



Universidade de Brasília

Faculdade de Economia, Administração, Contabilidade e Ciência da Informação

Departamento de Ciência da Informação e Documentação

Programa de Pós-Graduação em Ciência da Informação (PPGCIInf)

Mariana Baptista Brandt

Etiquetagem e Folksonomia:

uma análise sob a óptica dos processos de organização e recuperação
da informação na web

Brasília

2009

Mariana Baptista Brandt

Etiquetagem e Folksonomia:

uma análise sob a óptica dos processos de organização e recuperação
da informação na web

Dissertação apresentada à banca
examinadora como requisito parcial à
obtenção do Título de Mestre em Ciência da
Informação pelo Programa de Pós-Graduação
em Ciência da Informação do Departamento
de Ciência da Informação e Documentação
Universidade de Brasília.

ORIENTADORA: Dr.^a MARISA BRÄSCHER
BASÍLIO MEDEIROS

Brasília

2009

Brandt, Mariana Baptista.

Etiquetagem e folksonomia [manuscrito] : uma análise sob a óptica dos processos de organização e recuperação da informação na web / Mariana Baptista Brandt. – 2009.

142 f. : Il. color. ; 30 cm.

Dissertação (mestrado) – Universidade de Brasília, Departamento de Ciência da Informação e Documentação, 2009.

Orientadora: Prof.^a Dr.^a Marisa Bräscher Basílio Medeiros.

1. Etiquetagem. 2. Folksonomia. 3. Indexação Colaborativa. 4. Representação do conhecimento. 5. Organização e recuperação da informação na web. 6. Web 2.0. I. Título.

Ficha catalográfica elaborada pelo autor. CRB-1: 2.172.

MARIANA BAPTISTA BRANDT

Etiquetagem e Folksonomia: uma análise sob a óptica dos processos de organização e recuperação da informação na web

Dissertação apresentada ao Departamento de Ciência da Informação e Documentação, da Universidade de Brasília como requisito parcial para obtenção do título de Mestre em Ciência da Informação.

Aprovada em 15 de dezembro de 2009.

BANCA EXAMINADORA

Profa. Dra. MARISA BRÄSCHER BASÍLIO MEDEIROS
(Presidente)

Prof. Dr. MURILO BASTOS DA CUNHA
(Membro interno – UnB/PPGCinf)

Prof. Dr. JOSÉ AUGUSTO CHAVES GUIMARÃES
(Membro externo – UNESP/Marília)

Prof. Dr. FÁBIO JOSÉ DANTAS DE MELO
(Suplente – UnB/PPGCinf)

Agradecimentos

Mãe, pai e irmão (por tudo, sempre)

Marisa Bräscher, pela orientação, apoio e incentivo

Eliana Carlan e Fernanda Moreno, pelas dicas nas reuniões de orientação

Eduardo Dias, pelos infográficos e diagramação

Marcelo Pena, pela animação do slide do “termo”, usado em todas as apresentações de Power Point durante o mestrado.

Kelly Eustáquio, pelos artigos da BCE e pedidos no Comut

Osmar Arouck, por todas as dicas e pelo apoio

Martha e Jucilene, da secretaria da pós-graduação

English teachers (Alex, Andy, Juliano, Romes) que me fizeram gostar tanto da língua inglesa, sem a qual não teria sido possível realizar essa pesquisa!

Chris Coelho, pela concessão de horário para que eu pudesse terminar de escrever a dissertação

Ao pessoal dos *staffs* dos sites que responderam meus *e-mails*

E a todos que me ajudaram de alguma forma com esse trabalho.

With satellite television you can go anywhere

P. H. & D. E.

Resumo

A presente pesquisa estuda etiquetagem e folksonomia sob a óptica dos processos de organização e recuperação da informação na web. Para isso, definem-se elementos que devem estar presentes nesses processos, a saber: indexação, organização do conhecimento, usuários e aplicabilidade, permitindo assim uma validação da etiquetagem e da folksonomia nesse contexto. Utiliza a metodologia de revisão bibliográfica, além de estudo de usuários e pesquisa exploratória e descritiva de sítios da web. É realizada revisão bibliográfica na área de indexação, especificamente relacionada à indexação como processo cognitivo, indexação centrada no usuário, indexação de obras de ficção, indexação de imagens e indexação colaborativa ou democrática. Identifica elementos presentes nessas vertentes da indexação que compartilham características com o processo de etiquetagem, traçando um paralelo entre as duas práticas. A partir da mesma metodologia, estuda a folksonomia sob a óptica das abordagens em Organização do conhecimento nas quais a folksonomia encontra bases para ser analisada: abordagem baseada no usuário, cognitiva, social e da recuperação da informação. Ressalta os pontos positivos e negativos da folksonomia como esquema de representação do conhecimento, analisando também como essa estrutura pode ou não refletir conhecimento. Realiza estudo de usuários que participam do processo de etiquetagem, visando identificar sua motivação para a etiquetagem. Faz-se uma análise de como essa motivação se relaciona com os objetivos de organização e recuperação da informação. Por fim, verifica-se a aplicabilidade da etiquetagem e folksonomia na web. Descreve sítios que adotam etiquetagem e folksonomia com relação aos processos de organização e recuperação de informação. Conclui-se que, apesar de algumas limitações, a etiquetagem e a folksonomia possuem características suficientes para que sejam consideradas como modelo válido para a organização e recuperação da informação na web, com maior ou menor potencial, dependendo da natureza do sítio em que se insere e do objeto informacional com que se relaciona.

Palavras-chave: Etiquetagem. Folksonomia. Indexação Colaborativa. Representação do conhecimento. Organização e recuperação da informação na web. Web 2.0.

Abstract

This research studies tagging and folksonomy related to the process of organization and retrieval of information in the web. Some elements that must be in these processes are defined: index, knowledge organization, users and applications, allowing the validation of tagging and folksonomy in this context. It uses the methodology of bibliographic review, user study and exploratory and descriptive analysis of websites. A bibliographic research about index is presented, including cognitive index, user-centered index, fiction index, image index and democratic or collaborative index. It identifies elements presented in those kinds of indexes that share characteristics with the process of tagging, making possible a parallel between both practices. The same methodology is used to study folksonomy under the view of the approaches in knowledge organization in with folksonomy finds basis to be analyzed: user-based, cognitive, social and information retrieve approaches. This chapter highlights positive and negative aspects of folksonomy as a knowledge representation scheme and analyzes also how this structure can reflect knowledge or not. The research applies a study of the users that participate of the process of tagging, so as to find out the motivations for tagging. It analyses how this motivations are related to the objectives of organization and retrieval of information. The last chapter verifies the application of tagging and folksonomy in the web. It describes sites that use such systems relating with the process of information organization and retrieval. The research concludes that, despite of some limitations, tagging and folksonomy have enough characteristics to be considered a valid model for organization and retrieval of information in the web, with higher or lower potential, depending on the nature of the site in which its inserted and of the information package with which it's related.

Key words: Tagging. Folksonomy. Collaborative Index. Knowledge Representation. Organization and Retrieval of the information in the web. Web 2.0

Lista de ilustrações

Figura 1 – Diretórios do Yahoo!	32
Figura 2 – Coletivo x colaborativo (adaptado de WAL, 2008)	40
Figura 3 – Folksonomias abertas (adaptado de WAL, 2005)	44
Figura 4 – Lei da potência e cauda longa	45
Figura 5 – Folksonomia restrita (adaptado de WAL, 2005)	45
Figura 6 – Elementos da metodologia	47
Figura 7 – Processo cognitivo por trás da etiquetagem (adaptado de SINHA, 2005)	65
Figura 8 – Processo cognitivo por trás da categorização (adaptado de SINHA, 2005)	65
Figura 9 – Interação social e transmissão conceitual de etiquetas (adaptado de SINHA, 2006) ..	67
Figura 10 – Abordagens em organização do conhecimento	69
Figura 11 – Exemplo de <i>tag cloud</i> : Flickr em 15/06/08	71
Figura 12 – Formação de folksonomia	72
Figura 13 – Taxonomia das motivações para etiquetagem (adaptado de AMES; NAAMAN, 2007)	78
Figura 14 – Página inicial do Delicious	89
Figura 15 – <i>Tag bundles</i> com visualização em forma de nuvem	90
Figura 16 – Campo de busca principal	91
Figura 17 – Campo de busca por etiquetas	91
Figura 18 – Resultados da busca pela etiqueta ‘photo’	91
Figura 19 – Nuvem de etiquetas do Delicious	92
Figura 20 – Página inicial do Flickr	93
Figura 21 – Página da foto no Flickr	94
Figura 22 – Busca detalhada no Flickr	95
Figura 23 – Explorando as etiquetas do Flickr	96
Figura 24 – Página inicial do YouTube	97
Figura 25 – Página de descrição do vídeo	98
Figura 26 – Categorias de vídeos no YouTube	99
Figura 27 – Resultados de busca no YouTube	100
Figura 28 – Busca avançada no YouTube	100

Figura 29 – Página inicial do LibraryThing	101
Figura 30 – Registro de livro no LibraryThing	103
Figura 31 – Campos de busca no LibraryThing	103
Figura 32 – Informações da etiqueta	104
Figura 33 – Página inicial do Amazon.com	105
Figura 34 – Etiquetas no Amazon.com	105
Figura 35 – Etiquetas no produtos	105
Figura 36 – Taxonomia dos departamentos no Amazon	106
Figura 37 – Mecanismo de busca no Amazon	107
Figura 38 – Filtros de busca no Amazon	107
Figura 39 – Busca somente nas etiquetas	108
Figura 40 – Nuvem de etiquetas do Amazon	108
Figura 41 – Página inicial do Last.fm	109
Figura 42 – Guias de navegação do Last.fm	110
Figura 43 – Etiquetagem no Last.fm	110
Figura 44 – Busca no Last.fm	111
Figura 45 – Página da etiqueta “rock” no Last.fm	111
Figura 46 – Nuvem de etiquetas de artista no Last.fm	112
Figura 47 – Página inicial do Steve	113
Figura 48 – Etiquetagem de obras de arte no projeto steve.museum	114
Figura 49 – Etiquetagem no steve.museum	114
Figura 50 – Página inicial do Zigtag	116
Figura 51 – Etiquetagem no Zigtag	116
Figura 52 – Ferramenta do Zigtag no Firefox	117
Figura 53 – Guia “explore” do Zigtag	117
Figura 54 – Página de resultados de busca por uma etiqueta	118
Figura 55 – Gerenciador de etiquetas do Zigtag	118
Figura 56 – Qualificadores de etiquetas no Zigtag	119
Figura 57 – Definição de etiqueta no Zigtag.....	120

Sumário

1 Introdução	12
1.1 Definição do problema	13
1.2 Questão de pesquisa	13
1.3 Objetivo geral.....	14
1.3.1 Objetivos específicos	14
1.4 Justificativa e motivação da pesquisa	14
2 Revisão de literatura	16
2.1 Ciência da Informação	16
2.1.1 Organização da informação e Organização do conhecimento	17
2.1.2 Recuperação da informação	19
2.1.2.1 Indexação	21
2.1.2.2 Esquemas de representação do conhecimento	22
2.1.2.2.1 <i>Classificações</i>	22
2.1.2.2.2 <i>Tesauros</i>	23
2.1.2.2.3 <i>Taxonomias</i>	25
2.1.2.2.4 <i>Ontologias</i>	26
2.2 Internet e Web 2.0	28
2.2.1 Organização e recuperação da informação na web	31
2.2.2 Etiquetagem e Folksonomia	37
2.2.2.1 <i>Tipos de folksonomia</i>	44
3 Metodologia	47
3.1 Revisão bibliográfica	48
3.2 Usuários	48
3.2.1 Universo	49
3.2.2 Amostra	49
3.2.3 Coleta de dados	49
3.3 Aplicabilidade da folksonomia	49
4 Etiquetagem e indexação: um paralelo	51
4.1 Indexação como processo cognitivo	52

4.2 Indexação centrada no usuário	53
4.3 Indexação de obras de ficção	54
4.4 Indexação de imagens	56
4.5 Indexação colaborativa ou democrática	58
4.6 Etiquetagem	60
5 Folksonomia: esquema de representação do conhecimento?	62
5.1 Abordagens em organização do conhecimento	62
5.1.1 Abordagem baseada no usuário	63
5.1.2 Abordagem cognitiva	63
5.1.3 Organização social do conhecimento	66
5.1.4 Abordagem da recuperação da informação	67
5.2 Folksonomia como esquema de representação	69
5.2.1 Pontos positivos e negativos	69
5.2.2 Como esse esquema reflete ou falha em refletir conhecimento	71
6 Estudo de usuários de folksonomia: motivações para a etiquetagem	73
6.1 Contexto	73
6.2 Sobre os sítios pesquisados	73
6.3 Sobre motivação	75
6.4 Trabalhos relacionados	76
6.5 Análise dos dados	80
6.5.1 Dados demográficos	80
6.6 Motivações para etiquetagem	80
6.6.1 Delicious	80
6.6.2 Flickr	82
6.6.3 Resultado geral e comparação nos serviços	84
6.7 Análise dos resultados	85
7 Uso de folksonomia na web: diferentes aplicações e contextos	87
7.1 Delicious	88
7.1.1 Organização e recuperação da informação	90
7.2 Flickr	92
7.2.1 Organização e recuperação da informação	94
7.3 YouTube	97

7.3.1 Organização e recuperação da informação	99
7.4 LibraryThing	101
7.4.1 Organização e recuperação da informação	102
7.5 Amazon	104
7.5.1 Organização e recuperação da informação	106
7.6 Last.fm	108
7.6.1 Organização e recuperação da informação	109
7.7 Steve	112
7.7.1 Organização e recuperação da informação	115
7.8 Zigttag	115
7.8.1 Organização e recuperação da informação	117
7.9 Observações sobre os serviços descritos	120
8 Considerações finais e sugestões para pesquisas futuras	122
Referências	125
Apêndices	135

Capítulo 1

1 Introdução

A folksonomia¹ pode ser definida, com base nas idéias de Wal (2005, 2007, 2008), como o resultado do processo de etiquetagem, que consiste na atribuição de etiquetas a recursos informacionais na web. Essa atribuição de etiquetas é feita de forma livre e pessoal, não havendo regras ou restrições dos sistemas. A presente pesquisa buscou analisar folksonomia e etiquetagem e sua contribuição para a organização e recuperação da informação na web.

A pesquisa está inserida no contexto da Internet, no âmbito das aplicações da Web 2.0, em que a folksonomia surgiu e ganhou destaque, e que vem sendo utilizada em larga escala. Vários serviços de informação da web implementaram a prática, o que chamou atenção da comunidade de usuários da Internet, desenvolvedores e pesquisadores das áreas de computação, comunicação e informação.

A estrutura da pesquisa configura-se da seguinte forma: o primeiro capítulo traz a definição do problema de pesquisa, seus objetivos, bem como sua motivação e justificativa.

Após, o capítulo 2 foi dedicado à revisão de literatura sobre os temas pertinentes a este trabalho, buscando-se a elaboração de uma fundamentação teórica para a pesquisa. Assim, foi traçado um quadro que partisse da base teórica biblioteconômica e levasse ao tema principal deste trabalho, tratando-se do surgimento da Ciência da Informação com base no problema da explosão informacional, passando pelos processos de organização e recuperação da informação e esquemas de representação do conhecimento. Na segunda parte da revisão, foi abordado o contexto em que se insere a folksonomia, falando-se do surgimento da Internet e da web e sobre a recuperação da informação nesse meio, passando então para o ambiente atual no contexto de Web 2.0 e seus processos de organização e recuperação da informação. Enfim, chegou-se à temática da etiquetagem e folksonomia, identificando os autores e as abordagens principais desse contexto.

O capítulo 3 trata-se da metodologia da pesquisa, trazendo o tipo de pesquisa e detalhando cada etapa do trabalho.

A parte teórica da pesquisa concentra-se, além da revisão de literatura, nos capítulos 4 e 5. O capítulo 4, intitulado “Etiquetagem e indexação: um paralelo”, mostrou como o processo de etiquetagem guarda relações com o processo de indexação colaborativa, proposta por autores da área de Ciência da Informação nos anos 1990. Já o capítulo 5, “Folksonomia: esquema de representação do conhecimento?”, fez uma análise das abordagens em organização do conhecimento propostas por Hjørland (2007), identificando características dessas abordagens que estão presentes na folksonomia.

¹ Optou-se neste trabalho pela tradução dos termos: *folksonomy* para folksonomia, *tagging* para etiquetagem e *tag* para etiqueta. Demais questões terminológicas serão explicadas ao longo do trabalho, quando necessário.

O capítulo 6 analisou os dados de um estudo de usuários de folksonomia, realizado em outubro de 2008, que buscou identificar a motivação dos usuários para a etiquetagem. Por fim, o último capítulo do desenvolvimento da pesquisa descreveu alguns sistemas que utilizam folksonomia, mostrando assim um panorama geral do que vem sendo feito nessa área.

As considerações finais e sugestões para pesquisas futuras são apresentadas no capítulo 8. Os apêndices trazem os questionários aplicados no estudo de usuários, bem como os dados demográficos colhidos nesse estudo.

1.1 Definição do problema

Diante do grande volume de informação na web, encontram-se dificuldades de indexação de seu conteúdo pelos métodos tradicionais de indexação por especialistas, e os métodos automáticos ainda não são capazes de indexar satisfatoriamente todo tipo de material, como por exemplo, os recursos visuais. Essa situação gera dificuldades na organização e no gerenciamento da informação contida na web e faz com que o processo de recuperação de informação relevante ainda não tenha atingido níveis satisfatórios.

Verifica-se hoje a utilização em larga escala de métodos de classificação social ou indexação colaborativa, a chamada etiquetagem. O uso de etiquetas para recursos informacionais gera o que vem sendo denominado *folksonomia*. Este processo depende essencialmente da participação do usuário, mas o que o leva a realizá-lo? Como avaliar o potencial da etiquetagem e da folksonomia para organização e recuperação da informação e verificar se essa prática tem alguma validade? Qual a relação entre o que vem sendo feito na web atualmente e a base teórica de indexação e representação do conhecimento, tradicionais na Biblioteconomia e na Ciência da Informação?

Assim, essa tendência deve ser pesquisada com mais profundidade, com o intuito de analisar todas essas questões envolvidas e identificar elementos da Ciência da Informação presentes nessa prática atual.

1.2 Questão de pesquisa

A *folksonomia* seria uma alternativa válida para a organização e recuperação da informação na web?

1.3 Objetivos

Verificar se a *folksonomia* é uma alternativa válida para a organização e recuperação da informação na web

1.3.2 Objetivos específicos

- 1 – Analisar a etiquetagem como forma de indexação colaborativa dos recursos da web;
- 2 – Analisar a folksonomia sob a óptica das abordagens em organização do conhecimento;
- 3 – Investigar o que leva os usuários a executar a etiquetagem dos recursos da web e analisar quanto essa motivação se relaciona com o objetivo e recuperação da informação;
- 4 – Apresentar um panorama dos principais serviços de informação que utilizam etiquetagem e folksonomia, identificando seus elementos, características e abordagens, a fim de verificar sua aplicabilidade para diferentes tipos de objeto informacional.

