



# TEXTO DIGITAL

Revista de Literatura, Linguística e Artes

## O uso de elementos semânticos no processo de recuperação da informação em ambientes digitais

*The use of semantic elements in the process of recovering in digital environments*

José Eduardo Santarem Segundo<sup>a</sup>

<sup>a</sup> Universidade de São Paulo, Brasil – santarem@usp.br

**Palavras-chave:**

Repositórios Digitais. Ambientes Digitais. Elementos Semânticos.

**Resumo:** O presente artigo trata do uso de elementos semânticos no processo de armazenamento e recuperação de informação em repositórios digitais através do processo de Folksonomia Assistida e do modelo Representação Interativa.

**Keywords:**

Digital Repositories. Digital Environments. Semantic Elements.

**Abstract:** This article focuses on the use of semantic elements in processes and procedures for storing and retrieving information in digital repositories through Assisted Folksonomy and the Interactive Representation model.



## APRESENTAÇÃO

O capítulo que apresento neste livro é fruto da palestra, com mesmo título, apresentada no II Simpósio Internacional e VI Simpósio Nacional de Literatura e Informática, organizado pelo NUPILL - Núcleo de Pesquisas em Informática, Literatura e Linguística, na Universidade Federal de Santa Catarina.

A estrutura deste capítulo resume parte da pesquisa, em andamento, deste pesquisador que estuda a viabilização de técnicas e elementos semânticos na caracterização de uma estrutura semântica para repositórios digitais.

A pesquisa iniciou-se com o desenvolvimento de um modelo teórico que resultou na tese de doutorado “Representação Iterativa: um modelo para repositórios digitais”. A tese foi premiada como a melhor na área pela ANCIB (Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação), no ano de 2011. Recebeu ainda a menção honrosa no Prêmio Capes de Tese na área de Ciências Sociais Aplicadas, também em 2011.

O modelo teórico vem ganhando traços de implementação e novos testes para que seja homologado por completo em breve, porém ainda não há previsão de quando a pesquisa seja realmente finalizada com todos os itens descritos no modelo teórico.

A motivação originária para o desenvolvimento desta pesquisa nasceu da afinidade deste pesquisador com os repositórios digitais e principalmente pela forte relação com o uso da plataforma *Dspace*. Também foi motivador o contato com a Web Semântica, que vem sendo pesquisada e abordada desde o ano de 2004, quando ainda finalizava o mestrado usando este tema.

Portanto, a proposta apresentada neste capítulo tem como objetivo principal abordar o contexto dos ambientes digitais, com ênfase para os repositórios digitais, de forma que se busque evidenciar o uso de elementos semânticos como facilitador no processo de armazenamento e recuperação da informação, através do processo de Folksonomia Assistida e do modelo Representação Iterativa.

## INTRODUÇÃO

Grande parte do conteúdo textual e imagético de todo o mundo tem sido construído em formato digital. Com raras exceções, como as cadernetas que marcam o “fiado” da quitanda, receitas de bolo, algumas prescrições e prontuários médicos, contratos e um ou outro tipo de documento que ainda persiste no formato analógico. Armazenar documentos em formato não eletrônico deixou de ser uma tarefa trivial no cotidiano das pessoas.

Em um passado recente ainda era possível notar que grande parte dos documentos já eram produzidos de forma digital, mas ainda continuavam tramitando de forma impressa. Não era raro observar as pessoas carregando documentos que haviam sido elaborados em formato digital, mas que eram transformados em formato impresso para que pudessem chegar a seus destinatários.

Não há como negar a presença das tecnologias de informação e comunicação (TIC's) nos modelos de gerenciamento de informação atuais. A Internet tornou-se o principal meio condutor de informações e a informação digital tem se tornado muito mais rápida e efetiva, em comparação com a informação em suporte impresso.

A explosão do uso da Internet, do armazenamento em nuvem e principalmente o uso de dispositivos móveis, tem mudado de forma drástica a maneira de construir, armazenar e principalmente recuperar conteúdo informacional. Essa mudança nos processos que envolvem o fluxo informacional são muito representativas e nos apresentam um modelo totalmente diferente de lidar com a informação. Ela deixou de ser analógica e passou para o modelo digital.

