.

Outra característica da Representação Iterativa, baseada no Folksonomia, é criar efetivamente um relacionamento horizontal entre *tags* que descrevam o mesmo objeto digital. Esse processo constitui uma relação entre os termos individualmente citados, e, dentro de um contexto de domínio do conhecimento restrito, estabelece uma relação entre termos, de forma que possam ser recuperados posteriormente.

Essa relação horizontal, à medida que vai acontecendo sistematicamente durante os depósitos vai se solidificando e criando uma relação que possivelmente será objeto de busca e recuperação futura pelos usuários.

Portanto, toda vez que houver um mesmo relacionamento entre termos, deverá apenas ser incrementada a quantidade de vezes que o relacionamento acontece, sem a necessidade de se recadastrar a informação no banco de dados.

Essa estrutura de relacionamento de termos permite que se crie um grafo de *tags*, onde cada *tag* será representada por um vértice e a quantidade de relações entre as *tags* será representada visualmente pela largura da aresta que liga os vértices, formando uma rede de relacionamento entre os itens depositados.

É justamente a construção dessa rede, um dos novos fenômenos de estudo na Ciência da Informação, que permite estudos mais aprofundados dentro do contexto da Representação Iterativa.

Segundo Matheus e Silva (2009, p. 243),

Na análise de redes o foco do estudo é nos relacionamentos entre entidades. As entidades podem ser atores sociais, páginas web, neurônios do cérebro, dentro outras. Os relacionamentos podem dar-se por meio de trocas materiais (movimentação, proximidade) ou não materiais (informação, sinais elétricos). Em todo os casos, o relacionamento entre entidades pode ser modelados utilizando-se grafos.

Verifica-se assim que a construção do modelo em redes pode gerar frutos em relação à análise do domínio em questão.

Em relação à gravação das *tags*, para que os dados possam ser armazenados serão necessárias mais três tabelas que deverão ser acopladas ao modelo físico do repositório: *tags*, *tags2tags* e *tags2item*.

tags		tags2tags		tags2item	
codigo	numérico	tag1	numérico	tag	numérico
descricao	alfanumérico	tag2	numérico	item	numérico
quantidade	numérico	quantidade	numérico		

Figura 1 – Tabelas para armazenamento das *tags*

Fonte: Santarém Segundo (2010)

As tabelas *tags* e *tags2tags* (figura 1) serão utilizadas para armazenar os dados referentes às *tags* descritas no depósito. A tabela *tags* conta com os campos: **código**, que

definirá um indicador único para cada *tag*; **descrição**, que armazenará o texto real da *tag*; **quantidade**, que representará a quantidade de vezes que a *tag* foi utilizada no sistema. A tabela *tags2tags* indicará nos seus campos **tag1** e **tag2** os códigos referentes às *tags* que se relacionam, e o campo **quantidade** deverá informar a quantidade de vezes que isso acontece.

A tabela *tags2item* (figura 1) será utilizada para fazer a referência entre os itens (objetos digitais/recurso) armazenados no repositório e as *tags* que estão diretamente ligadas a eles.

Utilizando-se um conjunto de quatro artigos, sendo três deles publicados na revista *Datagramazero* e outro publicado na revista *Brazilian Journal Information Science (BJIS)*, demonstra-se como ficariam armazenadas as *tags* na estrutura proposta de tabelas.

Segue os artigos e suas respectivas palavras-chave, utilizadas como *tags* neste exemplo:

- Projeto de ontologia para sistemas de informação empresariais: delineando uma metodologia para desenvolver ontologias na área de telecomunicações, dos autores Beatriz Ainhize Rodriguez Barquín et al., que conta com as seguintes palavras-chave: Ontologia; Sistemas de Informação Empresariais; Web Semântica.
- Metadados e Web Semântica para estruturação da Web 2.0 e Web 3.0, dos autores Plácida Leopoldina Ventura Amorim da Costa Santos e Rachel Cristina Vesú Alves, com as seguintes palavras-chave: Informação e Tecnologia; Metadados; Web Semântica; Web 2.0; Web 3.0; Ambientes Informacionais.
- Semelhanças e Diferenças entre Tesouros e Ontologias, dos autores Rodrigo de Sales e Ligia Café, com as seguintes palavras-chave: Tesouro; Ontologia; Linguagem documentária; Representação do conhecimento.
- O nível do conhecimento e os instrumentos de representação: tesouros e ontologias, dos autores Alexandra Moreira, Lídia Alvarenga e Alcione de Paiva Oliveira, com as seguintes palavras-chave: Ontologia; Tesouros; Epistemologia; Representação do Conhecimento.

O conceito de publicação em que a Folksonomia Assistida atua não tem como característica apenas digitar as *tags* de documentos já publicados, mas evoluir com o processo de caracterização e inserção de *tags*. Porém, no caso deste exemplo, utiliza-se material já

publicado, mostrando, através das figuras 2 e 3, como ficariam registradas no banco de dados essas informações, de forma que possa dar entendimento à construção da estrutura de tabelas sugerida.

tags		
codigo	descricao	quantidade
1	Ontologia	3
2	Sistemas de Informação Empresariais	1
3	Web Semântica	2
4	Informação e Tecnologia	1
5	Metadados	1
6	Web 2.0	1
7	Web 3.0	1
8	Ambientes Informacionais	1
9	Tesouro	2
10	Linguagem Documentária	1
11	Representação do Conhecimento	2
12	Epistemologia	1

Figura 2 – Tabela *tags* populada  
Fonte: Santarem Segundo (2010)

Na figura 2, pode-se verificar que todas as *tags* foram registradas no banco de dados, sendo que algumas, como o caso de “ontologia”, “web semântica”, “tesouro” e “representação do conhecimento”, aparecem mais de uma vez.

tags2tags			tags2item	
tag1	tag2	quantidade	tag	item
1	2	1	1	1
1	3	1	2	1
2	3	1	3	1
3	4	1	3	2
3	5	1	4	2
3	6	1	5	2
3	7	1	6	2
3	8	1	7	2
4	5	1	8	2
4	6	1	9	3
4	7	1	10	3
4	8	1	11	3
5	6	1	1	4
5	7	1	9	4
5	8	1	11	4
6	7	1	11	4
6	8	1	12	4
7	8	1		
1	9	2		
1	10	1		
1	11	2		
9	10	1		
9	11	2		
10	11	1		
1	12	1		
9	12	1		
11	12	1		

Figura 3 – Tabelas *tags2tags* e *tags2item* populadas  
Fonte: Santarem Segundo (2010)

Através da figura 3, é possível registrar o armazenamento das relações, sendo que a tabela *tags2tags* leva a identificar que as *tags* codificadas como 1 e 9, que representam respectivamente “ontologia” e “tesauro”, estão relacionadas mais de uma vez, assim como 1 e 11, que são “ontologia” e “representação do conhecimento”, também relacionadas mais de uma vez. Essas relações citadas que contemplam mais de uma unidade de relacionamento acontecem, porque as mesmas palavras-chave são utilizadas em mais de um documento. Essa relação gera a estrutura em formato de rede.

A tabela *tags2item* (figura 3) representa a ligação que existe entre as *tags* e os documentos inseridos, lembrando que ela deve estar relacionada com a tabela *item* apresentada no modelo físico do repositório.

### **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 tesauro, que consiste em definir os limites de um domínio do conhecimento.

Santarem Segundo (2010, p. 187) 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.

Essa visão geral é detalhada na Figura 4 que apresenta os passos para que realmente aconteça o uso completo da Representação Iterativa.