1.4 Justificativa e motivação da pesquisa

O advento da *World Wide Web* mudou de forma significativa o mundo da informação. A web é hoje um grande repositório e, com isso, passou a ser a primeira opção de grande parte das pessoas com alguma necessidade de informação. Com seu crescimento vertiginoso, surgiu a necessidade de se utilizarem métodos de organização para a informação armazenada. Recorreu-se às técnicas tradicionais de organização de informação, já utilizadas pelas bibliotecas e centros de documentação, como o uso de metadados e dos princípios da classificação, empregados nas estruturas de diretório e nas taxonomias e ontologias.

Há, porém, uma grande dificuldade em se encontrar a informação necessária devido ao grande número de sites, repositórios e outros meios de armazenamento digital. Nesse contexto, surgiram novas tecnologias da informação desenvolvidas para a melhor organização dos conteúdos da web, além dos vários esforços no sentido de se buscar melhorias para a recuperação deste vasto universo informacional.

No âmbito das tendências da Web 2.0, em que se destacam as redes sociais e os ambientes de colaboração, surge o processo de etiquetagem dos recursos da web, resultando no que tem sido

denominado de *folksonomia*. Segundo O'Reilly (2005), as taxonomias tradicionais e estruturas em diretório são características da Web 1.0, enquanto a folksonomia se enquadra no contexto da Web 2.0. A atribuição de etiquetas aos recursos informacionais da web surgiu nos sítios Delicious.us (<http://delicious.com/>) em 2003 e, posteriormente, no Flickr (<http://www.flickr.com/>). Em pouco tempo, vários outros serviços de informação na web passaram a adotar tal prática.

A respeito da popularidade do uso de etiquetas na web, Voss (2007) comenta que tal fato levou alguns até mesmo a dizer que isso poderia superar os sistemas de classificação. O autor acredita que não há esse tipo de dicotomia: a etiquetagem apenas perpassa alguns aspectos da indexação de assuntos. Porém, acredita-se haver necessidade de um estudo mais aprofundado dos aspectos da indexação presentes na etiquetagem, o que é uma das propostas desse trabalho.

Outra questão que merece destaque em relação à folksonomia é sua inserção nas abordagens em Organização do conhecimento. Segundo Hjørland (2007b), tal conceito desafia o papel do profissional da informação para a organização do conhecimento, já que este é dispensado no processo de etiquetagem. É importante ressaltar que o autor utiliza o termo “organização do conhecimento” nesse contexto como sinônimo de organização da *informação*, e, dentre os processos de organização, inclui o de indexação. Hjørland (2007b) comenta ainda que a folksonomia deve ser considerada uma ferramenta semântica. Assim, pretende-se discutir também se a folksonomia pode ser vista como um novo esquema de representação do conhecimento, correlacionando elementos das abordagens teóricas da Organização do conhecimento.

Segundo Catarino e Baptista (2007), “A possibilidade de os próprios usuários participarem na organização desses conteúdos [conteúdos digitais na Web] é, em especial, uma questão que vale ser pesquisada e implementada.” Torna-se desejada, então, uma pesquisa diretamente com os usuários que participam desse processo.

Além disso, pesquisas nessa temática são bem vindas, visto que há pouca literatura sobre o assunto em língua portuguesa em Ciência da Informação. É possível encontrar trabalhos brasileiros sobre folksonomia em outras áreas como a Comunicação e a Ciência da Computação, mas na Ciência da Informação pouco tem sido pesquisado em relação ao tema.

Desta forma, torna-se necessária uma pesquisa que estude com mais profundidade esse tema, com o intuito de inserir a folksonomia no contexto da Ciência da Informação, destacando como os princípios teóricos da área podem contribuir para o estudo dessa nova prática. Pretende-se assim deixar uma base teórica sobre folksonomia e etiquetagem para a Ciência da Informação, que às vezes é negligenciada: por tradição costuma-se colocar a prática à frente da teoria – o que é compreensível, sendo a Ciência da Informação uma ciência que surgiu da prática (LE COADIC, 2004, p. 25).

Capítulo 2

2 Revisão de literatura

2.1 Ciência da Informação

O século XX deu início à chamada era da informação. A partir de seus anos 40 houve uma explosão documental, gerada pelo processo de especialização do saber. Como consequência, a informação tornou-se fator chave da economia e, nesse contexto, surge a necessidade de uma ciência que tenha por objeto de estudo a informação e seus processos de construção, comunicação e uso (LE COADIC, 2004, p. 2, 9).

Uma das definições mais citadas de Ciência da Informação na literatura da área é dada por Borko (1968, tradução nossa): “uma ciência interdisciplinar que investiga as propriedades e o comportamento da informação, as forças que governam o fluxo de informação, e os meios de processamento da informação para ótima acessibilidade e utilidade.”

Algumas décadas antes do surgimento da Ciência da Informação, o belga Paul Otlet, considerado um de seus precursores, já abordava questões relativas ao crescimento exponencial dos documentos e tentativas de soluções para sua recuperação.

Saracevic (1996) caracteriza o surgimento da Ciência da Informação a partir do problema da explosão documental, com base no artigo de Vannevar Bush: “o problema era (e, basicamente ainda é) a tarefa massiva de tornar mais acessível um acervo crescente de conhecimento.” (SARACEVIC, 1996, p. 42). Dai a necessidade de estudos na área da Recuperação da informação, que, segundo Saracevic, é o componente mais importante da Ciência da Informação. (SARACEVIC, 1996, p. 45).

Le Coadic (2004, p. 7) aponta duas importantes características que marcam o futuro da informação: sua explosão quantitativa e a implosão do tempo para sua comunicação. Dessa forma, têm-se fluxos informacionais cada vez maiores, fato relacionado ao crescimento acelerado dos periódicos científicos e, posteriormente, aos ambientes digitais utilizados para o armazenamento de informação em suportes imateriais: revista eletrônica, biblioteca eletrônica, livro eletrônico e portais na web.

Robredo (2005) também comenta as grandes mudanças por que vêm passando as ciências que lidam com a informação:

As bibliotecas e museus virtuais permitem consultar seus acervos digitalizados de qualquer lugar do planeta, ou mesmo do espaço; as informações circulam com a velocidade da luz e permitem não somente imprimir ao mesmo tempo, em lugares diferentes, as edições impressas dos grandes jornais e revistas, como também assistir ao desenrolar de quaisquer eventos ou notícias no instante e no lugar em que acontecem; o processamento cada vez mais rápido de enormes

volumes de dados e informações, permite estabelecer correlações entre eles quase em tempo real, aumentando seu valor e importância na tomada de decisões em todas as atividades e em todos os níveis; e ainda, mais importante, com o auxílio da tecnologia da informação e da comunicação, tornou-se possível o acesso imediato à cultura e ao conhecimento universais. (ROBREDO, 2005, p. 1).

A migração desse fluxo crescente de informação para a web gera uma necessidade cada vez maior de organização, definição de padrões e usos nos quais a informação é inserida no ambiente digital, além de criação de interfaces atrativas de informação. Assim, os métodos de organização da informação têm papel fundamental nesse contexto, na medida em que constituem uma forma de se possibilitar recuperação da informação.

2.1.1 Organização da informação e organização do conhecimento

Com os grandes fluxos de informação surgidos no século XX e crescendo exponencialmente desde então, tornam-se necessários processos de organização da informação e do conhecimento para que essa grande quantidade de informação seja organizada e possa ser então recuperada.

Para se discutir organização do conhecimento e da informação, deve-se primeiramente fazer uma breve distinção dos conceitos *conhecimento* e *informação*. Há muita discussão sobre o tema, mas para fins desta análise, entende-se conhecimento como o conjunto de conceitos (unidades do conhecimento) presentes em determinada área temática, e informação como o registro físico desse conhecimento. Posto isso, pode-se entender organização do conhecimento e organização da informação.

Para Dahlberg (2006, *apud* Hjørland, 2007c, tradução nossa), organização do conhecimento

[...] é a ciência de construir e arranjar sistematicamente unidades do conhecimento (conceitos) de acordo com seus elementos do conhecimento (características) inerentes e a aplicação de conceitos e classes de conceitos ordenados por sua forma de atribuição de conteúdos de referência válidos (objetos/sujeitos/assuntos) de todos os tipos.

Pode-se entender, então, que a organização do conhecimento está relacionada com um processo de análise conceitual de um domínio do conhecimento, e, a partir daí, sua estruturação, gerando uma *representação do conhecimento* de tal domínio. Dessa forma, obtém-se um instrumento – um esquema de representação do conhecimento – que será então usado para a organização da *informação* desse domínio de conhecimento produzida.

Taylor (2003) também faz essa discussão em relação ao que é organizado. Para a autora, organizamos informação quando a organizamos para o benefício de outras pessoas, ou seja, organizamos os objetos que contém informação. Taylor denomina tais objetos como *objetos*

informacionais: unidades de informação organizável. Para este trabalho, optou-se por utilizar a denominação *objeto informacional*, como proposto por Bräscher e Café (2008).

Dessa forma, tem-se a distinção entre organização do conhecimento e organização da informação: a partir da organização do conhecimento de um domínio, como definido por Dalhberg, pode-se organizar a informação, ou os objetos informacionais produzidos neste domínio, como entendido por Taylor. Para tanto, utilizam-se os sistemas, ou esquemas, de organização, ou representação, do conhecimento. De forma ampla, pode-se considerar os objetos informacionais como as unidades de organização do conhecimento. Há, porém, autores como Broughton *et al.* (2004) que afirmam que para cada abordagem da disciplina Organização do conhecimento, há uma unidade diferente a ser organizada.

Hjørland (2007c) discute que a definição de Dalhberg (2006) remete a um processo, mas defende que o estudo da Organização do conhecimento se dá em duas vertentes: os processos de organização do conhecimento e os sistemas de organização do conhecimento. Para Broughton *et al.*, no sentido estrito, organização do conhecimento se refere aos sistemas de organização do conhecimento como registros bibliográficos, sistemas de classificação, redes semânticas e tesouros. Já os processos de organização do conhecimento, segundo os autores, são a catalogação, a classificação, a indexação, e a análise de assunto. Tais processos são entendidos por Taylor como organização da *informação*, e não do conhecimento, conforme citado anteriormente.

Tem-se então que o termo *organização do conhecimento* é entendido por alguns autores como um processo (que para outros autores seria *organização da informação*), e que pode utilizar como auxílio para sua execução uma ferramenta elaborada para a *representação do conhecimento*, ou seja, um esquema de representação do conhecimento – que é a outra vertente da disciplina Organização do conhecimento, nessa perspectiva. Desta forma, acredita-se que a denominação *esquema de representação do conhecimento* seja mais apropriada e menos ambígua, sendo, portanto, adotada neste trabalho para se referir aos instrumentos elaborados no processo de organização do conhecimento de um domínio.

Tal denominação é também utilizada por Sowa (2000). O autor afirma que representação do conhecimento é um assunto multidisciplinar que aplica teorias e técnicas de três outras áreas: lógica, ontologia e computação. A representação do conhecimento, para este autor, é a aplicação da lógica e da ontologia nas tarefas de construção de modelos computadorizados para algum domínio. Essa visão é, de certa forma, restrita para a análise proposta neste trabalho.

Muito antes disso, Vickery (1986) também usa o termo “representação do conhecimento” e fala que essa vem sendo uma questão que preocupa o mundo da documentação desde sua origem. Afirma ainda que a questão passou a ser preocupação também de várias outras áreas como computação, linguagem, inteligência artificial e psicologia. Em todas essas áreas é necessário definir como o conhecimento vai ser representado para que as representações permitam sua manipulação. Para Vickery (1986):

Muitas técnicas de representação do conhecimento vêm sendo desenvolvidas em cada campo, as variações costumam se dar de acordo com os diferentes tipos de

manipulação que serão feitas. Mas há algumas convergências interessantes e parece que cada campo tem algo a aprender com os outros.

Como se pode notar, não há um consenso sobre o conceito de organização do conhecimento (independente do termo utilizado, *organização* ou *representação*) entre os autores. Em artigo recente, Bräscher e Café (2008) apresentam uma proposta conceitual preliminar para as áreas de organização da informação, organização do conhecimento, representação da informação e representação do conhecimento. As autoras se baseiam nos conceitos de Fogl sobre informação e conhecimento e nas definições citadas anteriormente de Taylor, Svenonius, Hodge e Dahlberg, além de outros autores como Shera, Egan e Soergel. Sobre organização e representação da informação, as autoras concluem então que:

A organização da informação é, portanto, um processo que envolve a descrição física e de conteúdo dos objetos informacionais. O produto desse processo descritivo é a **representação da informação**, entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico. (BRASCHER; CAFÉ, 2008).

E em relação à organização e representação do conhecimento:

Delineamos a organização do conhecimento como o processo de modelagem do conhecimento que visa a construção de representações do conhecimento. [...] A representação do conhecimento é feita por meio de diferentes tipos de sistemas de organização do conhecimento (SOC) que são sistemas conceituais que representam determinado domínio por meio da sistematização dos conceitos e das relações semânticas que se estabelecem entre eles. (BRÄSCHER; CAFÉ, 2008).

Para fins desta pesquisa, serão adotadas as abordagens em organização do conhecimento propostas por Hjørland (2007) e a proposta conceitual de Brascher e Café (2008) para organização do conhecimento, organização da informação, representação do conhecimento e representação da informação. Apenas a denominação das autoras para *Sistemas de organização do conhecimento* não será adotada, utilizando-se aqui, conforme explicado anteriormente, a denominação *Esquemas de representação do conhecimento*.

2.1.2 Recuperação da informação

Conforme já citado anteriormente, uma das grandes vertentes da Ciência da Informação é a disciplina da Recuperação da informação. Segundo Saracevic (1996), Mooers criou o termo recuperação da informação, que “engloba os aspectos intelectuais da descrição de informações e suas especificidades para a busca, além de quaisquer sistemas, técnicas ou máquinas empregados para o desempenho da operação.” (MOOERS, 1951, *apud* SARACEVIC 1996).

Saracevic (1996, p. 44) afirma ainda que, com a explosão informacional, “a recuperação da informação tornou-se uma solução bem sucedida encontrada pela Ciência da Informação e em processo de desenvolvimento até hoje.”

Essa solução tem por base dois processos básicos de organização da informação que possibilitam sua recuperação: a descrição física e a descrição de conteúdo. A descrição física se refere a aspectos físicos do documento e visa a localização do item fisicamente, ou seja, o acesso físico a determinado objeto informacional. O processo de catalogação corresponde a uma descrição física de um documento. Já a descrição de conteúdo é bem representada pelo processo de indexação, que consiste na atribuição de termos que representam o conteúdo semântico do documento.

A recuperação da informação ganhou nova importância no início dos anos 1990, com a introdução da web. Baeza-Yates e Ribeiro-Neto (1999, tradução nossa) afirmam que:

Apesar da sua maturidade, até recentemente, a recuperação da informação era vista como uma área restrita de interesse principalmente de bibliotecários e especialistas em informação. Essa visão tendenciosa prevaleceu por muitos anos, apesar da rápida disseminação, entre usuários de modernos computadores pessoais, de ferramentas de recuperação da informação para aplicações multimídia e de hipertexto. No início dos anos 1990, um único fato mudou toda essa perspectiva – a introdução da *World Wide Web*.

Para a recuperação da informação nesse ambiente, torna-se necessária uma ênfase maior nos processos de organização da informação, principalmente na descrição de conteúdo, já que o acesso físico ao documento se torna facilitado devido a imaterialidade do objeto informacional. Ou seja, o objeto não está restrito a um suporte e o documento eletrônico torna-se acessível por meio de diversos recursos. Na web, a descrição física da informação costuma se dar por meio do uso de esquemas de metadados.

Assim, para a recuperação do conteúdo semântico dos objetos informacionais, ganham mais destaque os processos de organização e representação do conhecimento, com auxílio de instrumentos para tanto. Segundo Gomes e Campos (2007), o sucesso da recuperação depende, basicamente, de três fatores:

É preciso haver um vocabulário-padrão; Os termos precisam ser atribuídos segundo critérios previamente estabelecidos, para assegurar consistência no tratamento dos dados; O *software* deve ser adequado às características de tal serviço de informação.

De acordo com Baeza-Yates e Ribeiro-Neto (1999, tradução nossa), a pesquisa atual em recuperação da informação inclui modelagem, classificação e categorização de documentos, arquitetura de sistemas, interface com usuários, visualização de dados, filtragem, linguagem entre outras.

A indexação e os esquemas de representação do conhecimento serão analisados com mais destaque nas seções seguintes.

2.1.2.1 Indexação

Indexação é o processo de atribuição de termos, que representam assuntos ou temas, a conteúdos informacionais. A Associação Brasileira de Normas Técnicas (ABNT), na norma NBR 12676 (1992), define indexação como o “ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação.”

Já a definição do Dicionário de biblioteconomia e arquivologia para indexação é: “representação do conteúdo temático de um documento por meio dos elementos de uma linguagem documentária ou de termos extraídos do próprio documento.” (CUNHA, 2009, p. 193).

O objetivo da indexação é extrair e representar o conteúdo dos documentos (objetos informacionais) para permitir sua recuperação posterior. Lancaster (2004, p. 1) afirma que o propósito principal da elaboração de índices (resultado do processo de indexação) é construir representações dos documentos para sua inclusão em bases de dados nas quais são armazenados esses documentos e onde é permitida sua recuperação. Hjørland (2007a, tradução nossa) deixa isso bem claro em sua definição de indexação: “é a representação de um documento (ou parte de um documento ou objeto informacional) em um registro ou em um índice para o propósito da recuperação.” A partir dessas definições, fica clara a importância do processo de indexação para a recuperação da informação.

Ainda em relação à importância da indexação, Gomes e Campos (2004) afirmam que “o uso consistente de princípios de indexação é um fator de qualidade e pode contribuir para a eficácia dos serviços, desde que a análise de assunto seja adequada.”

A indexação pode ser feita de forma manual (intelectual) isto é, realizada por agentes humanos especialistas, ou de forma automática, realizada por máquinas. Robredo (2005, p. 169) afirma que melhores resultados devem ser obtidos combinando os dois tipos de indexação.

A indexação resulta então na representação do conhecimento de um dado objeto informacional. Essa representação se dá por meio de termos, que são uma tradução dos conceitos que determinado documento abriga. É comum encontrar na literatura as etapas da indexação: análise conceitual e tradução. Outros autores abordam a indexação de forma mais detalhadas, com maior número de etapas. Lancaster (2004, p. 9) explica de forma simplificada que “a análise conceitual [...] implica decidir do que trata um documento – isto é, qual o seu assunto” e que a tradução “envolve a conversão da análise conceitual de um documento num determinado conjunto de termos de indexação.” (LANCASTER, 2004, p. 18). Na etapa da tradução, costumam-se utilizar instrumentos de padronização dos termos, as chamadas linguagens de indexação. Essas linguagens podem ser esquemas de representação do conhecimento. Tais ferramentas serão descritas na seção seguinte.