Atualmente é possível verificar que instituições e organizações variadas têm tomado decisões a respeito de processos que envolvem o uso da informação em formato digital. Nota-se, inclusive, que grande parte dessas instituições tem se preocupado em fazer com que os documentos que ainda persistem em continuar no formato papel voltem a ser convertidos para o formato digital através de processos de digitalização.

No que tange à produção de documentos em formato papel, ele se torna inviável em vários sentidos, desde a produção, quando exige alto custo de papel e *toner*/tinta para imprimir,

caracterizando-se contra os atuais modelos de sustentabilidade, passando pela guarda que exige cada vez mais espaço físico e indo até o processo de recuperação que em muitos casos envolve perda de tempo e principalmente impossibilita que o conteúdo de um documento seja compartilhado entre várias pessoas ao mesmo tempo (SANTAREM SEGUNDO; DONADI, 2013).

Esse cenário indica claramente que a Ciência da Informação, responsável por pesquisas que envolvem estruturas de dados e fluxos informacionais, tem um papel preponderante na construção de modelos que facilitem o acesso à informação nesse novo formato digital.

É por todo esse novo contexto, essa mudança de paradigma, que a organização e a recuperação da informação, principalmente quando esta se encontra na Internet, tem sido pauta de muitas das pesquisas que envolvem tecnologia da informação. Entretanto, esses desafios não param por aí. A partir do momento em que se produz uma quantidade exacerbada de informação em formato digital, é necessário que essas informações estejam não apenas disponíveis, mas também integradas com documentos e objetos digitais que têm relação direta com o conteúdo, através de relações semânticas.

Considera-se relação semântica entre dois dados distintos qualquer tipo de ligação realizada pelo cérebro humano que possa associar estes dados a partir de um determinado cenário.

Ressalta-se que, neste texto, o uso de elementos e processos que tratam da questão semântica serão sempre aplicados em ambientes digitais do tipo repositório.

A abordagem pelos repositórios é dada devido à inserção que esta ferramenta tem tido no contexto atual. As instituições têm optado por modelos de gestão de conteúdo digital que primam pelo uso de técnicas difundidas na Ciência da Informação, como a descrição, o uso de metadados para interoperabilidade, a indexação, o uso de vocabulários controlados, entre outros, e para isso se apoiam no uso de repositórios.

## **REPOSITÓRIOS DIGITAIS**

Repositórios digitais são ambientes de informação constituídos sobre ferramentas/plataformas tecnológicas, que também podem ser chamadas de sistema de informação, capazes de receber

depósitos de objetos digitais, em vários formatos, sejam eles realizados pelos próprios autores ou por equipes treinadas para este fim, de modo a armazenar e organizar estes objetos para que possam ser recuperados e principalmente preservados por longa data.

Os repositórios digitais, sejam eles temáticos ou institucionais, apresentam características semelhantes, possuem uma estrutura comum de submissão e acesso às informações e são desenvolvidos segundo padrões de interoperabilidade específicos, que potencializam o uso desses sistemas para agregação e divulgação da informação digital. Algumas das características ou observações quanto à estrutura destes repositórios são impostas pelos padrões que utilizam (SANTAREM SEGUNDO, 2010, tese).

De acordo com os objetivos propostos, um repositório digital pode contemplar uma infinidade de tipos de documentos, ou seja, ter uma tipologia variada de documentos. Mesmo com a concepção de que os repositórios foram desenvolvidos para divulgar documentos já publicados nos meios tradicionais, não há uma delimitação sobre os tipos de documentos que atualmente podem compor um repositório (CAFÉ, 2003).

Bekaert e Van de Sompel (2006) definem repositórios como:

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.

Difundidos principalmente na organização de material científico e utilizados em larga escala pelas universidades e institutos de pesquisa, conforme vimos nas definições apresentadas, os repositórios têm chamado também atenção de outros tipos de organização, devido à capacidade de gerenciar grandes volumes de dados facilitando o processo de armazenamento e recuperação da informação.

Para a implementação técnica são vários os *softwares*/plataformas 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<sup>1</sup>, GNU E-prints<sup>2</sup>, OPUS<sup>3</sup>, Open Repository<sup>4</sup>, DiVA<sup>5</sup>, Fedora<sup>6</sup>, Omeka<sup>7</sup>.

