

UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

José Valderlei da Silva

Gerenciamento do vocabulário do usuário em sistemas baseados em *tagging*

MARINGÁ

2009

José Valderlei da Silva

Gerenciamento do vocabulário do usuário em sistemas baseados em *tagging*

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Estadual de Maringá, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Área de concentração: Sistemas de Informação

Orientador: Prof. Dr. Sérgio Roberto P. da Silva

MARINGÁ

2009

"Dados Internacionais de Catalogação-na-Publicação (CIP)"
(Biblioteca Setorial - UEM. Nupélia, Maringá, PR, Brasil)

S436g Silva, José Vanderlei da, 1969-
Gerenciamento do vocabulário do usuário em sistemas baseados em *tagging* / José Vanderlei da Silva. – Maringá, 2009.
124 p. : il. (algumas color.)
Dissertação (mestrado em Ciência da Computação)--Universidade Estadual de Maringá, Dep. de Informática, 2009.
Orientador: Prof. Dr. Sérgio Roberto P. da Silva.
1. Sistemas de informação - Classificação livre - Método *tagging*. 2. Sistemas de informação - Persomania. 3. Gerenciamento de vocabulário. I. Universidade Estadual de Maringá. Departamento de Informática. Programa de Pós-Graduação em "Ciência da Computação".

CDD 22. ed. -005.15
NBR/CIP - 12899 AACR/2

UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

José Valderlei da Silva

Gerenciamento do vocabulário do usuário em sistemas baseados em *tagging*

MARINGÁ

2009

José Valderlei da Silva

Gerenciamento do vocabulário do usuário em sistemas baseados em *tagging*

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Estadual de Maringá, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Área de concentração: Sistemas de Informação

Orientador: Prof. Dr. Sérgio Roberto P. da Silva

MARINGÁ

2009

José Valderlei da Silva

Gerenciamento do vocabulário do usuário em sistemas baseados em *tagging*

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Estadual de Maringá, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Aprovado em 19 / 02 / 2009.

BANCA EXAMINADORA

Prof. Dr. Sérgio Roberto Pereira da Silva
Universidade Estadual de Maringá – DIN/UEM

Profa. Dra. Valéria Delisandra Feltrin
Universidade Estadual de Maringá – DIN/UEM

Profa. Dra. Lucia Vilela Leite Filgueiras
Escola Politécnica da USP

AGRADECIMENTOS

Agradeço e dedico este trabalho, em primeiro lugar, à minha querida esposa **Cristiane Miriam Carlos**, que muito me inspirou e me apoiou durante esta jornada de aprendizado, e ao meu filho, Christian Roberto da Silva, que nasceu durante essa fase.

Uma pessoa a qual tenho orgulho de chamar de meu irmão “de coração”, Roberto Pereira, foi quem esteve ao meu lado durante todo o trajeto do mestrado e com certeza sem ele e seu apoio, talvez, nem chegasse a esse ponto. Por essa razão eu escrevo em letras maiúsculas: OBRIGADO MEU IRMÃO ROBERT.

Meu especial agradecimento ao meu orientador Professor Sérgio, que me recebeu como orientando ensinando sobre a verdadeira “arte da pesquisa” e apoiando nos momentos em que precisei.

Uma turma não precisa ser grande em número, ela precisa de gigantes como membros. Eu cheguei até aqui por pertencer a uma pequena turma de gigantes. Obrigado queridos companheiros da turma de 2006. Nessa turma, duas guerreiras, Cris e Dani, merecem um destaque todo especial por serem minhas amigas e, também, por serem fiéis aos seus ideais.

Companheiros e parceiros de luta, aqueles que abandonariam o barco somente após seu afundamento total, Filipe Côgo e Júlio Marcelo Tófoli. Essas pessoas posso chamar de amigos.

Agradeço a querida Inês, um ser que nasceu para ajudar o próximo. Ela está sempre com o coração aberto para receber os novos alunos, seus futuros e eternos amigos. Agradeço a Ângela que ajudou a colocar ordem nas palavras, com o objetivo de facilitar a leitura dos futuros leitores e também à querida amiga Késsia, que sempre esteve presente.

E, por fim, estendo meu agradecimento a todos que, de alguma forma, colaboraram direta ou indiretamente nesta fase de estudos e crescimento.

“Seus filhos foram sua vida. Minha vida, eu devo a você, por isso todas as minhas realizações são em seu nome. Sinto saudades, querida mamãe.”

Gerenciamento do vocabulário do usuário em sistemas baseados em tagging

RESUMO

A diversidade de sistemas baseados em *tagging* existentes, cada um centrado em um tipo de objeto ou recurso, obriga os usuários a manterem seus vocabulários espalhados por diversos sistemas que adotam essa abordagem, o que prejudica a manutenção de um vocabulário de termos consistente e dificulta os processos de categorização e de recuperação de informação. Por essa razão, nesta dissertação definimos uma arquitetura para um sistema de gerenciamento das *tags* utilizadas nos sistemas de categorização. A arquitetura proposta tem como principal objetivo o de tornar-se um repositório central e único para o vocabulário de *tags* que o usuário utiliza para categorizar e recuperar conteúdos em sistemas baseado em *tagging*. O gerenciador de *tags* auxiliará o usuário na organização e no uso do seu vocabulário pessoal, reduzindo os problemas e a complexidade em lidar com diversos sistemas e vocabulários simultaneamente. A intenção é melhorar a qualidade das categorizações e, conseqüentemente, da recuperação de informação em sistemas baseados em *tagging*. Nesta dissertação são expostas as discussões teóricas e práticas sobre a ferramenta de gerenciamento de *tags* proposta, a arquitetura desse sistema e exemplos de testes que indicam a sua viabilidade.

Palavras-chave: *Tagging*. Personomia. Gerenciamento do vocabulário.

Management of the user's vocabulary in tagging-based systems

ABSTRACT

The diversity of existent tagging-based systems, each one centered in a distinct kind of object or resource, force users to maintain their vocabulary dispersed among several systems that adopt that approach, harming the maintenance of a well organized vocabulary of terms and hindering the processes of information categorization and retrieval. For this reason, in this dissertation we defined an architecture for a system to the management of tags used in categorization systems. The main goal of the proposed architecture is to become a central and unique repository for the vocabulary of tags that users apply to categorize and to retrieve contents in tagging-based systems. The tag manager will aid the users in the organization and in the use of their personal vocabulary, reducing the difficulties and complexity in simultaneously working with several systems and vocabularies. The intention is to improve the quality of the categorizations and, consequently, of the information retrieval in tagging-based systems. In this dissertation, it is exposed the theoretical and practical discussions about the proposed system from tags management, its architecture and text examples that indicate its viability.

Keywords: *Tagging*. Personomy. Vocabulary Management.

SUMÁRIO

1	Introdução	23
2	Tagging, Personomia e Folksonomia	31
2.1	Categorização	32
2.2	O Processo de <i>Tagging</i>	33
2.3	As Personomias.....	36
2.4	A técnica de folksonomia – <i>Tagging</i> coletivo	38
2.4.1	A folksonomia larga.....	40
2.4.2	A folksonomia estreita	41
2.4.3	O uso da folksonomia como ferramenta coletiva	43
2.4.4	Vocabulário não-controlado.....	44
2.4.5	Navegação social usando <i>tagging</i>	49
2.5	Os sistemas baseados em folksonomia	49
2.5.1	Alguns sistemas baseados em folksonomia.....	54
3	O Problema das Múltiplas Personomias e a Proposta de Solução.....	59
3.2	Problema versus solução.....	60
3.3	Gerenciamento do ciclo de vida de uma <i>tag</i>	62
3.4	Gerenciamento das <i>tags</i> das personomias de um usuário	66
3.5	O vocabulário de um usuário	67
3.6	As múltiplas personomias de um usuário	71
4	Proposta de Arquitetura para o Gerenciador de Tags.....	77
4.1	O Mapa conceitual (metamodelo) do gerenciador de <i>tags</i>	78
4.2	Necessidades do usuário	79
4.3	Requisitos.....	81
4.3.1	Requisitos funcionais	81
4.3.2	Requisitos não-funcionais.....	82
4.4	Casos de uso.....	85
4.5	Diagramas Atividade	89
4.6	Diagrama arquitetural	91
5	Aspectos de Implementação da Arquitetura Proposta	97
5.1	Tecnologias utilizadas na implementação da arquitetura	98
5.2	Codificação da arquitetura	99
5.2.1	Acesso a dados dos sistemas externos	99
5.3	Acessos externos ao gerenciador de <i>tags</i>	106
5.4	Testes de sincronização.....	106

6	Conclusão	111
7	Bibliografia.....	117
8	Anexo I – Tabelas Sistemas Baseados em Folksonomia.....	121

LISTA DE FIGURAS

<i>Figura 1 - Os três pivôs do processo de tagging</i>	34
<i>Figura 2 – O Usuário possui um vocabulário e utiliza parte dele como tags (tag A, tag B, tag D, tag G e tag I) que marcam os recursos de interesse pessoal.</i>	36
<i>Figura 3 – Modelo de um sistema de categorização colaborativa. Usuário, tag e recurso são elementos que compõem uma categorização</i>	37
<i>Figura 4 – Folksonomia larga – os usuário utiliza tag(s) para categorizar objeto(s).</i>	41
<i>Figura 5 - modelo de um sistema de folksonomia estreita (os diferentes tipos de seta indicam diferentes usuários e suas relações).....</i>	42
<i>Figura 6 - Distribuição do uso de tags em uma categorização. O efeito de “cauda longa” (long tail) caracteriza os níveis abruptos de decaimento no uso de cada tag.</i>	51
<i>Figura 7 - Visualização das tags em forma de grafo, mostrando as relações a partir de um nó central “complexity”</i>	53
<i>Figura 8 - Nuvens de tags (tag cloud), esquema de visualização de tags do usuário.</i>	53
<i>Figura 9 - Sistema Delicious - gerenciamento de bookmarks on-line</i>	55
<i>Figura 10 - Sistema Flickr - Possibilitando a publicação e organização de fotos, bem como rede de contatos.</i>	56
<i>Figura 11 - Sistema Youtube - Compartilhamento de vídeo baseado em tagging</i>	57
<i>Figura 12 - Sistema Slideshare - Compartilhando slides de apresentação</i>	58
<i>Figura 13 - A diversidade de sistemas baseado em folksonomia.....</i>	60
<i>Figura 14 - Camada de tags sobre os sistemas baseados em folksonomia.....</i>	61
<i>Figura 15 - Modelo de um tagging com ênfase no gerenciamento do ciclo de vida de uma tag</i>	63
<i>Figura 16 - Relação vocabulário e tag</i>	69
<i>Figura 17 - Perspectivas de um sistema de tagging - figura adaptada de Smith (2008).....</i>	72
<i>Figura 18 - Múltiplas Personomias de um usuário.....</i>	75
<i>Figura 19 - Mapa conceitual do gerenciador de tags.....</i>	78
<i>Figura 20 - Sistema Tag Manager</i>	85
<i>Figura 21 - Caso de uso “Cadastrar sistemas”</i>	86
<i>Figura 22 - Caso de uso “Sincronizar tags”</i>	87
<i>Figura 23 - Caso de uso “Sugerir tags”</i>	87

<i>Figura 24 - Caso de uso “Recuperar informação”</i>	88
<i>Figura 25 - Diagrama de atividades para o caso de uso “Cadastrar Sistemas”</i>	89
<i>Figura 26 - Diagrama de atividades para os casos de uso “Sincronizar tags” e “Recuperar Informações”</i>	90
<i>Figura 27 - Diagrama de atividades para o caso de uso “Recomendar Categorização”</i>	91
<i>Figura 28 - Visão da arquitetura do sistema gerenciador de tags</i>	92
<i>Figura 29 - Diagrama das tabelas necessárias para o gerenciamento de tags</i>	94
<i>Figura 30 - Fluxo de atividades para integração do Tag Manager com outro sistema baseado em folksonomia</i>	102
<i>Figura 31 - parte do código XML disponibilizados pela API do sistema Delicious</i>	104
<i>Figura 32 - Código fonte de uma página do sistema Delicious</i>	105
<i>Figura 33 - Login via web no gerenciador de tags</i>	107
<i>Figura 34 - Sincronização das tags</i>	108
<i>Figura 35 - Solicitação de senha ao usuário pelo gerenciador de tags</i>	108

LISTA DE QUADROS E TABELAS

<i>Tabela 1 - Cálculos de frequências de uso da tag em categorização, de busca, de acesso e de uso geral</i>	64
<i>Tabela 2- Enumeração dos casos de uso</i>	86
<i>Tabela 3 - Descrição essencial do caso de uso "Cadastrar sistemas"</i>	87
<i>Tabela 4 - Descrição essencial do caso de uso "Sincronizar tags"</i>	87
<i>Tabela 5 - Descrição essencial do caso de uso "sugerir tags"</i>	88
<i>Tabela 6 - Descrição essencial do caso de uso recuperar informação</i>	88
<i>Tabela 7 - Mapeamento do acesso as tags do usuário em sua personomia</i>	101

LISTA DE ABREVIATURAS E SIGLAS

URL	<i>Uniform Resource Locator</i>	23
API	<i>Application Programming Interface</i>	82
JSON	<i>JavaScript Object Notation</i>	83
XML	<i>Extensible Markup Language</i>	83
HTTP	<i>Hypertext Transfer Protocol</i>	84

Introdução

A partir de meados de 2002 o uso de rótulos de texto para categorizar informação na *Web* cresceu muito em popularidade (SMITH, 2008). Nesse período, muitos *websites* passaram a permitir que seus usuários efetuassem anotações ou utilizassem palavras simples (etiquetas, ou *tags*) na forma de um texto não-estruturado em um processo denominado *tagging*. Essas são anotações individuais utilizadas pelas pessoas para atribuírem algum “significado” ao conteúdo de uma URL (*Uniform Resource Locator*), que poderia ser uma página *Web*, uma foto, um vídeo, um *post* de um *blog* ou qualquer outro recurso (objeto) que pudesse ser referenciado por uma URL. A utilização do processo de *tagging* gera uma *personomia*: um conjunto de *tags* criadas e utilizadas por uma pessoa para a organização de recursos de interesse pessoal.

A prática de categorizar objetos popularizou-se com os serviços oferecidos pela *Web 2.0*, principalmente por sistemas como o *Flickr*¹ e o *Delicious*², que passaram a permitir que seus usuários indexassem e compartilhassem seus dados de forma prática e aberta (KIM, *et al.*, 2008). Esses sistemas adicionaram elementos sociais de compartilhamento de informação e de interação entre os usuários ao processo de *tagging*, caracterizando uma evolução (ou especialização) desse processo. Em 2004, Thomas Vander Wal (MATHES, 2004) utilizou o neologismo folksonomia (*folksonomy*, *folk* + *taxonomy*) para designar essa evolução, representando a técnica no qual as pessoas utilizam termos para categorizar, organizar ou atribuir significado aos recursos disponíveis na *Web* de forma compartilhada e livre.

A técnica de folksonomia é uma iniciativa para auxiliar na estruturação e na organização da informação na *Web* (PEREIRA, *et al.*, 2008b). Na prática, a folksonomia não possui nada de novo ou de inovador; a diferença fundamental entre sistemas que aplicam *tagging* e sistemas que aplicam folksonomia está no fato de o processo de *tagging* resultar em uma personomia do usuário P_u ³, enquanto a técnica de folksonomia é o resultado de um conjunto de personomias compartilhadas, que possibilita a emergência de um vocabulário social sobre determinada área ou assunto. Portanto, uma folksonomia nada mais é do que uma coleção de personomias (HOTHO, *et al.*, 2006). Para Cattuto *et al.* (2008), as personomias constituem o núcleo da folksonomia que, por sua vez, as trata como se fossem nós interligados por meio de *tags* e recursos. Conseqüentemente, para alcançar seus resultados e para gerar algum valor, a técnica de folksonomia fundamenta-se no compartilhamento e na emergência de comportamentos e de estruturas sociais, bem como em relacionamentos de estruturas lingüísticas e conceituais da cultura de uma linguagem.

¹ <http://flickr.com>

² <http://del.icio.us>

³ Personomia é definida por Hotho *et al.* (2006) como um conjunto de tuplas (T_u, R_u, Y_u) , sendo T_u as tags do usuário, R_u os recursos do usuário e Y_u relações entre as tags T_u e recursos R_u que forma um espaço utilizado pelo usuário para a categorização e organização de seus recursos (objetos) pessoais

A técnica de folksonomia representa o primeiro dos quatro pontos de tensão do processo de *tagging* citados por Smith (2008): o “Pessoal versus o Social⁴”. De acordo com o autor, os assuntos relacionados a tensão entre o paradigma individual versus o paradigma global levam a questões como: qual é o principal objetivo que norteia a categorização de objetos? Qual é a natureza social da categorização por meio de *tags*? Os usuários realizam categorizações para benefício próprio, ou eles são motivados pelo desejo de compartilhar informação, opiniões ou conhecimentos? Como alguém pode saber o que é o melhor para cada um dos paradigmas? As respostas para essas questões exigem uma análise detalhada sobre as aplicações e sobre os objetivos tanto do paradigma individual quanto do paradigma global.

Segundo Weinberger (2005), as *tags* estão se tornando uma nova camada de infraestrutura para a organização de informação na *Web*. Essa afirmação tem validade quando o foco está em apenas um sistema baseado em folksonomia, como o *Delicious*. A razão disso é que cada sistema que implementa o conceito de folksonomia cria, dentro de seus limites, uma camada de *tags* para o usuário. Essa camada de *tags* é local (isto é, pertence ao domínio do sistema) e, quando o usuário permite, suas *tags* e recursos são compartilhados com outras pessoas que utilizam o sistema. Essa camada não está disponível fora dos limites de um sistema, de modo que não é possível utilizar o mesmo vocabulário de *tags* em mais que um sistema baseado em folksonomia (Por exemplo: utilizar o vocabulário de *tags* do sistema *Delicious* no sistema *SlideShare*⁵). Normalmente, os sistemas baseados em folksonomia estão focados em um único objeto de interação (SILVA e DA SILVA, 2008): um *bookmark*, no caso do *Delicious*, uma foto, no caso do *Flickr*, ou um vídeo, no caso do *YouTube*⁶. Desse

⁴ Os três outros pontos são: 2 – Idiossincrasia versus Padronização, 3 – Liberdade versus Controle e 4 – Amadores versus *Experts*.

⁵ <http://www.slideshare.net>

⁶ <http://youtube.com>

modo, um usuário que utilize mais de um sistema baseado em folksonomia deverá, obrigatoriamente, possuir uma personomia para cada um dos sistemas utilizados.

Novamente, relacionado ao ponto de tensão do “Pessoal versus o Social” (SMITH, 2008), é possível observar algumas vantagens e desvantagens da aplicação da técnica de folksonomia. Se de um lado a liberdade concedida ao usuário na categorização e o compartilhamento dessas categorizações são responsáveis pelo sucesso dos sistemas baseados em folksonomia, de outro, essa liberdade pode trazer problemas, tais como: a ambigüidade originada da ocorrência de termos polissêmicos (quando uma palavra possui múltiplos significados), a sinonímia (devido ao fato de muitas palavras serem empregadas para representar o mesmo significado), o uso do plural e do singular para referenciar o mesmo tipo de objeto, erros de digitação na construção dos termos, junção de palavras com separadores (traços, sublinhados, iniciais maiúsculas), entre outros. Além disso, devido à heterogeneidade dos usuários de um sistema baseado em folksonomia, o compartilhamento das personomias desses usuários tende a causar uma sobrecarga de informação ao usuário quando o número de usuários e das informações sendo categorizadas começa a aumentar.

No contexto dos problemas mencionados acima, o fato do usuário ser levado a manter uma personomia independente em cada sistema baseado em folksonomia influencia negativamente na qualidade do vocabulário de termos utilizados e, conseqüentemente, das categorizações efetuadas por esse usuário — tanto do ponto de vista individual quanto do ponto de vista coletivo ou social. Isso ocorre porque o usuário terá dificuldades em gerenciar seu próprio vocabulário de termos utilizados nas categorizações, uma vez que esse vocabulário está disperso em diversos sistemas. Assim, há uma maior carga cognitiva e exige-se um maior esforço do usuário para que seja possível manter certa coerência e compatibilidade entre os termos utilizados e entre os significados concebidos para cada um. De forma genérica, pode-se dizer que a liberdade que facilita a categorização da informação

dificulta o processo de organização e de recuperação, uma vez que com um vocabulário não controlado e não centralizado, o usuário continuará com dificuldades para encontrar o conteúdo categorizado, já que terá dificuldade em manter um vocabulário organizado. O problema, quando considerado do ponto de vista coletivo, se torna mais evidente, pois a idiossincrasia das *tags*, ou seja, as particularidades do usuário colaboram para que sejam encontradas raridades, mas, também, dificulta a tarefa do usuário em identificar e escolher um termo adequado para seu propósito.

Guy e Tonkin (2006) afirmam que os usuários de sistemas que aplicam a técnica de *tagging* ou de folksonomia devem ser educados a adicionar *tags* melhores, ou seja, que reduzem os problemas mencionados anteriormente. Wu (2006) concorda com essa afirmação e expõe que é necessário desenvolver recursos que visem minimizar os efeitos negativos da folksonomia e oferecer apoio aos usuários tornando-os mais produtivos. Côgo e Da Silva (2008) propõem um sistema para a detecção de inconsistências em *tags*, representando uma iniciativa para auxiliar no controle e na organização das personomias. Já a proposta de Kim *et al.* (2008) consiste na criação de um vocabulário, denominado SCOT (*Social Semantic Cloud of Tags*), com o objetivo de representar o contexto de *tags* em ambientes sociais, provendo métodos para compartilhar e reutilizar recursos categorizados. Finalmente, o sistema *My.Tag*⁷ permite a recuperação de informação dos sistemas *Flickr*, *Youtube*, *Delicious*, *Connotea*⁸, *BibSonomy*⁹ e *BibTex*¹⁰. No *My.Tag*, um usuário pode manter um cadastro e recuperar seus objetos categorizados nos sistemas previamente mencionados (BRAUN, et al., 2008). Essa abordagem representa uma iniciativa de centralização e aproveitamento dos esforços executados pelos usuários no processo de categorização, entretanto, é focada apenas na recuperação de informação em algumas folksonomias.

⁷ <http://mytag.uni-koblenz.de/>

⁸ <http://www.connotea.org/>

⁹ <http://www.bibsonomy.org/>

¹⁰ <http://www.bibtex.org/>

Mesmo com as iniciativas mencionadas, ainda não existe um recurso que possibilite a centralização de todas as personomias de um usuário e o gerenciamento dos termos e dos recursos que compõem essas personomias. Baseado no fato de que a *tag* é o pivô central em todos os sistemas que utilizam a técnica de folksonomia, a presente pesquisa propõe a criação de um sistema para o gerenciamento das personomias originadas pelo trabalho de categorização dos usuários em diversos sistemas baseados em folksonomia. O gerenciador acessará a base de dados do usuário de todos os sistemas baseados em folksonomia que ele utiliza, trazendo para um repositório central suas *tags*. A partir desse ponto o gerenciador passará a controlar o uso do vocabulário do usuário de forma globalizada. Acredita-se que, além da praticidade de manutenção dos termos utilizados para a categorização de objetos, um recurso que centralize o vocabulário das personomias colaborará para aumentar a consistência e a padronização dos termos utilizados por um usuário e, conseqüentemente, para a emergência de uma folksonomia mais organizada e de maior qualidade.

Esta dissertação está organizada do seguinte modo: no Capítulo II são descritos os principais conceitos envolvidos em *tagging* e folksonomia, suas vantagens e suas desvantagens, além de algumas pesquisas que estão sendo realizadas sobre os sistemas que aplicam esses conceitos. No Capítulo III são discutidas questões referentes ao gerenciamento de informação em sistemas baseados em folksonomia, focando na importância do gerenciamento de *tags*. Nesse capítulo, também são expostos alguns dos problemas relacionados às *tags* dispersas em vários sistemas baseados em folksonomia e apresenta-se uma proposta de solução. No capítulo IV são apresentadas as necessidades do usuário para o gerenciamento de suas *tags*, em seguida, são levantados os requisitos para um sistema de gerenciamento de personomia e, a partir desses requisitos, é definida a arquitetura do sistema gerenciador proposto. No Capítulo V são apresentadas algumas considerações importantes a respeito da implementação da arquitetura da solução desenvolvida e expõem-se discussões

sobre o modelo proposto e sobre aspectos relacionados à sua utilização. No capítulo VI, apresentam-se as considerações finais e apontam-se direções e idéias para trabalhos futuros.

[Página intencionalmente deixada em branco]

Tagging, Personomia e Folksonomia

Com a grande quantidade de informação disponível na *Web* são necessários mecanismos de indexação de conteúdo para auxiliar o usuário na recuperação da informação. Essa indexação, atualmente, é executada por sistemas especializados como o Google¹¹, *Yahoo*¹², *Altavista*¹³, entre outros, os quais utilizam algoritmos complexos que executam automaticamente a tarefa de indexar as páginas da *Web*. O processo de *tagging* é uma alternativa a esse tipo de indexação tradicional (STURTZ, 2004). Ele é efetuado de forma coletiva, possibilitando a atuação de pessoas como agentes ativos no processo de organização e de estruturação dos conteúdos disponíveis na *Web*. A idéia fundamental que motiva a existência dos sistemas baseados em *tagging* é muito simples: em vez de empregar algoritmos executando o processo de categorização de conteúdos, são os próprios usuários da *Web* com suas características, opiniões, conhecimentos e interesses

¹¹ <http://google.com>

¹² <http://yahoo.com>

¹³ <http://www.altavista.com>

distintos que fazem esta categorização. As pessoas atribuem os termos que desejarem aos objetos que desejarem de uma forma compartilhada e aberta. Neste capítulo, apresentaremos algumas discussões conceituais sobre o processo de *tagging* no contexto de organização, armazenamento e gerenciamento de conteúdos pessoais e coletivos na *Web*.

2.1 Categorização

Podemos dizer que categorizar um objeto é diferente de conceituá-lo. Um conceito é usado para referir-se a uma representação mental de um objeto e uma categorização remete à formação de combinação dos significados dos conceitos pelas associações baseada na similaridade entre eles (LIMA, 2007). Na seqüência, apresentamos alguns detalhes sobre a questão da categorização.