2.1.2.2 Esquemas de representação do conhecimento

Os esquemas de representação do conhecimento são estruturas geradas a partir da organização do conhecimento de um determinado domínio temático e podem ser utilizados na etapa da tradução no processo de indexação. Quando utilizados para fins documentários, se prestam a dois objetivos básicos: controle terminológico (controlar sinônimos e diferenciar homógrafos) e agrupar termos afins (*clusters*). São utilizados para organizar materiais com o propósito de recuperação.

Hjørland (2007b, tradução nossa) utiliza a terminologia sistema de *organização* do conhecimento e os define de forma ampla como: “um termo genérico se referindo a, entre outras coisas, ferramentas que apresentam a interpretação organizada de estruturas do conhecimento. Esse significado corresponde a ‘ferramentas semânticas’”

Já Hodge (2000, tradução nossa) explica o termo Sistema de organização do conhecimento de forma mais detalhada:

O termo sistema de organização do conhecimento pretende englobar todos os tipos de esquemas para organização da informação e promover o gerenciamento do conhecimento. Sistemas de organização do conhecimento incluem esquemas de classificação e categorização que organizam materiais num nível geral (como livros em uma prateleira), cabeçalhos de assuntos, que fornecem acesso mais detalhado, e arquivos de autoridades, que controlam formas variantes de informações chave (como nomes geográficos e pessoais). Sistemas de organização do conhecimento incluem também vocabulários altamente estruturados, como tesouros, e esquemas menos tradicionais, como redes semânticas e ontologias. Por serem mecanismos para a organização de informação, são o coração de toda biblioteca, museu e arquivo.

A autora utiliza o termo de acordo com a definição cunhada pela Network Knowledge Organization Work Group em seu primeiro encontro, em 1998. Segundo Hodge (2000), os esquemas de representação do conhecimento possuem características importantes para serem utilizados na organização de bibliotecas digitais.

Os principais tipos de esquemas de representação do conhecimento são classificações, tesouros, taxonomias e ontologias. Tais estruturas serão comentadas a seguir.

2.1.2.2.1 Classificações

Costuma-se definir classificação como o ato de agrupar coisas de acordo com suas semelhanças e separá-las conforme suas diferenças. Um dos processos envolvidos na construção

de teorias, na acumulação do conhecimento e sua conseqüente representação é o processo de classificação (KWANSNIK, 1999).

No final do século XIX os principais esquemas de classificação bibliográfica foram criados como a Expansive Classification de Cutter, a Classificação Decimal de Dewey (CDD) e a Library of Congress Classification (LCC). Posteriormente, no início do século XX, surgem a Classificação Decimal Universal (CDU) de Otlet e La Fontaine, a Bibliographic Classification de Bliss e a Colon Classification de Ranganathan. Além das classificações citadas, destacou-se ainda a classificação alfabética de Cutter, com as listas de cabeçalhos de assuntos.

Todos os esquemas de representação do conhecimento baseiam-se, de alguma forma, em princípios da teoria da classificação. Gomes e Campos destacam três princípios básicos da classificação que podem ser usados na construção de taxonomias: *categorização*, que fornece bases para a apresentação sistemática, *cânones* para o trabalho no plano das idéias (construção de classes) e *princípios*, para a ordenação de classes e seus elementos (GOMES e CAMPOS, 2007, grifo nosso).

Uma grande evolução nesses sistemas foi a classificação facetada de Ranganathan, com sua proposta das categorias fundamentais PMEST (personalidade, matéria, energia, espaço e tempo). Esse tipo de classificação foi uma grande evolução em relação aos sistemas tradicionais, pois permite que uma determinada informação seja classificada com base em suas diversas facetas, o que o torna mais flexível e alcança um maior nível de representação de conteúdo.

No entanto, as classificações bibliográficas não atendem a todos os propósitos necessários para a representação desse conhecimento produzido em maior escala no século XX, pois foram desenvolvidas em outro contexto, para utilização em sistemas manuais pré-coordenados e que não previam uma representação multidimensional. São criados então, em decorrência da especialização da ciência, outros tipos de esquemas de representação, como os tesauros, as taxonomias e as ontologias. Estes serão descritos a seguir.

2.1.2.2.2 Tesauros

Um dos esquemas de representação do conhecimento mais difundidos na área da documentação é o tesouro. O termo tesouro é de origem grega e significa armazenagem/repositório ou tesouro e surgiu no dicionário de Peter Mark Roget, *Thesaurus of english words and phrases*, de 1852. Dodebei (2002, p. 60) explica a motivação e o questionamento de Roget para a criação do dicionário: “sendo dada uma palavra, pode-se encontrar seu significado ou a idéia que ela pretende trazer consigo, em um dicionário, mas, ao contrário, tendo-se uma idéia, qual/quais palavras podem melhor expressar seu significado?”. Assim, as palavras e frases foram ordenadas no tesouro de Roget de acordo com seu significado.

Posteriormente, já no século XX, a Ciência da Informação adota o termo e começa a desenvolver os primeiros tesauros documentários. Segundo Dodebei (2002, p. 66):

A partir de 1940, o termo tesouro começou a ser utilizado na esfera da Ciência da Informação e, em especial, no processo de recuperação da informação, como sendo um instrumento capaz de transportar conceitos e suas relações mútuas, tal como expressos na linguagem dos documentos, em uma língua regular, com controle de sinônimos e estruturas sintáticas simplificadas.

Gomes (1990, p. 14) afirma que “o tesouro documentário surgiu da necessidade de manipular grande quantidade de documentos especializados.” Essa necessidade tem origem na especialização do saber, a partir do crescimento da publicação de periódicos e da explosão documental, conforme discutido no início dessa revisão.

Robredo (2005, p. 157) define os tesouros com base em sua função e sua estrutura. De acordo com o autor, segundo a função, “um tesouro é um instrumento de controle terminológico que permite traduzir a linguagem natural dos documentos, dos indexadores e dos usuários, numa linguagem sistêmica mais rígida.” Já a definição de tesouro como estrutura estabelece que um tesouro é “um vocabulário controlado e dinâmico de termos relacionados semântica e genericamente, que cobre um campo específico de conhecimentos” (ROBREDO, 2005, p. 158).

As Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngues (AUSTIN, 1993, p. 14), documento base para a Norma Internacional de Elaboração de Tesouros Monolíngues (ISO 2788), define tesouro como “o vocabulário controlado de uma linguagem de indexação, formalmente organizado para explicitar as relações *a priori* entre conceitos (por exemplo, como genéricas e específicas).” O autor afirma que o controle de vocabulário é obtido de duas formas principais. A primeira é a definição dos significados dos termos: um termo se refere a apenas um significado, e sua aplicação se restringe deliberadamente aos significados selecionados. A segunda é o controle de sinônimos: “quando o mesmo conceito puder ser expresso por dois ou mais sinônimos, selecionar-se-á um destes como o termo preferido que será utilizado na indexação.” (AUSTIN, 1993, p. 19).

Há ainda outras definições de autores conceituados na área como Currás, Foskett, Vickery, Lancaster, além da definição técnica bastante citada oferecida pela UNESCO.

As duas grandes características dos tesouros são os conceitos, que são as unidades do tesouro, representados pelos termos (descritores) e as relações entre eles (GOMES, 1990). Tais relações podem ser hierárquicas, associativas e de equivalência. O principal objetivo do tesouro é o controle terminológico. Além disso, é uma linguagem que serve a duas importantes instâncias de um sistema de informação: a das necessidades de informação e a dos documentos disponíveis.

Um tesouro pode também ser utilizado na estruturação de informações em ambientes digitais, obtendo-se assim maior precisão na recuperação da informação, conforme apontam Gomes e Campos (2004):

[...] a questão que se coloca relativa ao tratamento e à recuperação de informações diz respeito à qualidade e à precisão no tratamento das informações e à adequação a uma dada solicitação de busca através de um controle terminológico que venha garantir precisão nas informações recuperadas em meio eletrônico.

Assim, a elaboração desse instrumento de representação do conhecimento e controle terminológico é de grande valia para a organização e recuperação da informação, tanto em meio analógico quanto digital.

2.1.2.2.3 Taxonomias

Outro tipo de esquema de representação do conhecimento é a taxonomia. O termo possui uma definição controversa na literatura: “não existe uma definição acordada para o termo taxonomia ou para os elementos que o compõem” (Conway, S.; Sligar, C. 2002, tradução nossa). Alguns autores afirmam que não passa de um modismo, sendo o mesmo que tesouros ou esquemas de classificação. O termo foi trazido das ciências naturais (Biologia), em que se refere à classificação dos seres vivos, proposta no século XVIII por Charles Linneu. Posteriormente, passou a ser usado na área da documentação e migrou para o ambiente digital, sendo então muito usado em Gestão do conhecimento e em Tecnologia da informação:

Recentemente, o uso de taxonomia tem sido adotado por permitir acesso através de uma navegação em que os termos se apresentem de forma lógica, ou seja, em classes, sub-classes, sub-sub-classes e assim por diante, em quantos níveis de especificidade sejam necessários, cada um deles agregando informação sobre os documentos existentes na base. [...] As taxonomias têm sido bastante empregadas em portais corporativos e bibliotecas digitais. Além dessas aplicações, o seu uso tem sido bastante difundido no contexto da Web Semântica. (CAMPOS e GOMES, 2007).

Não há uma definição única e precisa, pois se confunde com outros termos, mas Graef (2001) define taxonomias como “estruturas que promovem uma forma de classificar coisas – seres vivos, livros, produtos – numa série de grupos hierárquicos, facilitando desta maneira a localização, estudo ou identificação de uma dada informação.”

Sendo assim, taxonomias são instrumentos de organização, baseadas na teoria da classificação, ou seja, figuram entre os esquemas de representação do conhecimento. Segundo Edols (2001), as taxonomias são formadas por duas partes: estruturas e aplicações. Estruturas são categorias (ou termos) e as relações que os unem. Aplicações são ferramentas de navegação para ajudar usuários a encontrar informação. Pode-se dizer então que as taxonomias, vistas dessa forma, são algo mais que os tesouros. São baseadas em conhecimento bibliotecário, mas possuem outros elementos quando aplicadas à organização de informação na Web.

Os objetivos de uma taxonomia, de acordo com Terra *et al.* (2005, p.1), são:

representar conceitos através de termos; agilizar a comunicação entre especialistas e entre especialistas e outros públicos; encontrar o consenso; propor formas de controle da diversidade de significação; e oferecer um mapa de área que servirá como guia em processos de conhecimento.

Conway e Sligar (2002), dividem as taxonomias em três tipos: descritivas, vocabulários de gerenciamento de dados (também chamada de taxonomia compartilhada) e as navegacionais. As taxonomias descritivas fazem o controle do vocabulário da área tratada, tornando as buscas mais eficientes. São bastante utilizadas na área corporativa e funcionam como uma espécie de tesouro, explorando as relações entre os termos. Podem, neste caso, ser aproveitadas para uso em diferentes sítios, quando tratam do mesmo domínio de conhecimento. As taxonomias navegacionais buscam organizar a informação do sítio de forma mais flexível, de modo que o usuário possa ter uma visualização da estrutura taxonômica a partir da navegação. Esse tipo de taxonomia se aproxima mais do usuário, pois se baseia em sua navegação. A taxonomia navegacional consiste de etiquetas aplicadas a categorias de conteúdos baseados no modelo mental de conhecimento do usuário sobre como a informação é organizada (Conway, S.; Sligar, C. 2002). Diferentemente das taxonomias descritivas, as navegacionais não costumam ser reutilizadas, pois são modeladas de acordo com as necessidades específicas de determinado sítio. O terceiro tipo de taxonomia são as compartilhadas, ou vocabulários de gerenciamento de dados, muito usadas no campo dos negócios. Consiste em uma lista de termos autorizados sem estrutura hierárquica, que é usada para transações de negócios. Esse tipo de taxonomia tem sido separada da descritiva, mas há áreas de interseção entre ambas. (CONWAY, S.; SLIGAR, C. 2002).

Campos e Gomes (2007) sintetizam a importância das taxonomias: “(...) a organização das informações através do conceito de Taxonomia permite alocar, recuperar e comunicar informações dentro de um sistema de maneira lógica através da navegação.” Edols (2001) afirma que boas taxonomias, baseadas em classificação e vocabulário controlado, resultam em recuperação da informação mais eficiente.

2.1.2.2.4 Ontologias

Outra inovação que é aperfeiçoada para a organização da informação digital são as ontologias. Na filosofia, o termo se refere a questões de existência e do ser, mas, extrapolando esse campo, ganhou novo significado na área de representação do conhecimento. Segundo Le Coadic (2004, p. 64): “Surgiu, para a Rede, a ontologia, estabelecida sobre conhecimentos terminológicos e suas relações, conhecimentos sobre um domínio, aprovados por uma comunidade comprometida em facilitar seu compartilhamento.” Ontologia se refere a arquivos e documentos que estabelecem formalmente relações entre termos e conceitos, para então serem entendidos por máquinas.

O termo ontologia é definido por vários autores; segundo Almeida e Bax (2003, p. 8), “são diversas as definições apresentadas na literatura e existem contradições.” As ontologias podem ser conceituadas de forma diferente dependendo da área do conhecimento, tendo como principais a

Filosofia, a Ciência da Computação e a Ciência da Informação, e da abordagem acadêmica que pretende ser dada. Serão reunidas aqui as definições dos autores mais citados na literatura do assunto.

Gruber (1993, p.1, tradução nossa) define ontologia como uma “especificação explícita de uma conceitualização.” O autor afirma que a conceitualização, simplificação do mundo que está sendo representado, é a base de um corpo formal de representação do conhecimento, e que toda base de conhecimento, sistema baseado em conhecimento ou agente baseado em conhecimento é comprometido com uma conceitualização, implícita ou explícita.

Já para Guarino e Giaretta (1995), essa definição de Gruber é problemática, pois depende da definição dos termos “conceitualização” e “especificação.” Os autores discutem as várias definições de ontologia e estabelecem três sentidos para o termo: como sinônimo de teoria ontológica, como especificação de um compromisso ontológico e como sinônimo de conceitualização. Segundo Bax e Almeida (2003), após uma revisão da definição de conceitualização, Guarino apresenta uma interpretação mais satisfatória:

[...] ontologia se refere a um artefato constituído por um vocabulário usado para descrever uma certa realidade, mais um conjunto de fatos explícitos e aceitos que dizem respeito ao sentido pretendido para as palavras do vocabulário. Este conjunto de fatos tem a forma da teoria da lógica de primeira ordem, onde as palavras do vocabulário aparecem como predicados unários ou binários. (GUARINO, 1998 *apud* BAX e ALMEIDA, 2003).

Robredo (2005, p. 321) afirma que uma ontologia é, para a ciência da computação, “o resultado da formulação, tão rigorosa e completa quanto possível, de um esquema conceitual sobre um domínio.” O autor ressalta que a funcionalidade de um sistema baseado em ontologias depende completamente do *software* utilizado, ou seja, uma ontologia não funciona sozinha.

Segundo Gruber (1996, tradução nossa), as ontologias são usadas com o propósito de possibilitar compartilhamento e reuso do conhecimento, sendo assim uma especificação usada para fazer compromissos ontológicos: acordos para uso de vocabulário compartilhado de maneira coerente e consistente. O autor afirma que o vocabulário, mais que a forma, é o foco da especificação dos compromissos ontológicos. Uma ontologia possui uma taxonomia e um conjunto de regras de inferência. (GRUBER, 1996, tradução nossa).

Gomes e Campos (2007) afirmam que as ontologias:

[...] são meios poderosos de inter-relacionar sistemas e neste contexto elas possuem papel de destaque, como podemos observar através dos componentes que integram uma ontologia, ou seja: Termos e Definições; Classes e subclasses - que podem estar organizadas em uma taxonomia; Relações (*também chamadas de propriedades*), que devem representar os tipos de interação entre as classes de um domínio; Axiomas que são regras para determinar a verdade das sentenças; e Instâncias que são utilizadas para representar elementos específicos, ou seja, os próprios dados.

O principal propósito de uma ontologia é permitir o desenvolvimento de linguagens e tecnologias que tornem a informação legível tanto por máquinas quanto por humanos, sendo geralmente usadas em inteligência artificial e representação do conhecimento. Acredita-se que

a futura geração da web (Web 3.0), denominada por Tim Berners-Lee como Web Semântica, deverá fazer uso de ontologias como base de sua arquitetura de informação. Segundo Gomes e Campos (2007): “para que a Web Semântica venha a funcionar de forma efetiva, computadores têm que ter acesso às coleções estruturadas de informações e a conjuntos de regras de inferência que se consolidam através de mecanismos como as ontologias.”

2.2 Internet e Web 2.0

A Internet, rede mundial de computadores, teve sua origem no final dos anos 1960, a partir de uma iniciativa acadêmica e militar norte-americana que contou com o Massachusetts Institute of Technology (MIT) e o Departamento de Defesa dos Estados Unidos como instituições-chave. Tal iniciativa resultou posteriormente na criação da ARPANET, que contou também com a colaboração do RAND Corporation (EUA) e da National Physical Laboratory – NPL (Reino Unido):

No documento de apresentação do plano [de criação da ARPANET], intitulado *Redes para o compartilhamento de redes de computadores*, ficou explícito o objetivo de adquirir experiência na interconexão de computador e de melhorar a produtividade da pesquisa com auxílio do compartilhamento de recursos computacionais, apresentando como justificativa a insuficiência da tecnologia existente nos ambientes científicos e militares. (ROBREDO, 2005, p. 243).

Muito antes disso, porém, alguns autores como Otlet, Bush e Englebart, já visionavam o que seria essa rede mundial de informação. Em seu *Tratado de documentação*, de 1934,

Otlet anteviu um tipo de ‘máquina acadêmica’ ou de ‘estação de trabalho do pesquisador’ que permitiria selecionar as fichas das referências dos documentos, e projetar estes numa tela a grandes distâncias, mediante um ‘telescópio elétrico’ conectado a uma linha telefônica que asseguraria a transmissão de imagens facsimilares e o estabelecimento ‘das conexões que cada documento tem com todos os outros’, formando assim o que poderia ser chamado de ‘livro universal’. (ROBREDO, 2005, p. 237).

Da mesma forma e talvez inspirado em Otlet, Vannevar Bush propôs, em 1945, o Memex – memory extended, sistema que permitiria ao pesquisador guardar um grande volume de informações textuais e não-textuais, e que pudessem ser acessadas de forma rápida, além de permitir conexões com documentos afins. “Um memex é um dispositivo no qual cada indivíduo armazena livros, registros e comunicações, e que é mecanizado para que possa ser consultado com extrema velocidade e facilidade. É um suplemento íntimo da sua memória” (BUSH, 1945, p. 45, tradução nossa). Bush faz uma crítica ao meio como os documentos são indexados de forma tradicional em bibliotecas, diferente do cérebro humano, que trabalha por meio de associações.

Aparecem então as idéias de hipertexto e dos *links*, em que os documentos são interligados pelas associações presentes.