Durante esta pesquisa ficará evidente a necessidade de agregar elementos e recursos de Web 2.0 e de Web 3.0 nas ferramentas que implementam repositórios, visto que não é possível identificar esses elementos nas versões originais disponibilizadas.

Da necessidade de uso dos recursos de Web 3.0, principalmente os elementos, linguagens e tecnologias apresentadas como essenciais no “bolo de noiva” da Web Semântica, desenvolvido por Tim Berners-Lee, e da Folksonomia, elemento básico da Web 2.0, é que se constrói grande parte desta pesquisa.

## 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”.

A folksonomia é considerada elemento fundamental 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 que a Folksonomia se caracteriza como uma forma de inserir e relacionar recursos por meio da descrição dos mesmos através de palavras-chave, de forma aberta, que tem como

1 <http://www.dspace.org/>

2 <http://www.eprints.org/software/>

3 <http://opus.bath.ac.uk/>

4 <http://www.openrepository.com/>

5 <http://www.diva-portal.org>

6 <http://www.fedora-commons.org/>

7 <http://omeka.org>

principal objetivo facilitar o processo de gerenciamento e recuperação das informações em ambientes digitais.

O fato da 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*, 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 da Web 2.0, 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 formam 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).

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.

Como já fora mencionado, a pesquisa 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, 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 repositório 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 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.

O recurso de “sugestão” 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. Esta estrutura deverá ser caracterizada como uma ontologia.

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 ontologia, que são instrumentos que permitem uma busca hierárquica de termos em formato vertical e horizontal. Nesta pesquisa utilizar-se-á uma ontologia, descrita em linguagem *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 no qual 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 “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 os quais deposita, porém de forma que possa se amparar no próprio conhecimento já disponível no ambiente em que está utilizando.

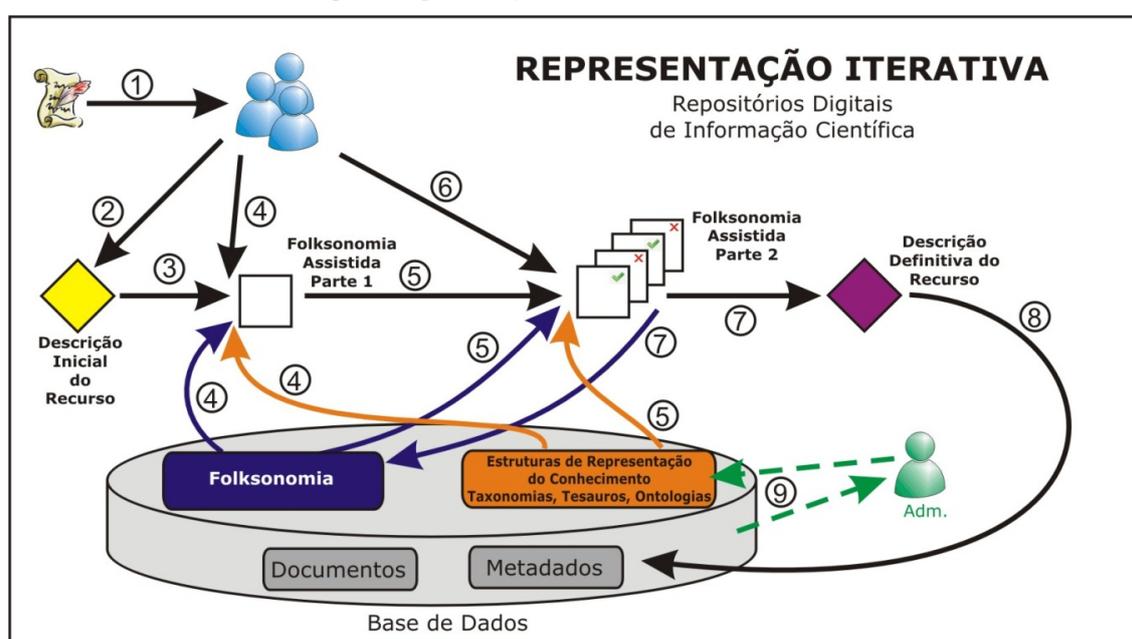
Ressalta-se que no ano de 2012 houve uma homologação do processo de folksonomia assistida, através do desenvolvimento de um modelo teórico-conceitual. Nesta homologação, utilizou-se um conjunto de palavras-chave de cinquenta artigos e fora possível verificar, entre outros resultados, uma diminuição de 12,79% das palavras-chave em relação ao depósito inicial. Os resultados homologam a vantagem de uso da folksonomia assistida, ao possibilitar a visualização da congregação dos relacionamentos entre as *tags* anteriormente dispersas pelas inconsistências. De forma cooperativa, desempenham um papel valioso no gerenciamento do vocabulário utilizado na representação informacional, em aspectos que contribuem para a não redundância de informação, e na confiabilidade da pesquisa e recuperação dos artigos pelos usuários do repositório digital (SANTAREM SEGUNDO; SIQUEIRA, 2013).