Em Rosch (1978) são apresentados os princípios da “economia cognitiva” e da “percepção da estrutura do mundo” que determinam a estrutura dos esquemas de categorização. O princípio da “economia cognitiva” contempla o fato de que a tarefa dos esquemas de categorização é o de prover o máximo de informação com o menor esforço cognitivo. O princípio da “percepção da estrutura do mundo” contempla o fato de que os objetos do mundo são percebidos com uma estrutura co-relacional alta, não apenas como um conjunto desestruturado de atributos co-ocorrentes. A vantagem do processo de categorização é a possibilidade de diferenciar as diversas propriedades dos objetos categorizados com um baixo esforço cognitivo, porém, pode-se desejar não diferenciar os atributos desses objetos quando as diferenças forem irrelevantes aos propósitos da categorização.

Outra questão levantada por Rosch (1978), referente à categorização, é a existência de um nível de abstração no qual os objetos de uma categoria são mais inclusivos. Este nível de abstração é formalizado em termos de métricas, tendo como princípio “a validade de sugestão” (*cue validity*). O **nível básico** é o nível mais inclusivo em uma categoria,

maximizando o número de atributos comuns entre os membros desta categoria e minimizando o número de atributos compartilhados com membros de outras categorias. Claramente, o termo que representa o nível básico de uma categoria é um potencial candidato a ser utilizado como *tag* em uma categorização. Ocorre que a variação no nível de proficiência sobre o assunto do qual o recurso a ser categorizado se refere, como também das peculiaridades culturais e sociais dos autores das *tags*, muitas vezes, acarreta em uma variação na concepção deste nível básico que, conseqüentemente, irá transmitir-se através do sistema por meio das categorizações realizadas pelos diferentes usuários em seus espaços pessoais. A variação do nível básico é percebida na medida em que o grau de especificidade de um termo utilizado para categorização faz diferença para a concepção do criador do termo. Sistemas *Web* utilizam a idéia de categorização para que o usuário classifique recursos *on-line* utilizando palavras do seu vocabulário.

2.2 O Processo de *Tagging*

A partir de meados de 2002 (SMITH, 2008) muitos *websites* passaram a permitir que seus usuários efetuassem anotações ou utilizassem palavras simples, também chamadas de *tags*, com o intuito de descrever, organizar ou atribuir algum tipo de significado aos conteúdos disponíveis na *Web*. O processo de realizar esse tipo de anotação (categorização) foi denominado de *tagging*. Normalmente, as anotações realizadas pelos usuários eram essencialmente individuais, pessoais e utilizadas para descrever ou atribuir algum “significado” aos conteúdos das URLs, foto, vídeo, *post* de um *blog*, etc.

O processo completo é essencialmente baseado em três pivôs (Figura 1): o usuário — que realiza a anotação; o objeto — que é categorizado; e as *tags* — que geram a categorização descrevendo o objeto (SMITH, 2008; RUSSELL, 2005). A fim de exemplificar esse processo, considere um usuário que tem como recurso uma foto. Ele poderia atribuir um conjunto de

termos livremente selecionados do seu vocabulário (*tags*) para efetuar a descrição ou anotação da foto, facilitando, dessa forma, a recuperação futura por ele ou por outras pessoas.

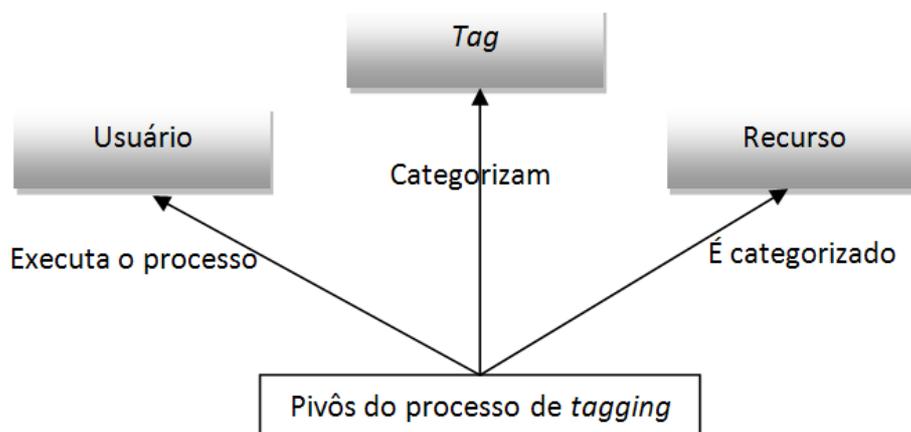


Figura 1 - Os três pivôs do processo de tagging

A utilidade das *tags* é muito ampla (RIDDLE, 2005). Quando aplicadas para a categorização de informação na *Web* elas podem auxiliar na recuperação de informações pessoais, descrevendo-as, organizando-as, estruturando-as e atribuindo significado. Existe um consenso entre os pesquisadores na área de *tagging* de que se utilize a palavra “categorização” no lugar da “classificação” (RUSSELL, 2005), (MATHES, 2004), (RIDDLE, 2005) e (STURTZ, 2004).

Segundo os autores, a palavra “categorizar” sugere um esquema menos rígido de se organizar algo. Segundo Sturtz (2004), também é possível estabelecer uma conexão com metadados atribuídos pelos usuário na categorização. A diferença é que os metadados tradicionais ficam escondidos dos usuários e são, geralmente, utilizados por mecanismos de busca apenas para melhorarem os resultados de uma consulta, enquanto que os termos utilizados para a categorização existem na superfície, **visíveis** e **úteis**, permitindo que os usuários arquivem, organizem e compartilhem seus conteúdos em diversos locais e de diversas formas simultaneamente. Desse modo, o processo de *tagging* permite a criação

dinâmica, distribuída e descentralizada de metadados, adicionando valor e realizando parte do trabalho que seria executado apenas por especialistas (RUSSELL, 2005).

No processo de *tagging*, uma das principais características é a liberdade quanto à quantidade e ao tipo de palavras-chaves que podem ser utilizadas para realizar uma categorização. A categorização é totalmente livre, ou seja, não existe nenhum controle ou esquema a ser seguido ou respeitado, de forma que os usuários utilizam as *tags* que desejarem não havendo hierarquias pré-definidas formalmente para o vocabulário de termos utilizado na categorização (MATHES, 2004). Também não há nenhuma especificação direta de relacionamento de descendência ou de irmandade entre os termos utilizados. As *tags* atribuídas são relacionadas automaticamente por co-ocorrência e se agrupam em *tags* baseadas em objetos comuns. Isso é muito diferente de uma taxonomia formal e de outros esquemas de classificação, nos quais existem múltiplos tipos de relacionamentos explícitos entre os termos. Essa liberdade resulta em sérias implicações sobre o esforço cognitivo exigido do usuário para efetuar categorizações, pois esse custo tende a ser muito inferior quando comparado ao esforço que o usuário precisa exercer para selecionar uma classificação específica dentro de uma estrutura hierárquica como uma taxonomia. Segundo Pereira e da Silva (2008b), a liberdade concedida ao usuário é um dos principais motivos pelos quais um sistema que aplica o conceito de *tagging* se torna simples de utilizar.

Com relação à recuperação de informação na *Web*, os sistemas que utilizam *tagging* possuem a vantagem de representar exatamente a opinião dos usuários (PEREIRA, et al., 2008b), pois não é um algoritmo que atribui *tags* aos objetos, são os usuários que estão fazendo isso — usuários com visões e opiniões diferentes, em contextos diferentes e com objetivos diferentes. Os sistemas de *tagging* permitem que sejam encontradas informações às quais os sistemas tradicionais talvez nunca conseguissem chegar.

Para finalizar as considerações sobre o processo de *tagging*, é interessante situá-lo diretamente dentro do escopo desta pesquisa. É exatamente no ponto de tensão “Pessoal versus o Social” (Smith 2008) que esta dissertação está focada. É importante saber quais são os impactos existentes quando o resultado de um processo de *tagging* é utilizado somente para fins pessoais (uma personomia) ou quando ele é compartilhado e utilizado de forma coletiva, social (uma folksonomia). Assim, considerando também os três pivôs do processo de *tagging* (usuário, objeto e *tag*), podemos dizer que a abordagem utilizada nesta dissertação foca-se no **pivô *tag*** e no **paradigma pessoal**, buscando fornecer meios para melhorar a qualidade dos resultados que podem ser obtidos do processo de *tagging*.

2.3 As Personomias

Uma personomia pode ser vista como a relação composta por um **usuário**, que utiliza seu vocabulário pessoal como *tags* para marcar os **recursos** (objetos) de seu interesse (Figura 2). Nesse sentido, uma personomia seria o conjunto de relações estabelecidas entre *tags* e objetos por um usuário específico. Em uma personomia, o pivô usuário é uma constante, enquanto os pivôs *tag* e objeto são variáveis.

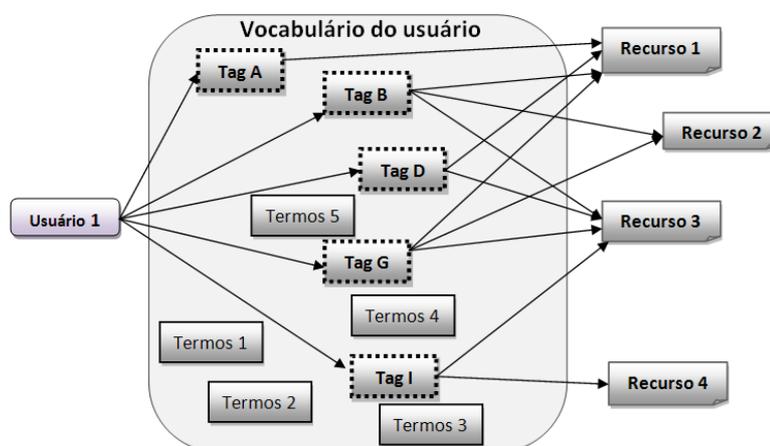


Figura 2 – O Usuário possui um vocabulário e utiliza parte dele como tags (tag A, tag B, tag D, tag G e tag I) que marcam os recursos de interesse pessoal.

Desse modo, uma personomia é criada para a organização da informação em um espaço pessoal. Esse espaço pessoal é oferecido pelos sistemas, que aproveitam todas as categorizações como matéria-prima para gerar o que chamamos de conhecimento coletivo. Diversos autores têm investigado a maneira como as pessoas escolhem *tags* em seu espaço pessoal, em que grau os membros destes sistemas dividem o mesmo vocabulário e quais tipos de *tags* estes usuários comumente utilizam para dar sentido aos recursos categorizados [AL-KHALIFA *et al.* 2007, CHOY *et al.* 2006, GOLDER *et al.* 2006, HALPIN *et al.* 2007, SEN *et al.* 2006, SHEN *et al.* 2005]. A seguir, apresentamos e discutimos as dinâmicas e as peculiaridades presentes nos sistemas que permitem a criação de uma personomia, mostrando como os termos utilizados nas categorizações evoluem e comportam-se nestes sistemas. Algumas das questões discutidas abaixo justificam a proposta de organização do vocabulário dos usuários apresentada neste trabalho.

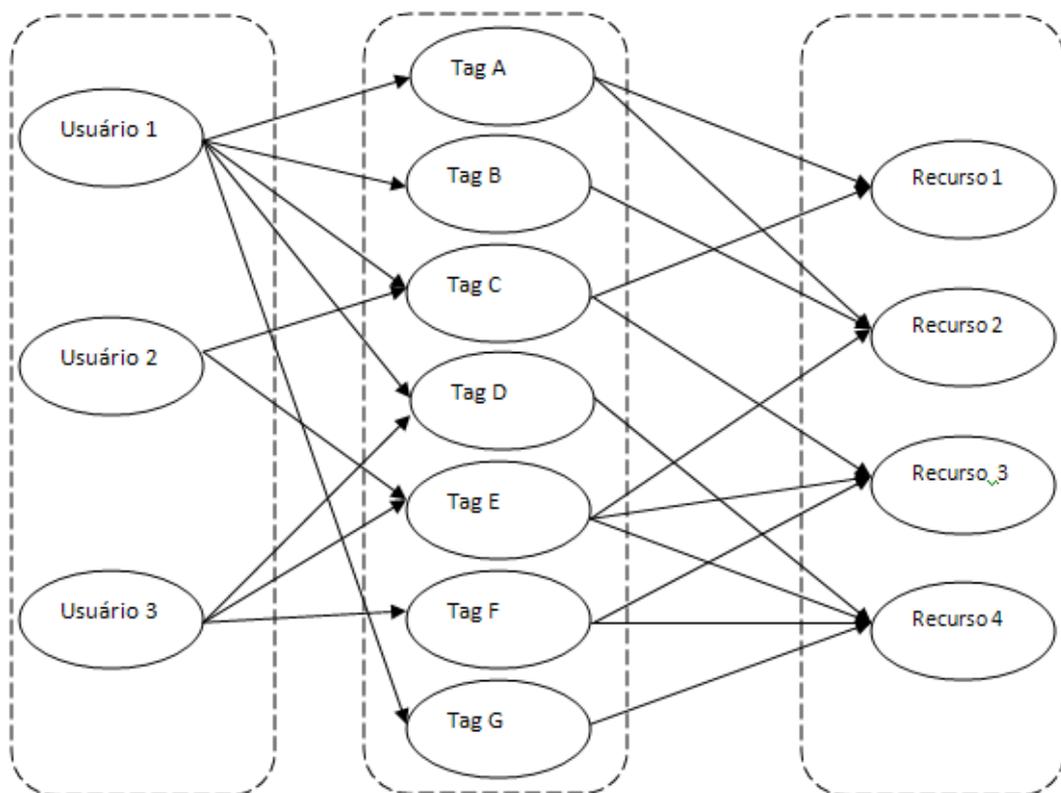


Figura 3 – Modelo de um sistema de categorização colaborativa. Usuário, tag e recurso são elementos que compõem uma categorização

O modelo conceitual de um sistema de categorização colaborativa genérico bem aceito é o modelo tripartite (MATHES, 2004), (SHEN, et al., 2005). Esse modelo, ilustrado na Figura 3 com os nós usuário, *tag* e recurso (os três pivôs do processo de *tagging*), é expresso por meio de um grafo tendo cada um de seus três conjuntos de vértices separados representando o conjunto de cada um destes elementos. No mesmo contexto do que foi exposto sobre os três pivôs, o primeiro conjunto representa os usuários (que realizam as categorizações), o segundo conjunto representa as *tags* (que são utilizadas para categorização) e o terceiro conjunto representa os objetos (que são categorizados), geralmente, representados por uma URI (*Uniform Resource Identifier*). Uma atribuição de *tag* a um objeto pode ser visualizada como duas arestas que ligam simultaneamente um usuário a uma *tag* e esta *tag* a um objeto. Cada usuário pode atribuir mais de uma *tag* a um mesmo objeto. Dessa forma, uma instância de categorização é denotada como um usuário, um conjunto de *tags* e um objeto.

A *tag* representa o nó central no modelo tripartite. Cada *tag* é utilizada por uma pessoa para marcar (anotar) um objeto, sendo utilizada, dentre outros fins, para recuperação desse objeto futuramente. A marcação do objeto é armazenada e passa a fazer parte da personomia do usuário que a efetuou.

2.4 A técnica de folksonomia – *Tagging* coletivo

O termo folksonomia é um neologismo criado por Thomas Vander Wal (MATHES, 2004) em 2004 com a junção das palavras “*folks*” (em inglês, pessoas) e “*taxonomy*” (taxonomia), resultando na palavra *folksonomy* — para a língua portuguesa a palavra *folksonomy* tem sido adaptada, principalmente, como “folksonomia”. Existe certo questionamento a respeito da precisão do termo folksonomia (MATHES, 2004) e alguns autores, como Golder et al. (2006) e Choy et al. (2006) preferem evitar o uso do termo

folksonomia, referindo-se ao conceito por meio do termo “*tagging* colaborativo”. Nesta dissertação utilizamos o termo “folksonomia” para referenciar essa técnica.

Uma folksonomia pode ser compreendida como uma coleção de personomias compartilhadas e interligadas. De acordo com Cattuto et al. (2008), as personomias constituem o núcleo da folksonomia ou, em outras palavras, pode-se dizer que a folksonomia é o resultado da socialização das personomias quando se adicionam os fatores sociais de compartilhamento e de interação entre os usuários.

Por ser originada das personomias, a folksonomia pode ser considerada como uma especialização do processo de *tagging*. Uma característica importante da folksonomia, consequência da liberdade oferecida por ela (LEW, 2006), é que seu sucesso é totalmente dependente dos usuários: se houver vários usuários efetivamente categorizando informações, então, ocorrerá uma construção de “sabedoria” coletiva — o trabalho individual de organização, de categorização e de atribuição de sentido aos objetos pelos usuários colabora para o coletivo, para o global. Por outro lado, se não houver usuários, não haverá sabedoria e nenhum outro tipo de informação a ser analisada. Pode-se afirmar, então, que o sucesso e a obtenção de benefícios em sistemas baseados em folksonomia dependem fortemente do modo como os usuários se sentem ao utilizar esses sistemas: os usuários precisam sentir-se seguros e guiados, pois se da mesma maneira que a liberdade é atrativa, uma liberdade total poderia deixá-los sem saber o que fazer ou como proceder. Tais requisitos apresentam fortes implicações para o *design* da interação desse tipo de sistema.

No que se relaciona ao *design* de interação de sistemas baseados em folksonomia, Wal (2005) explica que, de acordo com as decisões de *design* para os sistemas, a natureza da folksonomia produzida por eles pode ser bem diferente. Estas decisões podem levar a dois tipos distintos de folksonomia: folksonomia larga (*broad folksonomy*) e folksonomia estreita (*narrow folksonomy*). Essa classificação é exposta por alguns críticos tecnológicos como

Tomas Vander Wal¹⁴ e Clay Shirky¹⁵ em seus *blogs*, não possuindo ainda discussão formal na literatura.

2.4.1 A folksonomia larga

A folksonomia larga permite que diversos usuários atribuam *tags* a um mesmo objeto, utilizando termos do seu vocabulário. Dessa forma, os objetos podem ser recuperados por qualquer uma das *tags* atribuídas a eles. O sistema *Delicious*, por exemplo, aplica o conceito de folksonomia larga. Nestes sistemas qualquer usuário poderá utilizar-se do termo que desejar para categorizar determinada URL, assim, todos os usuários podem categorizar o que realmente quiserem. Com esse tipo de folksonomia consegue-se emergir termos comuns a um mesmo recurso, termos que são atribuídos em conjunto (co-ocorrência), bem como as *tags* mais utilizadas. Esse tipo de análise pode auxiliar na extração do senso comum a respeito do uso de termos para categorizar objetos.

¹⁴ <http://www.vanderwal.net/>

¹⁵ <http://shirky.com/>

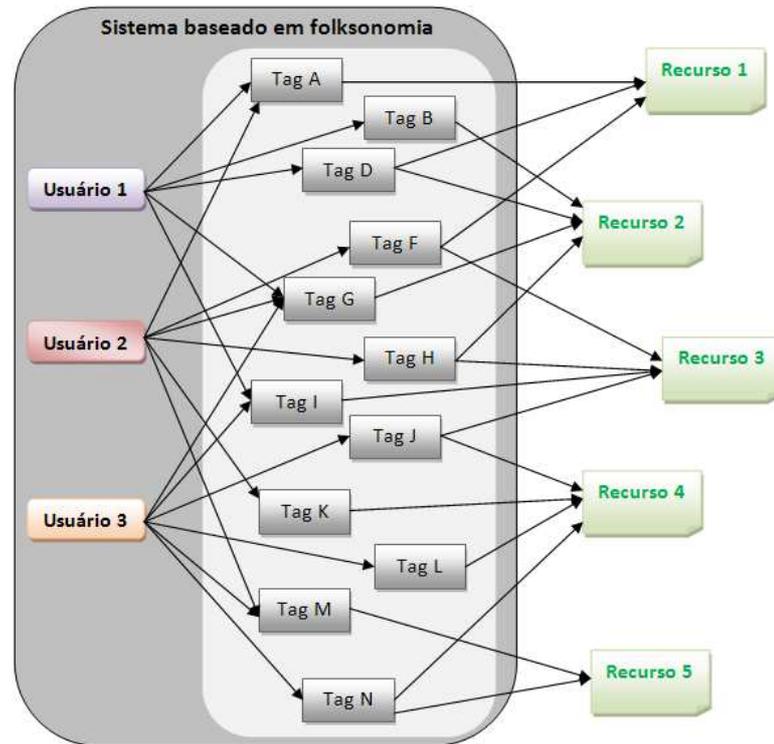


Figura 4 – Folksonomia larga – os usuário utiliza tag(s) para categorizar objeto(s).

A Figura 4 ilustra um esquema que representa a folksonomia larga. Nela, os três usuários (Usuário 1, Usuário 2 e Usuário 3) utilizam um conjunto de *tags* (Tag A, Tag B, ... , Tag N) para categorizar objetos (Recurso 1, ... , Recurso 5). É possível notar no esquema as interligações entre as personomias de cada usuário. Todo o esquema com seus componentes e suas interligações formam uma folksonomia.

2.4.2 A folksonomia estreita

A folksonomia estreita é gerada com o usuário atribuindo *tags* aos recursos produzidos e disponibilizados por ele mesmo, normalmente, algo pessoal como fotos, vídeos, *slides*, etc. As *tags* podem ser utilizadas por outros usuários para a recuperação desses recursos. A folksonomia estreita tem o caráter de posse, nesse tipo de folksonomia os objetos categorizados pertencem ao indivíduo e, assim, outros usuários não podem associar *tags* a ele. Como exemplos da aplicação desse tipo de folksonomia temos os sistemas *Flickr* e *Youtube*.

O sistema *Flickr* permite que o usuário poste suas fotos em um espaço na *Web* e organize-as com a aplicação de uma categorização pessoal. O sistema *Youtube* adota a mesma filosofia, nele o usuário disponibiliza seus vídeos para a comunidade e utiliza um conjunto de termos para organizar e compartilhar seus vídeos. Isso não impede que alguém utilize um sistema de folksonomia larga para categorizar uma URL que contenha a produção de um vídeo ou de uma foto de alguém, mas nesse caso, o objeto não será a foto ou o vídeo, mas a URL que os referencia.

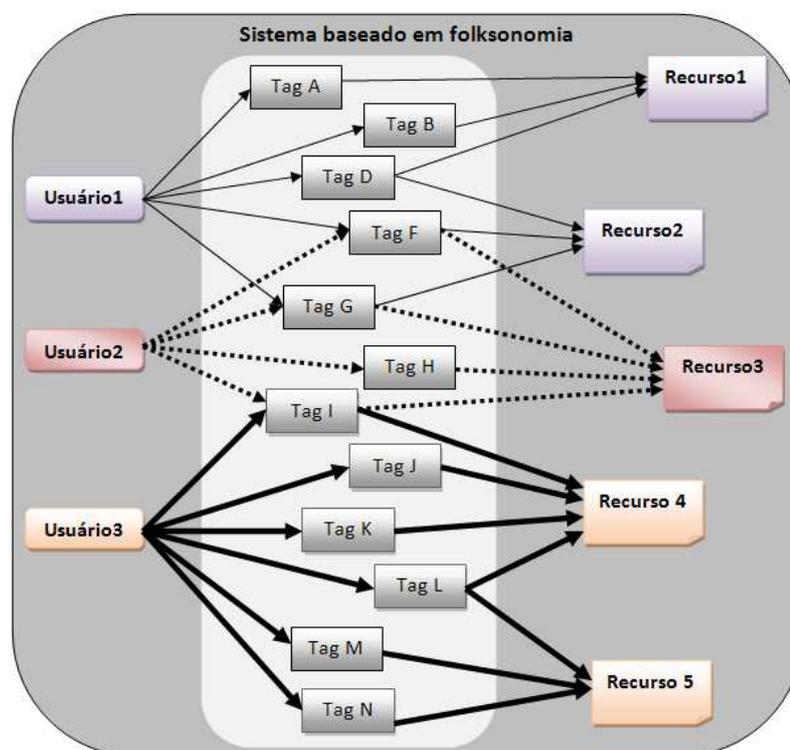


Figura 5 - modelo de um sistema de folksonomia estreita (os diferentes tipos de seta indicam diferentes usuários e suas relações)

A Figura 5 ilustra um esquema no qual o usuário utiliza *tags* para categorizar seus objetos. Uma *tag* pode ser utilizada por vários usuários, mas para categorizar objetos distintos. Embora o conceito de folksonomia estreita esteja fortemente relacionado com o conceito de *tagging*, principalmente quanto ao fato de ambos serem destinados à atribuição de *tags* para fins pessoais, existem os fatores sociais de compartilhamento da categorização e de interação entre os usuários que os diferencia. Nos sistemas que implementam o conceito de

folksonomia estreita, normalmente, um usuário não pode atribuir *tags* aos objetos de outros usuários, mas ele pode pesquisá-los, visualizá-los e analisar as *tags* utilizadas por outros usuários para categorizá-los.

2.4.3 O uso da folksonomia como ferramenta coletiva

Tradicionalmente, a tarefa de criação e de atribuição de metadados aos recursos envolvidos em um sistema pode ser abordada de duas maneiras distintas: i) a criação de metadados pode ser realizada por profissionais dedicados a esta tarefa ou ii) pelos próprios autores dos materiais (MATHES, 2004). A primeira abordagem remete aos problemas relativos à escalabilidade devido ao enorme volume de conteúdos sendo produzido e utilizado, principalmente, na *Web*, e à complexidade, pois os aparatos e as ferramentas utilizadas pelos profissionais dedicados a tarefa de criação de metadados são demasiadamente complicados para pessoas sem treinamento ou sem conhecimento especializados. A segunda abordagem se refere às iniciativas como o *Framework de Warwick* e o *Dublin Core Metadata Initiative* (BREITMAN, 2005), os quais podem auxiliar na redução dos problemas relacionados com a escalabilidade. No entanto, esta abordagem está sujeita à descrições inadequadas ou imprecisas por parte dos autores dos metadados, neste caso, autores também dos conteúdos a serem associados a eles.

Neste contexto, o uso de metadados criados por usuários aparece como uma alternativa, visando diminuir as dificuldades encontradas nas abordagens existentes para a tarefa de criação e de atribuição dos mesmos (MATHES, 2004). Os sistemas baseados em folksonomia permitem que os próprios usuários do sistema, em suas personomias, atribuam metadados ao conteúdo presente e, dessa forma, os sistemas transferem o ônus da categorização dos recursos aos usuários. Tal abordagem é especialmente útil quando não há a possibilidade de poucas autoridades atuarem na categorização das informações.