A idéia de hipertexto é comumente vinculada a Ted Nelson e seu projeto Xanadu, mas tal conceito pode ser identificado em trabalhos anteriores como o de Otlet. O termo foi cunhado pelo autor em 1965 e está relacionado às ligações entre textos que se interrelacionam: “o hipertexto é um documento digital constituído de nodos de informação interrelacionados por ‘links’.” (ROBREDO, 2005). O projeto Xanadu propunha uma rede que poderia reunir toda a produção escrita mundial estruturada por meio de *links* de hipertexto. Tal projeto acabou não sendo concluído (ROBREDO, 2005). O hipertexto traz uma forma de leitura não linear, em que se consegue chegar a informações que vão além do assunto principal do texto. Segundo Le Coadic (2004, p. 59), o que diferencia o hipertexto de outras formas de armazenamento digital é a estrutura associativa que reproduz a estrutura da memória humana e pode tornar-se seu complemento íntimo e ampliado.

Em meados dos anos 1980, foi criado o protocolo de comunicação TCP (Transfer control protocol), que substituiu o então NCP (Network control protocol) utilizado pela ARPANET. A utilização do TCP/IP (Transfer control protocol/Internet protocol) como protocolo padrão marca o nascimento da Internet e a define:

O TCP/IP fornece a tecnologia ponte para permitir às pequenas redes se conectarem à Internet muito mais rapidamente, assim como a qualquer um (pessoa física, associação, corporação, entidade pública ou privada) ter acesso a um sistema de comunicação e informação simples e de amplitude ilimitada. (ROBREDO, 2005, p. 246).

A web, aplicação mais difundida da Internet, surgiu com uma proposta de gerenciamento de informações feita por Tim Berners-Lee em 1989. Nessa época Berners-Lee trabalhava no Conselho Europeu para Pesquisas Nucleares (CERN) e elaborou um documento que propunha uma nova forma de armazenamento e organização dos documentos produzidos pelo CERN, já que o instituto estava tendo problemas de perda de informação importante nas pesquisas. A proposta tinha como solução o conceito de hipertexto para a organização dos documentos. Tal conceito veio a ser a base da *World Wide Web*. Devido a essa proposta, Berners-Lee é considerado o pai da web e hoje é o diretor do *World Wide Web Consortium*, o W3C, responsável pelo desenvolvimento de protocolos que promovam a evolução da web e garanta sua interoperabilidade, levando assim ao uso em seu maior potencial.

Segundo Robredo (2005, p. 248), a web é uma aplicação da Internet que permite a difusão e a transferência de informações e arquivos multimídias por meio da navegação entre *links* hipertextuais. O autor afirma ainda ser a web o fenômeno social mais importante do século XX.

Em artigo de 1998, Berners-Lee comenta o histórico da web, desde sua proposta inicial até o que estava acontecendo até então. O autor afirma que:

O sonho por trás da Web é um espaço comum de informação em que podemos nos comunicar por meio de compartilhamento de informação. Sua universalidade é essencial: o fato de que o hipertexto pode apontar para qualquer coisa,

seja ela pessoal, local ou global, seja ela um rascunho ou uma obra prima. Houve também uma segunda parte desse sonho, que depende da Web ter seu uso tão generalizado que se torne um espelho realístico (ou de fato sua principal personificação) das formas em que trabalhamos, nos divertimos e nos socializamos. (BERNERS-LEE, 1998, tradução nossa).

Essa segunda parte do sonho de Berners-Lee, que à época ainda não tinha ocorrido, hoje já pode ser vista como realidade, dentro da denominada Web 2.0. Tal conceito traz um grande leque de características que o definem e surgiu com uma sessão de *brainstorm* realizada por Tim O'Reilly e o International Media Center. Os autores notaram que várias novas aplicações e sítios estavam surgindo com regularidade, e assim denominaram essa nova fase como Web 2.0, nascendo então a Conferência Web 2.0.

Uma das principais características da Web 2.0 é a participação do usuário, tanto no desenvolvimento de aplicações quanto na geração de conteúdo. O'Reilly (2005, tradução nossa) afirma que o aproveitamento da inteligência coletiva foi o princípio central por trás do sucesso dos gigantes da Web 1.0 e que sobreviveram para liderar a Web 2.0. O autor afirma ainda que os efeitos da contribuição do usuário são a chave para o domínio de mercado na era da Web 2.0.

Dentro dessa arquitetura de participação proposta pela Web 2.0, um grande exemplo de colaboração em massa para geração de conteúdo é a Wikipédia. A enciclopédia virtual é totalmente gerada pelos usuários, que escrevem e editam os artigos e verbetes. Atualmente a Wikipédia em língua inglesa conta com quase três milhões de artigos, o que ultrapassa as enciclopédias tradicionais como a Britânica.

O crescimento dos *blogs* é também algo que deve ser destacado, assim como sua utilidade para fins diversos da proposta inicial de diário eletrônico. Na era da Web 2.0, os *blogs* são utilizados em substituição às páginas pessoais e possuem conteúdos diversos, como comentários a notícias, divulgação de empresas e organizações e para exposição de idéias em formato jornalístico. Segundo O'Reilly (2005, tradução nossa):

Se uma parte essencial da Web 2.0 é a coleta de inteligência coletiva, transformando a web em um tipo de cérebro global, a *blogosfera* é equivalente a uma constante vibração mental no cérebro (*forebrain*), uma voz que ouvimos em nossas cabeças. Isso pode não refletir a estrutura profunda do cérebro, que está inconsciente, mas em vez disso equivale ao pensamento consciente. E como reflexo do pensamento e atenção conscientes, a *blogosfera* começou a ter um efeito poderoso.

Aliada aos *blogs* e a outros serviços de Rede está a tecnologia RSS (Really Simple Syndication), que permite ao usuário receber as atualizações dos *blogs* e outros sítios assim que elas ocorrem, por meio de assinatura. Outro componente chave da Web 2.0 é o AJAX, uma junção de tecnologias que funcionam em conjunto em uma nova forma poderosa. Várias aplicações como Flickr, Gmail e Orkut a utilizam.

Devido a essa forte característica de colaboração do usuário na Web 2.0, ela é também chamada de Web Social e as redes sociais de interação são marcantes nessa nova era: “Tendo amadurecido ao longo dos anos como um meio estático de apresentação, a web é agora a base para novas

formas dinâmicas de comunidade e expressão criativa” (TAPSCOTT e WILLIAMS 2007, p. 51). As principais redes sociais são Facebook, Orkut, MySpace, Second Life, e outras como o Twitter, plataforma de *microblogging* bastante aclamada recentemente que permite ao usuário acompanhar o dia-a-dia dos contatos presentes em sua lista de relacionamento, chamados de “seguidores.”

Dessa forma, além da característica tecnológica, o conceito “2.0” se referindo à colaboração em massa é aplicado em vários outros campos da sociedade atual. Segundo Tapscott e Williams (2007, p. 9), “[...] as mudanças profundas na natureza da tecnologia, da demografia e da economia global estão fazendo emergir novos e poderosos modelos de produção baseados em comunidade, colaboração e auto-organização, e não em hierarquia e controle.”

Toda essa nova era da web também necessita de formas de organização de seu conteúdo, que cresce ainda mais com a colaboração do usuário. A organização da informação na web será abordada no tópico a seguir.

2.2.1 Organização e recuperação da informação na web

O que mais se discute hoje no contexto da recuperação da informação são os ambientes digitais e seus mecanismos de busca de informação. A web é um grande repositório de informação e constitui uma fonte amplamente recorrida por grande parte dos usuários com alguma necessidade de informação. Enfrentam-se, porém, problemas com a informação recuperada, principalmente em relação à relevância. Conceito importante na Ciência da Informação, a relevância é definida por Le Coadic (2004, p. 62) como a correspondência entre documento e consulta. O problema da recuperação da informação já era conhecido, discutido e estudado pelo profissional bibliotecário, conhecedor de vários modelos de recuperação da informação em bases de dados bibliográficas. Recentemente, a dificuldade em se encontrar informação útil na web atraiu interesse para essa área de pesquisa, desenvolvendo-se temas como indexação automática, processamento da linguagem natural, mineração de texto, entre outros.

Feitosa (2006, p. 33) explica que os serviços de recuperação da informação na web são categorizados, segundo suas características tecnológicas, em diretórios de pesquisa e mecanismos de busca. Ambos são utilizados para a recuperação da informação na web, mas de forma diferente um do outro. Os mecanismos de busca criam índices de recuperação automaticamente por meio de *softwares* robôs, os chamados *spiders* ou *crawlers*, buscando a todo o momento informações presentes nos sites. Já os diretórios elaboram seus índices por meio de descrições de páginas fornecidas pelos seus criadores no momento de submetê-las a um cadastro. (FEITOSA, 2006).

As pesquisas [nos diretórios] baseiam-se nas informações fornecidas, e não no conteúdo real das páginas. Assim, a indexação dos mecanismos de busca é feita de maneira diferente daquela dos diretórios e os fatores importantes para a classificação das páginas também diferem. (CAMPOS; BAX, 2000).

Os diretórios foram a primeira tentativa de organização da informação na web, precursores dos mecanismos de busca. De acordo com a Wikipédia, os diretórios de sítios são índices geralmente organizados hierarquicamente por categorias e subcategorias. Dessa forma, o usuário pode recuperar informação por meio de classes ou categorias de assuntos ao invés de usar palavras-chave como em um mecanismo de busca. Em exemplo de diretório sempre citado é o do Yahoo!, que na época de seu surgimento utilizou os diretórios como base da sua organização, conforme ilustrado na figura 1:

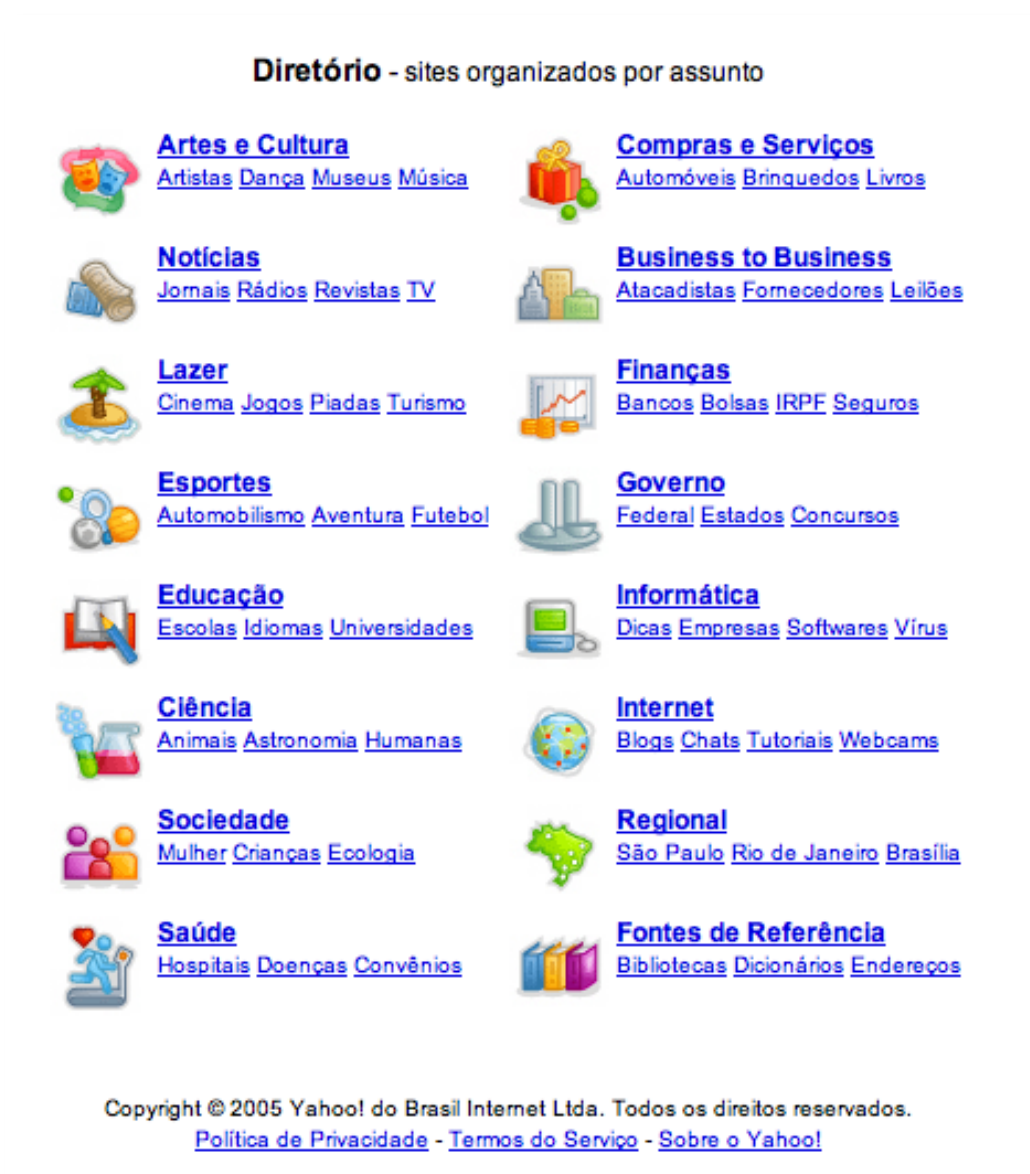


Figura 1 – Diretórios do Yahoo!

Os mecanismos de busca começaram a surgir nos anos 1990. Pretende-se, nesta parte da revisão, destacar a evolução tecnológica dos mecanismos de busca, citando assim os que foram surgindo ao longo do tempo e que representaram alguma mudança de tecnologia na forma de recuperação da informação.

De acordo com Sonnenreich (1997), o primeiro foi o Archie, criado em 1990, antes mesmo do surgimento da web. Apesar da WWW ainda não existir nessa época, a Internet já possuía muitos arquivos em sua vasta rede. O Archie foi desenvolvido para recuperar arquivos depositados via File Transfer Protocol (FTP), método primário de arquivar e recuperar arquivos, por meio de coleta de dados que unia listas de arquivos depositados por FTP, ou seja, era feita uma indexação de todos os arquivos encontrados, permitindo assim recuperá-los. Em 1993, surgiu o VERONICA (Very Easy Rodent-Oriented Netwide Index to Computerized Archives), com o mesmo propósito do Archie, mas para documentos depositados via Gopher (um servidor como o FTP, mas para documentos ao invés de arquivos). Pouco depois surgiu, também para documentos via Gopher, o Jughead (Jonzy's Universal Gopher Hierarchy Excavation and Display).

O primeiro robô de busca na web foi o Wanderer, desenvolvido para acompanhar o crescimento da web. No início eram contados somente os servidores de web, e logo depois as URLs também. Essas URLs coletadas deram origem ao Wandex, primeira base de dados da web (SONNENREICH, 1997). Outro mecanismo que surgiu para a web foi o ALIWEB (Archie-Like Indexing of Web), equivalente ao Archie, mas para a web. O sistema não utilizava robôs de busca automática, mas dados indexados pelos webmasters das páginas, o que acabou resultando em uma base de dados pequena, já que nem todos mandavam a descrição da página.

Com o crescimento da web, começaram a aparecer mais mecanismos baseados em robôs, os *spiders*, como são chamados, executando as buscas com base também nos *links* das páginas. Surgiram então no final de 1993 o JumpStation, o *World Wide Web Worm*, e o Repository-Based Software Engineering (RBSE). O RBSE foi o primeiro a ordenar os resultados, implementando um sistema de *ranking* baseado na relevância da cadeia de palavras-chave. Fundado inicialmente com o nome de Architext em 1993, o Excite tinha por base o uso de análise estatística de relacionamentos entre as palavras para fornecer buscas mais eficientes.

Já em 1994, surgiu o Galaxy, o primeiro mecanismo de busca baseado em diretórios da web. Por ser um diretório, o Galaxy foi organizado em categorias hierárquicas, e todas as URLs submetidas aos diretórios foram categorizadas manualmente, o que resulta numa busca de maior qualidade, porém há menos páginas para se buscar. Logo depois, em abril de 1994, surgiu o Yahoo!, criado originalmente como um diretório de documentos preferidos de dois candidatos ao PhD na Universidade de Stanford, David Filo e Jerry Yang. Segundo Sonnenreich (1997), com o crescente número de *links* e as páginas de Filo e Yang recebendo milhares de visitas por dia, eles criaram formas de melhor organizar os dados, tornando o Yahoo! então um diretório pesquisável por meio de um simples mecanismo de busca de base de dados. O autor afirma ainda que pelo fato do Yahoo! ser categorizado manualmente, ele não é classificado como um mecanismo de busca, mas geralmente considerado um diretório pesquisável. Apesar disso, o autor comenta que, desde a automação do Yahoo!, essa distinção entre mecanismo de busca e diretório não é mais tão clara.

Com a evolução tecnológica, os robôs de busca passaram a indexar o texto completo de uma página web, e, baseado nessa tecnologia, surgiu o WebCrawler, primeiro mecanismo que

permitia busca em texto completo. O mecanismo foi desenvolvido como projeto de um aluno da Universidade de Washington, posteriormente comprado pela American Online (AOL) e depois pelo Excite. Logo depois surgiram os concorrentes Lycos, Infoseek e OpenText. Uma das inovações do Lycos foi, além de fornecer os resultados por ordem de relevância, incorporar às buscas a correspondência de prefixos e proximidade de palavras.

Em 1995 surgiu o AltaVista com várias inovações como busca em linguagem natural e operadores booleanos. Além disso, foi o primeiro a permitir aos usuários adicionarem ou apagarem URLs aos resultados do índice. A interface do usuário também trouxe novidades, como dicas de busca e sistema de representação gráfica dos resultados de busca. Em 1996 foi lançado o HotBot, até então o mais poderoso motor de busca, com robôs *spiders* que podiam indexar 10 milhões de páginas por dia. Além disso, o HotBot começou a fazer uso extensivo da tecnologia dos *cookies*, pequenos arquivos que armazenam informação sobre preferências pessoais de busca, o que traz maior personalização nos resultados (SONNENREICH, 1997).

Com essa variedade de mecanismos de busca e peculiaridades de cada um, as buscas realizadas traziam resultados completamente diferentes quando executadas nos diversos mecanismos, ou seja, a revocação não era confiável. Pensando nisso, Eric Selburg, da Universidade de Washington, desenvolveu em 1995 o MetaCrawler, mecanismo que buscava nos principais buscadores já existentes simultaneamente, tornando assim a busca mais completa. Outra ferramenta que surgiu nesses moldes de metabusca foi o SavvySearch.

Em 1998 surgiu o Google, projeto de Larry Page e Sergey Brin, da Universidade de Stanford. Inicialmente os fundadores do Google criaram um buscador chamado BackRub, que analisava os *links* apontados para um dado site, usando a noção de citação, conceito popular no meio acadêmico. Segundo WALL (2007?), na web, os *links* funcionam como citações. A tecnologia responsável por esse tipo de análise é baseada no algoritmo PageRank, criado por Sergey Brin e posteriormente utilizado no Google. A ordenação dos resultados de busca com base nesse algoritmo gera uma maior relevância dos resultados, o que é um dos pontos fortes do Google. Também em 1998 a Microsoft lançou o MSN Search, e posteriormente, em 2006, o Live Search.