## ITERATIVIDADE, A RETROALIMENTAÇÃO DA INFORMAÇÃO

A implementação do processo de folksonomia assistida é a base para a consolidação da Representação Iterativa, que deve ser retroalimentada, sempre baseada no contexto de uma estrutura de representação do conhecimento, através de uma ontologia, que consiste em definir os limites de um domínio do conhecimento.

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

Fig. 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 realizada 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 tesauros) 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 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 a retroalimentação da estrutura de representação do conhecimento fique a cargo de um usuário administrador, 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.

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 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 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 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 das quais 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 co-autorias e co-citações. Porém, a mesma ideia utilizada neste conceito se aplica 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 co-autoria 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 gráfico, que é a estrutura matemática e computacional escolhida para representar as redes.

A apresentação do gráfico possibilita algumas variações. 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.

A implementação do processo de recuperação ainda não está finalizado, porém a indicação é do uso das estruturas de gráfico e linguagem *Sparql* para recuperação da informação através do uso de ontologias.

## **CONSIDERAÇÕES**

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 modelo conceitual ainda está em processo de implementação e vem evoluindo conforme o tempo. Destaca-se que já é possível perceber novas pesquisas que se apoiam no modelo proposto para concepção de outros trabalhos com características parecidas.

Dessa maneira, entende-se que tanto o processo, já homologado, de folksonomia assistida, quanto todo o modelo nomeado Representação Iterativa, cumprem com o propósito de qualificar a produção e recuperação de informação em contextos semânticos aplicados em repositórios digitais.

Essas características apresentadas indicam claramente a possibilidade de apoio de estruturas como ontologias para a concepção de modelos semânticos em repositórios digitais, facilitando o processo de organização da informação e principalmente da recuperação desta através de relações semanticamente estruturadas.

## 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: 3 set. 2013.

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 set. 2013.

CAFÉ, L. et al. Repositórios institucionais: nova estratégia para publicação científica na Rede. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 26. 2003, Belo Horizonte. **Anais...** Belo Horizonte: INTERCOM, 2003. Disponível em: <[http://www.intercom.org.br/papers/nacionais/2003/www/pdf/2003\\_ENDOCOM\\_TRABALHO\\_cafe.pdf](http://www.intercom.org.br/papers/nacionais/2003/www/pdf/2003_ENDOCOM_TRABALHO_cafe.pdf)>. Acesso em: 2 set. 2013.

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: 15 set. 2013.

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.

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.

SANTAREM SEGUNDO, J. E.; DONADI, R. O profissional da informação no contexto das TICS's: aplicação do conceito de WIKI em empresas de desenvolvimento de software.

In: CASTRO FILHO, C. M. (Org.). **Olhares sobre o profissional da Ciência da Informação**. São Paulo: Editora Todas as Musas; 2013. p. 163-184.

SANTAREM SEGUNDO, J. E.; SIQUEIRA, C. S. Aplicación teórico-conceptual de folksonomías asistidas para la recuperación de información. **Scire**, Zaragoza, v. 19, n. 2, p. 77-82, 2013.

WAL, T. V. **Folksonomy definition and wikipedia**. Disponível em: <<http://www.vanderwal.net/random/entrysel.php?blog=1750>>. Acesso em: 2 set. 2013.