Considerando que este cenário é característico da *Web* atual, esta é uma das razões pelo qual a categorização colaborativa tem crescido em popularidade.

O ato de colecionar e ordenar objetos (úteis e interessantes) é natural quando se utiliza *tags* e disponibilizam-se publicamente as *tags* associadas aos objetos categorizados. Isso pode revelar aspectos sobre a paisagem informacional e cultural do indivíduo que categorizou e disponibilizou o objeto publicamente. É possível aprender a partir da estrutura destas paisagens e, muitas vezes, não é necessário um alto investimento cognitivo. Os usuários dos sistemas baseados em folksonomia são úteis para a organização dos dados, pois são grandes agregadores e atratores de conhecimento, podendo repassá-los por meio do compartilhamento de suas personomias. É possível aproveitar o conhecimento social, gerado nos sistemas baseados em folksonomia a partir da exploração das contribuições individuais de cada usuário, ou seja, do conhecimento emergente das personomias compartilhadas no sistema. A exploração destas contribuições individuais, a fim de capturar o conhecimento social emergente, detém aplicações para a personalização dos usuários (SZOMSZOR, et al., 2008), para organização do vocabulário dos usuários (CÔGO, et al., 2008), para melhoria dos propósitos de recuperação de informações nestes sistemas (HOTHO, et al., 2006) e para o desenvolvimento de recursos que ofereçam apoio aos usuários nos sistemas baseados em folksonomia.

2.4.4 Vocabulário não-controlado

Mathes (2004) afirma que da mesma forma como as pessoas utilizam os caminhos que outras pessoas utilizaram anteriormente para se guiarem, pode-se usar as *tags* mais populares ou comumente utilizadas para desenvolver um vocabulário que, de fato, reflita a linguagem da maioria dos usuários dos sistemas baseados em *tagging*. O autor afirma ainda que a folksonomia, com seu crescimento orgânico e sua natureza não-controlada, tem a capacidade

de se adaptar rapidamente às mudanças e às necessidades dos vocabulários dos usuários, o que não acontece em nenhum outro tipo de sistema. Mathes (2004) e Sturtz (2004) sugerem que, talvez, o benefício mais importante da folksonomia seja essa capacidade de refletir o vocabulário dos usuários. Em um sistema de recuperação de informação, há vários vocabulários envolvidos, entre eles: do usuário, do sistema, do projetista, do autor do material, do responsável pela definição das classificações, e assim por diante. Os sistemas baseados em folksonomia refletem diretamente as ações realizadas pelos usuários, suas escolhas, terminologias e precisões — os usuários constroem e escolhem o vocabulário que será utilizado.

Entretanto, o uso de um vocabulário não-controlado é um dos pontos de conflitos na folksonomia (STURTZ, 2004) pois essa característica peculiar é considerada uma fonte de benefícios e também de muitos aspectos considerados desvantajosos na utilização da técnica. Um vocabulário controlado pode melhorar e auxiliar no momento da recuperação de informação (*findability*), embora apresente um custo cognitivo maior para o usuário no momento da categorização.

A folksonomia permite que sejam realizadas grandes descobertas não intencionadas por meio de relacionamentos, categorizações, termos utilizados, etc., todavia as *tags* não funcionam bem para medidas de precisão e de recuperação. Russel (2005) argumenta que os sistemas baseados em folksonomia permitem que sejam encontradas informações que provavelmente nunca seriam apresentadas aos usuários — Mathes (2004) e Sturtz (2004) caracterizam isso como **serendipismo** e afirmam que a folksonomia favorece a descoberta de conhecimento e de informações úteis e interessantes.

No entanto, apesar de favorecer o serendipismo, uma grande quantidade de termos pode causar uma sobrecarga de metadados ruidosos¹⁶ que são sem utilidade e sem relevância

¹⁶ Ruído refere-se a termos que não serão entendidos pelos usuários do sistema baseado em *tagging*

para os usuários. Possivelmente, muitos dos termos utilizados pelos usuários em suas categorizações detenham potenciais falhas para a recuperação e gerenciamento das informações. Estas falhas contribuem para a necessidade de um maior esforço cognitivo para se recuperar informações ou impedem que um determinado recurso possa ser recuperado por meio de um termo pertinente, tanto para quem os utilizou quanto para outros usuários interessados em uma navegação utilizando-se deste vocabulário.

É importante considerar que nem sempre as categorizações efetuadas por um usuário trarão benefício para o coletivo. Os termos com caráter pessoal para um usuário podem ser adequados apenas a ele e vir a ser considerados ruídos pelo restante dos usuários. Além disso, como mencionado anteriormente, a liberdade de trabalho concedida aos usuários é a responsável por grande parte dos problemas encontrados na folksonomia. Abaixo, são explicados os fenômenos da **polissemia** e da **sinonímia** e os motivos pelos quais estes fenômenos são considerados problemáticos:

1. A **polissemia** ocorre quando um mesmo termo representa muitos significados. Do ponto de vista dos sistemas de folksonomia, a polissemia pode ser percebida como a utilização de uma mesma *tag* para categorizar recursos com conteúdo e significado diferentes. A polissemia prejudica os resultados de uma consulta, pois ela retorna itens relacionados com a palavra, porém, potencialmente inaplicáveis ao significado idealizado. Por exemplo, uma consulta por “processo” poderia retornar recursos associados aos diferentes significados do termo, os quais, eventualmente, dependendo do algoritmo de busca selecionado, ocupariam lugar de destaque nos resultados mesmo sendo inaplicáveis ao significado pretendido. Em termos práticos, isso gera ambigüidades e implica em um problema de precisão: muitos resultados que são retornados não satisfazem os critérios reais da busca.

2. Os **sinônimos** são palavras diferentes que detém o mesmo significado. Por exemplo, um recurso cujo conteúdo refere-se à forma como proceder na execução de uma determinada técnica pode ser buscado por meio do termo “processo”, enquanto um usuário eventualmente o tenha categorizado apenas com o termo “método”. Do ponto de vista dos sistemas de folksonomia, são considerados sinônimos *tags* cujos termos utilizados denotam o mesmo significado associado a um recurso, porém, a forma destes termos é diferente. Também não há concordância quanto ao uso do separador de palavras compostas, nem no uso de plural ou singular dos termos, o que pode acarretar em problemas semelhantes a sinonímia. Há ainda a existência de erros ortográficos como, palavras com letras trocadas e palavras com vírgula no final, que podem nunca mais serem encontrados caso o mesmo erro não seja cometido na busca. Em termos práticos, isso implica em num problema de recuperação, pois não há como assegurar que todos os itens relevantes foram retornados.

Como já mencionado, o uso de um vocabulário não-controlado em sistemas baseados em folksonomia implica tanto em benefícios quanto em malefícios. Para melhor entendimento desta questão, devem-se considerar dois momentos distintos nos quais se pode fazer uso da folksonomia: o momento de categorizar recursos por meio de *tags* e o momento de recuperar o recurso por meio de *tags*. No primeiro momento, a ausência de controle dos termos a serem associados aos recursos proporciona maior facilidade na categorização, pois o uso de *tags* permite descrever o recurso a ser armazenado por meio de vários de seus atributos, sem sobrecarga cognitiva ao usuário e de maneira flexível. Ao categorizar um novo recurso, o usuário atribui o conjunto de termos que julga adequado naquele momento, sem garantir, no entanto, que estes termos também serão adequados no momento em que for recuperar este

mesmo recurso. Diversas questões exercem influência direta sobre a definição de termos durante a categorização, entre elas o conhecimento prévio e as atividades recentes realizadas pelo usuário, ou seja, o contexto de trabalho do usuário ao categorizar o recurso, podendo acarretar em termos que não são de uso freqüente no vocabulário deste usuário, ou que não trazem muito sentido em longo prazo (PEREIRA, et al., 2008b). Desta forma, não há como garantir que o usuário poderá relembrar ou identificar os termos utilizados em suas categorizações, a menos que o sistema possua mecanismos efetivos que o auxiliem nessa tarefa. Assim sendo, torna-se difícil a manutenção de um vocabulário organizado, consistente e bem estruturado. No segundo momento, a polissemia (geradora de ambigüidades), a sinonímia e a falta de padronização nos termos utilizados atrapalham o processo de recuperação dos objetos. A característica de liberdade que auxilia o usuário no armazenamento e no compartilhamento de seus objetos pode atrapalhar no momento de recuperar estes mesmos objetos, pois, com um vocabulário desorganizado, o próprio usuário poderá ter dificuldades em encontrar o conteúdo categorizado. Estas questões são ainda mais evidentes quando consideradas do ponto de vista de outros usuários interessados na recuperação destes recursos: a idiosincrasia das *tags* também colabora para o serendipismo, no entanto, dificulta a atuação do usuário em tentar identificar um termo adequado para realizar uma busca.

Nos sistemas baseados em folksonomia, a variação do nível de especificidade de um termo utilizado para a categorização pode acarretar problemas, pois não é possível determinar qual o nível utilizado pelo autor da *tag*. Por exemplo, “fusca”, “carro” e “automóvel” são maneiras razoáveis de categorizar um mesmo objeto e podem ser distintamente utilizadas de acordo com o interesse, o nível de proficiência ou alguma outra necessidade de quem realiza a categorização do objeto. No entanto, usuários interessados em uma busca por objetos

relacionados a este assunto não são capazes de prever quais termos foram utilizados e, dessa forma, acabam utilizando o termo condizente puramente com suas expectativas pessoais.

2.4.5 Navegação social usando *tagging*

No processo de categorização a *tag* faz a ligação entre o usuário e o objeto. Assim, as *tags* podem ser ferramentas poderosas para a navegação social uma vez que elas rotulam (descrevem) a informação e auxiliam os usuários no compartilhamento e na descoberta de novas informações categorizadas por outros usuários (SHEN, et al., 2005). A navegação social é uma alternativa aos métodos prevalentes das metáforas de navegação (FORSBERG, et al., 1998). Ao invés de se basear apenas em representações abstratas do espaço, a navegação social utiliza marcas deixadas pelos usuários que já utilizaram o sistema para auxiliar na navegação dos usuários posteriores. Estas marcas podem ser suas experiências, comentários ou opiniões, expressadas de alguma maneira dentro do sistema (CHALMERS, et al., 2004). No caso dos sistemas baseados em folksonomia, estas marcas podem ser pensadas como o conjunto de *tags* utilizadas pelos usuários para categorizar seus objetos. A navegação social e os sistemas baseados em folksonomia se relacionam, pois a navegação nestes sistemas é realizada a partir do conjunto de termos utilizados por todos os usuários para categorização dos objetos armazenados. Este conjunto de termos é modificado diversas vezes no tempo, refletindo a natureza dinâmica dos vocabulários presentes no sistema.

2.5 Os sistemas baseados em folksonomia

A técnica de folksonomia vem sendo empregada por diversos sistemas, sendo o *Delicious* (folksonomia larga) e o *Flickr* (folksonomia estreita) os dois maiores exemplos (SHEN, et al., 2005) (STURTZ, 2004) (RIDDLE, 2005). O comportamento dos sistemas

baseados em folksonomia é comparado ao de sistemas complexos (GOLDER, et al., 2006), (HALPIN, et al., 2007), pois cada *tag* que categoriza um objeto é repetida um número de vezes por diferentes usuários, gerando, portanto, uma distribuição de frequência associada a estas *tags*. Outras características dos sistemas baseados em folksonomia, que são associadas a sistemas complexos, são: o grande número de usuários, a falta de coordenação central e as dinâmicas não-lineares. Os sistemas complexos são conhecidos por produzir, com o passar do tempo, um tipo de distribuição conhecida como lei de potência (*power law*). A distribuição de lei de potência é relacionada ao Princípio de Pareto (NEWMAN, 2006), no qual uma grande parte dos efeitos são provocados por uma pequena parcela das causas, que acarretam numa pequena parcela dos efeitos. Uma importante característica desta distribuição é que ela é livre de escala, ou seja, não importa quão grande o sistema se torne, a forma da distribuição continua a mesma, ou seja, estável, significando que eventualmente existe um conjunto de *tags* que descrevem bem um recurso em questão e estas mesmas *tags* são reiteradas pela maioria dos usuários do sistema.

Em um sistema de categorização que utiliza a técnica de folksonomia é possível observar a existência do fenômeno de lei de potência, gerando na distribuição de *tags* e de cada objeto, um efeito conhecido como cauda longa (*long tail*) (ANDERSON, 2006). Esse efeito é que caracteriza os níveis abruptos de decaimento no uso de cada *tag*, ou seja, representa o fato de que um número relativamente pequeno de *tags* é empregado com grande frequência enquanto uma grande quantidade de *tags* é empregada com pouca frequência nas categorizações. A Figura 6 mostra um histograma da distribuição de *tags* em um objeto categorizado num sistema baseado em folksonomia, ilustrando o efeito da cauda longa. O efeito pode ser justificado pela influência que o comportamento coletivo dos membros de uma comunidade causa no comportamento individual dos outros membros desta mesma comunidade, pois os usuários tendem a observar o ato (comportamento) de outros usuários e

passam a acreditar que aquela é a forma correta de agir (GOLDER, et al., 2006) (SEN, et al., 2006).

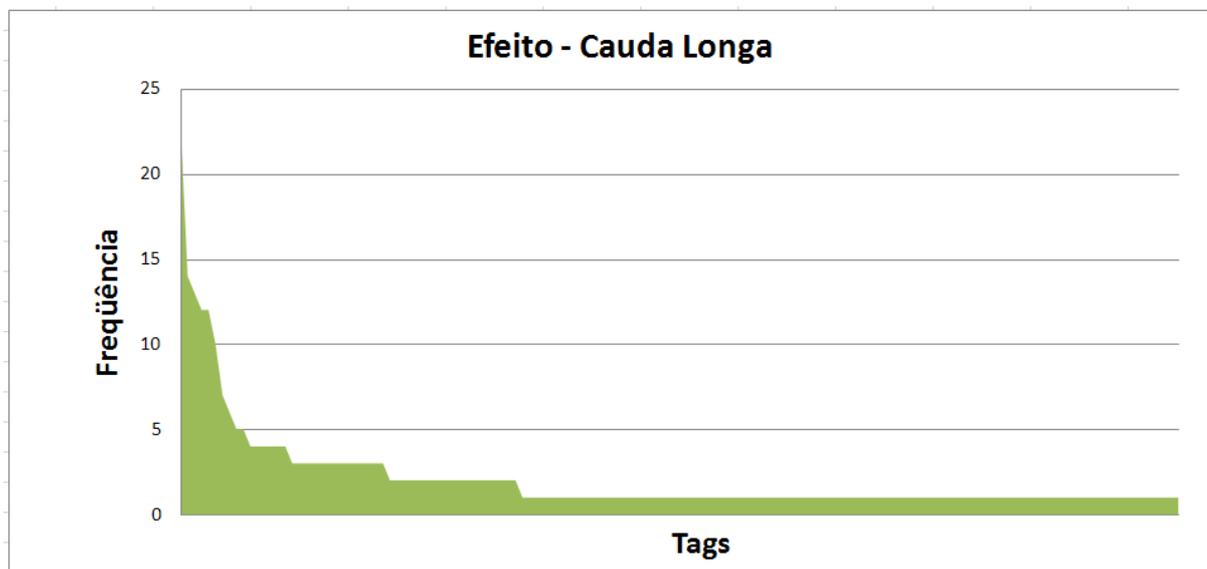


Figura 6 - Distribuição do uso de tags em uma categorização. O efeito de “cauda longa” (long tail) caracteriza os níveis abruptos de decaimento no uso de cada tag.

Alguns autores como Al-Khalifa *et al.* (2007), Golder *et al.* (2006) e Sen *et al.* (2006) classificam em seus trabalhos os diferentes tipos de *tags* comumente utilizadas em sistemas baseados em folksonomia. De forma geral, estes tipos caracterizam-se por meio das diferentes intenções no uso do recurso no momento da categorização. Assim, as *tags* podem ser classificadas como:

- **Tags fatuais:** são *tags* que identificam fatos a respeito do recurso como, por exemplo, “*blog*”, “*java*” e “*programação*”;
- **Tags subjetivas:** são aquelas *tags* que exprimem a opinião do usuário a respeito do recurso como, por exemplo, “*interessante*” e “*divertido*”;
- **Tags pessoais:** são *tags* relacionadas com uma necessidade pessoal do criador da *tag*, muitas vezes, utilizadas para referências próprias ou para o gerenciamento de suas tarefas. Um exemplo deste tipo são as *tags*, “*toread*”, “*todo*” e “*meusite*”.

Alguns dos principais fatores apontados como estimulantes para o sucesso dos sistemas baseados em folksonomia, além de sua propriedade social, são o *feedback* imediato que eles proporcionam, assim como a simplicidade, a facilidade que as *tags* apresentam para os usuários armazenarem e organizarem suas coleções de recursos (objetos) (CHOY, et al., 2006), e para recuperá-los posteriormente, pois, diferente dos sistemas de arquivos usuais nos quais os recursos são armazenados em diretórios hierárquicos o uso de *tags* permite que estes recursos sejam armazenados em um único local com as *tags* atuando como *links*, podendo ser utilizadas em locais diferentes.

Os artigos de Halpin *et al.* (2007) e Begelman *et al.* (2006) discutem a estrutura informacional mais simples que podem ser derivadas dos processos de categorização nos sistemas baseados em folksonomia: a estrutura de co-relação de *tags*. A partir das *tags* utilizadas nas diversas categorizações pode ser calculada uma medida de distância entre um par de *tags* co-ocorrentes quaisquer. Esta medida de distância captura o grau de co-ocorrência entre as *tags* e pode ser interpretada como uma métrica de similaridade entre os conceitos que estas *tags* representam. A partir destas medidas de similaridade é possível a construção de grafos de co-ocorrência de *tags*, como mostrada na Figura 7, apresentada em (HALPIN, et al., 2007). Nesta figura, o tamanho dos nós é proporcional à frequência absoluta de cada *tag*, enquanto a distância entre cada par de *tags* é inversamente proporcional à medida da frequência de co-ocorrência. Todas as *tags*, mostradas na Figura 07, relacionam-se com a *tag* “*complexity*”. A análise da topologia dos grafos obtidos permite perceber como as pessoas utilizam as *tags* e como estruturas semânticas emergem desta utilização.

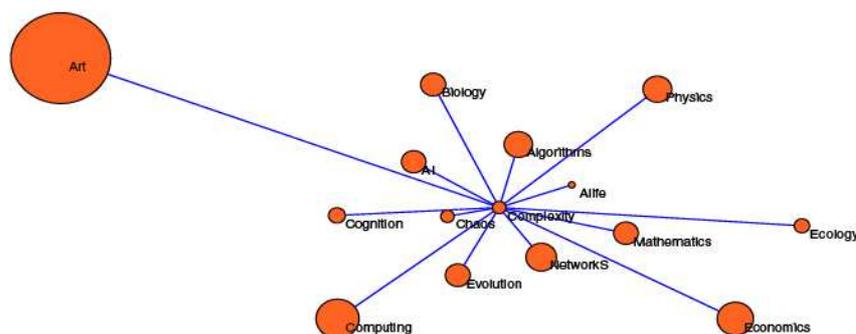


Figura 7 - Visualização das tags em forma de grafo, mostrando as relações a partir de um nó central “complexity” (HALPIN, et al., 2007).

Alguns autores como Gruber (2008), Wu *et al.* (2006) e Hak *et al.* (2008) propõem a evolução de ontologias a partir de folksonomias e outros propõem a exploração destes grafos para auxílio da navegação dos usuários nos sistemas baseados em folksonomia (SIMPSON, 2008) alternativamente às listas de tags ou nuvem de tags (*tag cloud*), como a mostrada na Figura 8 que são comumente empregadas.



Figura 8 - Nuvens de tags (*tag cloud*), esquema de visualização de tags do usuário.

O *tagcloud* é um recurso de visualização de informação que se tornou popular juntamente com os sistemas baseados em folksonomia. Ele representa um recurso alternativo para demonstrar e dar ênfase à informação de acordo com a sua importância.

2.5.1 Alguns sistemas baseados em folksonomia

Dentre os sistemas baseados em folksonomia existentes, os sistemas *Delicious* e *Flickr* são considerados os mais maduros e são os principais exemplos citados (STURTZ, 2004) (SHEN, *et al.*, 2005) (RIDDLE, 2005) (RUSSELL, 2006). A seguir descreveremos alguns dos principais sistemas baseados em folksonomia.

O sistema *Delicious* (YAHOO, 2003) pode ser definido como um gerenciador de favoritos (*bookmarks*) social que permite que URLs, de qualquer página da *Web*, sejam adicionadas, classificadas com *tags*, acessadas em qualquer máquina e, principalmente, compartilhadas com outros usuários (Figura 9). Ele é considerado como o exemplo mais completo da aplicação da técnica de folksonomia atual. O *Delicious* implementa o conceito de folksonomia larga, na qual os usuários cadastrados categorizam seus recursos particulares e podem categorizar qualquer outro recurso disponível na *Web*. Segundo dados divulgado pelos administradores do sistema, em novembro de 2008 foi ultrapassada marca de 5,3 milhões de usuários, bem como ultrapassada a marca de 180 milhões de URLs diferentes postadas (Delicious, 2008). A principal diferença entre o *Delicious* e os demais gerenciadores de *bookmarks* é justamente a ênfase no usuário adicionando as *tags* como uma construção organizacional fundamental, o que o caracteriza como um sistema baseado em folksonomia.

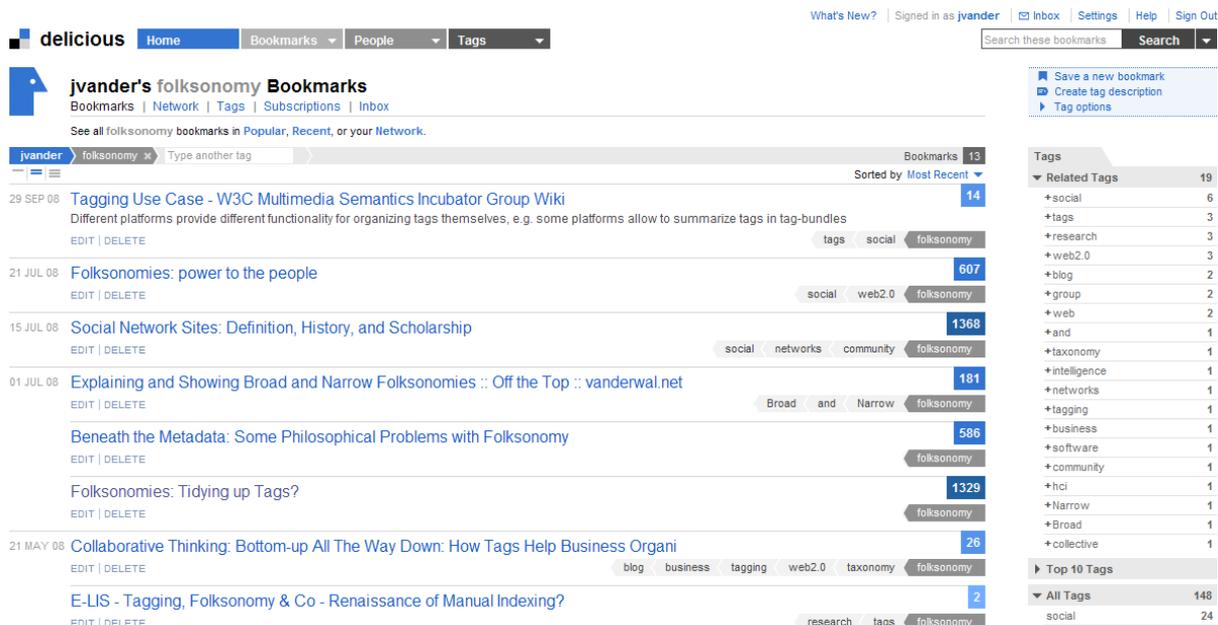


Figura 9 - Sistema Delicious - gerenciamento de bookmarks on-line

Na outra perspectiva da utilização da técnica de folksonomia, aplicando o conceito de folksonomia estreita (*narrow folksonomy*), está o sistema *Flickr* (YAHOO, 2004), ver Figura 10. O *Flickr* é um compartilhador de fotos pessoais que permite aos seus usuários a inserção de *tags* para categorizar e organizar suas imagens. Inicialmente, o *Flickr* permitia apenas ao proprietário da imagem a atribuição de *tags* para categorizá-la, mas, em seguida desenvolveu um recurso no qual o proprietário da imagem pode definir um ou mais usuários que terão permissão para categorizá-la também. A anotação de imagens por meio de *tags* permite que mecanismos de busca e outras ferramentas possam trabalhar de forma mais eficaz essas imagens e seus significados. Comparando com outros tipos de conteúdos, normalmente, as imagens ficam restritas a poucas informações de texto relacionadas a elas, sendo mais difícil de se realizar um processamento ou uma análise com essas imagens. Os administradores do *Flickr* pretendem permitir que qualquer usuário categorize qualquer foto, o que deixará de caracterizar o sistema como de folksonomia estreita para caracterizá-lo como de folksonomia larga.



Figura 10 - Sistema Flickr - Possibilitando a publicação e organização de fotos, bem como rede de contatos.

O Youtube¹⁷, apresentado na Figura 11 (GOOGLE, 2005), também é considerado como um sistema baseado em folksonomia que implementa o conceito de folksonomia estreita. Nele, o usuário faz *upload* de um vídeo, atribui *tags* a esse material e o disponibiliza para a comunidade de usuários do sistema.