Já entre os diretórios, surgiu em 1998 o Open Directory Project (também conhecido como DMOZ), que é o maior diretório da web, elaborado em sua maioria por editores voluntários. Existe também o Librarian's Internet Index, diretório organizado por bibliotecários. Além desses, surgiram também alguns diretórios especializados em áreas do conhecimento, como o Business.com, que é um diretório de sítios de negócios. (WALL, 2007?).

Com uma nova abordagem para a recuperação da informação na web, surgiu em 2009 o Wolfram Alpha, mecanismo mais voltado para operações científicas e matemáticas que pretende, em longo prazo, fazer todo o conhecimento sistematizado imediatamente acessível por qualquer pessoa. O sítio pretende trazer respostas definitivas para questões factuais, e não se considera um mecanismo de busca, mas um mecanismo de conhecimento computacional.

Foram citados os principais mecanismos de busca encontrados na literatura da área, porém, além desses, existem muitos outros que surgiram e desapareceram com a mesma rapidez, ou não foram tão expressivos a ponto de serem aqui citados. Apesar do desenvolvimento da tecnologia e da evolução dos mecanismos de busca, a recuperação de informação relevante na Internet ainda não atingiu um bom nível de qualidade.

Além dos mecanismos de busca independentes, outra forma de recuperação de informação na web são os mecanismos dentro dos próprios sítios, que começaram a surgir devido ao crescimento de informação e conseqüente demora na indexação das páginas pelos mecanismos de busca externos. Um exemplo disso pode ser encontrado em sítios de comércio eletrônico e em grandes portais.

Outra tecnologia que surgiu para a organização e recuperação da informação na web foi o chamado *clustering*. Tal técnica consiste em agrupar objetos semelhantes, no caso da web, informação semanticamente próxima. Segundo Feitosa (2006, p. 41), “a técnica do *clustering* difere de outras como a classificação, a construção de taxonomias, e a marcação de dados, uma vez que é totalmente automatizada, dispensando a intervenção humana.” Um dos sistemas que utiliza tal técnica é o Vivísimo, apresentando os resultados da busca em forma de *clusters*. Há também outros sistemas de busca que inovam na apresentação dos resultados, como o KartOO, que traz mapas gráficos interativos, o ProFusion, que organiza as fontes de informação por categorias de assuntos e o Scirus, especializado em pesquisa científica, que incorpora o controle terminológico nas buscas. (FEITOSA, 2006, p. 41-43).

O problema da recuperação da informação na web está basicamente na enorme quantidade de documentos nela depositados. Mesmo com a alta tecnologia dos mecanismos de busca, os resultados ainda não oferecem uma precisão muito alta quando se trata de buscas de assuntos menos triviais. Há algumas formas de melhorar esses resultados. Há ainda muito que se estudar e esperam-se grandes melhorias nessa área, como recuperação de alta qualidade, técnicas de indexação mais rápidas e resultados de busca mais precisos. Comenta-se que a próxima geração da web, a chamada Web Semântica, poderá proporcionar tais melhorias.

Dois elementos que “ajudam” o sistema de indexação automática dos buscadores são os metadados e os *links*. Como já citado anteriormente, o Google utiliza os *links* como critério para ordenar seus resultados. Já os metadados se referem a termos descritivos da informação contida em determinado documento, ou seja, dados sobre o documento. Segundo Feitosa (2006), o uso de metadados é uma tendência para a organização da web. De acordo com Taylor (2003), há várias definições para o termo metadado, das mais simples às mais complexas, mas o que todas têm em comum é a noção de que metadado é informação estruturada que descreve atributos de recursos informacionais com o propósito de identificação, descoberta, e, às vezes, administração. (TAYLOR, 2003, p. 139, tradução nossa). As principais características dos metadados são:

- Interoperabilidade: capacidade dos vários sistemas de interagir uns com os outros, minimizando assim a perda de informação devido a diferenças tecnológicas. A interoperabilidade se dá em nível sintático, semântico e estrutural.
- Flexibilidade: possibilidade dos criadores de metadados de incluírem o máximo ou o mínimo de detalhes desejados em um registro, seguindo ou não as regras de catalogação ou listas de autoridades.
- Extensibilidade: uso de elementos adicionais de metadados e qualificadores conforme necessário para as várias comunidades. (TAYLOR, 2003, p. 143, tradução nossa).

Entre os principais esquemas de metadados que podem ser utilizados para a organização da web destaca-se o Dublin Core, criado em 1995 pela Online Computer Library Center (OCLC). O Dublin Core é um padrão de metadados que objetiva a descrição de recursos na web. De acordo com o sítio do projeto, a iniciativa está comprometida em desenvolver padrões de metadados *online* interoperáveis que suportem uma ampla gama de propósitos e modelos de negócios. A missão do Dublin Core Metadata Initiative é fornecer padrões simples para facilitar a busca, o compartilhamento e o gerenciamento da informação. (DCMI, 2009).

Há ainda outros tipos de padrões de metadados utilizados para os recursos da Rede, inclusive em domínios específicos, como o TEI (Text Encoding Initiative) para textos antigos literários e acadêmicos, o VRA (Visual Resources Association) para recursos visuais, e o Archival APPM (Arquivos, papéis pessoais e manuscritos) para a descrição de materiais de arquivo. Esses padrões podem ser codificados por computadores por meio das chamadas linguagens de marcação, que permitem a construção de padrões públicos e abertos, facilitando assim a troca de informação na web. Os documentos são constituídos de conteúdo, estrutura e estilo, e as linguagens de marcação separam essas três instâncias. As linguagens mais conhecidas são SGML e HTML, além das mais recentemente aclamadas XML e RDF. (BAX, 2001, p.32, 34).

As linguagens de marcas trabalham com o conceito de *tags* ou *meta tags* para descrever as informações em sua estrutura. Bax (2001) utiliza o termo como sinônimo de marca, nesse contexto. A *tag*, nesse caso, refere-se a uma marcação do texto que aparece na estrutura da linguagem, como por exemplo <title>, <head> e <author>. Segundo Bax (2001, p. 33), “elas indicam a função (o propósito) da informação no documento, em vez de como ela deve ser apresentada, ou seja, sua aparência física. A idéia básica é a de que o conteúdo do documento deve estar separado do estilo usado em sua apresentação.” É importante ressaltar que essa *tag* (marca) não se confunde com a *tag* (etiqueta) da folksonomia, que tem um conceito diferente e será apresentado na próxima seção desta revisão.

Outra ferramenta que pode ser utilizada para a organização da web são os esquemas de representação do conhecimento, já descritos anteriormente nesta revisão. Segundo Hodge (2000, tradução nossa), “há um interesse significativo no uso de sistemas de organização do conhecimento para organizar e buscar material na Internet.” O uso de esquemas de classificação ajuda a elaborar um diretório, uma taxonomia pode ser construída para um sítio de comércio

eletrônico, e as ontologias são ditas como a base da Web Semântica. Os sistemas de organização do conhecimento existentes, como as classificações, têm sido reconhecidos como importante fonte de vocabulários estruturados e formalizados que podem ser explorados para dar suporte ao desenvolvimento da Web Semântica (SLAVIC, 2005, tradução nossa).

Uddin e Janecek (2007, p. 219, tradução nossa) também afirmam que uma estrutura de classificação influencia a arquitetura de informação de um sítio, especificamente no processo de organização, busca e *browsing*, o que ajuda o usuário a encontrar e acessar informação. Os autores propõem o uso da classificação facetada para arquitetura de sítios, pois as estruturas de classificação tradicionais, apesar de úteis, possuem limitações:

a classificação tradicional enumerativa monodimensional hierárquica, quando aplicada a uma arquitetura de sítio, trata o conhecimento do sítio como um todo integrado que é dividido e subdividido em grupos específicos, como uma estrutura de árvore. Em uma estrutura hierárquica, um objeto é tipicamente alocado em uma categoria, não sendo possível expressar as propriedades e relacionamentos multidimensionais dos objetos digitais [...] Por essa razão, noções tradicionais de simples hierarquias de classificações e taxonomias precisam ser aumentadas com métodos mais poderosos. (BROUGHTON, 2001 *apud* UDDIN e JANECEK, 2007; UDDIN e JANECEK, 2007, p. 220, tradução nossa).

Dessa forma, o uso de uma estrutura de classificação facetada consegue superar essas limitações, com o uso de categorias múltiplas como na proposta da classificação facetada de Ranganathan. O conteúdo de um sítio deve ser analisado para que se definam as facetas mais adequadas para a recuperação da informação no sítio.

A partir da Web 2.0, outras formas de organização da informação começaram a surgir e ganhar destaque. Entre essas novidades, o que vem crescendo no ambiente web é o uso de etiquetas (*tags*) para descrever objetos informacionais, o que pode ser usado posteriormente para a recuperação da informação. Esse processo gera as chamadas folksonomias, tema principal desta pesquisa e que será abordado a seguir.

2.2.2 Etiquetagem e Folksonomia

Etiquetagem (em inglês, *tagging*) é uma forma de indexação em que as próprias pessoas, no caso, os usuários da informação, classificam os documentos. Golder e Huberman (2006, p. 198, tradução nossa) explicam que:

Marcar conteúdo com termos descritivos, também chamados palavras-chave ou etiquetas, é uma forma comum de organizar conteúdo para futura navegação, filtragem ou busca. Apesar disso, organizar conteúdo eletrônico dessa forma não é algo novo, uma forma colaborativa desse processo que vem sendo chamada de etiquetagem pelos seus proponentes, está ganhando popularidade na web.

Da mesma forma, Wal (2007) afirma que a história da etiquetagem é muito mais profunda e antiga do que se pensa: um dos primeiros produtos que incluíram etiquetagem ou anotações foi o Lotus Magellan, que surgiu em 1988 e permitia anotações de documentos e objetos do disco rígido do usuário. Já em 2001 é apontado o uso de etiquetagem para qualquer mídia que contém URL. Para o autor, o valor da etiquetagem aparece quando o sistema de etiquetagem traz um objeto comum etiquetado e a identidade de quem o etiquetou. É o que começa a acontecer em 2003, quando o sítio de *social bookmarking* Delicious é criado. (WAL, 2007, tradução nossa)

Wal (2005) define etiquetagem como uma maneira não-hierárquica de organizar informações ou objetos na web. A informação é encontrada por meio de correlações associativas e inferências. Segundo Wal, as hierarquias eram uma maneira eficiente de organizar informação antes do surgimento da computação, mas com os computadores há outras formas que podem funcionar tão bem quanto elas. (WAL, 2005, tradução nossa).

O resultado do processo de etiquetagem gera o que vem sendo denominado de folksonomia. O termo surgiu em 2004 no contexto da Internet, onde a observação de uma prática crescente – a atribuição de etiquetas a conteúdos informacionais despertou o interesse de pessoas que acompanham os fenômenos que ocorrem na web. Tal interesse suscitou um tópico sobre o assunto em uma lista de discussão e daí uma necessidade de denominação de tal fenômeno.

A criação do termo *folksonomy* é atribuída ao arquiteto da informação Thomas Vander Wal, que o define como:

[...] o resultado da atribuição livre e pessoal de etiquetas a informações ou objetos (qualquer coisa com URL), visando à sua recuperação. A atribuição de etiquetas é feita num ambiente social (compartilhado e aberto a outros). A etiquetagem é feita pelo próprio consumidor da informação. (WAL, 2004).

Segundo o autor, o termo vem da sugestão de outro membro da lista, Eric Scheid: *folk classification* (classificação por pessoas, ou social). O prefixo folk é então usado por Wal para substituir o ‘tax’, de taxonomia, gerando o termo folksonomia. O autor ressalta que folksonomia não é colaborativa, não é “colocar coisas em categorias” e não está relacionada à taxonomia (estaria mais próxima de uma antítese da taxonomia) e explica que o termo foi cunhado em torno da idéia de consumidores do conteúdo colocando etiquetas livremente na informação ou objeto para seu próprio uso, bem como compartilhando essa etiquetagem abertamente. (WAL, 2005a, 2005b, 2007, tradução nossa)

Relata-se na literatura que os primeiros sítios a permitirem que seus usuários atribuíssem etiquetas aos recursos informacionais foram o Delicious, em 2003, seguido do Flickr. Posteriormente, vários outros serviços de informação da web adotaram a prática, implementando a etiquetagem em seus sítios.

O assunto “folksonomia” é algo recente e é um fenômeno que surgiu com a própria prática, encontrando-se assim poucos estudos teóricos que tratam do tema. Dessa forma, não há uma vasta literatura de base teórica na área, visto que o fenômeno tem um caráter muito mais pragmático do que teórico. As pesquisas envolvendo o assunto são, em sua grande maioria, aplicações práticas e

análises da folksonomia em diferentes ambientes e formatos. Assim, esta parte da revisão de literatura buscará reunir os primeiros artigos sobre o surgimento da folksonomia como conceito e objeto de estudo, trazendo as principais definições, idéias e opiniões propostas pelos autores. Uma das definições mais vistas na literatura é a de Thomas Vander Wal, já citada anteriormente neste trabalho. Mas há ainda várias propostas de definições por outros autores, que serão abordadas a seguir.

O editorial do periódico eletrônico *Webology*, de junho de 2007, trata de folksonomia, definindo-a com uma visão diferente de Wal. Segundo o editorial, a folksonomia é uma taxonomia gerada por usuários para categorizar e recuperar conteúdo como recursos da web, fotos *online* e *links* da web, utilizando etiquetas. É um sistema de classificação de conteúdos da web que permite aos usuários etiquetarem seus recursos favoritos com suas palavras e frases escolhidas, selecionadas da linguagem natural. Folksonomia é uma classificação dos usuários, pelos usuários e para os usuários. (NORUZI, 2007, tradução nossa).

De forma simplificada, Tapscott e Williams (2007, p. 57-58) explicam etiquetagem e folksonomia: “quando as pessoas ‘etiquetam’ (colocam *tags*) um conteúdo de maneira colaborativa, cria-se uma ‘folksonomia’, que é essencialmente uma taxonomia orgânica de baixo pra cima que organiza o conteúdo da web.” A questão da colaboração é destacada pelos autores, e, no caso, significa dizer que as etiquetas colocadas por um usuário podem ser compartilhadas com outros usuários: na maioria dos sistemas que permitem etiquetagem, todos os usuários podem ver as etiquetas utilizadas pelos outros para um determinado recurso. Além disso, a formação de folksonomias só é possível devido a essa coletividade, já que a nuvem de etiquetas (*tag cloud*), que é a representação gráfica da folksonomia, é formada a partir das etiquetas mais utilizadas pelos usuários.

Quintarelli (2005, tradução nossa) define folksonomia como uma “classificação gerada pelo usuário, surgindo por meio de um consenso de baixo para cima (*bottom-up*).” A autora também destaca a questão da colaboração, enfatizando o poder das pessoas para a criação de folksonomias e a questão do ambiente de classificação social:

Folksonomias demandam pessoas para associar palavras-chave a conteúdos [...] é uma manifestação das pessoas afastando-se dos esquemas hierárquicos autoritários [...] as pessoas começam a associar suas próprias etiquetas aos itens que querem colecionar e compartilhar. Em um ambiente social distribuído, compartilhar suas próprias etiquetas é uma forma inovadora de mapear significado e deixar os relacionamentos surgirem naturalmente. [...] O relacionamento do termo com o significado emerge por meio de um contrato implícito entre os usuários. (QUINTARELLI, 2005, tradução nossa).

Para essa autora, o poder da folksonomia está relacionado com a capacidade de agregar a informação que os usuários fornecem, ultrapassando o limite da criação de etiquetas para uso pessoal e atingindo um nível de interação social.

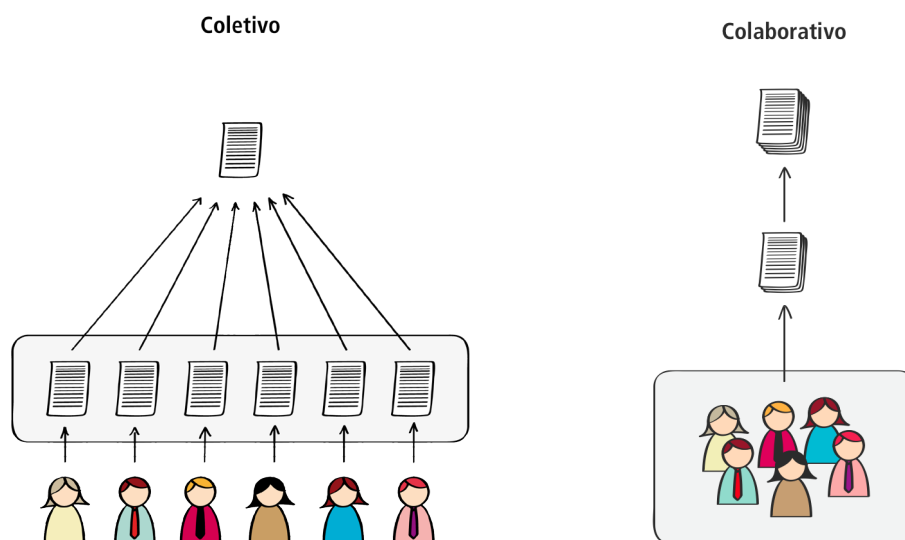
Já para o criador do termo, a folksonomia não é algo **colaborativo**, e sim **coletivo**. Wal insiste nessa discussão, afirmando que colaborativo e coletivo são coisas diferentes, mas muitas

vezes tratadas como sinônimos pelos autores, e no ambiente de *software* social há uma grande diferença entre os termos. Wal explica que o coletivo é quando várias pessoas fazem anotação, colocam etiquetas ou discutem sobre um objeto, de forma individual. O objeto é o foco do coletivo e as vozes e anotações individuais são separadas, com cada indivíduo trabalhando individualmente. Essas anotações e contribuições individuais podem ser coletadas e agregadas. Isso é o que permite o surgimento de uma folksonomia, como no Delicious. Para Wal, a falta de entendimento da abordagem coletiva significa falta de entendimento de folksonomia. Já na abordagem colaborativa, várias pessoas contribuem para o entendimento de um objeto e o fazem trabalhando juntos, agregando várias perspectivas para construir um entendimento único. Um exemplo de colaboração são as wikis, em que vários usuários participam na criação dos artigos, ou seja, há um esforço colaborativo para a definição de um tópico de assunto. (WAL, 2008).

Figura 2 – Coletivo x colaborativo (adaptado de WAL, 2008)

Hammond *et al.* (2005, tradução nossa) consideram a folksonomia:

uma abordagem de classificação com estruturação livre, com usuários designando seus próprios rótulos, o que tem sido denominado folksonomia, classificação por pessoas (*folk classification*), etnoclassificação, classificação distribuída ou classificação social, além de outros termos como etiquetagem livre e hierarquia facetada.



Os autores se dizem inclinados a utilizar o termo *classificação social* ou *classificação distribuída*, pois acreditam ser os que descrevem de forma mais próxima a natureza da atividade, apesar de reconhecerem que a palavra *folksonomia* tem ganhado frequência considerável e poucos estão fugindo disso. Percebe-se que os autores não distinguem o processo de etiquetagem ou *classificação social* do resultado gerado, ou seja, a *folksonomia*.