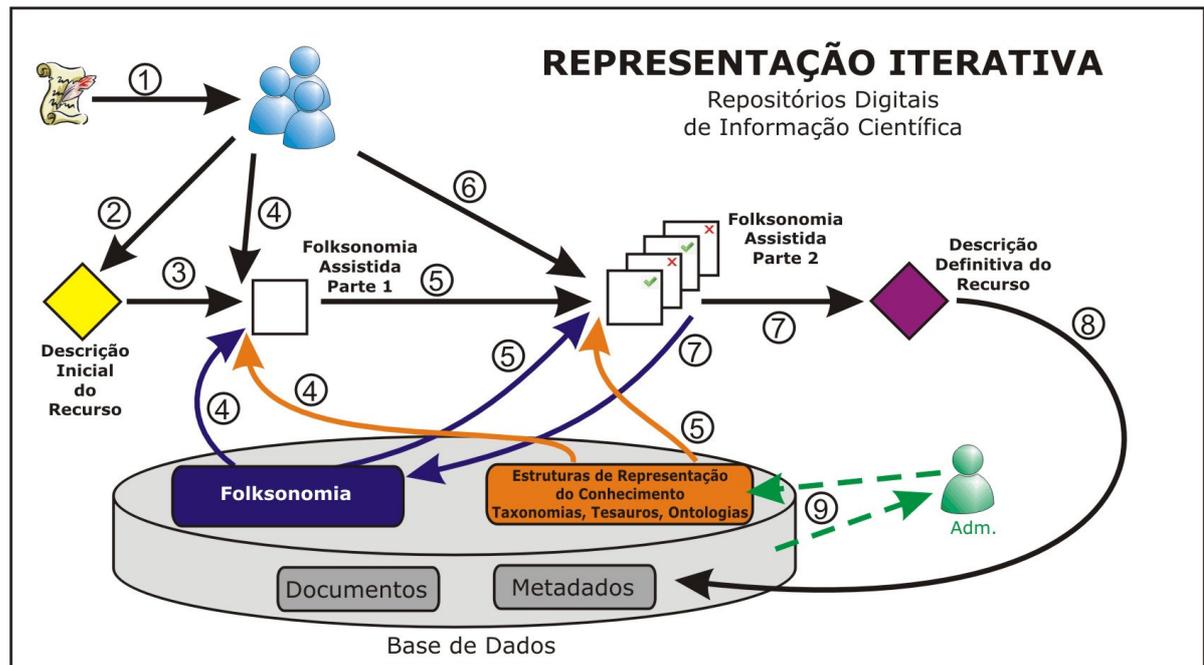


Figura 4: Representação Iterativa – Visão Detalhada  
Fonte: Santarem Segundo (2010)

A construção do modelo nomeado Representação Iterativa, apresentado de forma detalhada na Figura 4, 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 em 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 em 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 o 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/gestor 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 do conjunto de dados gerado pela 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 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. 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, p. 47),

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) afirma que o processo iterativo é também evolutivo, assim como acontece no modelo proposto neste trabalho, 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 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, ou seja, uma nova estrutura da rede com novos relacionamentos. 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 servir 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 através da Rede de *Tags*

A construção e a aplicação da Representação Iterativa altera a estrutura dos repositórios digitais e permite que seja revisto o conceito de recuperação utilizado nesse tipo de ambiente, portanto, o modelo vem no sentido de colaborar justamente com a recuperação da informação, independente do formato em que ela estiver.

Segundo Buckland (2006, p.6),

A técnica de pesquisa por seqüências de caracteres de texto funciona muito bem, mas nem sempre e não perfeitamente, porque recursos de texto não são inteiramente homogêneos. Algumas palavras possuem vários significados (polissemia, por exemplo, mouse); às vezes palavras diferentes utilizam a mesma seqüência de caracteres, mas com outros significados (homógrafos, por exemplo, pane significa painel de vidro em inglês, mas não em português); e palavras diferentes podem ser utilizadas com o mesmo significado (sinônimos, por exemplo, câncer e neoplasma).

A estrutura da Representação Iterativa permite criar um novo sistema de recuperação da informação dentro dos repositórios. O novo modelo não deve substituir os tradicionalmente utilizados, mas sim agregar mais um tipo de pesquisa e interação do usuário com o ambiente.