¹⁷ <http://br.youtube.com/>

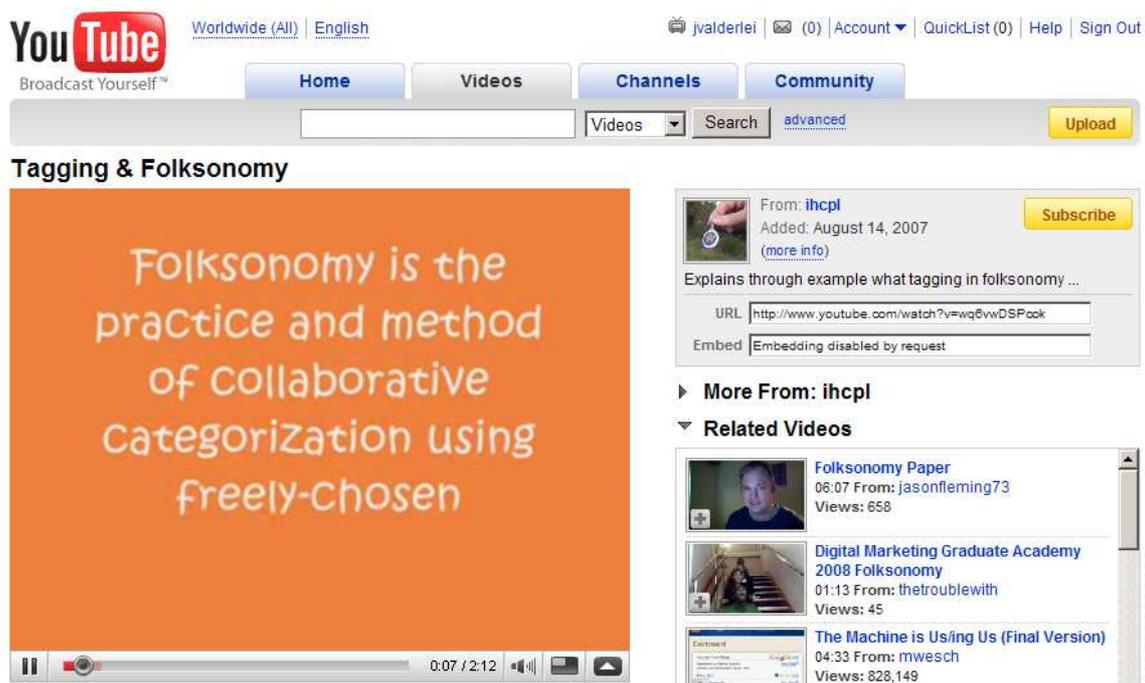


Figura 11 - Sistema Youtube - Compartilhamento de vídeo baseado em tagging

Outro sistema que vem se destacando é o *SlideShare* (SLIDESHARE, 2007) (Figura 12), que também implementa uma folksonomia estreita e tem funcionamento semelhante ao sistema *Youtube*. Entretanto, ao invés de permitir a organização e a disponibilização de vídeos, ele trabalha com o compartilhamento de *slides* de apresentações entre seus membros e a comunidade que o utiliza.



Figura 12 - Sistema Slideshare - Compartilhando slides de apresentação

Os sistemas citados são alguns exemplos que demonstram a tendência da disponibilização de recursos que incentivam a utilização de *tags*, a participação dos usuários e a utilização dos benefícios obtidos das informações geradas pelos sistemas. Com esses exemplos, podemos verificar que a categorização de objetos por meio de *tagging* vem sendo aplicada em diversos sistemas. Entretanto, a necessidade de criar uma personomia exclusiva para cada sistema eleva a dificuldade e a carga de trabalho exigida do usuário para que ele mantenha um vocabulário de *tags* coerente e bem organizado. Essa dificuldade colabora para a geração de personomias desorganizadas e com dados ruidosos que emergem para as folksonomias destes sistemas. No próximo capítulo, discutimos os problemas gerados com as múltiplas personomias de um mesmo usuário que ficam armazenadas em diferentes sistemas de categorização e apresentamos uma proposta de solução para esses problemas.

O Problema das Múltiplas Personomias e a Proposta de Solução

Todo sistema baseado em folksonomia exige que seus usuários criem e mantenham um vocabulário de termos utilizados para a categorização de seus objetos. Assim, um usuário que utilize vários sistemas baseados em folksonomia, obrigatoriamente, deverá manter um vocabulário em cada sistema, independentemente desses termos serem os mesmos ou completamente diferentes. Atualmente não há nenhum meio de centralizar e gerenciar os termos que um usuário utiliza para efetuar suas categorizações, o que impacta negativamente na facilidade de uso, na consistência e na qualidade desses termos. Neste capítulo, discutiremos a questão da existência de múltiplas personomias devido à impossibilidade do uso de um vocabulário já existente para um sistema baseado em *tagging* em outro sistema. Apresentaremos, também, a proposta de um gerenciador de *tags* para atuar nessas múltiplas personomias, centralizando o vocabulário de termos de um usuário em um único local.

3.2 Problema versus solução

A arquitetura do sistema gerenciador de *tags* deverá solucionar o problema que um usuário enfrenta ao utilizar vários sistemas baseados em *tagging* (SILVA e DA SILVA, 2008). Conforme ilustrado pela Figura 13, o usuário deverá acessar cada sistema individualmente para utilizar o seu vocabulário de *tags* ou para fazer alguma manutenção ou busca nesses sistemas. Uma solução para o problema de busca em diversos sistemas é disponibilizada pelo sistema *My.Tag*, que indexa recursos espalhados em alguns sistemas, tais como *Flickr*, *Youtube*, *Delicious*, *Connotea*, *BibSonomy* e *BibTex*. No entanto, esse sistema não centraliza essas *tags*, não possibilita a sua manutenção, não permite a realização de novas categorizações, não controla a frequência e nem o contexto de uso das *tags*.

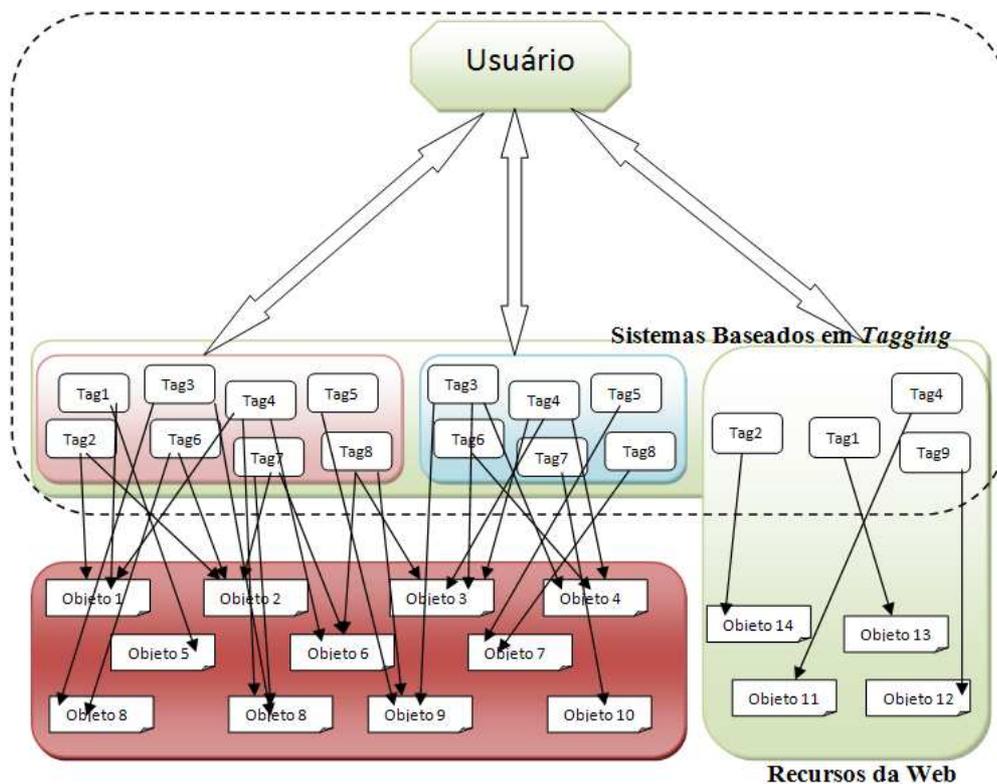


Figura 13 - A diversidade de sistemas baseado em folksonomia

Uma arquitetura para o gerenciamento das *tags* do usuário deverá possibilitar a criação de uma camada de *tags* (Figura 14) com as *tags* advindas das personomias dos usuários. Esse agrupamento de *tags* possibilitará o gerenciamento das *tags* dessas personomias, ajudando o

usuário na reutilização de termos e na estruturação de um vocabulário comum para os diversos sistemas de categorização que ele utilizar. Para utilizar as *tags*, o usuário não necessitará acessar diretamente os sistemas baseados em *tagging*, mas sim o sistema gerenciador de *tags* que fará o controle de uso do seu vocabulário pessoal. Deste modo, o usuário fará o acesso transparente aos sistemas suportados pelo gerenciador de *tags*.

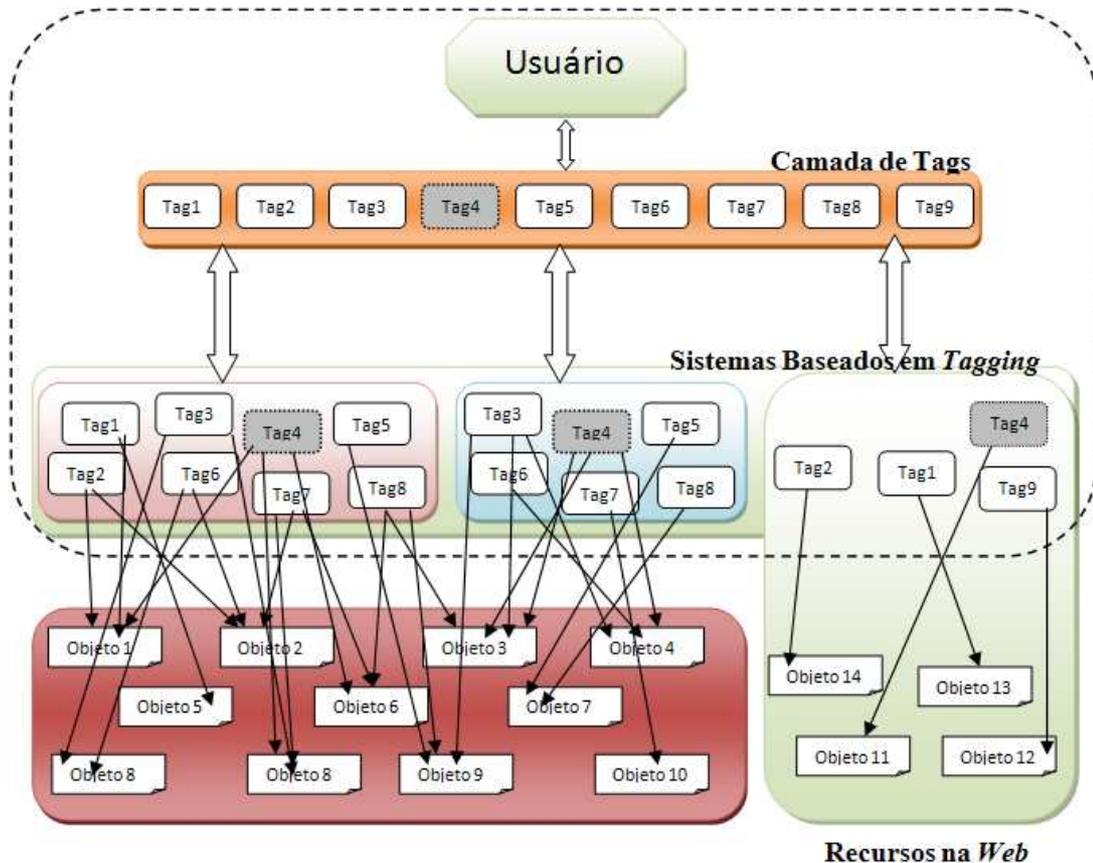


Figura 14 - Camada de tags sobre os sistemas baseados em folksonomia

O sistema gerenciador de *tags* formará uma camada de *tags* sobre os dados do usuário nos diversos sistemas baseado em *tagging* (Figura 14) que ele utiliza. Esta camada de *tags* permitirá ao gerenciador manter um controle sobre o uso e o desuso (ciclo de vida) de termos pelo usuário, além de agrupar os objetos categorizados nos diversos sistemas baseados em *tagging*. Dessa forma, integrando a ele conhecimento associado a um determinado termo espalhado pelos diversos sistemas, conforme destacado na Figura 14 com o uso da “tag4”.

3.3 Gerenciamento do ciclo de vida de uma *tag*

Quando uma categorização é efetuada por um usuário, ele utiliza termos de seu vocabulário para anotar o objeto (recurso). Na arquitetura do gerenciador de personomias foram adicionados alguns metadados a cada termo do vocabulário do usuário para o acompanhamento do seu **ciclo de vida** de uma *tag*. Consideramos como ciclo de vida de uma *tag* o momento do seu primeiro uso em uma categorização até o momento em que ela não é mais utilizada para novas categorizações, e nem mesmo para recuperação de informações.

O mapa conceitual da Figura 15 ilustra como estão relacionados os conceitos relativos ao monitoramento do ciclo de vida de uma *tag*. No modelo, pode-se perceber que o uso do vocabulário é considerado em dois contextos distintos, sendo eles: seu uso em categorizações e seu uso para recuperação de informação (busca). Um conjunto de marcadores é atualizado em cada contexto de uso, sendo eles: a **data da utilização** e os **contadores** para cada contexto de uso. Já no momento de uma nova categorização, além da atualização desses marcadores, também se atualiza a data da utilização da palavra como uma *tag* pela primeira vez. É interessante observar que, no acompanhamento do ciclo de vida implementado, a arquitetura não considera a utilização global de um termo, pois o objetivo do gerenciador é o gerenciamento das *tags* da personomia de um usuário.

O término do ciclo de vida de uma *tag* poderá acontecer de duas formas: como uma sugestão do sistema ou como uma escolha pessoal do usuário. Em ambos os casos, a decisão final cabe ao usuário. Encerrar o ciclo de vida de uma *tag* significa marcar a sua inatividade no vocabulário do usuário. Dessa forma, garante-se que esse termo ainda faça parte do vocabulário do usuário, podendo voltar a ser utilizado futuramente como *tag*, dependendo apenas da decisão do usuário. Os metadados (Atributos) que serão colocadas nos termos do vocabulário utilizado geram um conjunto de frequências de uso e que consideram apenas as personomias, em razão de cada usuário manter uma frequência particular de acesso ao

sistema, bem como uma frequência particular para o uso de seu vocabulário para novas categorizações e para recuperação de informação.

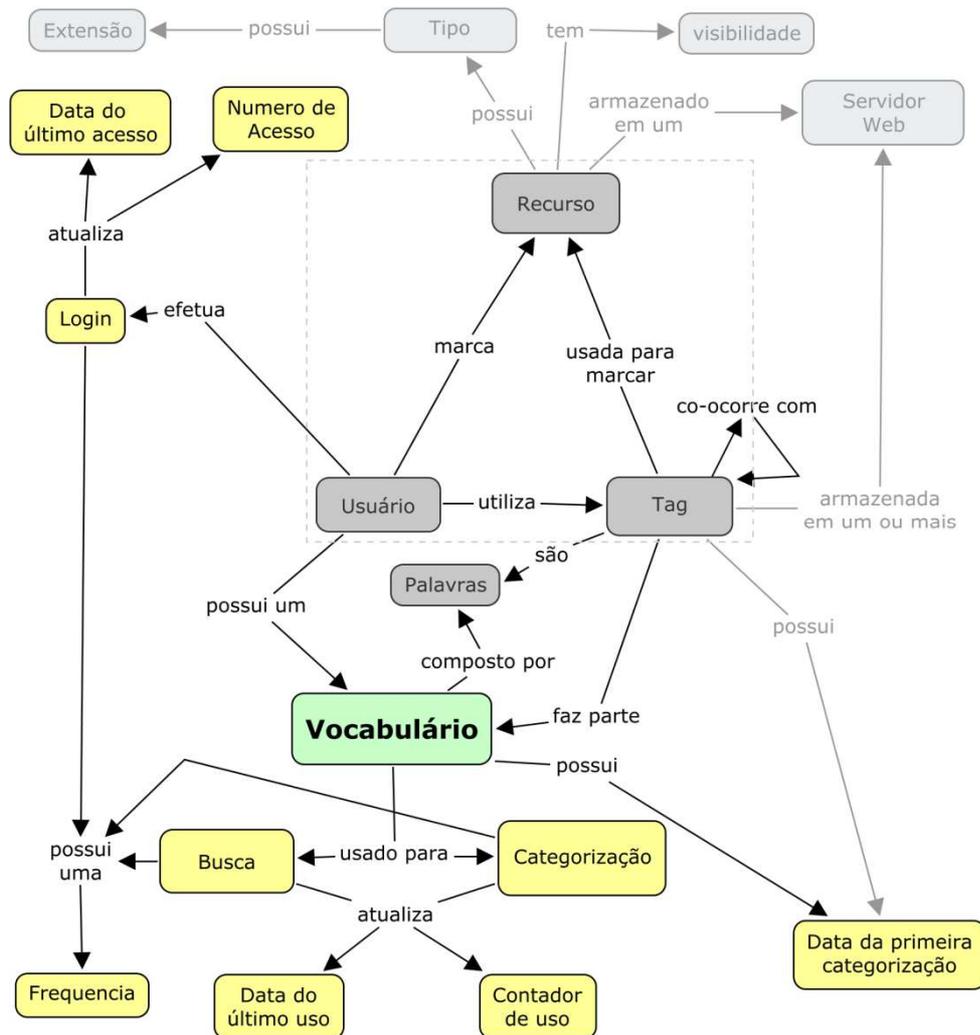


Figura 15 - Modelo de um tagging com ênfase no gerenciamento do ciclo de vida de uma tag

O modelo mostrado na Figura 15 apresenta o processo de *tagging* e os elementos que o compõem. Os pivôs de um processo de *tagging* (usuário, *tag* e recurso) estão destacados no quadro pontilhado. Cada pivô possui suas particularidades, que facilitam distingui-lo e controlá-lo. O usuário é o pivô ativo no processo de *tagging*, ele constrói e organiza suas personomias. Para isso, ele faz acesso ao sistema por meio de um *login* e uma senha e, nesse momento, é marcada sua data de acesso e atualizado o seu contador de números de acessos. O pivô recurso é o objeto de interesse do usuário e a ele estão associados: a visibilidade desse

recurso para outros usuários; o local em que ele está hospedado (servidor *Web*), ou seja, o seu endereço real, e o tipo do recurso (URL, Música, Arquivo texto, etc.), ao qual está associado uma extensão de arquivo (.doc, .xls, .ppt, .pdf, etc.). Ao pivô *tag* é associado: a data na qual a *tag* foi utilizada pela primeira vez; um, ou mais, servidores *Web* nos quais ela está hospedada; e também o vocabulário do usuário a qual está associado. Ao vocabulário do usuário foram associados dois tipos de usos: para busca e para categorização; em cada um destes usos é gravada a data da utilização e atualizado um contador para números de vezes que o termo foi utilizado em buscas ou o em novas categorizações com o objetivo de servir para cálculos da frequência de uso do vocabulário.

Dessa forma, o ciclo de vida de uma *tag* será acompanhado pelo sistema utilizando os registros de datas do uso dos termos e os contadores associados com a utilização do vocabulário do usuário em buscas ou novas categorizações, considerando ainda o número de acessos ao sistema pelo usuário em relação ao tempo de acesso, desde o seu cadastro no sistema, visando determinar a periodicidade de uso do sistema pelo usuário. Assim, o monitoramento do ciclo de vida de uma *tag* será baseado nas datas do último uso nos contextos de nova categorização ou de busca, e nos contadores atualizados em cada um dos contextos de uso. A seguir mostramos na Tabela 1 os cálculos das frequências utilizadas para acompanhar o ciclo de vida da *tag*.

Tabela 1 - Cálculos de frequências de uso da tag em categorização, de busca, de acesso e de uso geral

Cálculos indicativos para o ciclo de vida de uma <i>tag</i> .	
Frequência Acesso (FA)	$FA = \{ [1 / ((DA - D_{cad}) + 1)] * [1 / ((DA - DUA) + 1)] * CA \}$
Frequência Categorização (FC)	$FC = \{ [1 / ((DA - D_{cat}) + 1)] * [1 / ((DA - DUC) + 1)] * CC \}$
Frequência Busca (FB)	$\{ [1 / ((D - DB) + 1)] * [1 / ((DA - DUB) + 1)] * CB \}$
Frequência Uso (FU)	$FUT = (FC * 7) + (FB * 3) / FA$

DA = data de acesso; DCad = Data de cadastro; DUA = data do último acesso; CA = contador de acesso; Dcat = data da categorização; DUC = data da última categorização; CC = contador de categorização; DB = data da primeira busca; DUB = data da última busca;

A frequência de uso de uma *tag* está associada diretamente ao seu uso em categorizações. Isso nos levou a considerar inicialmente somente a frequência das categorizações efetuadas pelo usuário. No entanto, adicionamos a frequência de busca ao cálculo da frequência de uso em razão de uma *tag* poder ter um baixo índice na frequência de categorização, mas ser muito utilizada para recuperar informação. Isso implica que seu índice de uso deve estar diretamente relacionado, também, ao índice da frequência de busca, pois a busca é um tipo de uso da *tag*. Um peso maior foi atribuído à categorização porque ela reflete uma decisão do usuário em empregar um termo qualquer para descrever ou organizar determinado recurso, deixando evidente que, pelo menos no contexto momentâneo, o termo em questão apresenta relevância e está em foco na mente do usuário. A busca utilizando um termo também indica a relevância do mesmo, entretanto, a tendência é que ela ocorra com maior frequência do que a categorização. Os casos nos quais a frequência de categorização for maior que a frequência de busca de determinado termo podem indicar algum comportamento atípico do usuário: um termo que não seja significativo ou uma forma de caracterizar, classificar ou avaliar os recursos categorizados.

Acreditamos que, tanto na categorização quanto na busca, a ordem na qual o termo foi empregado (o primeiro ou segundo termo na categorização, a primeira ou segunda tentativa na busca) exerce algum nível de influência sobre a importância de um termo para o usuário. Entretanto, este tipo de análise, assim como a verificação dos pesos utilizados nos cálculos das frequências deve ser desenvolvida posteriormente. A distribuição de pesos atuais (“7” e “3”) foi feita de forma intuitiva baseada apenas em testes empíricos. O comportamento desses pesos será monitorado e alterado, se for julgado necessário, quando o sistema possibilitar a análise de dados com usuários reais.

Também consideramos importante associar a frequência de acesso do usuário no sistema ao cálculo da frequência de uso de uma *tag*. Com essa frequência, podemos balancear o uso da *tag* em razão do uso efetivo do sistema pelo usuário, ou seja, o quanto o usuário realmente utiliza o sistema. Por exemplo, pode haver usuários com baixas frequências de acesso e, conseqüentemente, com baixas frequências no uso do sistema e de seu vocabulário para categorizar ou para efetuar buscas; e também pode haver usuários com altas frequências de acesso, implicando em frequências altas na utilização do vocabulário para categorizações e buscas. Assim, o emprego da frequência de acesso no cálculo de uso da *tag* tem como objetivo normalizar o índice de uso dessa *tag* com relação à quantidade de uso que o usuário efetivamente faz do sistema como um todo.

Conhecidos os problemas gerados com o vocabulário de um usuário disperso nos diversos sistemas baseados em folksonomia e a solução proposta por meio de um sistema de gerenciamento de *tags*, no próximo capítulo apresentamos as necessidades de um usuário, a especificação dos requisitos funcionais e não funcionais do gerenciador de *tags*, e finalizando, apresentamos a arquitetura para o gerenciador de *tags*.

3.4 Gerenciamento das *tags* das personomias de um usuário

A solução para o problema da existência de múltiplas personomias para um único usuário é a criação de um espaço na *Web* que centralize e gerencie as *tags* provenientes dessas personomias, facilitando a reutilização e controlando o ciclo de vida de cada *tag*. Centralizar e gerenciar as *tags* visa garantir um vocabulário unificado, mais consistente e com maiores possibilidades de reuso, facilitando a criação de relações entre os recursos dos diversos sistemas que utilizam as personomias. Outro ponto a ser considerado é que outros sistemas que apliquem a técnica de folksonomia poderão tirar proveito do vocabulário, mais consistente, de um usuário para sugerir termos para a categorização de recursos em seu espaço

de atuação, pois o usuário deve possuir acesso total e irrestrito aos termos pertencentes ao próprio vocabulário.

Um ponto importante tratado pelo gerenciamento das *tags* é o gerenciamento do ciclo de vida de cada *tag*. O ciclo de vida de uma *tag* corresponde ao seu início ou nascimento, no momento em que ela é utilizada pela primeira vez para uma categorização, até o momento no qual ela não é mais utilizada para novas categorizações e nem mesmo para recuperação de informação por um usuário — isso foi discutido e detalhado na sessão 3.2. Gerenciar o ciclo de vida de uma *tag* é fundamental para uma melhor recomendação de termos para os usuários no momento da categorização ou da recuperação de informação, para a recomendação de objetos e para a manutenção do vocabulário de termos utilizados.

Como já mencionado no capítulo anterior, a viabilidade da solução proposta nesta pesquisa é fundamentada nas discussões de Smith (2008) sobre o ponto de tensão do “Pessoal versus Social ou Coletivo” existente nos sistemas que aplicam o conceito de *tagging*. A questão considerada nesse ponto de tensão é: as pessoas categorizam objetos principalmente em benefício próprio? Ou as pessoas são motivadas pelo desejo para compartilhar informação com um grupo, pelo desejo de ser vistas como instruídas, ou por outros fatores sociais? O gerenciador de *tags*, ao controlar o vocabulário pessoal, promoverá uma melhoria nas categorizações individuais, propagando-a para o âmbito coletivo que utiliza um vocabulário semelhante para categorizar e/ou recuperar informação, bem como para toda a folksonomia de uma forma geral.

3.5 O vocabulário de um usuário

As buscas tradicionais na *web* são feitas utilizando uma ou mais palavras chaves que fazem parte do vocabulário do usuário ou que estão sendo incorporadas a ele. Segundo Lyman (2003), quando um usuário não encontra aquilo que deseja na primeira tentativa,

normalmente, ele adiciona outras palavras chaves às que já foram utilizadas ou troca os termos da consulta. Isso é repetido até que se chegue ao que está sendo procurado ou a algo próximo do que interesse ao usuário. Desse modo, quando o usuário encontra um recurso (objeto) de seu interesse é possível que isto desperte nele o interesse em guardar este objeto para uma recuperação futura de maneira mais rápida e direta.