Mathes (2004) trata as etiquetas como um tipo de metadado compartilhado e define *folksonomia* como “um conjunto de termos utilizados por um grupo de usuários para etiquetar conteúdos.” O autor destaca a simplicidade como aspecto importante da *folksonomia*: não

há hierarquia nem relacionamentos diretos específicos entre os termos. Apesar disso, é possível a identificação de *clusters*, que podem ser gerados automaticamente por meio de etiquetas relacionadas, ou seja, atribuídas à mesma URL. Para Mathes, apesar de o termo classificação ser bastante utilizado no contexto da folksonomia, o mais apropriado seria categorização, pois para o autor, a classificação envolve maior rigor que categorização. Dessa forma, para Mathes, a categorização se aproximaria mais do que ocorre nas folksonomias. Além disso, afirma que nos esquemas de classificação há apenas uma classificação possível para cada item, são muito hierárquicos e com relacionamentos claros. Numa folksonomia não há relacionamentos claramente definidos entre os termos. (MATHES, 2004, tradução nossa).

Acredita-se que tal distinção não se faz necessária, pois a classificação, em sentido amplo e geral, diz respeito a qualquer processo de agrupamento de coisas semelhantes e separação de coisas diferentes, independente de um esquema formal de classificação. Além disso, o processo de categorização também envolve uma decisão de se alocar um item em um sistema maior já definido, baseado em regras e características apropriadas. Ou seja, não justifica a distinção, e acredita-se ainda que o termo classificação seja mais adequado por ter um sentido mais genérico, abrangendo todo e qualquer processo de organização de conteúdos de acordo com suas características.

Em novembro de 2007, o *Bulletin of the American Society for Information Science and Technology* (ASIS&T) publicou uma edição especial sobre folksonomia. Em seu editorial, Neal (2007) destaca a folksonomia como uma das grandes tendências da Internet atualmente, fazendo parte do contexto da Web 2.0. A autora explica que uma folksonomia é criada quando usuários de um sítio adicionam etiquetas para descrever itens no sítio. O usuário escolhe suas próprias etiquetas, pouca ou nenhuma restrição é imposta em suas escolhas. Os termos não são escolhidos de um vocabulário controlado já existente, nem de uma taxonomia estrita ou qualquer outro tipo de método de descrição bibliográfica. Com base na definição de Thomas Vander Wal, a autora afirma que a folksonomia é uma ferramenta extremamente útil para o gerenciamento pessoal de informações, pois os termos utilizados são os do usuário, e não os do sistema. Além disso, destaca o benefício do aspecto social como o uso de etiquetas de outros usuários para recuperar conteúdos ainda desconhecidos e a criação de comunidades baseadas em determinadas etiquetas. Estas funcionam como uma representação do conhecimento coletivo dos usuários. (NEAL, 2007, pp. 7-8, tradução nossa).

Também publicado nessa edição do *Bulletin* da ASIS&T, um artigo sobre organização de recursos visuais destaca a folksonomia como tendência em indexação de imagens na web: “Parece que a web criou sua solução pessoal, usando seus próprios métodos de indexação de imagens, como fazem com a etiquetagem colaborativa, última tendência na indexação de imagem” (MÉNARD, 2007, p. 23, tradução nossa). Ménard utiliza o termo “etiquetagem colaborativa”, e afirma que esta começa com qualquer usuário designando sua própria palavra-chave para documentos textuais, o que foi rapidamente implementado para a indexação de imagens. A autora destaca as vantagens da folksonomia para a recuperação de recursos visuais:

Instintivamente, os usuários parecem incluir o que eles acham que é significante e imperativo empregar para alguém que precise recuperar a imagem. A etiquetagem colaborativa parece muito sedutora, pois tem uma grande relação com usuários reais e a forma como vêem e descrevem coisas. Além disso, neologismos e todos os novos termos criados são rapidamente integrados à indexação colaborativa. [...] Nos sistemas de indexação colaborativa, a indexação pode ser feita em uma ou várias linguagens, o que facilita a recuperação. (MÉNARD, 2007, p. 23, tradução nossa).

Beaudoin (2007) também faz uma análise da etiquetagem para a recuperação de imagens, aplicando um estudo no Flickr para a identificação de padrões nas etiquetas. A autora acredita que a etiquetagem é útil na recuperação de recursos visuais, e destaca as implicações desse sistema para os profissionais da informação. Ressalta que esses devem desenvolver novos métodos e técnicas pra ajudar as pessoas na etiquetagem e recuperação de seus sempre crescentes acervos de materiais visuais, além de estudar o comportamento de quem coloca as etiquetas: saber o que o usuário está preocupado em recuperar pode atingir uma maior audiência. (BEAUDOIN, 2007, p. 29, tradução nossa).

Ainda nesta edição do *Bulletin*, Morrison afirma que as folksonomias são geralmente utilizadas para organizar a informação e permitir que ela seja recuperada. O autor acredita que a folksonomia é mais provável de ter sucesso quando os objetivos dos sítios ou sistemas de informação coincidem com os objetivos e motivações dos usuários. (MORRISON, 2007, p. 12, tradução nossa). Já Guy e Tonkin definem a folksonomia como um sistema de classificação distribuída, geralmente criada por um grupo de pessoas, tipicamente os usuários do recurso e, de forma um tanto imprecisa, definem etiqueta como palavra-chave, categoria, nome ou metadado, ou simplesmente um grupo de palavras livremente escolhidas. (GUY; TONKIN, 2006).

Marlow *et al.* (2006) comentam sobre a popularização da etiquetagem a partir dos sistemas de *social bookmarking*, que permitem ao usuário guardar seus sítios favoritos de forma acessível em qualquer computador com Internet, e não somente nos favoritos de seu navegador, e ainda marcá-los com etiquetas para auxiliar na recuperação. A popularização desses sistemas deu origem aos chamados sistemas de etiquetagem social (*social tagging systems*): “essas ferramentas têm crescido em popularidade à medida que elementos de interação social têm sido introduzidos, conectando atividades individuais de marcação de favoritos a uma rica rede compartilhada de etiquetas, recursos e usuários.” Os autores definem folksonomia como uma “taxonomia popular (*folk taxonomy*) de conceitos importantes e emergentes para um grupo de usuários” e propõem um modelo conceitual dos sistemas de etiquetagem com base em três elementos: recursos, usuários e etiquetas. (MARLOW *et al.*, 2006, tradução nossa).

Catarino e Baptista (2007) fazem uma revisão de literatura sobre o tema, e, segundo as autoras, a folksonomia pode ser considerada como um novo paradigma para a organização dos conteúdos dos recursos digitais na web:

Trata-se de um novo conceito que tem sido utilizado por diversos profissionais e estudiosos da área de informação. No entanto, parece não haver ainda um consenso na área, quer sobre a utilização desse termo, quer sobre seu significado. Há

os que preferem utilizar outros termos como, por exemplo, classificação social ou *tagging*. (CATARINO; BAPTISTA, 2007).

Nesse ponto, as autoras utilizam a denominação de folksonomia de forma genérica ao afirmar que alguns autores utilizam termos como classificação social ou *tagging*. No caso, a folksonomia, pela definição de Wal, pode ser entendida como o **resultado** do processo de classificação social ou *tagging*, ou seja, são coisas distintas e não há essa ambivalência entre os termos. As autoras concluem:

Portanto, Folksonomia é o resultado da etiquetagem dos recursos da *Web* num ambiente social (compartilhado e aberto a outros) pelos próprios usuários da informação visando a sua recuperação. Destacam-se, portanto três fatores essenciais: 1) é resultado de uma indexação livre do próprio usuário do recurso; 2) objetiva a recuperação *a posteriori* da informação e 3) é desenvolvida num ambiente aberto que possibilita o compartilhamento e, até, em alguns casos, a sua construção conjunta. (CATARINO; BAPTISTA, 2007).

Há realmente certa confusão entre os termos folksonomia e etiquetagem, *tagging* ou classificação social, sendo muitas vezes utilizadas como sinônimos. Autores como Golder e Huberman evitam usar o termo folksonomia, já que ainda há discussões terminológicas e conceituais em torno dele, e preferem utilizar a terminologia “etiquetagem colaborativa” para se referir ao processo de marcação de itens com etiquetas na web. (GOLDER; HUBERMAN, 2005). Para fins deste trabalho, será utilizada a noção de etiquetagem (*tagging*, classificação social) como um **processo** que tem como **resultado** uma folksonomia, conforme definição e discussões de Thomas Vander Wal.

A maioria dos autores que tratam sobre folksonomia identificam suas vantagens e desvantagens para a organização e recuperação da informação. A principal desvantagem citada é a falta de controle de vocabulário, ou seja, problemas como termos homônimos, sinônimos e uso de singular/plural, gerando inconsistência, redundância e falsas associações, o que prejudica a precisão da informação recuperada. Afirma-se também que a folksonomia é um estrutura sem controle: a mesma etiqueta pode ser usada para recursos diferentes, com sentidos semânticos diferentes, criando associações não relacionadas entre os recursos. Como vantagens os autores apontam o baixo custo de implementação, o vocabulário próprio do usuário e não o do sistema, a inserção em tempo real de temas emergentes, o cunho social e colaborativo, a formação de comunidades em torno de assuntos de interesse, a liberdade de expressão do usuário, a flexibilidade do sistema, a descoberta de informação nova, e, no caso de recursos visuais, a folksonomia possibilita a descrição do conteúdo visual e as etiquetas funcionam como uma espécie de metadado para a recuperação da informação por um motor de busca. (BRUCE, 2008; WALKER, 2005 *apud* MARLOW *et al.*, 2006; WAL, 2007; CATARINO; BAPTISTA, 2007; NORUZI, 2007; QUINTARELLI, 2005; SPITERI, 2007; GOLDER; HUBERMAN, 2005; GUY; TONKIN, 2006).

2.2.2.1 Tipos de folksonomia

Wal distingue dois tipos de folksonomia: *broad folksonomy* e *narrow folksonomy*. Os termos *broad* e *narrow* são comumente traduzidos para geral e específico, respectivamente. Porém, essa terminologia já é própria da área de linguagens documentárias, e se refere ao termo (termo geral e termo específico). Já no caso das folksonomias, o adjetivo se refere ao tipo de ambiente em que se insere a folksonomia e ao processo por meio do qual esta é produzida, e não aos termos ou às etiquetas. Além disso, não foi estabelecido um consenso entre os autores de língua portuguesa que tratam do tema, encontrando-se as traduções aberta/fechada, ampla/estreita e larga/estreita. Por esse motivo, optou-se pela tradução dos termos para folksonomia aberta, no caso de *broad*, e restrita no caso de *narrow*, pois acredita-se que descrevem melhor o conceito pretendido: as *broad folksonomies* são abertas a qualquer usuário e as *narrow folksonomies*, restritas ao usuário que detém o objeto ou, no máximo, a outros usuários permitidos pelo detentor.

Nas folksonomias abertas, várias pessoas colocam etiqueta no mesmo objeto e cada pessoa pode colocar sua própria etiqueta, utilizando seu vocabulário. Dessa forma, um mesmo objeto pode receber etiquetas de milhares de pessoas diferentes. É o que acontece nos sistemas de *social bookmarking*, como o Delicious. As folksonomias abertas são ilustradas por Wal da seguinte maneira:

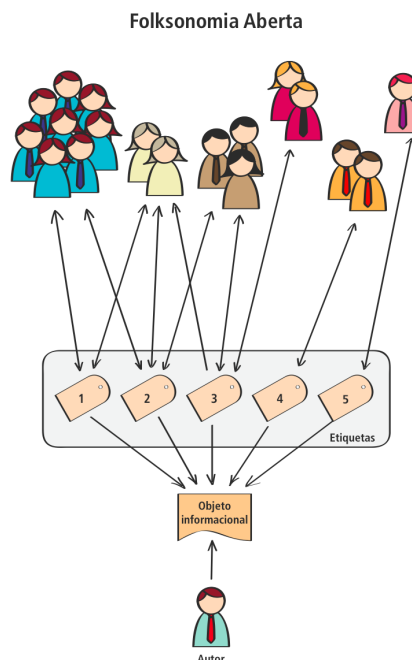


Figura 3 – Folksonomias abertas (adaptado de WAL, 2005)

Nesse caso, duas características são identificadas: a lei da potência e cauda longa (*power law* e *long tail*) ilustradas na figura 4. A lei da potência significa dizer que muitas etiquetas se repetem, pois muitos usuários utilizam o mesmo termo para descrever determinado objeto. Ao mesmo tempo, há muitas etiquetas usadas poucas vezes, formando a chamada cauda longa:

“É onde se encontra uma pequena minoria de pessoas que denominam o objeto por um termo” (WAL, 2005, tradução nossa). Essas duas características são importantes: a primeira permite a identificação de tendências, ou seja, destaca as etiquetas mais populares, e a segunda permite que o objeto seja recuperado por outros usuários que utilizem o mesmo vocabulário das minorias que o classificaram com etiquetas menos populares. Segundo Quintarelli (2005, tradução nossa, grifo nosso): “O poder real das folksonomias abertas está na riqueza das massas, nas pessoas expondo sua forma de definir e descrever coisas, o que leva à lei da potência e à cauda longa. Esses efeitos estão ausentes em *personomias*, isto é, etiquetagem individual para o próprio conteúdo.”

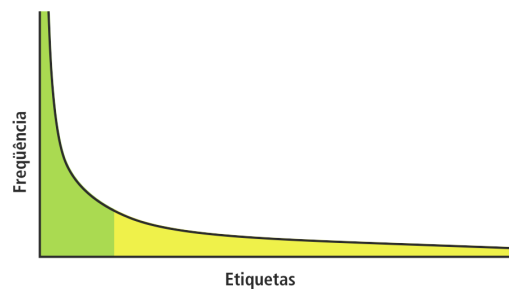


Figura 4 – Lei da potência e cauda longa

Já nas folksonomias restritas, um grupo menor de pessoas, ou às vezes somente uma pessoa, coloca etiqueta em um objeto, como no caso de fotos no Flickr. O usuário que postou a foto pode permitir que algumas pessoas também adicionem etiquetas a suas fotos, ou não permitir que ninguém adicione, e, neste caso, somente ele coloca as etiquetas. Outro serviço que representa bem a folksonomia restrita é o YouTube, onde só é permitido que o próprio usuário que postou o vídeo coloque etiquetas. Dessa forma, os objetos receberão uma quantidade bem menor de etiquetas como ilustrado na figura 5:

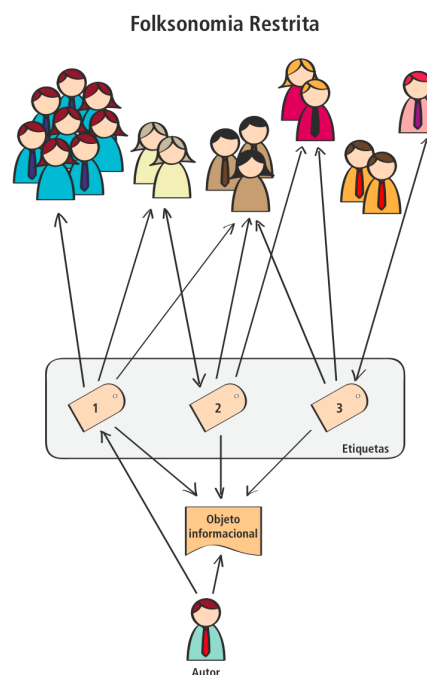


Figura 5 – Folksonomia restrita (adaptado de WAL, 2005)

Esse tipo de folksonomia não possui as propriedades da lei da potência e da cauda longa, porém, ainda são úteis na descrição de objetos visuais como fotos e vídeos, em que as etiquetas funcionam como metadados descritivos, permitindo assim a recuperação do objeto informacional por meio de um motor de busca. Para Wal (2005, tradução nossa): “estamos melhores com as folksonomias do que sem elas, mesmo que seja usada uma folksonomia restrita.”

Wal conclui que, nas folksonomias abertas, ocorre o benefício social e vocabulários interdisciplinares e interculturais podem ser descobertos, enquanto nas restritas a vantagem é fornecer etiquetas ao objeto, facilitando sua recuperação. O autor ainda identifica um tipo de folksonomia pessoal, que seria o que ocorre, por exemplo, no Gmail, que permite ao usuário colocar etiquetas em seus emails e organizá-los para seu uso pessoal. Wal acredita que as pessoas tendem a aprender novas habilidades se perceberem seu grande valor e que as etiquetas serão o passo para a próxima geração de gerenciamento de informação pessoal. (WAL, 2005, tradução nossa).

Capítulo 3

3 Metodologia

Por meio de pesquisa exploratória, buscou-se compreender o que vem sendo feito para a organização e recuperação da informação na web utilizando etiquetagem e folksonomia e sua contribuição nesse contexto, objetivo geral desta pesquisa.

Para isso, tentou-se chegar a um mínimo de elementos que permitam considerar se algo, no caso desta pesquisa, etiquetagem e folksonomia, podem contribuir para a organização e recuperação da informação na web, permitindo assim uma validação. Os elementos identificados estão representados na figura 6 a seguir.



Figura 6 – Elementos da metodologia

A escolha desses elementos considerou os seguintes aspectos: os dois primeiros estão relacionados com a base teórica da organização e recuperação da informação dentro da Ciência da Informação: indexação e esquemas de representação do conhecimento, conforme visto na revisão de literatura desta pesquisa. O terceiro elemento é fundamental nos processos estudados, acreditando-se então ser importante o estudo de usuários nessa pesquisa. Já o quarto está relacionado com a possibilidade de implementação dos processos nos diferentes contextos e ambientes, bem como para tipos de objetos informacionais distintos. Acredita-se então que a presença dos quatro elementos citados permite, de alguma forma, validar um processo de organização e recuperação da informação.

A pesquisa apoiou-se em investigação teórica e prática. Para os elementos 1 e 2, relacionados aos objetivos específicos 1 e 2, a saber: analisar a etiquetagem como forma de indexação colaborativa dos recursos da web e analisar a folksonomia sob a óptica das abordagens em organização do conhecimento, a base de pesquisa foi teórica, apoiada em revisão na literatura da

área. Já para os objetivos 3 e 4 (respectivamente: investigar o que leva os usuários a executar a etiquetagem dos recursos da web e analisar quanto essa motivação se relaciona com o objetivo e recuperação da informação e apresentar um panorama dos principais serviços de informação que utilizam etiquetagem e folksonomia, identificando seus elementos, características e abordagens, a fim de verificar sua aplicabilidade para diferentes tipos de objeto informacional) relacionados aos elementos 3 e 4, foi feita pesquisa prática, com estudo de usuários e análise descritiva dos próprios sítios que se inserem no contexto estudado. Os diferentes percursos metodológicos serão detalhados nas próximas seções.

3.1 Revisão bibliográfica

Para os objetivos específicos 1 e 2, que pretenderam identificar um conjunto de elementos relacionados com a base teórica da Ciência da Informação, mais precisamente na área de organização, representação e recuperação da informação e do conhecimento, utilizou-se o método de pesquisa bibliográfica.