O modelo apresentado a seguir é de uma rede de *tags*, que deverá oferecer ao usuário um conjunto de ligações, em formato de conexões “cerebrais”, onde os pontos são os termos mais pertinentes dentro do ambiente digital. Esse formato de rede caracteriza-se por ter o mesmo tipo de relacionamento entre seus itens que os relacionamentos criados em redes sociais.

Segundo Wasserman e Faust (1994, p.9),

o termo ‘rede social’ se refere ao conjunto de atores e suas ligações entre eles. Assim, a análise de rede tem por objetivo modelar as conexões entre os atores, a fim de retratar, descrever e representar a estrutura de um grupo, quer seja composto por países, instituições ou pessoas.

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.

Analisar as redes construídas através da estrutura da Representação Iterativa deverá 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 nas tabelas *tags* e *tags2tags* deverão formar uma matriz de adjacência que possibilite a construção do grafo, que é a estrutura matemática e computacional escolhida para representar as redes.

Utilizando-se do exemplo citado anteriormente que relaciona quatro artigos para dispor as palavras-chaves no banco de dados, para compor as tabelas físicas do repositório, será aproveitado aqui para compor a rede de *tags*.

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 indique a quantidade de relacionamentos existentes entre as *tags*.

A matriz de adjacências gerada com base no exemplo é apresentada na figura 5.

	Ontologia	Sistemas de Informação Empresariais	Web Semântica	Informação e Tecnologia	Metadados	Web 2.0	Web 3.0	Ambientes Informacionais	Tesouro	Linguagem Documentária	Representação do Conhecimento	Epistemologia
<b>Epistemologia</b>	1								1		1	
<b>Representação do Conhecimento</b>	2								2			
<b>Linguagem Documentária</b>	1								1			
<b>Tesouro</b>	2											
<b>Ambientes Informacionais</b>			1	1	1	1	1					
<b>Web 3.0</b>			1	1	1	1						
<b>Web 2.0</b>			1	1	1							
<b>Metadados</b>			1	1								
<b>Informação e Tecnologia</b>			1									
<b>Web Semântica</b>	1	1										
<b>Sistemas de Informação Empresariais</b>	1											
<b>Ontologia</b>												

Figura 5 – Matriz de adjacências de quatro artigos utilizados como exemplo.

Fonte: Santarem Segundo (2010)

Baseado na matriz de adjacências construída é possível construir o grafo de *tags*, nomeada neste trabalho de rede de *tags*.

A apresentação visual da rede de *tags* possibilita algumas variações, utilizam-se neste trabalho os vértices com tamanhos diferentes, de forma proporcional, de modo que se define um grupo de níveis em que as *tags* serão representadas.

Assim sendo cada nível terá um tamanho diferente, ficando os termos mais populares com o maior diâmetro e os menos populares com menor diâmetro, conforme figura 6.

As arestas que são responsáveis por conectar os círculos (vértices) e que representam a quantidade de ligações existentes entre cada uma das *tags* também deverão seguir um padrão que é definido da seguinte forma: a quantidade de relacionamentos existentes entre as *tags* será representada visualmente pela largura da ligação entre as arestas, e a largura das relações deverá ser construída com o emprego de níveis pré-estabelecidos, ou seja, quando maior o peso entre dois termos mais larga será a linha que une os termos, e quanto menor o peso mais fina será a linha, conforme pode ser visto na figura 6.

No plano de visualização da rede pelo usuário, é inviável que seja apresentada toda a rede de *tags*, logo 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*, contudo, à medida que o usuário vai navegando na rede, o nó principal passa a ser trocado e então mudam a profundidade e largura, para que novos vértices do grafo passem a fazer parte da visualização. O procedimento de apresentação é calculado através do procedimento matemático de busca em largura e busca em profundidade em grafos.

A distancia ( $d$ ) entre termos é a quantidade de nós que se deve 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 dar-se-á através do termo que é mais citado no repositório, iniciando a rede sempre por esse termo.