Considere uma situação na qual uma determinada palavra (“termo1”) foi utilizada em uma busca, resultando em uma recuperação bem sucedida, da qual o usuário deseja guardar determinado recurso para futuras consultas. É bastante razoável pensar que a palavra “termo1” deveria ser utilizada como *tag* para categorizar este recurso. Para que isso aconteça, inicialmente, a palavra deveria ser adicionada ao vocabulário do usuário e esse mesmo termo deveria ser oferecido como primeira opção no momento da categorização do recurso, seguido de outros termos relacionados. Caso o usuário não encontre o que procura utilizando somente o “termo1”, ele adicionará outros termos que julgue necessários. Caso ocorra uma categorização do recurso encontrado, a ordem na qual o usuário foi adicionando outros termos deve ser seguida para a sugestão de termos. Na categorização, essa recomendação da primeira palavra “termo1” que foi utilizada para a busca é justificada pelo fato de que essa foi a primeira palavra que surgiu na mente do usuário ao procurar pelo recurso em questão e, deste modo, existe uma tendência natural do usuário utilizar o mesmo termo quando ele deseja recuperar novamente esse mesmo recurso no sistema no qual efetuou a categorização — Lima (2007) e Artêncio (2007) discutem essa questão com maiores detalhes.

A Figura 16 mostra que as *tags* empregadas em sistemas de categorização fazem parte do vocabulário do usuário e que o vocabulário de uma pessoa é composto por um conjunto de palavras que compõe a linguagem natural desses usuários. No entanto, no vocabulário de um usuário poderão existir termos, ou palavras, que farão sentido somente para ele (1) as quais poderão ser utilizadas como *tags* (2). Com relação aos termos utilizados para categorizar, ou

seja, as *tags*, poderá ainda acontecer de existirem termos que não pertençam ao vocabulário do usuário e a nenhuma linguagem conhecida (3) (esta situação aconteceria para aqueles termos particulares criados pelo usuário e que estão sendo utilizados pela primeira vez também para os erros de digitação), e também de algumas *tags* não estarem no vocabulário do usuário, mas pertencerem ao conjunto de palavras dos diversos idiomas (4).

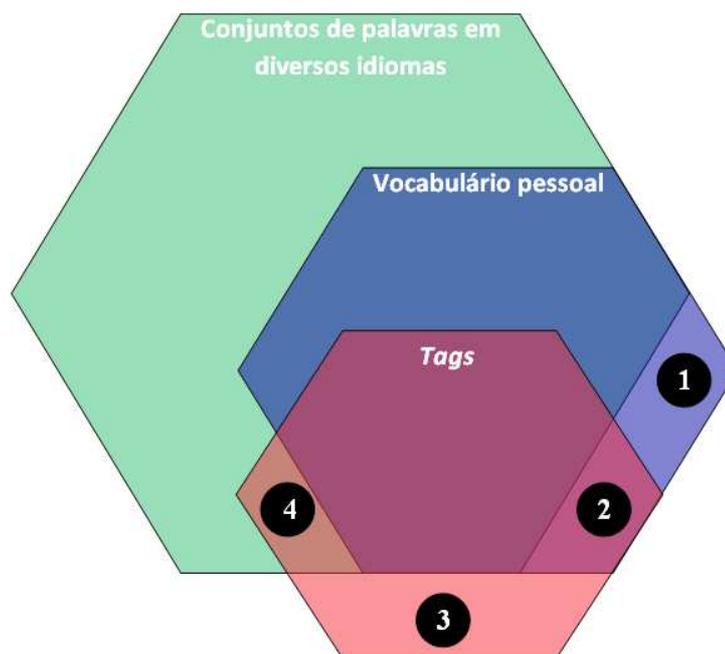


Figura 16 - Relação vocabulário e tag

Esses termos pertencem ao seu vocabulário por serem utilizados com frequência e são responsáveis pelas categorizações que normalmente ficarão restritas ao seu espaço pessoal. Quando certas *tags* são compartilhadas com outros usuários, elas colaboram para o serendipismo e representam a riqueza do vocabulário do usuário, suas particularidades, conhecimentos, preferências, etc.

Os termos utilizados como *tags*, que não pertencerem ao vocabulário de termos da linguagem natural do usuário e nem ao conjunto global de termos existentes nos diversos idiomas (*tags* do tipo 3), devem ser eliminados ou reduzidos a um número mínimo. Esses termos não fazem sentido nem para o usuário e nem para a comunidade (em razão de não ajudarem na recuperação de informação) sendo, provavelmente, *tags* ruidosas ou

extremamente idiossincráticas. Isso poderá ser contemplado empregando-se um corretor ortográfico sobre as *tags* e monitorando-se o uso dos termos, identificando termos que sejam utilizados uma única vez e depois ficam esquecidos e não são mais aplicados em novos processos de categorização ou em processos de recuperação de informação.

O fato de o usuário utilizar termos que não fazem parte do seu vocabulário, embora estes pertençam ao conjunto de idiomas (4), pode levar ao esquecimento do termo pelo pouco uso. A solução aplicável para esse caso seria adicionar esse termo ao vocabulário do usuário e monitorar seu uso para sugerir sua retirada caso ele esteja somente em uma categorização e seja esquecido pelo usuário em buscas futuras.

É interessante discutir a respeito do uso do vocabulário para a categorização de objetos. O elemento *tag* pode ser explorado no sentido de criar sobre ele novas dimensões ou instâncias de uso que deverão ser aplicadas em benefício do usuário. As relações de uso de termos, considerando os já utilizados para categorizar e o vocabulário do usuário serão melhoradas, quer seja para categorizar novos objetos, quer seja para localizar algum recurso já categorizado. Neste sentido, esta pesquisa considera que as *tags* devem ser monitoradas em três instâncias de uso: a **pessoal**, a da **comunidade** escolhida pelo usuário, e a **global** em que são considerados todos os usuários que utilizam o sistema de *tagging*. Assim temos:

1. A instância **pessoal** corresponde ao vocabulário pessoal e são os termos que pertencem à personomia do usuário. Esses termos também pertencem tanto ao vocabulário global quanto ao vocabulário das comunidades que o usuário faz parte. Esse espaço pessoal é o primeiro a ser considerado nas buscas e, também, é a partir dele que devem ser feitas as principais sugestões de *tags* para as novas categorizações.
2. A instância da **comunidade**, definida pelo usuário, corresponde ao vocabulário de termos que as pessoas pertencentes a uma determinada comunidade (rede de usuários) utiliza para categorizar seus objetos. No caso dos sistemas baseados em folksonomia,

é possível que existam termos específicos que somente serão utilizados por um grupo de pessoas embora eles pertençam também ao vocabulário global. Assim é interessante ter um espaço limitado para a busca que facilite a recuperação da informação, bem como a de sugestão de *tags* para novas categorizações.

3. Finalmente, a instância **global** do espaço de *tags* pode ser considerada como o vocabulário global das palavras utilizadas por todos os usuários de um sistema baseado em folksonomia.

O conceito de comunidade deve ser contemplado na arquitetura do sistema, de modo que o usuário possa escolher e definir os demais usuários que podem compartilhar e utilizar os recursos por ele categorizados. O trabalho de Pereira (2008) sobre o conceito de *Folkauthority* será posteriormente incorporado ao sistema proposto neste trabalho e tornará possível a identificação de comunidades ou grupos de usuários.

3.6 As múltiplas personomias de um usuário

De acordo com Smith (2008), um sistema de *tagging* pode ser visto na intersecção de três perspectivas, sendo elas: a da informação pessoal, a do software social e a da arquitetura da informação, conforme ilustra a Figura 17. Essas perspectivas se encontram fortemente conectadas, de tal forma que ao fortalecermos uma delas provocamos um aumento nas demais, o contrário também poderá acontecer, se uma delas perder qualidade as demais tendem a perder também.

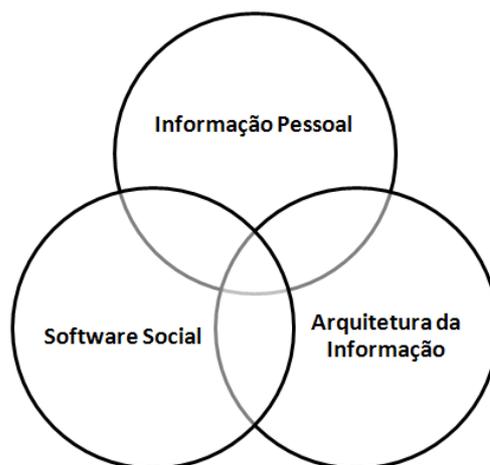


Figura 17 - Perspectivas de um sistema de tagging - figura adaptada de Smith (2008)

As perspectivas ilustradas na Figura 17 podem ser decompostas da seguinte forma: os sistemas baseados em folksonomia são softwares sociais¹⁸, uma categorização efetuada nestes sistemas exige apoio de uma arquitetura de informação bem projetada, a qual deve atender às necessidades de gerenciamento de informação pessoal dos usuários. As informações pessoais geram uma organização de dados na forma de uma personomia que podem colaborar com o coletivo, ou seja, com uma folksonomia. A colaboração e o compartilhamento da informação entre os usuários do sistema tornam-se possíveis por meio de mecanismos presentes no software social, que possibilitam a integração das personomias. Dessa forma, um software social que implementa a técnica de folksonomia compartilha e expõe os objetos categorizados e os termos utilizados para as categorizações. Assim, nestes sistemas, todos os usuários têm acesso tanto aos objetos quanto aos termos utilizados nas personomias dos outros usuários, podendo visualizar estes objetos e/ou reutilizar os termos em novas categorizações, contribuindo, assim, com a organização coletiva dos objetos (recursos).

¹⁸ Smith (2008) utiliza o termo “software social” para designar os aplicativos que possibilitam a comunicação, colaboração e compartilhamento on-line.

Como já mencionado anteriormente, cada sistema baseado em *tagging*, normalmente, é destinado a um único tipo de objeto, gerando e armazenando as personomias criadas pelos seus usuários no momento da categorização desses objetos. Essas personomias contêm informações pessoais de cada usuário, os vocabulários de termos utilizados nas categorizações e os objetos categorizados. Portanto, um usuário termina tendo várias personomias, uma em cada sistema baseado em folksonomia que ele utilizar para categorizar informações.

Em princípio, as *tags* empregadas nos sistemas baseados em *tagging* devem pertencer ao vocabulário natural do usuário, pois seu propósito primário é auxiliar na organização e na recuperação de informação pessoal. Assim, os termos utilizados como *tags* e que fazem parte do vocabulário do usuário poderiam ser reutilizados em outros sistemas baseados em folksonomia que o mesmo usuário utilize. Essa reutilização de termos deveria ser natural, pois o usuário cria um vocabulário de termos que são utilizados para seu processo de categorização, e não exclusivamente para descrever um único objeto. No contexto atual, a falta de integração entre as personomias de um mesmo usuário dificulta a organização e a recuperação de conhecimento a partir de suas *tags*, de seus objetos e dos relacionamentos de categorização existente entre ambos. Por causa dessa limitação pode-se perder informações ricas sobre o comportamento, preferências e conhecimento do usuário, que não podem ser obtidas analisando cada personomia separadamente.

Como um caso prático, considere a situação na qual uma pessoa participa de uma conferência. Nesta conferência, ela fotografa, grava um vídeo de algum momento interessante e efetua anotações. Esses conteúdos são todos publicados: as fotos são publicadas no sistema *Flickr* (com o objetivo de organizar e mostrar para seus amigos), o vídeo no sistema *Youtube* e as anotações são publicadas num *blog* e este *post* é salvo como um *bookmark* no *Delicious*. A organização de conteúdos nesses sistemas é possível com a utilização de *tags* inseridas pelo usuário, as quais são utilizadas para organizar, descrever ou qualificar o conteúdo de modo a

facilitar sua recuperação pelo próprio usuário ou pelos demais usuários desses sistemas. Essa mesma pessoa ao navegar na *Web* encontra conteúdos publicados em *blogs* e em outras páginas, que se referem à mesma conferência que ela participou e utiliza o sistema *Delicious* para salvar essas páginas como *bookmarks*. Esses *bookmarks* serão indexados e organizados com *tags* escolhidas pelo usuário, o qual teria seu trabalho facilitado se pudesse utilizar um conjunto padrão de termos, ou seja, termos compatíveis e diretamente relacionados (provavelmente o mesmo termo ou conjunto de termos) com os quais ele já utilizou para descrever suas fotos (*Flickr*), seus vídeos (*Youtube*) e suas postagens sobre a conferência. Como cada personomia é gerada separadamente, a não ser que o usuário acesse os sistemas previamente mencionados para lembrar quais foram os termos utilizados para categorizar os objetos que se encaixariam num mesmo modelo mental de organização, ele dificilmente conseguirá manter um padrão no vocabulário e uma consistência nos termos utilizados para efetuar suas categorizações. Nesta situação, a geração de problemas de sinonímia e polissemia já começa a ocorrer dentro das personomias, ou seja, no conjunto de informação do próprio usuário e, conseqüentemente, causará impacto também na folksonomia. Além disso, no que diz respeito a organização dessas informações e a recuperação de conhecimento, uma integração entre essas personomias auxiliaria a identificar que, apesar de distintos, os objetos previamente mencionados compartilham um núcleo comum de atributos e de contexto para o usuário.

Para que a organização do vocabulário de um usuário que categoriza objetos em diversos sistemas seja facilitada, as *tags* utilizadas em suas diferentes personomias devem ser visíveis ao usuário (Figura 18). Como não foi encontrado nada na literatura atual que indique a existência de tal facilidade, propõe-se nesta dissertação o desenvolvimento de uma ferramenta que possibilite gerenciar o vocabulário de *tags* utilizadas nas diversas personomias de um usuário, centralizando-as e oferecendo recursos para sua manutenção.

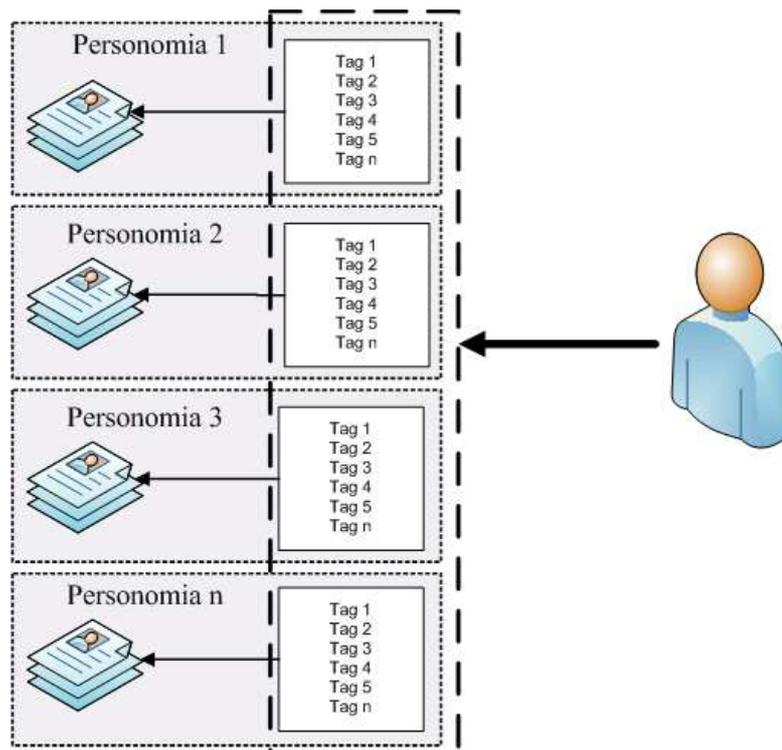


Figura 18 - Múltiplas Personomias de um usuário

Conforme discutido no capítulo anterior, boa parte dos problemas relacionados aos termos utilizados como *tags* enfrentados pelos sistemas baseados em folksonomia é reflexo da falta de gerenciamento das personomias e da falta de ferramentas que apoiem os usuários. Embora sistemas como o *Delicious* ofereçam facilidades de sugestão de *tags*, isto não é suficiente, pois não possibilita nenhum nível de integração com as personomias de outros sistemas. Um dos desafios de um sistema gerenciador de *tags* é oferecer mecanismos que auxiliem o usuário no momento da categorização de um objeto, procurando criar relações entre os objetos ou recursos que foram categorizados nos diversos sistemas. Assim, as relações entre os recursos serão estabelecidas com o uso de termos de um vocabulário centralizado, podendo facilitar o trabalho do usuário na recuperação de suas informações e refletindo nos sistemas que esse usuário utiliza.

[Página intencionalmente deixada em branco]

Proposta de Arquitetura para o Gerenciador de *Tags*

A existência de diversos sistemas baseados em folksonomia faz com que um usuário que deseje categorizar objetos diferentes tenha que manter uma personomia em cada sistema utilizado. Sendo assim, cabe a ele estabelecer um conjunto de termos em cada um desses sistemas, o que dificulta a padronização e a consistência dos termos utilizados e a criação de um vocabulário único. Conforme discutido no capítulo anterior, faz-se necessário gerenciar o vocabulário do usuário em suas personomias. Neste capítulo, descrevemos os principais pontos considerados para a definição da arquitetura do gerenciador de *tags*. Primeiramente apresentamos um modelo conceitual para o gerenciamento de *tags*, em seguida expomos as necessidades de um usuário, o levantamento de requisitos, os diagramas de casos de uso e os diagramas de atividades. Na seqüência, mostramos o modelo de dados e a arquitetura criada para o gerenciador de *tags*.

4.1 O Mapa conceitual (metamodelo) do gerenciador de *tags*

O mapa conceitual, ilustrado pela Figura 19, mostra a visão dos conceitos envolvidos no problema de gerenciamento de *tags* e suas relações com o sistema de gerenciamento de *tags* proposto. O gerenciador de *tags* deve ter acesso a um, ou mais, sistemas baseados em folksonomia e identificar o usuário, seus *logins* e suas senhas nesses sistemas. Um usuário deste gerenciador possui um conjunto de *tags* e um conjunto de objetos em um ou mais sistemas baseados em folksonomia, ou seja, ele pode possuir várias personomias. Deste modo, o gerenciador deve possibilitar tanto a recuperação e a alteração desses dados quanto a inserção de novas categorizações.

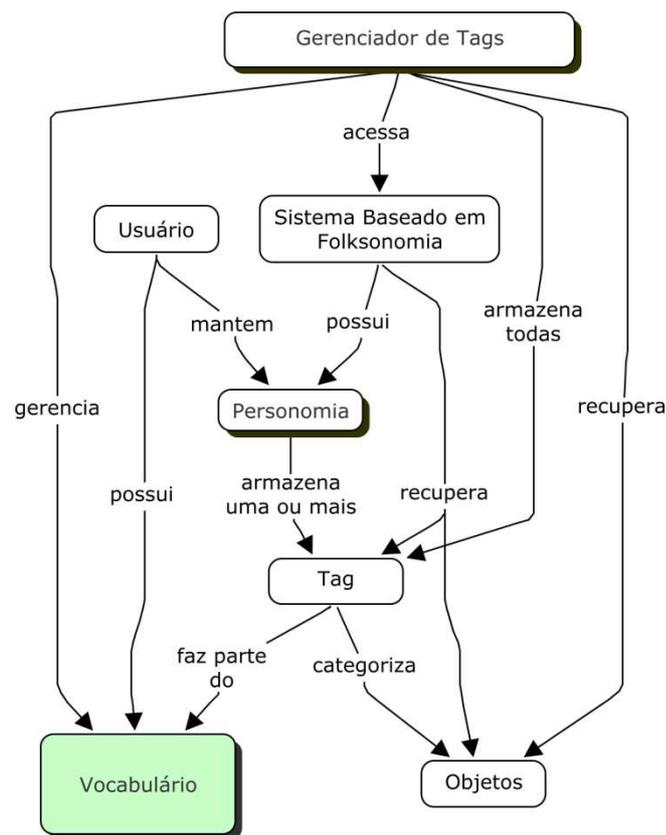


Figura 19 - Mapa conceitual do gerenciador de *tags*

No mapa conceitual acima, o conceito de **vocabulário** está associado a um **usuário**, cujas *tags* devem fazer parte desse vocabulário. Isso implica que o conjunto de palavras do

vocabulário desse usuário deve ser maior ou igual ao número de termos que ele utilizou na categorização de seus recursos (**objetos**). Com base nestes estudos as necessidades do usuário foram estabelecidas.

4.2 Necessidades do usuário

Ao utilizar mais de um sistema baseado em folksonomia para a categorização de informação o usuário tem seu vocabulário de termos, empregados nestas categorizações, distribuído nas personomias geradas em cada um dos sistemas. Este problema nos leva ao requisito principal de nosso gerenciador: o de manter um vocabulário unificado para o usuário, pois a descentralização dificulta a padronização e a manutenção dos termos utilizados, prejudicando a reutilização de *tags* para a categorização e para a recuperação de informação.

Como parte do processo de um sistema baseado em folksonomia, os usuários necessitam que sejam oferecidas sugestões de termos quando eles estão categorizando novos recursos. Como discutido no capítulo 3, estas sugestões deveriam ser baseadas em um único vocabulário, pois os usuários podem desejar categorizar objetos que possuam relação entre si como, por exemplo, fotos, vídeos e *blogs* que se referem ao mesmo assunto. Em geral, os sistemas baseados em folksonomia permitem que os usuários escolham termos para organizar seus objetos em suas personomias. Como, neste caso, os objetos possuem uma relação entre si, as *tags* que deveriam ser utilizadas nas categorizações pelos usuários deveriam manter uma relação que os ajudassem na recuperação futura desses objetos, ou seja, os usuários deveriam estabelecer uma relação entre eles utilizando as mesmas *tags*, ou *tags* que sejam relacionadas de alguma forma.

O monitoramento do uso do vocabulário do usuário, utilizado em buscas e em categorizações, poderá dar indicações ao sistema para a sugestão de termos a serem utilizados

nas novas categorizações que o usuário efetuar. Essas indicações passam a ser globais, pois contemplam o uso do vocabulário em diversos sistemas baseados em folksonomia.

Quando um usuário deseja recuperar informações, ele normalmente acessa o sistema baseado em folksonomia no qual categorizou o objeto que deseja localizar. Caso ele precise recuperar outros tipos de objetos relacionados, ele deverá acessar outros sistemas e fazer uma nova busca para cada sistema acessado. O ideal para este usuário é que fosse possível realizar uma busca simultânea em todos os sistemas que ele utiliza para categorizar seus objetos, resultando em uma redução de esforços e na possibilidade de obter informações mais pertinentes a uma busca.

Além da centralização do vocabulário do usuário e do monitoramento do ciclo de vida das *tags*, o gerenciador também auxiliará na manutenção dos termos utilizados nas categorizações. O fato de um usuário possuir várias personomias dificulta o processo de verificação e correção de *tags* quanto à sua coerência ortográfica ou semântica, o que eleva as chances da existência de termos ruidosos. Conforme abordado por Côgo e Da Silva (2008), se os termos ruidosos forem removidos do vocabulário de um usuário e se os problemas de polissemia e sinonímia forem reduzidos, os termos situados na cauda longa, gerada pela distribuição do uso das *tags* que permanecerem no vocabulário do usuário, representarão as preferências ou as particularidades desse usuário e poderão ser aplicados para a recomendação de informação ou de outros termos. Como exemplo de um problema que pode ser amenizado com esse tipo de gerenciamento considere um usuário que utiliza uma *tag* “X” para categorizar recursos em diversos sistemas baseados em folksonomia e, por alguma razão, ele resolve trocar a *tag* “X” pela *tag* “Y”. Neste caso, esse usuário teria que acessar cada sistema baseado em folksonomia, procurar pela *tag* “X” e renomeá-la para “Y”. O sistema de gerenciamento de *tags* pode facilitar o trabalho do usuário efetuando essa alteração automaticamente em todas as personomias que o usuário deseja alterar.

Diante dessas necessidades levantadas para o usuário que mantém múltiplas personomias, na próxima sessão definimos os requisitos funcionais e não funcionais.

4.3 Requisitos

Definidas as necessidades dos usuários, descrevemos a seguir os requisitos para a ferramenta gerenciadora do vocabulário de termos utilizados para categorização. A seguir, mostramos os requisitos funcionais e não funcionais que serão considerados no projeto dessa ferramenta denominada de *Tag Manager* (TM).

4.3.1 Requisitos funcionais

Os requisitos funcionais descrevem o comportamento do sistema, ou seja, suas ações para cada possível entrada de dados, demonstrando as funcionalidades que ele deverá dispor para satisfazer as necessidades dos usuários. Os requisitos funcionais levantados são:

- Permitir o cadastro de usuários;
- Permitir o cadastro de *logins* para diferentes sistemas baseados em *tagging*;
- Buscar as *tags* utilizadas pelo usuário nos diversos sistemas baseados em *tagging* que ele utilizar;
- Sincronizar as *tags* utilizadas pelos usuários nos sistemas baseados em *tagging* utilizados, trazendo elas para uma base comum e refletindo a alteração em um *tag* nos diversos sistemas;
- Sugerir *tags*, considerando o vocabulário do usuário no momento da categorização;
- Adicionar ao vocabulário do usuário os termos utilizados para recuperação de informação;
- Efetuar buscas nos sistemas homologados junto ao gerenciador;

- Monitorar os sistemas baseados em folksonomia que estiverem homologados junto ao gerenciador. Esse requisito refere-se em refletir as alterações efetuadas nos sistemas baseados em folksonomia para o gerenciador de *tags*.

4.3.2 Requisitos não-funcionais

Os requisitos não-funcionais, normalmente, são aspectos de qualidade de um sistema, estando relacionados à confiança do usuário no sistema. A confiança está relacionada à disponibilidade, confiabilidade, segurança e proteção (SOMMERVILLE, 2003), acrescentamos ainda os requisitos não-funcionais de desempenho e usabilidade.

Disponibilidade: Esse requisito está relacionado à capacidade do sistema de disponibilizar serviços aos seus usuários. O gerenciador o contemplará disponibilizando ferramentas *on-line* diretamente no *website* do gerenciador de *tags*, por meio de acesso direto à página do sistema gerenciador. Outra forma de disponibilização de serviços será por meio de um *plugin* que funcionará diretamente no *browser* do usuário, evitando, assim, que o seu uso esteja condicionado ao acesso via página *Web*. Além disso, o vocabulário do usuário, as estatísticas sobre o uso de termos e outras funcionalidades, serão expostos por meio de uma *API (Application Programming Interface)* com serviços e dados globais, públicos e restritos ao usuário com cadastro no sistema gerenciador.