Dessa forma, foi feito levantamento da literatura sobre indexação e organização do conhecimento, por meio de pesquisa nas diversas fontes de informação, buscando identificar os principais autores e abordagens da área. Após leitura e análise da bibliografia levantada, buscou-se elaborar uma fundamentação teórica que permitisse relacionar as abordagens propostas com o processo de etiquetagem e a geração de folksonomias.

3.2 Usuários

A partir do objetivo específico 3, que pretendeu investigar a motivação do usuário para a etiquetagem, foi aplicado um estudo de usuários. Segundo Pinheiro (1982, p. 1) os estudos sobre usuários da informação são importantes para o conhecimento do fluxo de informação científica e técnica, de sua demanda, da satisfação do usuário, dos resultados ou efeitos da informação sobre o conhecimento, do uso, do aperfeiçoamento, além de outros aspectos relacionados ao fluxo informacional, podendo incluir inclusive pesquisas sobre o comportamento do usuário. Este último aspecto foi o foco do estudo de usuários realizado neste trabalho. A autora (1982, p. 2) afirma ainda que “o estudo de usuários, uma vez que envolve comportamento, transcende a meras técnicas biblioteconômicas e documentárias, estendendo-se a estudos psicológicos e sociológicos.” O estudo de usuários feito aqui considerou aspectos psicológicos em sua concepção, visto que tratou de um comportamento do usuário muito abordado na psicologia, que é a motivação.

O estudo de usuários realizado neste trabalho teve como enfoque um estudo centrado no usuário e utilizou como base metodológica a pesquisa qualitativa, com pergunta aberta, portanto sem uso de técnicas de amostragem.

3.2.1 Universo

O universo da pesquisa foi composto pelos usuários de dois serviços de informação da web: Delicious e Flickr. O número de usuários cadastrados nesses serviços já ultrapassa a casa dos milhões, e está em constante crescimento. Desta forma, configura-se uma população infinita.

3.2.2 Amostra

Para seleção da amostra, buscou-se uma forma aleatória, por conveniência, de escolha dos usuários. Devido a diferenças entre os serviços estudados, não foi possível que a amostra fosse selecionada da mesma maneira nos dois serviços.

A seleção da amostra no Flickr foi feita a partir das próprias etiquetas dos usuários, por meio do recurso “explore tags”, presente no Flickr. Tal recurso mostra as etiquetas mais frequentes nas últimas 24 horas, na última semana e de todos os tempos. Utilizou-se a opção de etiquetas mais frequentes na última semana pela possibilidade de atingir usuários ativos e que utilizam etiquetas. Desta forma, a partir do acesso a uma etiqueta frequente, tem-se as fotos marcadas com tal etiqueta e chega-se assim aos usuários que a utilizaram. A partir de então, selecionaram-se aleatoriamente os usuários para envio do questionário. Foram utilizadas várias etiquetas frequentes e, ao todo, foram enviados questionários para 100 usuários do serviço.

Não houve seleção da amostra no Delicious. Utilizou-se o fórum de discussão do próprio sistema para se disponibilizar a pesquisa. Sabe-se que esta não é a melhor forma de seleção, pois atinge somente os usuários do fórum, o que pode ser considerado um viés: um público mais seletivo e menos abrangente. Porém, foi a única forma viável encontrada para se atingir os usuários desse serviço.

A seguir, será explicada como foi feita a coleta de dados nos sítios estudados.

3.2.3 Coleta de dados

O método utilizado para coleta de dados foi um questionário com 6 questões, sendo 5 delas de dados demográficos e, por último, o alvo do estudo: uma questão aberta sobre a motivação do usuário para a etiquetagem (ver apêndice A). Como os serviços estudados são originalmente em inglês, optou-se por elaborar o questionário nesta língua, para que se pudesse atingir o maior número de usuários, já que estes foram selecionados aleatoriamente.

O questionário foi elaborado no Google Docs², ferramenta que permite a criação de vários documentos, e, entre eles, questionários em linha. O serviço coleta os dados e os armazena em uma planilha, que pode ser exportada em formato Excel e compartilhada com outras pessoas. Foi feito pré-teste do questionário com cinco usuários de Internet antes da aplicação nos sistemas estudados.

O período de coleta de dados foi de 14/10/08 a 31/10/2008. A coleta de dados realizada no Flickr foi feita com o envio de mensagens pelo FlickrMail, forma de comunicação interna entre os usuários do sistema. As mensagens continham um texto explicativo da pesquisa e o *link* para o questionário de coleta de dados.

Já para o sítio Delicious, devido à impossibilidade de contato direto com o usuário, optou-se pela utilização do fórum de discussão disponibilizado no sistema, e, posteriormente, uma lista de *e-mails* sobre o Delicious, onde foram enviadas mensagens com o conteúdo da pesquisa e o *link* para o questionário *online*. Como última tentativa de se obterem mais respostas nesse sistema, o questionário foi enviado diretamente para pessoas que sabidamente o utilizam. Tentou-se aplicar o questionário por outro meio, utilizando-se um recurso de indicação de sítio favorito para membros do serviço. Foi feito contato com um membro interno do Delicious para se consultar sobre esta possibilidade, porém, sem sucesso, pois o membro considerou que esta seria uma forma de *spam*.

3.3 Aplicabilidade da folksonomia

Esta etapa da pesquisa teve cunho prático, exploratório e descritivo. Foi feita uma análise descritiva de diferentes serviços de informação na web (sítios) que permitem a etiquetagem de objetos informacionais distintos, gerando folksonomias. A metodologia para essa etapa consistiu no levantamento dos sítios que utilizam folksonomia com posterior seleção dos sítios analisados, que teve por base o critério da tipologia documental, no caso, o tipo de objeto informacional que é foco do serviço. Esse critério possibilitou a constatação da aplicabilidade ampla da etiquetagem e folksonomia, identificando o quarto elemento de validação do sistema.

A partir da escolha dos sítios que seriam estudados, o procedimento consistiu basicamente em consultar os sítios para identificar como é feito o processo de etiquetagem, quais os principais aspectos, as características específicas e como é gerada a folksonomia nesses ambientes, atingindo-se assim o objetivo específico 4 e mostrando um panorama geral do que vem sendo feito na web com base nessa proposta de organização.

² Google Docs™, <<http://docs.google.com/>>

Capítulo 4

4 Etiquetagem e indexação: um paralelo

Conforme visto anteriormente na revisão de literatura, indexação é o processo de representação do conteúdo temático dos objetos informacionais para permitir sua recuperação posterior. Este processo guarda relações estreitas com a etiquetagem dos recursos da web: a atribuição de etiquetas pode ser considerada uma forma de indexação, já que partem do mesmo processo e objetivam o mesmo propósito. O que ocorre na web, porém, é um processo coletivo, pois há várias pessoas “indexando” o mesmo objeto informacional. Essa proposta já havia sido feita por alguns autores, muito antes do surgimento dos sistemas de etiquetagem da web, para a indexação de imagens e de obras de ficção.

A indexação colaborativa está na vertente dos processos de indexação orientados ao usuário. Lancaster (2004) comenta que métodos colaborativos ou democráticos de indexação são recomendados para o caso da indexação de imagens, e que são mais viáveis em ambientes de bibliotecas digitais, em que o usuário pode interagir diretamente com os registros. É o que ocorre com a etiquetagem na web, ou seja, os objetos informacionais estão disponíveis para serem marcados com etiquetas, funcionando da mesma forma que o processo de indexação colaborativa.

Para Wal, como discutido anteriormente na revisão de literatura, o processo que gera a folksonomia é coletivo, e não colaborativo. A denominação indexação colaborativa ou democrática refere-se ao fato dos registros não mais serem indexados somente por bibliotecários e profissionais da informação, mas também pelos usuários, que estariam colaborando nesse processo e o tornando mais democrático. Dessa forma, para o processo de indexação, não se faz necessário o preciosismo terminológico proposto por Wal. Sendo assim, denomina-se aqui a indexação como colaborativa pois esse termo já é de uso comum na terminologia da área de Ciência da Informação.

Segundo Voss (2007), a popularidade da etiquetagem na web fez com que ressurgisse o interesse pela indexação manual. Neste capítulo serão abordados os tópicos pertinentes à análise da indexação que permitam delinear um paralelo com o processo de etiquetagem: indexação como processo cognitivo, indexação centrada no usuário, indexação de obras de ficção, indexação de imagens e indexação colaborativa ou democrática.

4.1 Indexação como processo cognitivo

A indexação envolve processos cognitivos, pois lida com conceitos, análise temática, categorização e atribuição de termos. Relata-se na literatura que o processo de indexação exige uma combinação de precisão na leitura, entendimento, categorização, armazenamento e concentração e, por isso, indexar é um trabalho cansativo que envolve uma atividade cerebral muito grande. (ABBOTT; CALVERT, 2007).

Segundo Farrow (1995), a emergência da psicologia cognitiva nos anos 1970 proporcionou o desenvolvimento de um campo de pesquisa que pode ser aplicado para auxiliar na compreensão do processo de indexação. O autor aborda essa temática em alguns de seus artigos e delinea o processo de indexação com base na psicologia cognitiva (FARROW, 1991; 1995). Farrow afirma que o primeiro passo para modelar um processo cognitivo é a descrição das tarefas, e identifica três estágios no processo de indexação: exame do documento e estabelecimento do assunto, identificação dos principais conceitos presentes no assunto e expressão dos conceitos selecionados com termos da linguagem de indexação. No processo de leitura do texto, os autores concordam que o ideal seria uma leitura atenta e aprofundada, porém, isso seria inviável em termos práticos. Dessa maneira, o que costuma ser feito é o chamado *scanning*, ou seja, uma leitura rápida ou passada de olho pelo texto para identificação do assunto e busca de conceitos. Esse processo é tanto conceitual (*top-down*) quanto perceptivo (*bottom-up*). No primeiro caso, o indexador busca identificar, com base em seu conhecimento prévio, categorias maiores em que se encontram os assuntos do texto. Já no segundo, ocorre a identificação de assuntos com base somente no próprio texto. Para Farrow (1995), a indexação será mais efetiva quanto maior for o conhecimento do indexador sobre o assunto indexado.

Frohmann (1990) também segue essa abordagem e acredita haver regras implícitas no modelo mental humano que guiam o indexador:

A figura do mentalismo aplica-se facilmente à indexação. De acordo com ele, a operação intelectual fundamental da indexação é, em princípio, explicável por regras internas e conhecimento tácito que geram uma frase de indexação de um dado texto. Parece então que devem haver algumas regras guiando a atividade mental dos indexadores, de outra maneira seria impossível explicar *como* eles são capazes de escrever frases de indexação para um texto. O problema é descobrir a forma precisa dessas regras. (FROHMANN, 1990, p. 84, tradução nossa).

No processo de indexação, a escolha do termo para designar o assunto é essencial. Segundo Coates (2002, p. 15, tradução nossa) “a seleção do termo consiste no processo de entendimento da informação do texto e na então articulação dessa informação na forma de cabeçalhos principais e sub-entradas.” Para que essa seleção seja feita da melhor forma, algumas habilidades cognitivas são fundamentais. Para Coates (2002), uma boa seleção de termo depende de habilidades cognitivas como compreensão de texto, capacidade de conceitualização e habilidade

de articular conceitos. Essas habilidades cognitivas não podem ser ensinadas, e melhorá-las não está no escopo de um curso de indexação. Ou seja, a atribuição de termos, ou etapa da tradução da indexação, é um processo mental individual que não pode ser aprendido pelas pessoas. A autora conclui que o ensino de seleção de termo, para ser bem sucedido, requer uma certa habilidade cognitiva do aluno (COATES, 2002, p. 17, tradução nossa).

Além de todo esse processo cognitivo em que o indexador deve estar atento ao documento, há ainda que se pensar que o trabalho de indexação tem por objetivo a recuperação dos documentos pelos usuários da informação. O indexador deve então, além de descrever com precisão o documento, ter em mente sempre que esse processo visa atender a necessidades de informação do usuário. A indexação deve então ter esse foco, como será visto na seção a seguir.

4.2 Indexação centrada no usuário

A abordagem da indexação centrada no usuário contrapõe-se à indexação orientada ao documento e parte do princípio de que a indexação deve ser feita com base nos pedidos dos usuários, ou seja, os descritores de assunto devem coincidir com os termos de busca do usuário no momento da recuperação.

Fidel (1994) comenta que a melhor forma de poupar tempo no momento da recuperação de um documento é antecipar os pedidos dos usuários e checar cada documento no momento da entrada em uma base de dados com uma lista de pedidos anteriores. “Quando uma coincidência ocorre (ou seja, um documento é provável de satisfazer um pedido da lista), o documento foi indexado de acordo com o pedido” (FIDEL, 1994, p. 574, tradução nossa).

Hjørland (2001) também concorda que a indexação deve ser orientada ao pedido do usuário. Para o autor, o assunto de um documento está próximo da questão para a qual este deve fornecer uma resposta:

Pelo fato de que qualquer documento pode, em princípio, fornecer respostas a uma infinidade de questões, a análise de assuntos deve estabelecer prioridades baseadas em grupos específicos de usuários atendidos. [...] O assunto de um documento é então relativo ao objetivo de um serviço de informação específico. (HJØRLAND, 2001, p. 776, tradução nossa).

O autor define assunto como “os potenciais epistemológicos ou informativos dos documentos” e afirma que “a melhor análise temática é a que faz o melhor prognóstico quanto ao uso futuro do documento.” (HJØRLAND, 2001, p. 776, tradução nossa).

Da mesma forma, Bates (1998) trata da indexação centrada no usuário. Para a autora, tal abordagem é vista como um desafio para o indexador, que deve “tentar antecipar quais os termos que as pessoas com lacunas de informação de vários tipos procurariam nos casos em que o

registro de que dispõem, de fato, fica a meio caminho de satisfazer a necessidade de informação do usuário.” A indexação centrada no usuário deve responder “quais os tipos de necessidade de informação que as pessoas teriam e as levariam a querer informações que o registro de fato contém?” (BATES, 1998, p. 1187, tradução nossa). Farrow (1995) também enfatiza esse aspecto, afirmando que o processo de indexação precisa levar em conta as exigências dos usuários, e que a indexação é baseada em seus pedidos (FARROW, 1995, p. 246, tradução nossa).

Esse tipo de indexação é algo complicado para o indexador, pois este possui o documento e deveria apenas fazer o registro de algo conhecido, ou seja, da informação que tem em mãos, caso a indexação fosse orientada ao documento. Já no caso do usuário, este possui apenas dúvidas, incertezas, necessidades de informação, e sua tarefa é descrever algo que possivelmente desconhece. Assim, não há uma relação simétrica entre as partes, no processo de indexação e recuperação, do ponto de vista dos agentes envolvidos: “uma lacuna de experiência é formada entre o usuário e o indexador, o que quase garante algumas falsas coincidências entre os termos de busca do usuário e os termos de indexação do registro.” (BATES, 1998, p. 1187, tradução nossa).

Mai (2001) também ressalta a dificuldade de se indexar todo o conteúdo que possa ser de interesse do usuário. O autor afirma que “seria quase impossível prever com precisão quais das muitas possíveis idéias e significados que estivessem associados ao documento seriam especificamente úteis para os usuários ou dariam ao documento alguma utilidade duradoura” (MAI, 2001, p. 606, tradução nossa).

Lancaster (2004, p. 11) afirma que para certos tipos de materiais, como imagens e obras de ficção, a indexação orientada ao usuário é até mais importante do que para artigos científicos e livros técnicos. Ou seja, quando a subjetividade é maior. Outra observação importante feita pelo autor é que, para esse tipo de indexação, o indexador deve estar “inteiramente a par dos interesses da comunidade atendida e das necessidades de informação de seus membros”, e que deve atentar para o fato de que “em relação a determinado acervo de documentos e determinado grupo de usuários, qualquer conjunto ideal de termos de indexação será ideal *somente em determinado ponto do tempo*.” (LANCASTER, 2004, p. 12). Dessa forma, esse tipo de indexação demandaria atualização constante para acompanhar as necessidades e demandas dos usuários, o que não é viável em termos práticos nos serviços de informação tradicionais.

4.3 Indexação de obras de ficção

Pouca importância costuma ser dada à indexação de obras de ficção nas bibliotecas e centros de informação tradicionais, e sua base teórica também não foi muito trabalhada pelos autores da área. Segundo Pejtersen e Austin (1983, p. 203, tradução nossa) muito foi gasto na melhoria dos métodos e ferramentas de recuperação para não-ficção, mas quando se trata de ficção, os

bibliotecários ainda se baseiam nos recursos subjetivos e incertos da própria memória para guiar os usuários. Apesar disso, acredita-se haver uma relevância significativa neste tipo de trabalho: as obras ficcionais representam cerca de 50% do acervo das bibliotecas públicas e respondem por mais da metade de seus empréstimos (PEJTERSEN, 1992 *apud* LANCASTER, 2004, p. 204).

A indexação de obras de ficção pode ser de dois tipos: indexação de assuntos em nível genérico e indexação para a criação de um índice de final de livro. A indexação de ficção pode se basear em aspectos relacionados a gênero, cronologia e geográfica, como num esquema de classificação tradicional, abarcando gênero literário, ano de publicação e local de publicação ou de origem do autor. Esse tipo de indexação é mais objetiva e simplificada, perpassando aspectos mais relacionados à catalogação, e não incluem a descrição de assunto do conteúdo da ficção. Já a indexação dos assuntos das obras de ficção propriamente ditos apresenta questões de interpretação e caráter subjetivo. Neste tipo de indexação, alguns autores fazem uma divisão entre elementos denotativos ou factuais – como cenários, personagens e elementos factuais do roteiro – e conotativos ou imaginativos – que consiste de elementos interpretativos da obra de ficção, ou ainda em factuais e expressivos, em que o factual questiona *de que trata* enquanto o expressivo se refere a *sobre o que trata* determinada obra de ficção. (RANTA, 1991; SHATFORD, 1986 *apud* SAARTI, 1999, tradução nossa)

Lancaster (2004) dedica um capítulo de sua obra a esse tema, ressaltando as diferenças entre a indexação tradicional (material técnico) e a indexação de obras de ficção. O autor explica que dizer que um documento “é sobre” alguma coisa de forma técnica é diferente de identificar o mesmo para uma ficção, devido aos diferentes objetivos que possuem:

Os dois tipos [ficção e não-ficção] são criados com objetivos diferentes: o primeiro, fundamentalmente, para entreter ou suscitar emoções, o segundo, fundamentalmente, para veicular informações. [...] Se atribuirmos o termo AGRICULTURA a um filme documentário ou a um artigo de periódico sobre este assunto, implicamos que estes itens veiculam certas informações sobre agricultura, e que os usuários de um índice procurariam por este termo porque querem aprender algo sobre este assunto específico. Por outro lado, se alguém procura, sob o termo AGRICULTURA, num índice de obras de ficção, por exemplo, filmes, com certeza não o faz com o propósito de encontrar informações sobre agricultura. (LANCASTER, 2004, p. 200).