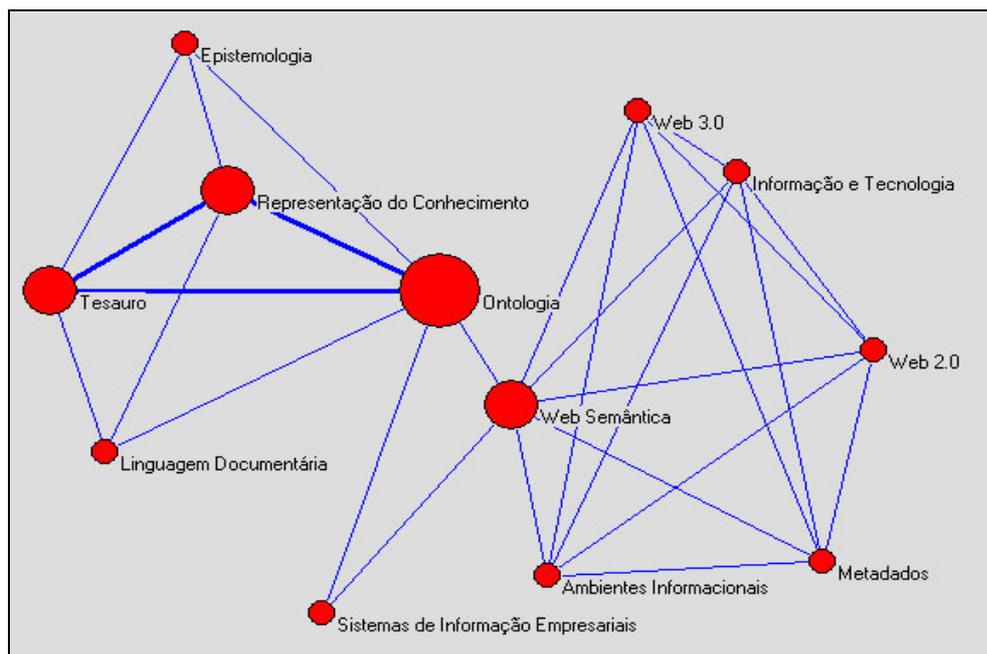


Figura 6 – Rede de *Tags* de quatro artigos utilizados como exemplo.  
Fonte: Santarem Segundo (2010)

A rede de *tags* resultante da matriz apresentada na figura 5 pode ser visualizado na figura 6, no entanto, dado o suporte de apresentação deste trabalho, não é possível realizar o deslocamento, que deverá ser implementado através de técnicas de programação visual no ambiente digital.

Ressalta-se que a rede de *tags* apresentada na figura 6 foi gerada através de um exemplo contendo 4 artigos já previamente publicados e escolhidos pelo autor, portanto, não há repetição de palavras-chave parecidas, plurais entre outras inconsistências que podem ser geradas em uma rede de *tags*, e que deverão ser suprimidas com o uso da Folksonomia Assistida.

Outro item a considerar na figura 6 é a estrutura estática da informação, contudo a característica da Representação Iterativa é justamente que a rede seja dinâmica, sendo alterada a cada novo depósito, deste modo é possível que uma determinada palavra-chave seja o centro da rede em um determinado período e em outro faça parte de uma região mais periférica, dependendo do contexto dos documentos apresentados.

A análise desta rede indica que a palavra Ontologia é a mais citada nos 4 trabalhos utilizados como exemplo, sendo possível perceber também uma ligação mais forte entre as palavras ontologia, tesauro e representação do conhecimento.

A Representação Iterativa caracteriza-se, portanto, como um modelo que tem como objetivo evidenciar uma postura diferente do usuário que faz o depósito de documentos em ambientes digitais, de forma que possa evoluir com o processo de recuperação da informação, principalmente através de novos recursos como as redes de *tags*.