Confiabilidade: A confiabilidade do gerenciador de *tags* será alcançada por meio de políticas de utilização e armazenamento de dados transparentes ao usuário. Os objetivos do compartilhamento de informações com outros membros da comunidade e a forma como esse compartilhamento é realizado também deve ser claramente definidas e especificadas, assim como a disponibilização dos serviços e dados especificados para os usuários.

Segurança: o sistema de gerenciamento manterá uma cópia dos dados necessários para o gerenciamento do vocabulário do usuário. Caso haja perda dos dados referentes às *tags*

no gerenciador, provenientes de outros sistemas baseados em folksonomia, essa base de dados poderá ser refeita de forma rápida e sem muitos prejuízos, em razão desses termos estarem disponíveis nos sistemas hospedeiros de suas categorizações. Todo esse processo de reestruturação de seus dados será transparente para o usuário.

Proteção: Sistemas *Web* estão sujeitos a ataques constantes. O gerenciador de *tags* contará com a proteção de rede do próprio servidor no qual estiver instalado. Quanto aos dados particulares dos usuários, esses são protegidos por políticas de autenticação de acesso. O acesso ao gerenciador, mediante o cadastro do usuário, permitirá que sejam cadastrados os *logins* e senhas dos sistemas baseados em folksonomias que o usuário já utiliza para categorizar informação. A garantia dada para proteção e privacidade de seus dados será estabelecida pelos termos de uso disponibilizado publicamente no *website* e por políticas de privacidade e segurança quanto ao uso da informação incorporadas ao gerenciador de *tags*. Ao se cadastrar no sistema de gerenciamento de *tags*, o usuário deverá tomar ciência e concordar com os termos para adquirir o direito de utilizar os serviços disponibilizados.

Desempenho: O tráfego de dados na *Internet* é um fator crítico que afetará o gerenciador de *tags*, uma vez que todo o vocabulário de um usuário deverá estar presente no local em que essa categorização ocorrerá. Para minimizar esse problema, o formato a ser adotado é o JSON (*JavaScript Object Notation*)¹⁹ e XML (*Extensible Markup Language*) utilizados para trocas de informações, sendo estes de fácil compreensão e com um custo baixo para trafegar na *Internet*.

Usabilidade: O gerenciador de *tags* deve ser simples de utilizar, prático e de fácil compreensão. O principal objetivo de um recurso que gerencie e centralize as personomias dos usuários é facilitar o trabalho desses usuários, tornando-os mais produtivos e permitindo que eles aproveitem o máximo possível dos benefícios propiciados pela técnica de

¹⁹ <http://json.org>

folksonomia. Desse modo, é essencial que o sistema gerenciador simplifique os processos de categorização e de recuperação de informação que atualmente os usuários precisam executar em vários sistemas distintos. Para satisfazer este requisito, deve-se efetuar o *design* de interação do produto, testes e validações com protótipos, aplicações de técnicas para a avaliação da usabilidade e da comunicabilidade da ferramenta, além de testes com usuários reais.

É preciso reforçar a questão da confiança no que se refere às políticas de privacidade dos *logins* e senhas que os usuários utilizam nos sistemas baseados em folksonomia e que serão solicitadas pelo gerenciador. O fato de o usuário ter que informar seu *login* e senha utilizados em outros sistemas pode levá-lo a não querer utilizar o gerenciador. Além dos termos de compromisso entre as partes (sistema gerenciador de *tags* x usuário) essas questões devem ser analisadas e outras abordagens relacionadas à privacidade devem ser adotadas para a segurança do usuário e suas informações presentes em outros sistemas. Deste modo, o gerenciador permitirá que o usuário deixe em branco o campo senha do sistema baseado em folksonomia cadastrado e informe-a no momento em que fizer uma requisição de sincronização de *tags*.

A Figura 20 ilustra uma abstração da estrutura necessária para o gerenciamento de *tags*. A interface *Web* (visão de dados pelo usuário) para acesso é basicamente formada por dados estruturados e prevê a utilização da linguagem XML e JSON. O sistema *Tag Manager* é o componente que gerencia o acesso a sistemas externos, processa respostas dos sistemas baseados em folksonomia e prepara os dados a serem mostrados para o usuário. Duas conexões HTTP (*Hypertext Transfer Protocol*) são exigidas, sendo uma de interação do usuário com o sistema e outra de interação do sistema com as personomias do usuário.

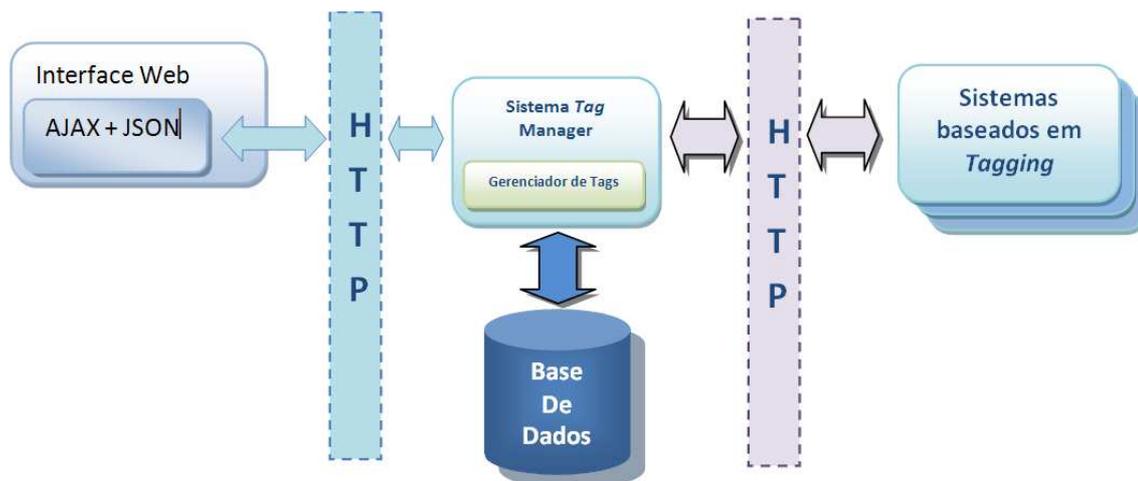


Figura 20 - Sistema Tag Manager

Para contemplar essa definição inicial, abaixo são apresentados os principais casos de uso em conjunto com os diagramas de atividades, representando a interação do usuário com a interface *Web*. Para uma modelagem mais detalhada, são apresentados os diagramas arquiteturais com as abstrações da arquitetura proposta: diagramas com a visão de casos de uso; diagramas de atividades utilizando a linguagem UML (*Unified Modeling Language*) e um diagrama com uma visão de alto nível dos elementos que compõem a arquitetura. Finalmente, apresentamos um diagrama entidade-relacionamento das tabelas que compõem a base de dados do sistema gerenciador de *tags*, destacando as tabelas necessárias para o gerenciamento do vocabulário do usuário.

4.4 Casos de uso

A visão de casos de uso descreve o comportamento do sistema e como ele deve ser visto pelos usuários finais e engenheiros de software nas fases de análise, implementação e testes (BOOCH, RUMBAUCH, & JACOBSON, 2005). Na Tabela 1, estão os casos de uso do software que fará o gerenciamento de *tags*.

Tabela 3 - Descrição essencial do caso de uso "Cadastrar sistemas"

Usuário	Gerenciador de tags
1. Fornece sua identificação	2. Identifica o <i>Usuário</i> 3. Oferece opções de cadastro de sistemas baseados em <i>tagging</i>
4. Fornece as informações necessárias sobre o sistema a ser cadastrado (<i>login</i> - obrigatório e a senha - opcional)	5. Persiste as informações

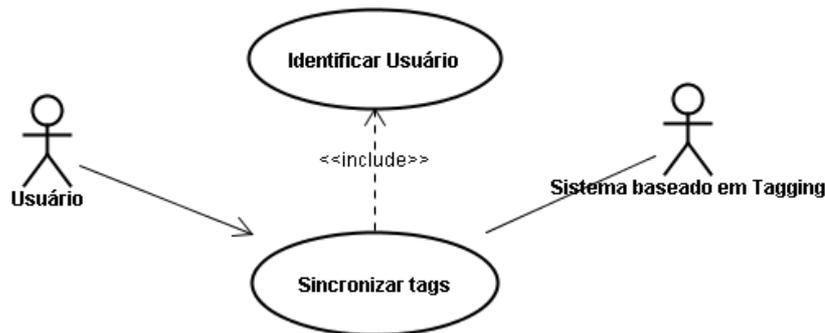


Figura 22 - Caso de uso "Sincronizar tags"

Tabela 4 - Descrição essencial do caso de uso "Sincronizar tags"

Usuário	Gerenciador de tags
1. Fornece sua identificação	2. Identifica o <i>Usuário</i> 3. Fornece opções
3. Solicita sincronização das <i>tags</i>	4. Fornece opções de escolha dos sistemas a serem sincronizados
5. Escolhe um ou mais sistemas ou pode marcar todos.	5. Comunica com Sistemas baseados em folksonomia para sincronização em <i>background</i>

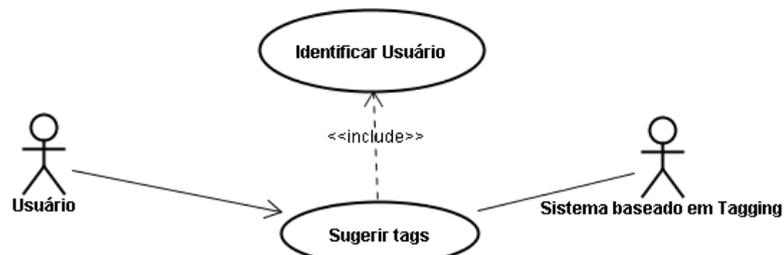


Figura 23 - Caso de uso "Sugerir tags"

Tabela 5 - Descrição essencial do caso de uso “sugerir tags”

Usuário	Gerenciador de tags
1. Abre interface para categorizar um objeto em um dos sistemas baseado em folksonomia que ele tenha conta.	2. Oferece termos do vocabulário do usuário para serem utilizados como tags.
3. Escolhe as tags do seu vocabulário global ou criar novas tags.	4. Grava as categorizações para um determinado objeto. 5. Marca a utilização dos termos para categorização.

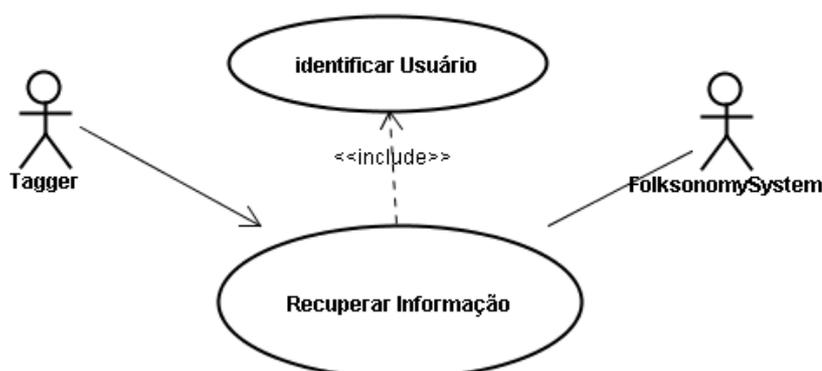


Figura 24 - Caso de uso “Recuperar informação”

Tabela 6 - Descrição essencial do caso de uso recuperar informação

Usuário	Gerenciador de tags
1. Insere termos no mecanismo de busca;	2. Marca que o termo foi utilizado para recuperação. 3. Efetua a busca de acordo com os parâmetros. 4. Apresenta o resultado para o usuário.
5. Analisa os resultados apresentados, caso seja o que procura, utiliza a informação, caso contrário volta a interação 1.	

Os diagramas de atividades modelam a interação do usuário com o sistema de gerenciamento de tags contemplando os casos de uso definidos. Nesse tipo de diagramas, abstrai-se uma seqüência para que o fluxo da lógica possa ser visualizado e segmentado (PENDER, 2004). Abordamos os principais casos de interação na Figura 25. Conforme pode ser observado, quando um usuário acessar o sistema ele deverá informar seu login ou efetuar o cadastro no sistema para, então, cadastrar os sistemas baseados em folksonomia que ele utiliza.

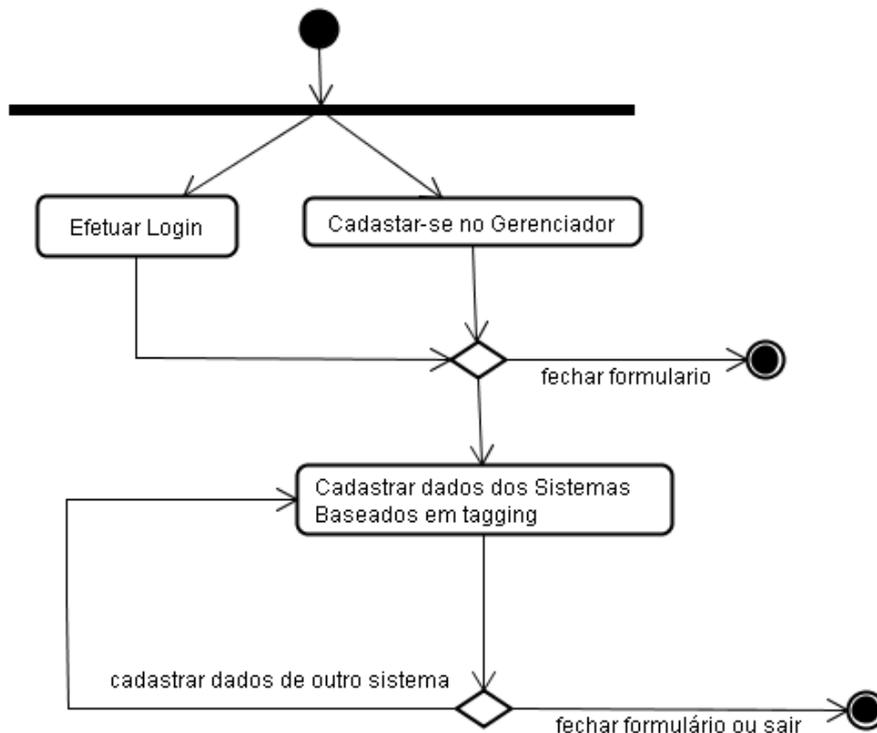


Figura 25 - Diagrama de atividades para o caso de uso “Cadastrar Sistemas”

4.5 Diagramas Atividade

A Figura 26, a seguir, mostra a interação do usuário ao acessar o sistema, ele poderá explorar o sistema no ambiente público oferecido ou autenticar-se para utilizar o sincronizador de *tags*, para recuperar recursos pessoais bem como para efetuar categorizações utilizando os termos de seu vocabulário.

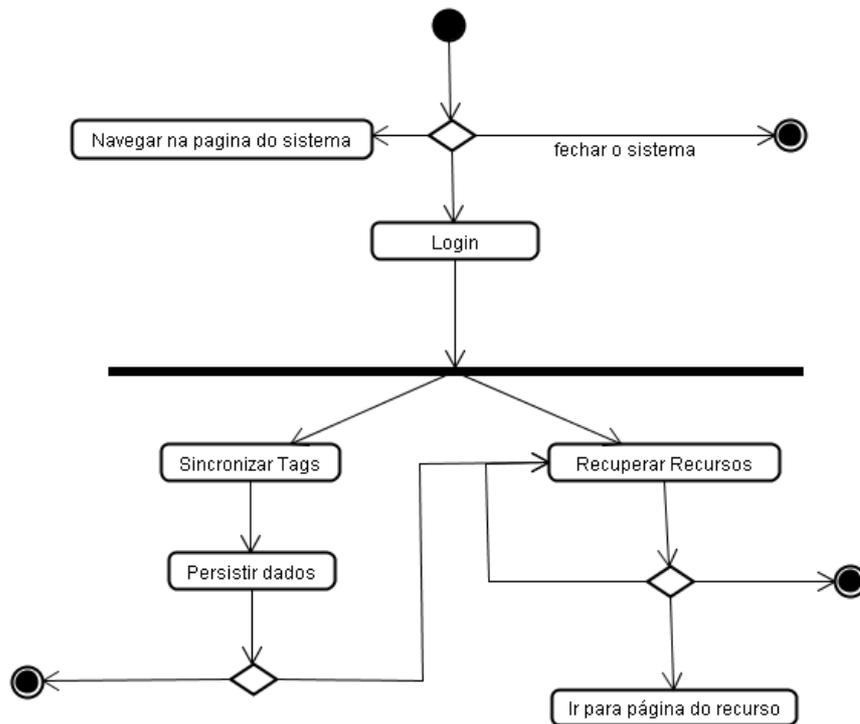


Figura 26 - Diagrama de atividades para os casos de uso “Sincronizar tags” e “Recuperar Informações”

A Figura 27 considera um usuário autenticado no sistema e que pode tanto fazer buscas nos sistemas baseados em folskonomia quanto acessar qualquer sistema cadastrado em seu perfil para efetuar categorizações de objetos. Quando o usuário categorizar um objeto, ele poderá solicitar a intervenção do gerenciador de *tags* para oferecer-lhe sugestões de *tags* e atualizar as *tags* escolhidas no vocabulário mantido pelo gerenciador de forma automática.

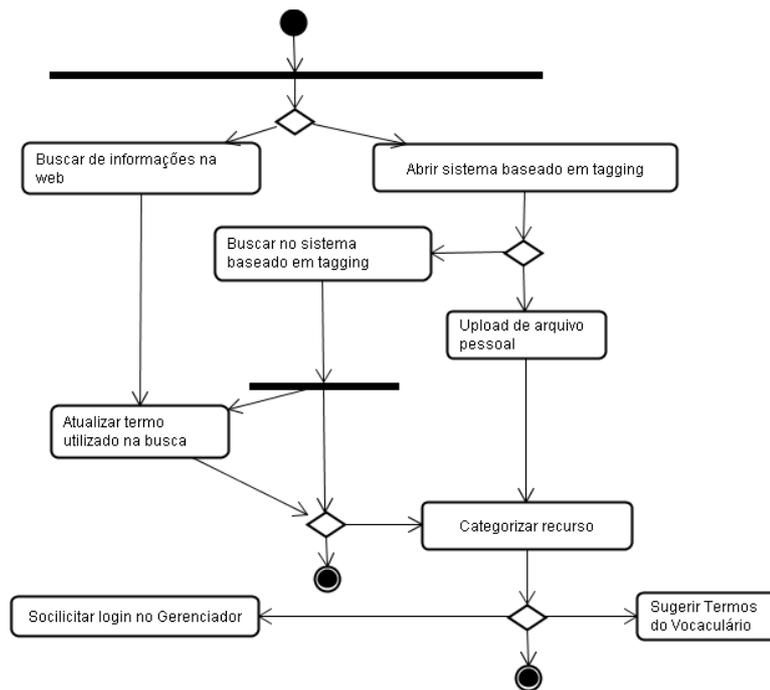


Figura 27 - Diagrama de atividades para o caso de uso “Recomendar Categorização”

A recomendação de termos para uma categorização será feita no momento de uma categorização com base no vocabulário unificado dos sistemas baseados em folksonomia e, também, com base nos termos utilizados em buscas efetuadas pelo usuário.

4.6 Diagrama arquitetural

Dado os requisitos e os casos de usos já expostos, o sistema de gerenciamento de *tags* foi dividido em cinco camadas acopladas pelas interfaces de programação que trabalham juntas para satisfazer as necessidades do usuário. A Figura 28 ilustra a arquitetura em camadas que tem como objetivo principal a divisão de responsabilidade e a criação de um isolamento entre elas visando facilitar a codificação do sistema e sua futura manutenção.

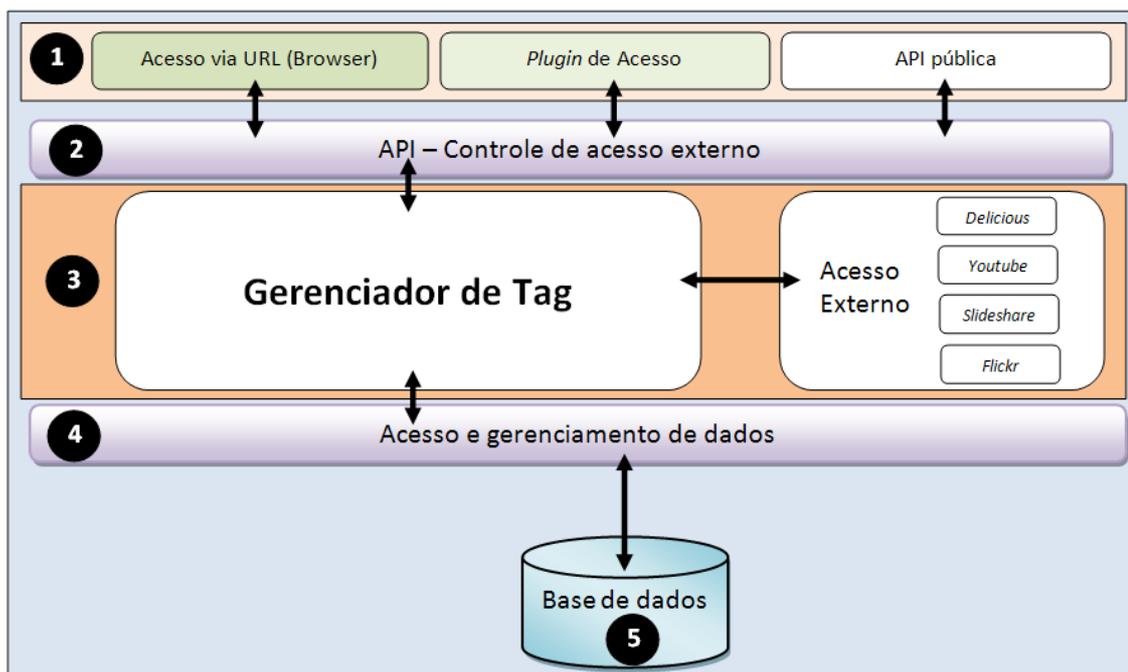


Figura 28 - Visão da arquitetura do sistema gerenciador de tags

Para facilitar a visualização e a compreensão cada camada foi numerada, sendo detalhada abaixo:

1. Camada de interação com o usuário

Nesta camada ocorrerem às interações do usuário com o sistema. No detalhe “1”, da Figura 28, são previstas: a interação via navegador do usuário diretamente no *website* do sistema gerenciador de *tags*, um *plugin* a ser instalado no navegador para facilitar o acesso a dados e a disponibilização desses dados via API pública por meio de um documento JSON ou XML. Outras formas de interação poderão ser implementadas para as mais diversificadas tecnologias, bastando apenas prover uma forma de acesso a camada de controle.

2. Camada de controle

Esta é a camada que recebe e responde todas as solicitações feitas pelo usuário ao gerenciador de *tags*. Essa camada é integrada com a camada de interação para receber as requisições e com a camada de gerenciamento que é responsável pelo processamento dos dados e pelo atendimento das solicitações. Dessa forma, toda e qualquer solicitação vinda de

qualquer dispositivo conectado a *Internet* deverá interagir com essa camada antes de se dirigir ao gerenciador propriamente dito.

3. Camada de negócios

Esta é a camada de lógica de negócios do gerenciador de *tags*, a qual é responsável pelo processamento e pela resposta das solicitações efetuadas pelo usuário por meio da camada de controle. A lógica de acesso aos sistemas baseados em *tagging* fará parte dessa camada juntamente com os *parsers* que extrairão as informações advindas dos sistemas externos, prepararão esses dados e os enviarão para a persistência no sistema *Tag Manager*. Isolando a camada de negócios e atribuindo a ela a responsabilidade de preparação dos dados para a interface, independente de que tipo de dispositivo ou local de onde venha a requisição deixamos transparente a interação do usuário com o sistema facilitando a construção das lógicas do sistema que farão o gerenciamento das *tags* do usuário.

4. Camada de persistência de dados

Esta é a camada responsável pela lógica de persistência de dados, inicialmente, um SGDB (Sistema de Gerenciamento de Banco de Dados), porém futuramente poderão ser utilizadas outras estruturas para a persistência. Essa camada é a responsável por fazer o mapeamento objeto-relacional, controlando, assim, toda a transação efetuada sobre esses dados. Dessa forma, essa camada isola e abstrai o acesso a dados armazenados pelo sistema, isso nos permitirá efetuar mudanças de banco e do tipo de persistência dos dados bem como, mudanças do local em que esses dados são persistidos.

5. Camada de dados

Esta é a camada responsável pelo armazenamento dos dados. Essa camada pode suportar diversas estruturas de armazenamento em disco, como banco de dados, XML e texto e outras estruturas. O seu isolamento do sistema permitirá que ela possa estar em qualquer lugar, desde que o sistema possa interagir com ela. O diagrama apresentando pela Figura 29

demonstra a estrutura de tabelas consideradas necessárias para o gerenciamento das *tags* e para a composição de um serviço de *tagging*.