### **Considerações finais**

O modelo Representação Iterativa nasceu da necessidade de melhorar a recuperação da informação em repositórios digitais informacionais. Através do modelo foi possível desenvolver um formato de recuperação da informação denominado rede de *tags*, que segue a linha de pesquisas da Ciência da Informação que se utiliza de redes de informações para construir conhecimento.

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 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 recuperação no formato de rede de *tags* favorece a recuperação da informação, principalmente porque dá ênfase na relação entre os termos destacados como *tags* no ambiente digital. Dessa forma não só o processo de recuperação é favorecido, mas também é possível dar foco aos estudos dos relacionamentos apresentados pela rede.

Verifica-se que o modelo em rede apresenta tendências em suas ligações com destaque para as conexões que efetivam a forte ligação entre *tags*, mas principalmente nos laços mais fracos, que evidenciam a ligação entre grupos de *tags*, ou seja, alguns pontos da rede servem como pontes para ligação entre grupos de *tags*.

Notou-se que efetivar ligações em rede dentro do contexto da Ciência da Informação torna um ambiente digital passível de muitos estudos e abordagens e que dessa forma o modelo Representação Iterativa pode ser aplicado também a outros ambientes digitais e não apenas em repositórios como apresentado neste trabalho.

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.

## Referências

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

BARQUÍN, B. A. R. et al. Projeto de ontologia para sistemas de informação empresariais: delineando uma metodologia para desenvolver ontologias na área de telecomunicações. **Brazilian Journal of Information Science**, Marília, v.2, n. 2, p. 17-34, jul./dez. 2008.

BUCKLAND, M. K. Description and search: Metadata as infrastructure. **Brazilian Journal of Information Science**, Marília, v. 0, n.0, p. 3-15, jul./dez. 2006. Disponível em <<http://www.bjis.unesp.br>>. Acesso em: 28 ago. 2009.

GUY, M.; TONKIN, E. Folksonomies: tidying up tags? **D-Lib Magazine**, Reston, v.12, n.1, Jan. 2006. Disponível em: <<http://www.dlib.org/dlib/january06/guy/01guy.html>>. Acesso em: 13 fev. 2009.

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.

MATHEUS, R. F.; SILVA, A. B. O. Fundamentação básica para análise de redes sociais: conceitos, metodologia e modelagem matemática. In: POBLACIÓN, D. A.; MUGNAINI, R.; RAMOS, L. M. S. V. C. **Redes sociais e colaborativas em informação científica**. São Paulo: Angellara, 2009. cap. 7, p. 239-287.

MOREIRA, A.; ALVARENGA, L.; OLIVEIRA, A. P. O nível do conhecimento e os instrumentos de representação: tesouros e ontologias. **DataGramZero: Revista de Ciência da Informação**, v.5, n. 6, dez. 2004. Disponível em: <[http://dgz.org.br/dez04/Ind\\_art.htm](http://dgz.org.br/dez04/Ind_art.htm)>. Acesso em: 3 dez. 2009.

O'REILLY, T. **What is web 2.0**: design patterns and business models for the next generation of software. 30 Sept. 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.

SALES, R. de; CAFÉ, L. Diferenças entre tesouros e ontologias. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.14, n.1, p.17-98, jan./ abr. 2009. Disponível em: <<http://www.eci.ufmg.br/pcionline/index.php/pci/article/view/646/541>>. Acesso em: 3 jan. 2010.

SANTAREM SEGUNDO, J. E. **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.

SANTOS, P. L. V. A. C.; ALVES, R. C. V. Metadados e Web Semântica para estruturação da Web 2.0 e Web 3.0. **DataGramZero**: Revista de Ciência da Informação, v.10, n. 6, dez. 2009. Disponível em: <[http://www.datagramzero.org.br/dez09/Art\\_04.htm](http://www.datagramzero.org.br/dez09/Art_04.htm)>. Acesso em: 03 dez. 2009.

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

WASSERMAN, S.; FAUST, K. **Social network analysis**: methods and applications. Cambridge: Cambridge University Press, 1994.

Artigo submetido em: 28 mar. 2011

Artigo aceito em: 15 abr. 2011