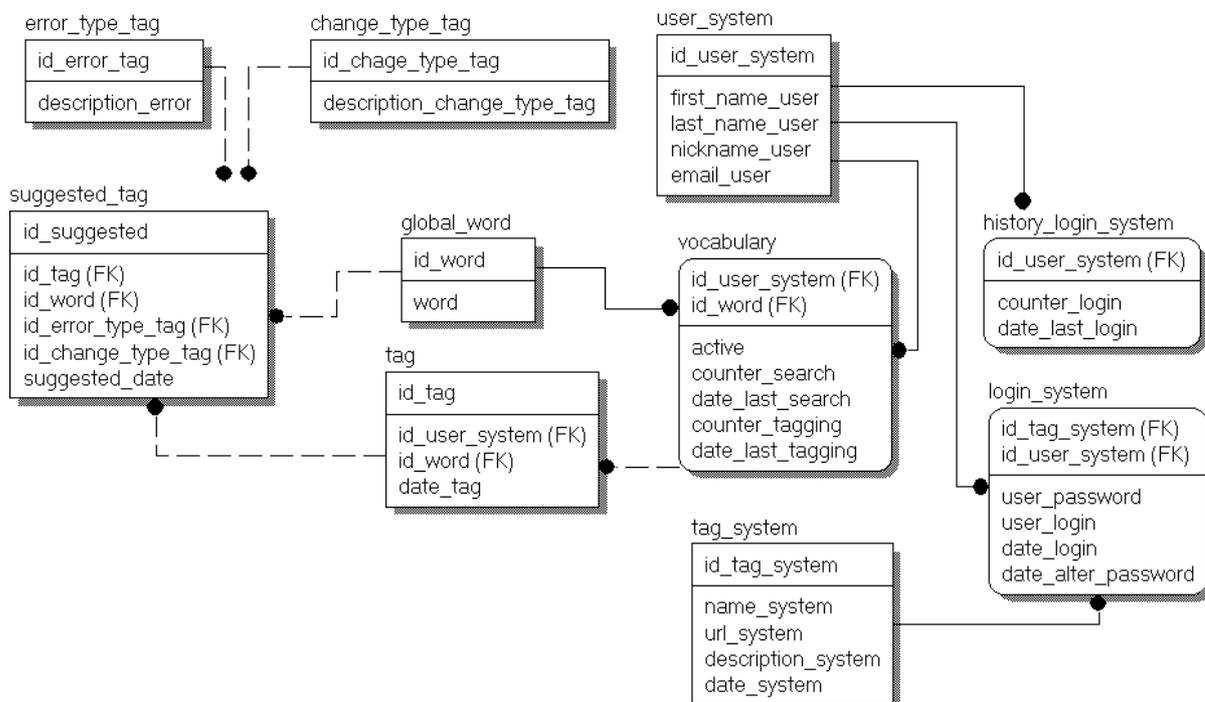


Figura 29 - Diagrama das tabelas necessárias para o gerenciamento de tags

Destacamos como as principais tabelas para o agrupamento das *tags*: a “*global_word*”, na qual são armazenados todos os termos do sistema, funcionando como um grande repositório de palavras (vocabulário global). A tabela “*vocabulary*” que guardará as relações entre o usuário e os termos da “*global_word*”, ou seja, cada palavra relacionada com o tipo de uso (busca ou categorização), as datas do último uso e o contador de uso pelo usuário desse termo ou para busca ou para nova categorização. Outra tabela importante é a tabela “*tag*” que guardará a relação entre a data que foi utilizada como *tag*, os termos da “*global_word*” e também qual o usuário do gerenciador que está associado a essa *tag*. Duas tabelas foram incorporadas ao gerenciador de tags com o objetivo de manter os históricos de usos do vocabulário em novas categorizações ou em buscas feitas pelo usuário. Foram criadas tabelas que possibilitarão o controle de correção/alteração de uma *tag* pelo usuário, para este fim temos as tabelas: “*error_type*” com os tipos de erros mais frequentes; a “*change_type*”

que apenas manterá as descrições das mudanças e a “*suggested_tag*” que terá a responsabilidade de guardar a mudança sugerida pelo sistema.

Caso o sistema de *tagging* integrado ao gerenciador de *tags* permita será feita a sincronização dos recursos categorizados pelo usuário com o objetivo de organizá-los juntamente com as termos associados a eles. Para este fim, no gerenciador de *tags* foram adicionadas as tabelas de “*resource*”, “*resource_description*”, “*extension_file*” e “*type_resource*” para organizar o recurso e a tabela “*categorization*” que guardará as relações entre o usuário do gerenciador, a *tag*, o recurso e o sistema hospedeiro da personomia do usuário que estará na tabela “*tag_system*”.

Com a arquitetura para o sistema gerenciador de *tags* já definida, no próximo capítulo serão discutidos assuntos referentes à codificação do gerenciador de *tags* bem como a descrição de alguns testes realizados para validar a arquitetura e os resultados obtidos.

[Página intencionalmente deixada em branco]

Aspectos de Implementação da Arquitetura Proposta

A arquitetura definida no capítulo anterior tem como meta agrupar as *tags* de diversos sistemas baseados em folksonomia, visando minimizar o esforço do usuário no uso do seu vocabulário utilizado para categorização e para a recuperação de recursos, além de facilitar a manutenção dos termos que compõem esse vocabulário. Outro ponto considerado importante no gerenciamento é o controle do ciclo de vida de um termo utilizado por um usuário, desde o momento de sua criação ou apropriação até o momento em que ele deixa de ser utilizado para novas categorizações, bem como para buscas. Neste capítulo serão discutidos alguns aspectos importantes para a codificação da arquitetura proposta, e também serão analisados alguns pontos relacionados aos testes realizados para validá-la.

5.1 Tecnologias utilizadas na implementação da arquitetura

A arquitetura proposta para o gerenciador de *tags* foi implementada utilizando a linguagem de programação Java devido a algumas APIs disponibilizadas pelos sistemas baseados em folksonomia estudados (*Delicious*, *YouTube*, *Flickr* e *SlideShare*) serem implementadas nesta linguagem. Outro ponto que nos levou a decidir sobre o uso da linguagem Java é sua versatilidade e a existência de *frameworks* que facilitam e minimizam o tempo de desenvolvimento de algumas atividades essenciais como, por exemplo, requisições via HTTP, persistência dos dados, processamento de texto, entre outros. Abaixo estão listados os principais *frameworks* utilizados e suas respectivas responsabilidades no contexto do sistema *Tag Manager*.

*HttpClient*²¹ é um projeto iniciado no ano de 2001 pela *Apache Software Foundation*²². O *framework HttpClient* provê as funcionalidades do *Hyper-Text Transfer Protocol (HTTP)* para serem utilizadas pelas aplicações. No gerenciador de *tags*, esse *framework* foi utilizado para realizar o acesso às informações dos usuários nos sistemas baseados em folksonomia. Nele, é encapsulado o acesso via protocolo *HTTP* ou *HTTPS*, permitindo a autenticação do usuário no sistema acessado.

*Jericho*²³ (*Parse HTML*) é uma biblioteca Java que permite analisar e manipular documentos *HTML*. Com esta biblioteca pode-se extrair informações com mais facilidade de um documento *HTML* disponibilizado na *Web*. Portanto, o *framework Jericho* foi utilizado para facilitar a atividade de *scrapping* em páginas com códigos *HTML*, *CSS*, *JavaScript*, e outros, no sistemas baseados em folksonomia que não disponibilizam APIs.

²¹ <http://hc.apache.org/httpclient-3.x/>

²² <http://www.apache.org/>

²³ <http://jerichohtml.sourceforge.net>

API do *Delicious*, versão 1.14, utilizando a linguagem Java que permite: buscar, renomear ou excluir *tags*; buscar todas as categorizações para um usuário; buscar categorizações para uma *tag* e, também, alterar uma categorização. No projeto *Tag Manager* ela foi alterada para efetuar requisições com um intervalo de 10 segundos entre uma requisição e outra em virtude do sistema *Delicious* bloquear a conexão do IP (*Internet Protocol*) que fizer muitas requisições de forma seqüencial e ininterrupta no sistema.

O sistema gerenciador de base de dados utilizado é o *PostgreSQL*²⁴, escolhido pelo fato de ser gratuito e possuir funcionalidades que permitem a expansão do projeto.

5.2 Codificação da arquitetura

Escolhidos os *frameworks* e a linguagem a ser utilizada para codificação do sistema *Tag Manager*, descrevemos a seguir alguns aspectos importantes relacionados aos modelos utilizados na implementação e sugeridos nesta pesquisa para a integração futura do gerenciador com outros sistemas que aplicam a técnica de folksonomia para a categorização de informação.

5.2.1 Acesso a dados dos sistemas externos

O acesso a dados em um sistema *Web* depende dos serviços de rede disponíveis aliados às permissões de acesso que estes sistemas disponibilizam para os usuários. Normalmente, a recuperação de dados é feita por meio de requisições *HTTP* ou *HTTPS* e, dependendo de como esses dados são disponibilizados, o acesso pode ser feito via API ou por meio de *scraping* da página que contenha a informação desejada. No *Tag Manager* implementamos as seguintes formas de acesso aos sistemas baseados em folksonomia:

²⁴ <http://www.postgresql.org/>

- Via API: Empregado quando o sistema disponibiliza uma API para o acesso aos dados do sistema, normalmente no formato XML. Uma API estável e bem definida permite que *frameworks* sejam codificados para facilitar esse acesso. Esse tipo de acesso, geralmente, possui restrição na quantidade de informação disponibilizada. No entanto, quando se exige uma autenticação do usuário, tem-se acesso a dados mais restritos e pode-se alterá-los utilizando-se de recursos disponibilizados pela própria API.
- Via JSON. Empregado quando o sistema pode disponibiliza seus dados em um formato JSON, o qual deverá ser processado para a extração da informação desejada. Assim, esses dados podem ser acessados via *Web*, normalmente sem a necessidade de autenticação.
- *Via Scraping HTML*: Empregado quando o sistema não disponibiliza uma forma estruturada de acesso aos dados. É acessada à página *Web* efetuando a extração dos dados manualmente. Neste modo, é necessário utilizar-se de algoritmos para detecção dos padrões nos quais os dados estão dispostos numa determinada página. Um dos problemas é a quantidade limitada de dados que estarão disponíveis na página e o outro é o trabalho necessário para a manutenção dos algoritmos que fazem a extração dos dados, pois eles devem ser alterados sempre que a estrutura da página for alterada.

O gerenciador *Tag Manager*, atualmente, acessa os dados dos sistemas *Delicious*, *Flickr*, *Youtube* e *Slideshare*. Abaixo, apresentamos uma tabela contendo um quadro comparativo resumido as possibilidades de acesso aos dados nesses sistemas. A análise completa encontra-se no anexo 01 desta dissertação.

Tabela 7 - Mapeamento do acesso as tags do usuário em sua personomia

Sistemas	<i>Delicious</i>	<i>Flickr</i>	<i>Youtube</i>	<i>Slideshare</i>
Serviços				
Acesso as tags	SIM	SIM	SIM	SIM
Modificação	SIM	NÃO	NÃO	NÃO
Exclusão	SIM	NÃO	NÃO	NÃO
Formas Acesso	API, JSON e <i>scraping</i> .	API, JSON, <i>WebService</i> e <i>scraping</i> .	API, JSON e <i>scraping</i> .	API, JSON e <i>scraping</i> .

A Tabela 7 demonstra uma relação dos sistemas com os serviços por eles disponibilizados. Este mapeamento permitiu-nos decidir alguns aspectos relevantes sobre a tecnologia a ser utilizada para a concretização do projeto. Conforme mostra a Tabela 7, o sistema *Delicious* é um dos mais abertos para o acesso aos dados. Ele disponibiliza uma API estável que depende apenas do *login* e da senha do usuário para a manipulação dos dados da personomia do usuário em questão. Esse sistema também permite o acesso aos dados via *JSON* não solicitando a autenticação do usuário, embora a quantidade de dados capturada seja limitada. Os demais sistemas não permitem alteração nos dados e para acesso as categorizações é preciso um código para desenvolvedor que pode ser solicitado mediante formulário.

Para a integração de outros sistemas baseados em folksonomia com o gerenciador de tags, algumas etapas devem ser seguidas, conforme mostra o diagrama da Figura 30.

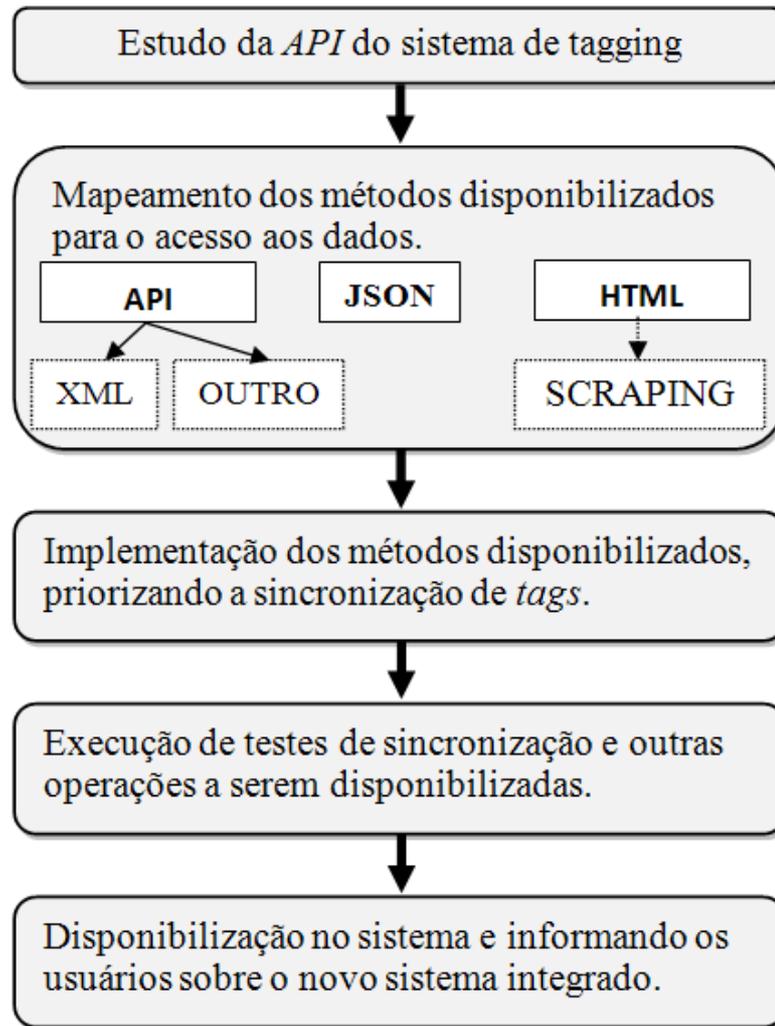


Figura 30 - Fluxo de atividades para integração do Tag Manager com outro sistema baseado em folksonomia

Passo 1: Estudo da API do sistema de tagging a ser integrado

Este estudo visa descobrir a forma como esses dados são disponibilizados para o usuário. O sistema *Delicious* apresenta um “help” em sua página para desenvolvedores mostrando as opções de acesso, o qual no *Delicious* pode ser feito por meio da API, JSON, HTML ou pelo uso de *scraping*.

Passo 2: Mapeamento dos métodos disponibilizados para o acesso aos dados

O estudo efetuado na primeira etapa ajuda a identificar o formato dos dados e a mapear o modo como esses dados são disponibilizados. Para o sistema *Delicious* as opções de

acesso foram identificadas na fase anterior e, em seguida, efetuou-se um estudo de como utilizar cada um dessas formas de acesso.

Quando os dados são disponibilizados no formato *JSON*, uma chamada via *HTTP* ou *HTTPS* deve ser efetuada usando uma URL para o acesso aos dados. No exemplo abaixo é possível observar a forma como os dados referentes ao vocabulário de *tags* de um usuário do sistema *Delicious* (no caso, *jvander*) são retornados.

URL acessada: <http://feeds.delicious.com/v2/json/tags/jvander>

Dados disponibilizados pelo sistema *Delicious* no format *JSON*:

```
{ "academic":1, "and":1, "AOP":1, "artigo":1, "artigos":1, "audio":1, "Barabasi":2, "blog":5, "blogs":3, "book":1, "books":2, "Broad":1, "buscador":1, "business":3, "cataloging":1, "categorizacao":1, "classification":2, "clustering":1, "coelho":1, "cognitivo":1, "collaboration":5, "collaborative":1, "collaborative_filtering":1, "collaborativefiltering":1, "collective":3, "color":1, "community":4, "complex":1, "conceituais":1, "concept":1, "conceptual":1, "conference":2, "copy":1, "cor":1, "deep":1, "design":5, "di":1, "dissertation":1, "download":1, "education":1, "engine":1, "engineer":2, "escrita":1, "extension":1, "filesystem":1, "filtering":1, "firefox":1, "firefox:rss":1, "folksonomy":13, "generator":1, "gerenciamento":2, "google":1, "gqm":1, "graphic":1, "group":2, "hci":3, "history":1, "ia":1, "ihc":4, "image":1, "informacao":1, "informatica":1, "information":3, "insects":1, "intelligence":2, "interface":1, "internet":3, "introduction":1, "java":2, "jena":1, "jericho":1, "journal":1, "language":1, "librarything":1, "Linked":1, "linux":1, "logo":1, "long_tail":1, "longtail":1, "mapas":1, "maps":2, "mcp":1, "media":1, "menos":1, "mozilla":1, "mytag":1, "Narrow":1, "navegation":1, "nepomuk":1, "network":4, "networks":3, "ontology":2, "orientador":1, "paper":3, "people":2, "pim":4, "plugin":1, "Podcast":1, "Podcasting":1, "powerpoint":1, "presentations":1, "productivity":1, "programming":1, "prosat":1, "receptor":1, "recommendation":1, "reference":3, "research":11, "retrievel":1, "scot":1, "search":3, "searchengine":1, "semantic":2, "semanticweb":3, "shared":1, "slides":1, "sociais":1, "social":24, "social_search":1, "social_software":2, "socialnetworks":2, "socialsearch":1, "software":13, "ss":1, "stanford":1, "Stigmergy":4, "tag":3, "tagging":7, "tags":6, "task":1, "taxonomy":2, "technology":1, "tese":1, "testes":1, "texto":1, "Tidying":1, "tools":3, "tutorial":2, "ubiquity":1, "uem":1, "UML":1, "usabilidade":3, "usability":3, "visualization":1, "web":10, "web2.0":16, "webdesign":1, "wike":1, "workshop":2 }
```

Uma API normalmente retorna os dados no formato XML, conforme mostrado abaixo na Figura 31, também em um exemplo de uso no sistema *Delicious*. O acesso aos dados, que precisam ser processados, é feito normalmente via *HTTPS*.

URL acessada: <https://api.del.icio.us/v1/tags/get>.

```

<?xml version="1.0" encoding="UTF-8" ?>
- <tags>
  <tag count="1" tag="AOP" />
  <tag count="2" tag="Barabasi" />
  <tag count="1" tag="Broad" />
  <tag count="1" tag="Linked" />
  <tag count="1" tag="Narrow" />
  <tag count="1" tag="Podcast" />
  <tag count="1" tag="Podcasting" />
  <tag count="4" tag="Stigmergy" />
  <tag count="1" tag="Tidying" />
  <tag count="1" tag="UML" />
  <tag count="1" tag="academic" />
  <tag count="1" tag="and" />
  <tag count="1" tag="artigo" />
  <tag count="1" tag="artigos" />
  <tag count="1" tag="audio" />
  <tag count="5" tag="blog" />
  <tag count="3" tag="blogs" />
  <tag count="1" tag="book" />
  <tag count="2" tag="books" />
  <tag count="1" tag="buscador" />
  <tag count="3" tag="business" />

```

Figura 31 - parte do código XML disponibilizados pela API do sistema Delicious

Quando utilizamos a técnica de *scraping* sobre uma página *HTML* o código da página deve ser processado por um *parser* que extrai dela as informações necessárias. Veja o exemplo a seguir:

URL acessado via HTTP: <http://delicious.com/jvander>.

O código-fonte apresentado deve ser estudado (ver Figura 32) para se identificar padrões ou estruturas que permitam localizar os dados que devem ser capturados. Em seguida, deve-se codificar um *parser* capaz de ler esse código e capturar os dados desejados.

```

<a title="social (24)" href="/jvander/social;_ylt=A0wNB9Sjs3hJw3MA4NhFrh54;_ylv=3"><span class="m">social<em>24</em></span></a>
</li>
<li class="">
<a title="web2.0 (16)" href="/jvander/web2.0;_ylt=A0wNB9Sjs3hJw3MA4NhFrh54;_ylv=3"><span class="m">web2.0<em>16</em></span></a>
</li>
<li class="">
<a title="software (13)" href="/jvander/software;_ylt=A0wNB9Sjs3hJw3MA4NhFrh54;_ylv=3"><span class="m">software<em>13</em></span></a>
</li>
<li class="">
<a title="folksonomy (13)" href="/jvander/folksonomy;_ylt=A0wNB9Sjs3hJw3MA4NhFrh54;_ylv=3"><span class="m">folksonomy<em>13</em></span></a>
</li>
<li class="">
<a title="research (12)" href="/jvander/research;_ylt=A0wNB9Sjs3hJw3MA4NhFrh54;_ylv=3"><span class="m">research<em>12</em></span></a>
</li>
<li class="">
<a title="web (10)" href="/jvander/web;_ylt=A0wNB9Sjs3hJw3MA4NhFrh54;_ylv=3"><span class="m">web<em>10</em></span></a>
</li>
<li class="">
<a title="tagging (7)" href="/jvander/tagging;_ylt=A0wNB9Sjs3hJw3MA4NhFrh54;_ylv=3"><span class="m">tagging<em>7</em></span></a>
</li>
<li class="">
<a title="tags (6)" href="/jvander/tags;_ylt=A0wNB9Sjs3hJw3MA4NhFrh54;_ylv=3"><span class="m">tags<em>6</em></span></a>
</li>
<li class="">
<a title="blog (5)" href="/jvander/blog;_ylt=A0wNB9Sjs3hJw3MA4NhFrh54;_ylv=3"><span class="m">blog<em>5</em></span></a>
</li>
...

```

Figura 32 - Código fonte de uma página do sistema Delicious

Quando for oferecida mais de uma opção para acesso aos dados, deve-se escolher por utilizar uma delas e, preferencialmente, que a opção escolhida possibilite a manipulação dos dados pelo usuário. Para os sistemas *Delicious*, *Slideshare*, *Youtube* e *Flickr* os acessos foram executados por meio da API de cada um deles sendo a principal razão a existência de um *framework* implementado em Java que facilitando assim a integração dos sistemas com o gerenciador de *tags*.

Passo 3: *Implementação dos métodos disponibilizados priorizando a sincronização de tags*

Conhecendo os formatos de dados disponibilizados e as regras impostas para a captura desses dados, um novo *parser* deverá ser codificado no sistema, ou seja, na camada de negócios, na qual já estão implementados os outros *parsers*.

Passo 4: *Execução de testes de sincronização e de outras operações a serem disponibilizadas*

Após a implementação do método de acesso e sua integração com o sistema gerenciador de *tags* alguns testes deverão ser executados antes de disponibilizar aos usuários o novo sistema já integrado ao *Tag Manager*. Estes testes visam descobrir falhas na implementação, bem como verificar o desempenho dos *parsers*. Na próxima seção apresentamos a seqüência utilizada para os testes reais efetuados no sistema *Delicious*.

Passo 5: *Execução de testes de sincronização e de outras operações a serem disponibilizadas*

Somente após os testes de acesso aos dados de uma personomia e a sincronização com o sistema *Tag Manager* é que deve ser liberado o uso desse sistema via gerenciador e informado aos usuários sobre a integração de um novo sistema.

5.3 Acessos externos ao gerenciador de *tags*

A camada de controle, apresentada no capítulo anterior, disponibilizará os dados para o acesso externo ao gerenciador de tags pelo usuário ou outros sistemas que queiram utilizar-se do gerenciador. Ainda não foram codificadas ferramentas que assegurem o acesso externo e venham a satisfazer o caso de uso de "Sugestão de *Tags*" quando os termos estiverem sendo utilizados nos sistemas baseados em *tagging*. No entanto, já foi definido que os dados serão disponibilizados utilizando os formatos JSON e XML, facilitando assim, que outros sistemas possam adotar o *Tag Manager* como gerenciador do vocabulário de seus usuários. Também serão implementados *plugins* e disponibilizados para que os usuários do *Tag Manager* os instalem em seus navegadores. A um *plugin* caberá a tarefa de interceptar o uso dos sistemas baseados em *tagging* associados ao gerenciador de *tags*, possibilitando assim o uso do *Tag Manager* sem o acesso direto a página do gerenciador.

5.4 Testes de sincronização

Para efetuar testes na arquitetura proposta foram implementados alguns processadores de dados para o sistema *Delicious*. Com isso, foram codificadas as classes para receber os dados e gravá-los em uma base de dados definida exclusivamente para esse fim, conforme foi discutido no capítulo anterior. Estes testes visaram identificar a viabilidade da arquitetura proposta sob um aspecto prático de utilização.

O acesso ao sincronizador é realizado mediante o cadastro do usuário no gerenciador e mediante a indicação dos sistemas que esse usuário utiliza para categorizar seus recursos.

Essa indicação implica em o usuário fornecer o seu *login*, sendo opcional a senha de acesso ao sistema. Se a senha não for informada ela será solicitada no momento em que a sincronização for acionada pelo usuário. A execução trará por meio do uso da API do sistema *Delicious*, todas as categorizações efetuadas pelo usuário nesse sistema, armazenando-as na base de dados local do sistema *Tag Manager*. A execução da sincronização ocorre sempre em segundo plano, por meio de *threads*, liberando o sistema para que seus usuários continuem a utilizá-lo normalmente.

Os testes efetuados demonstraram que o modelo funciona como um repositório de *tags* para o usuário, criando uma lista com todas as palavras utilizadas no sistema sincronizado. Essas palavras são vinculadas ao vocabulário do usuário que pode ser utilizado no momento da recuperação e da categorização de objetos nos sistemas com os quais o gerenciador se comunica.

A seqüência apresentada abaixo mostra a sincronização com o sistema *Delicious*, sendo ele o único até o momento a solicitar a senha no momento da sincronização. Os demais sistemas não solicitam a senha devido a política adotada para compartilhar os dados com desenvolvedores por meio do uso de uma chave de acesso que é liberada a partir do cadastro do desenvolvedor e a solicitação dessa chave.



Figura 33 - Login via web no gerenciador de tags

Para o uso do sistema *Tag Manager*, as pessoas envolvidas nos testes fizeram seus cadastros na ferramenta e informaram seus respectivos nomes de usuário do sistema *Delicious*

sem informar a senha de acesso correspondente. Após isso, cada usuário acessou o sistema *Tag Manager* (Figura 33).



Figura 34 - Sincronização das tags

Ao acessar o sistema, o usuário escolhe a opção de “Sincronização de *Tags*” destacada com um retângulo na Figura 34. Ao ser solicitada a sincronização, o gerenciador fornece uma tela para a escolha dos sistemas a serem sincronizados para que o usuário informe quais sistemas ele deseja sincronizar, ou marque a opção “todos” para sincronizar os dados de todos os sistemas que ele tiver cadastrado. Feito isso, o *Tag Manager* efetua a busca das *tags* do usuário nos sistemas escolhidos. Para os testes realizados, apenas o sistema *Delicious* foi considerado.



Figura 35 - Solicitação de senha ao usuário pelo gerenciador de tags

Como citado acima, ao fazer o cadastro dos sistemas baseados em folksonomia, o usuário cadastra seu *login* e senha, sendo que esta última não é obrigatória. Caso ele cadastre a senha o processo de sincronização será executado diretamente quando o botão “Sincronização de *Tags*” for acionado, caso contrário, o sistema solicitará que o usuário informe a sua senha (ver o destaque do retângulo na Figura 35) no momento da sincronização.

Após os três usuários terem efetuado as sincronizações, foram obtidos os seguintes dados:

- Na tabela “*global_word*” são armazenadas todas as palavras, ou *tags*, distintas conhecidas pelo gerenciador. Após a execução da sincronização dos dados dos três usuários obteve-se um total de 490 palavras;
 - Para o usuário 1 foram recuperadas 147 *tags*.
 - Para o usuário 2 foram recuperadas 155 *tags*.
 - Para o usuário 3 foram recuperadas 233 *tags*.
- O total de palavras na tabela “*global_word*” (490) difere da soma das *tags* dos usuários (535) em virtude do gerenciador não gravar, nesta tabela, palavras repetidas, significa que os usuários mantêm um total de 45 *tags* em comum entre eles.

Isso demonstra que a arquitetura projetada e codificada consegue executar a sincronização de dados, trazendo para a base local os termos utilizados pelos usuários no sistema *Delicious*. Os sistemas *Youtube*, *Slideshare* e *Flickr* foram integrados seguindo esses mesmo princípios.

[Página intencionalmente deixada em branco]

Conclusão

As categorizações efetuadas por um usuário têm como objetivo principal a organização, descrição e atribuição de significado à informação manipulada por este usuário. Para categorizar um recurso, um usuário deve utilizar termos que pertençam ao seu vocabulário, caso contrário, ele terá dificuldades para realizar a recuperação dessa informação e de usufruir dos benefícios que a atribuição de *tags* pode oferecer.

A existência de diversos sistemas baseados em folksonomia, focados em tipos distintos de objetos, leva um usuário a utilizar mais de um sistema e, conseqüentemente, a criar uma personomia em cada um deles. Isso implica em uma dispersão do vocabulário de *tags* utilizadas pelo usuário e prejudica o reuso dos termos em novas categorizações, além de dificultar a recuperação de informação, exigindo que o usuário acesse múltiplos sistemas tanto para categorizar quanto para recuperar a informação desejada. Isso nos levou a propor a criação de um sistema para o gerenciamento das *tags* do usuário originadas pelo trabalho de categorização de informação em sistemas baseados em folksonomia.

Nesta dissertação, apresentamos a arquitetura de um gerenciador de *tags* (*Tag Manager*) para resolver o problema relacionado à complexidade de gerenciamento e de manutenção de um vocabulário de termos consistentes quando se possui diversas personomias em diferentes sistemas baseados em folksonomia. Como o principal objetivo de um usuário deste tipo de sistema é a organização da informação de seu interesse, acreditamos que o sistema *Tag Manager* possa auxiliar o trabalho de categorização, de recuperação e de manutenção dos termos utilizados pelo usuário, possibilitando a centralização das personomias e o gerenciamento das mesmas de uma forma unificada. Essa é uma iniciativa de controle e gerenciamento do vocabulário do usuário, que colabora para a criação de uma camada de *tags* sobre os dados do usuário.

Com um repositório central para as *tags*, o usuário terá toda a liberdade do uso do seu vocabulário nos diversos sistemas baseados em folksonomia. Para o gerenciamento e controle do uso dos termos destacamos os indicadores adotados (datas e contadores), os quais facilitarão o controle do uso dos termos, seja como *tag* em uma categorização ou como palavra chave em uma busca. Esses indicadores possibilitarão cálculos de frequências do uso em contextos de buscas ou de categorizações nas quais o termo foi utilizado e, dessa forma, darão indicativos sobre o ciclo de vida de uma *tag*.

Foi discutido ao longo desta dissertação questões referente ao aspecto “pessoal” em contraste com o aspecto “social” promovido por um sistema que aplica o conceito de *tagging*. O fato de uma folksonomia ser formada por uma coleção de personomias permite afirmar que uma melhora na qualidade das categorizações pessoais de um usuário (sua personomia) não melhorará apenas a recuperação de informação do usuário em questão, mas também colabora com uma melhoria para todo o grupo de usuários que utiliza o sistema (a folksonomia). Assim, o gerenciador de *tags*, auxiliando o usuário a melhorar a qualidade dos termos em

suas personomias, colaborará indiretamente para uma melhoria na qualidade das categorizações nos sistemas baseados em folksonomia de interesse desse usuário.

Para a melhoria dos resultados das buscas obtidos em sistemas baseados em folksonomia, consideramos importante a proposta de Wu (2006) para a criação de ontologias. Por essa razão acreditamos que pesquisas sobre a geração de ontologias a partir de *tags* de uma folksonomia e, também, das personomias de um usuário como a proposta por Basso e da Silva (2008), devem continuar, criando assim novas alternativas para o relacionamento dos termos. Acreditamos que o gerenciador de *tags*, ao funcionar como um repositório central de *tags*, constitui um ambiente para o controle do uso de vocabulário de um usuário, e as informações presentes no gerenciador poderão ser utilizadas como entrada de dados para a geração de uma ontologia. Assim, a geração de uma ontologia poderá considerar todo o vocabulário de um usuário e os relacionamentos entre todos os seus termos.

Conforme mencionado anteriormente, o sistema *My.Tag* (<http://mytag.uni-koblenz.de>), proposto por Braun et al. (2008), faz buscas em sistemas baseados em folksonomia. No entanto, esse sistema não se destina ao agrupamento dessas *tags* em um espaço pessoal centralizado. O tipo de abordagem executada no sistema *MyTag* também poderia tirar proveito de uma camada de *tags* de um determinado usuário para explorar melhor a recuperação de informação em seu espaço pessoal e considerando a relação de termos na personomia do usuário. Porém, no sistema *MyTag* não há preocupação com o gerenciamento dos termos utilizados pelo usuário para categorizar e para recuperar recursos e este pode ser considerado o principal diferencial de nossa proposta com relação a tal sistema.

Com a definição da arquitetura para o gerenciador de *tags* atendendo os sistemas baseados em folksonomia *Delicious*, *Flickr*, *Youtube* e *SlideShare*, o próximo passo é a integração do gerenciador com outros sistemas baseados em folksonomia, como o

*Bibsonomy*²⁵, o *Bibtex*²⁶ e outros, bem como a melhoria e a estruturação de componentes que farão uso da camada de *tags*, como novos *parsers* mais seguros e ágeis. A expansão do gerenciador para atender outros sistemas baseados em folksonomia é necessária para acompanhar as necessidades dos usuários em utilizar sistemas para a categorização de diferentes tipos de objetos e para permitir que os mesmos mantenham um vocabulário centralizado e gerenciado.

O agrupamento das *tags* pelo gerenciador favorecerá a integração com os projetos relacionados que estão sendo desenvolvidos pelo Grupo de Sistemas Interativos Inteligentes (GSII), como o Sistema de Autoridade Cognitiva proposto por Pereira (2008a) e o recurso para a limpeza e estruturação de *tags* proposto por Côgo (2008). No contexto dos três pivôs de um sistema de *tagging*, tanto a proposta desta dissertação quanto a proposta de Côgo (2008) são direcionadas ao pivô *tag*, enquanto a proposta de Pereira (2008a) é direcionada ao pivô usuário. Estes projetos trabalham em conjunto visando melhorar a qualidade das categorizações efetuadas e das informações recuperadas em sistemas baseados em folksonomia.

Uma questão importante da arquitetura proposta é a disponibilização de uma API de acesso público, a qual poderá fornecer os dados da personomia globalizada do usuário para seu uso pessoal. O objetivo da API é prover os recursos básicos para categorizações — os termos que darão sentido aos recursos produzidos, disponibilizados e/ou existentes, e organizar esses recursos com a preocupação de uma recuperação futura eficiente. Para a disponibilização das informações em uma API pública algumas questões devem ser avaliadas como, por exemplo: Qual a linguagem que irá encapsular os dados (XML, JSON, ou outra)? Qual a estrutura que irá compor esses dados? Quais as informações relevantes para serem

²⁵ <http://www.bibsonomy.org/>

²⁶ <http://www.bibtex.org/>

utilizadas externamente? Que tipo de privacidade deve ser imposta a esses dados? entre outras.

Uma necessidade imediata para o gerenciador de *tags* é um estudo dos aspectos de interação do usuário, pois o acesso ao vocabulário via gerenciador deverá ser transparente ao usuário, funcionando como uma interface para múltiplos sistemas. A utilização do *Tag Manager* dependerá do mesmo possuir interfaces de interação muito bem definidas e projetadas, pois quando for solicitada alguma ação pelo usuário a interface deverá fazer com que a tarefa seja executada de forma natural e transparente. Os custos em termos de esforços para a utilização do sistema *Tag Manager* devem ser baixos, para que o usuário opte por utilizá-lo em vez de continuar utilizando apenas os sistemas baseados em folksonomia com os quais já está familiarizado.

Outra necessidade será a implementação de *plugins* que possibilitarão o acesso externo aos dados pelo usuário sem que seja preciso entrar na página do gerenciador de *tags*. Com um *plugin* interceptando o uso dos sistemas baseado em *tagging* o gerenciador terá condições de auxiliar o usuário na escolha de termos a serem utilizados como *tags* em suas categorizações. No entanto, estudos devem ser feitos para definir uma política de sugestão de *tags* que melhore a qualidade dos termos sugeridos como *tags*, levando em conta as palavras utilizadas em buscas, o vocabulário já construído pelo usuário, a comunidade da qual ele faz parte e os termos emergentes.

Outro trabalho futuro é a implementação de um controle das relações entre termos que apresentam ambigüidade. Esse tipo de estudo deverá trazer indicativos para uma tomada de decisão (automatizada ou pelo usuário) no momento da recuperação da informação. Como exemplo, imagine o usuário recuperando informações utilizando a palavra “*ant*”, ele estaria se referindo a formiga, ao algoritmo *ant system* ou então a ferramenta de anotação usada pela

linguagem Java. Se uma boa relação fosse criada entre os termos seria possível oferecer recursos para que o usuário identificasse o sentido desejado para o termo utilizado.

A maior dificuldade encontrada durante a codificação do sistema foi o acesso às personomias do usuário em outros sistemas. As dificuldades são de acesso a dados, pois além de serem muito restritos, os sistemas acabam limitando as alterações nas categorizações efetuadas pelo usuário quando elas são efetuadas por algum recurso externo ao sistema.

Para finalizar, acreditamos que o sistema *TagManager* deverá auxiliar o usuário monitorando sua personomia global, que é criada utilizando os dados oriundos dos diversos sistemas baseados em folksonomia do usuário. Esse monitoramento deverá trazer melhorias que serão refletidas nas personomias dos vários sistemas que um usuário utiliza. Como consequência da melhoria dos termos em uma personomia de sistema baseado em folksonomia, a recuperação da informação nesse sistema será melhorada tanto no contexto pessoal de busca quanto no contexto global para todos os usuários desse sistema.

Bibliografia

ANDERSON, C *The Long Tail*. New York : Hyperion Books, 2006. - Vol. I. 2006.

ARTÊNCIO, L. (2006). **Princípios de Categorização nas Linguagens Documentárias**. *Dissertação de Mestrado* . USP, SP, São Paulo.

BASSO, C.A.M.; DA SILVA, S.R.P. **Uma Proposta para a Evolução de Ontologias a partir de Folksonomias**. Webmedia 2008 - Workshop de teses e dissertações. - Vila Velha - ES : ACM, 2008.

BOOCH, G., RUMBAUCH, J., JACOBSON, I. (2005). **UML: Guia do Usuário**. Campus / Elsevier.

BRAUN, M. et al. *Personalized Search and Exploration with MyTag*. International World Wide Web Conference, Proceeding of the 17th international conference on World Wide Web. - Beijing, China : [s.n.], 2008. - pp. 1031-1032.

BREITMAN, K. K. **Web Semântica: A Internet do Futuro**. Rio de Janeiro : LTC Editora, 2005. - Vol. 1.

CATTUTO, C. et al. *Network Properties of Folksonomies* - 2008.

CHALMERS, M.; DIEBERGER, A.; HÖÖK, K.; RUDSTRÖM, Å. *Social Navigation and Seamful Design, Cognitive Studies*. Sept. 2004 : Vols. pp. 1-11.

CHOY, S.; LUI, A. *Web Information Retrieval in Collaborative Tagging Systems*. Web Intelligence, 2006. WI 2006. IEEE/WIC/ACM International Conference on. - Hong Kong : IEEE, 2006. - pp. On page(s): 352-355. - ISBN: 0-7695-2747-7.

CÔGO, F.; DA SILVA, S.R.P. **Uma Proposta de Organização do Vocabulário de Tags de Usuários de Sistemas Baseados em Folksonomia**. XIII Simpósio Brasileiro Sobre Fatores Humanos em Sistemas Computacionais. - Porto Alegre - RS : ACM, 2008.

CONSTANTINE, L.L.; LOCKWOOD, L.A.D. *Software for use: a practical guide to the models and methods of usage-centered design*. Addison Wesley, 1999.

DA SILVA, J.V.; DA SILVA, S.R.P. **Gerenciamento do Vocabulário das Tags dos Usuários de Sistemas Baseados em Folksonomia**. Webmedia 2008 - Workshop de Teses e Dissertações, Vila Velha - ES. Webmedia 2008 - Workshop de Teses e Dissertações. v. 2. 201-204 (2008).

Delicious Blog *delicious blog*. - *Yahoo*, 2008. - 03 de dezembro de 2008. -

<http://blog.delicious.com/blog/2008/11/delicious-is-5.html>.

ECHARTE, F. et al. ***Ontology of Folksonomy: A New Modeling Method***. Proceedings of the Semantic Authoring, Annotation and Knowledge Markup Workshop (SAAKM-2007). - Canada : [s.n.], 2007.

FILHO, A.M.S. **Serendipismo: Descoberta por Acidente e Sagacidade**. Maringá : Espaço Acadêmico, 2002. - 2 : Vol. 13.

FORSBERG, M.; HÖÖK K.; SVENSSON, M. ***Design principles for social navigation tools***. Workshop on Personal and Social Navigation in Information Space. - Stockholm, Sweden : [s.n.], 1998.

GAMMA, E. et al. **Padrões de Projeto - Soluções Reutilizáveis de Software Orientado a Objetos**. Porto Alegre : Bookman, 2000. - Reimpressão 2002. - ISBN 85-7307-610-0.

GOLDER, S.; HUBERMAN, B. ***The Structure of Collaborative Tagging Systems***. HP Labs Technical Report, 2006.

GRUBBER, T. TagCommons wiki. ***TagCommons / prod. Peoples. - 2007***. - 27 de 04 de 2008. - http://tagcommons.org/wiki/Tag_Data_Conceptualization.

GUY, M.; TONKIN, E. ***Folksonomies - Tidying up tags?***. D-Lib Magazine. - D-Lib Alliance, 2006. - 05 de 2008. - <http://www.dlib.org/dlib/january06/guy/01guy.html>.

HALPIN, H.; ROBU V.; SHEPERD H. ***The Complex Dynamics of Collaborative Tagging***. WWW2007. - Alberta : IW3C2-ACM, 2007.

HOTHO, A.; JÄSCHKE, R; SCHMITZ, C. ***Information retrieval in folksonomies: Search and ranking***. Springer. - June de 2006. - pp. volume 4011 of LNCS, pages 411- 426.

KIM, H. et al. ***Review and Alignment of Tag Ontologies for Semantically-Linked Data in Collaborative Tagging Spaces***. In Proceedings of the 2nd International Conference on Semantic Computing. - San Francisco, USA : [s.n.], 26 de 06 de 2008.

KIM, H. et al ***Review and Alignment of Tag Ontologies for Semantically-Linked Data in Collaborative Tagging Spaces***. 2nd International Conference on Semantic Computing. - San Francisco, USA : [s.n.], 2008. - p. In Proceedings.

LEW, M. S. ***Content-based Multimedia Information Retrieval: State of the Art and Challenges***. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP). - Fevereiro de 2006. - pp. v.2 n.1, p.1-19.

LIMA, G. A. B. **Categorização como um processo cognitivo**. Ciência e Cognição. - Belo Horizonte : [s.n.], 25 de 07 de 2007. - Vol. 11. - pp. 156 - 167. - ISSN 1806 5121.

LYMAN, P. (2000). *How Much Information? University of California. USA*. Acesso em 14 de Maio de 2007, disponível em <http://www.sims.berkeley.edu/research/projects/how-much-info/how-much-info.pdf>

MATHES, A. *Folksonomies - Cooperative Classification and Communication Through Shared Metadata*. University of Illinois Urbana-Champaign. - 2004. - 14 de Maio de 2007. - http://blog.namics.com/archives/2005/Folksonomies_Cooperative_Classification.pdf.

NEWMAN, M. E. J. *Power laws, Pareto distributions and Zipf's law*. Statistical Mechanics. - Contemporary Physics 46, 323-351, 29 de 05 de 2006. - 11 de 12 de 2007. - <http://arxiv.org/abs/cond-mat/0412004>. - 10.1080/00107510500052444.

O'REILLY, T. *What Is Web 2.0?*. O'REILLY. - 30 de 09 de 2005. - 13 de 04 de 2008. - <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>.

PEREIRA R.; DA SILVA, S.R.P. **Folksonomias: Uma Análise Crítica Focada na Interação e na Natureza da Técnica**. XIII, Simpósio Brasileiro Sobre Fatores Humanos em Sistemas Computacionais. - Porto Alegre - RS : ACM, 2008b.

PEREIRA, R. *Folkauthority: A Aplicação do Conceito de Autoridade Cognitiva por meio de Folksonomia*. Dissertação de Mestrado em Ciência da Computação. – Universidade Estadual de Maringá. - Maringá : [s.n.], 2008a.

RIDDLE, P. *Tags: What are They Good For?*. School of Information; University of Texas; USA. - 2005. Disponível em <<http://www.prentissriddle.com/papers/riddle-2005-tags.pdf>>. Acesso em 30/09/2008.

ROSCH, E. **Principles of Categorization**. 1988 - University of California, Berkeley.

RUSSELL, T. Contextual Authority Tagging: Cognitive Authority Through Folksonomy. School of Information and Library Science. University North Carolina. - 2005. - 14 de Maio de 2007. - <http://www.terrellrussell.com/projects/contextualauthoritytagging/conauthtag200505.pdf>.

SEN, S. et al. *Tagging, communities, vocabulary, evolution*. - CSCW 06 : ACM - Computer Supported Cooperative Work, 2006.

SHEN, K.; WU, L. *Folksonomy as a Complex Network*. Department of Computer Science. - Shangai : Fudan University, 2005.

SIMPSON, E. *Clustering Tags in Enterprise and Web Folksonomies - HP Labs Technical Reports*. International Conference on Weblogs & Social Media. - 2008.

SLIDESHARE, *Slideshare*. Online. Slideshare. - 2007. - 2008. - www.slideshare.net.

SMITH, G. *Tagging: People-Powered metadata for the Social Web*. Berkeley : New Riders, 2008.

SOMMERVILLE, I. *Engenharia de Software*. - São Paulo : Addison Wesley, 2003. - 6.

STURTZ, D. N. *Communal Categorization: The Folksonomy*. Content Representation, 2004.

SZOMSZOR, M.; CANTADOR, I.; ALANI, H. *Correlating User Profiles from Multiple Folksonomies*. Em Proceedings of the Fineteenth ACM Conference on Hypertext and Hypermedia ,pp. 33-43. ACM. - [s.l.] : ACM, 2008. - pp. pp. 33-43. ACM.

WAL, T. V. *Folksonomy*. Explaining and Showing Broad and Narrow Folksonomies. - vanderwal.net, 21 de February de 2005. - 12 de January de 2008. - <http://www.vanderwal.net/random/entrysel.php?blog=1635>.

WAL, T. V. *Folksonomy. Online Information*. vanderwal.net. - 18 de January de 2005. - 12 de January de 2008. - <http://www.vanderwal.net/random/entrysel.php?blog=1622>.

WEINBERGER, D. *Taxonomies to tags: From trees to piles of leaves*. Release 1.0, 23, no. 2, 1-33, February de 2005.

WU, H. *Harvesting Social Knowledge from Folksonomies*. 17th conference on Hypertext and Hypermedia, 2006.

Anexo I – Tabelas Sistemas Baseados em Folksonomia

Os quadros a seguir foram elaborados para verificar as possibilidades de acesso aos dados do usuário em suas personomias.

Sistema: Delicious

Tipo de acesso: API (XML)

Informações: <http://delicious.com/help/api>

Outras informações relevantes sobre o acesso e o sistema:

- A utilização dependerá apenas do usuário e senha.

Entidades	Dados Esperados	Satisfação
<i>Tag</i>	<i>Disponibilizar as tags</i>	Sim
	<i>Data da criação da tag</i>	Sim
	<i>Hits de uso da tag</i>	Sim
	<i>Atualização da tag</i>	Sim
	<i>Remoção da tag</i>	Sim
<i>Tagging</i>	<i>Permitir buscas</i>	Sim
	<i>Data do tagging</i>	Sim
	<i>Tags associadas</i>	Sim
	<i>Atualização do tagging</i>	Sim
	<i>Criação de novos tagging</i>	Sim
	<i>Remoção de taggings</i>	Sim

Outras informações relevantes: O acesso é feito via *http* e foi implementado uma API usando a linguagem Java, denominado "*delicious.jar*" que implementa as ações permitidas pela API.

Sistema: *Flickr*

Tipo de acesso: *Web API* (REST, XML-RPC, SOAP, JSON, PHP)

Informações: <http://flickr.com/services/api/>

Outras informações relevantes sobre o acesso e o sistema:

- É necessário criar uma “*ApiKey*” para ter acesso a API.
- Existem alguns frameworks, em várias linguagens, para ajudar no acesso a API.

Entidades	Dados Esperados	Satisfação
<i>Tag</i>	<i>Disponibilizar as tags</i>	Sim (precisa o ID do usuário ou da foto categorizada)
	<i>Data da criação da tag</i>	Não
	<i>Hits de uso da tag</i>	Sim, por meio do método <i>getHotList</i> , especificando um período e a <i>ApiKey</i> .
	<i>Atualização da tag</i>	Sim, mas a <i>tag</i> deve ser removida (<i>removeTag</i>) e uma nova adicionada (<i>addTags</i>).
	<i>Remoção da tag</i>	Sim
<i>Tagging</i>	<i>Permitir buscas</i>	Sim (método <i>search</i>)
	<i>Data do tagging</i>	Sim (método <i>getInfo</i> retorna a data do <i>upload</i> da foto e da última alteração)
	<i>Tags associadas</i>	Sim (método <i>getInfo</i>)
	<i>Atualização do tagging</i>	Sim,
	<i>Criação de novos tagging</i>	Sim (a API fala em “ <i>upload</i> de fotos”, o que tem o mesmo efeito de um <i>tagging</i>).
	<i>Remoção de taggings</i>	Sim. Pode ser feita a substituição da foto bem como das <i>tags</i> , comentários, etc. Para cada tipo de alteração, utiliza-se um método diferente.

Outras informações relevantes:

Sistema: *SlideShare*

Tipo de acesso: *Web API (XML)*

Informações:

- <http://www.slideshare.net/developers>
- <http://www.slideshare.net/developers/documentation>

Outras informações relevantes sobre o acesso e o sistema:

- É necessária a criação de uma *ApiKey* e um *SharedSecret* para acesso.
- Existem frameworks em várias linguagens para auxiliar no acesso a API.

Entidades	Dados Esperados	Satisfação
<i>Tag</i>	<i>Disponibilizar as tags</i>	Sim, mas somente junto dos <i>SlideShows</i> .
	<i>Data da criação da tag</i>	Não
	<i>Hits de uso da tag</i>	Não
	<i>Atualização da tag</i>	Permite modificar as <i>tags</i> utilizadas em um <i>SlideShow</i>
	<i>Remoção da tag</i>	Idem
<i>Tagging</i>	<i>Permitir buscas</i>	Sim, recebe uma <i>string</i> , ou <i>tags</i> ou o nome do usuário.
	<i>Data do tagging</i>	Sim, apenas a data de <i>upload</i> .
	<i>Tags associadas</i>	Sim
	<i>Atualização do tagging</i>	Sim, existe uma função para alterar os dados como título e <i>tags</i> , mas não permite alterar o <i>SlideShow</i> . Para isso, precisaria remover e criar um novo.
	<i>Criação de novos tagging</i>	Sim, permite fazer o <i>upload</i> de um <i>SlideShow</i> com os respectivos atributos, incluindo as <i>tags</i> .
	<i>Remoção de taggings</i>	Sim, permite remover um <i>SlideShow</i>

Outras informações relevantes: Permite obter *thumbnails* dos *SlideShows*. Permite Saber quantas vezes um *SlideShow* foi baixado e acessado.

Sistema: *YouTube*

Tipo de acesso: *Web API (RSS / Atom/ XML)*

Informações:

http://code.google.com/intl/pt/apis/youtube/developers_guide_protocol.html

http://code.google.com/intl/pt/apis/youtube/developers_guide_java.html

Outras informações relevantes sobre o acesso e o sistema:

Necessita uma “Developer Key” e um “Client ID”.

Possui *framework* em várias linguagens.

Entidades	Dados Esperados	Satisfação
<i>Tag</i>	Disponibilizar as <i>tags</i>	Sim, mas apenas junto com todas as informações de um vídeo. São chamadas de <i>KeyWords</i> .
	Data da criação da <i>tag</i>	Não
	Hits de uso da <i>tag</i>	Não
	Atualização da <i>tag</i>	Pode ser alterada somente alterando o post (precisaria remover a <i>tag</i> e adicionar outra).
	Remoção da <i>tag</i>	Sim, mas apenas alterando o post.
<i>Tagging</i>	Permitir buscas	Sim. Cada página pode ter até 50 resultados, podendo começar a mostrar a partir de um index para obter vídeos mais antigos.
	Data do <i>tagging</i>	Sim, data de publicação e da última edição
	<i>Tags</i> associadas	Sim, mas com o nome de <i>KeyWords</i> .
	Atualização do <i>tagging</i>	Sim para informações como <i>tags</i> , título, descrição, etc. Não tem informações se o vídeo também pode ser alterado.
	Criação de novos <i>taggings</i>	Sim. Aqui também é chamado de “novo upload”.
	Remoção de <i>taggings</i>	Sim.

Outras informações relevantes: Permite obter os *thumbnails* dos vídeos.