Outra grande diferença ressaltada por Lancaster, conforme já comentado, é o alto grau de subjetividade, já que as obras literárias e cinematográficas têm um caráter mais interpretativo que outras. Além disso, seu campo é essencialmente aberto, ou seja, tem um caráter mais geral para a representação temática. Isso implica na etapa da tradução para a linguagem de indexação: “o vocabulário usado na indexação deve também ser aberto, pois constantemente estão sendo feitos filmes e romances que tratam de personalidades, acontecimentos e lugares que não foram anteriormente abordados por esses meios” (LANCASTER, 2004, p. 202).

Pejtersen e Austin (1983; 1984) também comentam o problema da subjetividade, afirmando que isso poderia facilmente afetar a análise documental e a indexação do romance com

a mesma extensão com que afeta os pedidos dos usuários. As autoras desenvolveram um sistema de indexação para romances e outras obras de ficção relativamente complexo, o que demandaria um grande trabalho intelectual do indexador.

Saarti (1999) acredita que houve um aumento de interesse na indexação de obras de ficção no final dos anos 1990 por uma necessidade de recuperação de material ficcional e pela criação de sistemas de recuperação para ficção devido ao crescimento da automatização. Ainda assim, a maioria do material não indexado continua sendo o de ficção. O autor ressalta a importância da indexação dessas obras, afirmando que, sem esse tratamento de classificação e indexação, torna-se difícil recuperar material, e outra razão é o fato de que busca livre em texto completo não traz resultados eficientes para material de ficção (SAARTI, 1999, tradução nossa).

4.4 Indexação de imagens

Além dos documentos tradicionais em texto, outro objeto informacional importante e que traz peculiaridades para a indexação é o recurso visual. Dois tipos de indexação de imagens costumam ser apontados na literatura: indexação baseada em conceitos (ou de nível alto) e indexação baseada em conteúdo (ou de nível baixo). A primeira se refere à representação textual de uma imagem ou sua tradução em palavras, ou seja, *do que trata* a imagem. Já a segunda se refere aos aspectos intrínsecos da imagem como cor, textura e forma (LANCASTER, 2004, p. 214). Assim, a indexação de imagens é bastante importante, pois não há outro meio ou forma automática de se fazer recuperação baseada em conteúdo temático ou em conceitos: “Diferentemente dos textos, imagens não descrevem elas próprias. Se temos que identificar uma figura, tudo, com exceção de cor, forma e textura, deve ser descrito em palavras” (BERISTEIN, 1999, tradução nossa).

Layne (1994, p. 583) faz uma divisão diferente, pois, para a autora, a indexação de imagens deve possibilitar duas coisas: o acesso a imagens baseado nos atributos dessas imagens e o acesso a grupos úteis de imagens, e não somente acesso a imagens individualmente. Os atributos a que a autora se refere podem ser de quatro tipos: bibliográficos (dados relacionados a origem, data, local, título, donos, entre outros), de assunto (se refere à indexação de conceitos), exemplificadores (atributos usados para servir de exemplo, características dos objetos) e de relacionamento (imagens relacionadas a outras imagens). Já o acesso a grupos úteis de imagens refere-se a indexar imagens de forma a criar *clusters* baseados em algum critério de semelhança entre as imagens, que podem ser de acordo com atributos da imagem, com atributos do que a imagem representa ou com a proveniência da imagem. Dessa forma, pode-se recuperar imagens baseadas em alguma característica comum e entre elas escolher a mais apropriada para determinado fim.

As formas automáticas de indexação de imagens somente permitem a recuperação dos elementos do chamado nível baixo, ou seja, são abordagens baseadas no conteúdo. Esses sistemas trabalham com similaridade de formas, espectro de cores e texturas, e, mesmo assim, ainda não atingiram um nível satisfatório de funcionamento. Além disso, é pouco provável que os usuários façam buscas com esse tipo de critério: “É importante reconhecer, contudo, que a maioria dos usuários de bases de dados de imagens provavelmente não fará buscas sobre aspectos mais abstratos como cor, forma e textura, embora possam empregá-los para limitar uma busca.” (LANCASTER, 2004, p. 220).

Lancaster (2004, p. 223-228) cita vários estudos no campo da recuperação de imagens com diferentes propostas e abordagens. Alguns autores utilizam a técnica da retroalimentação como forma de indexar uma base de dados de imagens, outros fazem buscas por similaridade. Para imagens inseridas em conteúdo textual, como em um jornal, por exemplo, é possível utilizar o próprio texto da legenda e do conteúdo próximo da imagem como matéria indexável para a imagem. De forma semelhante, há propostas de utilização de legendas de vídeos para a indexação das cenas, no caso de imagens em movimento. Outra proposta para a indexação de conceitos seria a utilização dos elementos de nível baixo para agrupar imagens semelhantes, e, a partir daí, os conjuntos de imagens receberiam os mesmos termos que teriam sido previamente atribuídos por meio de indexação humana.

Apesar dos esforços que têm sido despendidos nessa área, acredita-se que a eficácia desses métodos ainda não tenha sido atingida de forma satisfatória. Uma abordagem mista, que reúna características da indexação de conteúdo (automática) e indexação de conceitos (humana), parece ser a melhor opção para a recuperação efetiva dos recursos visuais:

Em conclusão, o método ideal de recuperação de imagens talvez seja aquele que combine acesso convencional por meio de texto (termos de indexação ou narrativa descritiva) com o cotejo de imagens. Assim, uma busca com palavras (batalha, ataque, luta) recuperaria uma imagem de determinado tipo de cena e esta, por sua vez, poderia ser usada como insumo para localizar outras iguais. (LANCASTER, 2004, p. 237).

Outra questão complexa em relação à indexação de imagens por conceitos é seu caráter subjetivo. As imagens são difíceis de se descrever pois possuem múltiplos significados e podem ser interpretadas de maneiras diferentes por pessoas diferentes. Aspectos culturais, valores e crenças fazem com que uma imagem tenha significados totalmente diferentes, dependendo de quem analisa. Além disso, uma imagem é ao mesmo tempo genérica e específica: “uma imagem da ponte do Brooklyn pode, para determinada pessoa, ser útil por se tratar daquela ponte especificamente, ou, para outra pessoa, por ser de uma ponte qualquer, num sentido genérico.” (LAYNE, 1994, p. 584, tradução nossa).

Segundo Berinstein (1999) as imagens contêm mais significados diversos que os textos porque são tanto *de* algo como *sobre* algo: “Uma foto no jornal de um rapaz jovem e bonito não é nada até que você descubra que o rapaz jovem e bonito é um assassino em série brutal. A foto

é **de** Joe Blow, mas é **sobre** um assassino em série procurado pelo FBI” (BERINSTEIN, 1999, tradução nossa, grifo nosso). Layne também enfatiza essa distinção: “uma imagem pode ser de um leão, mas sobre orgulho ou uma imagem de uma pessoa chorando pode ser sobre tristeza” e afirma que dizer que uma imagem é *de* algo é talvez mais concreto e objetivo enquanto dizer que uma imagem é *sobre* algo, mais abstrato e subjetivo (LAYNE, 1994, p. 584, tradução nossa).

4.5 Indexação colaborativa ou democrática

Conforme visto, a indexação de obras de ficção e a indexação de imagens apresentam questões mais complexas que a indexação de materiais tradicionais como textos técnicos e artigos de periódicos científicos, pois enfrentam o problema da subjetividade. Devido a essa questão, alguns autores fizeram propostas de diferentes abordagens de indexação, nas quais o usuário destaca-se como colaborador na indexação desse tipo de recurso em que o fator interpretativo está presente em maior grau. A indexação colaborativa ou democrática é uma abordagem centrada no usuário e está relacionada com a idéia dos usuários colaborarem com seus próprios termos para indexar determinado recurso: “O princípio da indexação democrática é baseado no conceito de que os indivíduos terão sua interpretação própria e potencialmente diferente de uma imagem. A diferença pode ser manifestada em diferentes focos em partes da imagem e em uso de diferentes vocabulários para descrever a imagem.” (HIDDERLEY; RAFFERTY, 1997, tradução nossa).

Brown e Hidderley (1996) apresentam o projeto de um sistema democrático de indexação de imagens:

Indexação democrática ou baseada no usuário tem a pretensão de ser utilizada em sistemas dinâmicos que permitem ao usuário contribuir com o processo de recuperação, indo além do modelo tradicional que assume que os mecanismos de recuperação são construídos somente por bibliotecários/indexadores. (BROWN; HIDDERLEY, 1996, p. 109, tradução nossa).

Os autores apontam como principais problemas da indexação de imagens o acesso físico ao documento, como no caso de coleções espalhadas por uma grande área geográfica e quando o material é frágil (o que não ocorre quando todo o acervo está em meio digital), e a dificuldade em antecipar os pedidos dos usuários para a recuperação, já que estes têm propósitos diferentes para o uso do recurso visual. O problema da subjetividade também é destacado: “Essencialmente, o significado em imagens é difícil pois está muito relacionado com a interpretação individual.” (BROWN; HIDDERLEY, 1996, p. 108, tradução nossa). Nesse projeto houve ainda a previsão de elaboração de um tesouro a partir dos termos utilizados pelos usuários, de forma que os termos estariam sempre atualizados pois seriam alimentados de acordo com seu uso pelas pessoas que colaboram com o sistema.

Outra abordagem para o uso de indexação democrática foi feita pelos mesmos autores anteriores também para a indexação de obras de ficção. Eles acreditam que as várias leituras de uma obra de ficção devem ser exploradas e incorporadas em um sistema de indexação, e a abordagem democrática permite a criação desse tipo de sistema:

A indexação democrática pode ser utilizada para textos ficcionais como uma ajuda para a interpretação. Poderia também ser usada em bibliotecas ou sistemas de bibliotecas para indexar uma coleção de ficção, tornando o acesso mais fácil e mais interessante, pois poderia identificar respostas emocionais bem como informações factuais. [...] Nosso sistema usaria este princípio, mas além de resenhas, o leitor seria encorajado a atribuir termos de indexação ou descritores para ajudar os outros em suas buscas e identificação de ficção legível. Indexação democrática torna isso possível para um sistema a ser construído, que poderia ser utilizado para identificar o que a maioria dos leitores acreditam ser um livro feliz, ou romântico, ou que contém personagens notadamente malvados, temperamentais ou magníficos. Como resultado, a escolha e o acesso do usuário seriam enriquecidos. (HIDDERLEY; RAFFERTY, 1997, p. 105, tradução nossa).

Assim, para recursos informacionais em que prevalece o caráter subjetivo, a proposta de indexação colaborativa ou democrática surge como alternativa para ampliação e melhora do acesso a esses conteúdos. Hidderley e Rafferty sugerem também que a indexação democrática poderia ser utilizada para a recuperação de filmes de ficção, que estão abertos a múltiplas leituras.

Outra aplicação da indexação colaborativa que pode ser citada é a de um projeto realizado em 2003 para a indexação dos jornais comunitários de Vashon Island (Washington). O projeto consistia na indexação de oito jornais semanais dessa comunidade publicados no último século (1907 até os dias atuais), ou seja, um grande volume de informação a ser indexada. Para isso, foi criado um protocolo de indexação com formato padrão de entrada dos dados e uso de um vocabulário controlado, e, com isso, os bibliotecários ofereceram treinamento para voluntários na indexação dos jornais, criando-se uma forma colaborativa de indexação. (WEAVER, 2006).

As principais aplicações da indexação colaborativa ou democrática seriam então para imagens (estáticas ou em movimento), obras de ficção, e quando há um grande volume de informação a ser tratado, o que demanda um grande número de pessoas e nem sempre pode-se contar com especialistas da área. Tais usos têm sido aplicados no contexto da Internet, com a chamada etiquetagem, que será tratada a seguir.

4.6 Etiquetagem

O processo de etiquetagem consiste, assim como a indexação, na atribuição de termos a objetos informacionais. Nesse caso, os termos são chamados de etiquetas, que podem ser atribuídas por qualquer pessoa ou usuário da informação. Conforme já apontado anteriormente,

esta prática vem sendo bastante utilizada na web, fazendo parte da abordagem colaborativa característica da Web 2.0.

Voss (2007, p. 3) afirma que a etiquetagem apenas perpassa alguns aspectos da indexação de assunto, porém, acredita-se que a semelhança é um pouco mais significativa do que o sugerido pelo autor. Após o exame dos tópicos anteriores referentes à indexação, é possível identificar características comuns entre as práticas. O processo básico é o mesmo: atribuição de termos que representem conceitos presentes nos objetos informacionais.

Com base na abordagem cognitiva, Farrow (1991) explica que a informação é organizada no cérebro dentro de redes semânticas. “Quanto um conceito é ativado, conceitos relacionados com a mesma rede semântica também são ativados, e o grau de ativação depende da distância semântica dos conceitos ativados” (FARROW, 1991, p. 244, tradução nossa). De forma semelhante, Sinha (2005) descreve o processo cognitivo da etiquetagem. A autora explica que o primeiro estágio corresponde a comparação de similaridade entre um item e conceitos candidatos. Imediatamente, conceitos semânticos relacionados são ativados, e, a partir daí, os conceitos são escritos em forma de etiquetas (ou termos) que os representam.

A etiquetagem é também um processo mental, porém, seu custo cognitivo é menor, pois se dá de forma livre e não envolve o processo de decisão que está presente em uma categorização formal. Não há o comprometimento com um esquema de representação pré-definido, o que elimina o gasto mental conceitual da determinação de uma categoria maior em que deve se alocar o conceito. Acredita-se que na etiquetagem, o processo perceptivo se sobressai, ou seja, ocorre na abordagem *botton-up*. Não existe a preocupação com um sistema maior em que as etiquetas devem estar, pois isso acontece naturalmente com as folksonomias que são geradas posteriormente. O processo cognitivo da etiquetagem será analisado com mais detalhes no capítulo 5.

A importância do usuário para a etiquetagem é fundamental, já que é ele o responsável por sua criação. Dessa forma, a etiquetagem deve ser enquadrada na abordagem de indexação baseada no usuário. Os usuários colocam etiquetas que são de seu uso comum, e serão utilizadas posteriormente no momento da recuperação dos objetos informacionais marcados. Como existe a formação de comunidades na web em torno de assuntos ou interesses comuns, as etiquetas mais frequentes serão destacadas nas folksonomias e vários usuários poderão de beneficiar desse tipo de indexação no momento da recuperação:

A informação torna-se algo subjetivo (mas não necessariamente de forma individual: uma comunidade pode compartilhar o mesmo ponto de vista subjetivo). O mesmo objeto carrega então diferentes informações para diferentes usuários. Isso também implica que diferentes sistemas de informação devem interpretar e representar documentos para diferentes tarefas e comunidades de usuários. (HJØRLAND, 2001, p. 777, tradução nossa).

A etiquetagem tem também um grande valor quando se trata da indexação de obras de ficção, pois, conforme citado anteriormente, a indexação desse tipo de material é extremamente subjetiva. Segundo Saarti (1999) o maior problema da indexação desse tipo de obra é depender

do ponto de vista do indexador, baseado na interpretação: “Isso é especialmente verdadeiro no caso de experiências emocionais, que não pertencem à obra mas sim ao leitor” (SAARTI, 1999, p. 87). Nesses casos, a indexação feita pelo usuário teria um grande valor, já que as etiquetas mais utilizadas em determinado conteúdo de ficção seriam destacadas na folksonomia, o que funcionaria dentro das comunidades virtuais, corroborando a citação anterior de Hjørland. Outra vantagem de se indexar obras de ficção por meio da etiquetagem é que os usuários já vivenciaram a experiência da obra, seja o livro ou o filme, ele já deve ter sido apreciado pelo usuário de forma completa. No caso de uma indexação profissional, não é viável a leitura completa das obras de ficção, por exemplo.

Da mesma forma, o uso da etiquetagem para a indexação de imagens é bem vindo, e vários autores afirmam que essa é uma tendência para a indexação dos recursos visuais na web. O Flickr é um bom exemplo disso: o sistema foi um dos primeiros a utilizar a etiquetagem e a tornou conhecida. O uso de etiquetas para imagens é de grande importância já que não há outra maneira de se recuperar esse tipo de recurso em termos conceituais. Layne (1994, p. 583) afirma que a indexação de imagens deve fornecer acesso a grupos úteis de imagens, o que pode ser gerado automaticamente pelos sistemas de etiquetagem, que criam *clusters* em torno de etiquetas mais utilizadas formando um grupo de imagens relacionadas. Isso permite que o usuário escolha entre as várias imagens semelhantes recuperadas, a que melhor representa seu interesse.

Pelo exposto, percebe-se que a etiquetagem é a realização, na web, do que os autores sugeriram nos anos 1990, ou seja, a indexação colaborativa ou democrática. Rafferty e Hilderley retomam essa temática apontando a indexação democrática como uma abordagem alternativa para a recuperação de imagens baseada em conceitos, usando como exemplo também o Flickr. Os autores acreditam que o significado do documento não é ontológico, mas que muda com o tempo, e a abordagem democrática permite uma atualização do significado, com contingência histórica da visão pública anexada ao documento indexado (RAFFERTY; HIDERLEY, 2007).

O ambiente web permite que os usuários participem de forma mais direta do processo informacional, tanto em sua produção, quanto em seu tratamento e disseminação. Essa aproximação é importante pois a etiquetagem traz a garantia do usuário em termos de recuperação. Segundo Bates (1998, p. 1187, tradução nossa) “não seria surpreendente, então, se o usuário e o indexador utilizarem terminologia diferente para descrever o registro, e mais genericamente, conceitualizarem a natureza e o caráter do registro de forma diferente.” Já na abordagem democrática, a autoridade é determinada por um acordo entre os usuários: essa garantia vem da interpretação construtiva dos usuários (RAFFERTY; HIDERLEY, 2007).

Dessa forma, deve-se supor que a etiquetagem teria uma maior eficiência na recuperação pois a terminologia usada seria a do usuário. Além disso, retomando a citação de Lancaster de que “um conjunto ideal de termos de indexação será ideal somente em determinado ponto do tempo”, encontra-se na etiquetagem uma atualização automática das etiquetas mais frequentes, que são as que se destacam nas folksonomias. Identifica-se então mais uma característica desejada para a indexação na etiquetagem.

Capítulo 5

5 Folksonomia: esquema de representação do conhecimento?

Este capítulo foi baseado em revisão bibliográfica e faz uma análise das abordagens em Organização do conhecimento propostas pelo pesquisador dinamarquês Birger Hjørland, identificando as principais características de cada uma e apontando correlações com a folksonomia.

5.1 Abordagens em Organização do conhecimento

A partir da sistematização proposta por Hjørland (2007a), pode-se traçar um pequeno histórico das abordagens em Organização do conhecimento. No final do século XIX, surgiram os primeiros sistemas de classificação bibliográfica, os enumerativos, como a CDD, proposta do Melvin Dewey em 1876. Hjørland (2007a) considera as classificações enumerativas como abordagens tradicionais em Organização do conhecimento.

Já no século XX, surgiram outras abordagens, como o modelo facetado (analítico-sintético), bastante difundido por Ranganathan, contrapondo os princípios tradicionais da classificação enumerativa. A partir dos anos 1950, a introdução dos computadores trouxe uma grande mudança: o uso das máquinas para a recuperação da informação. Hjørland (2007d) considera a recuperação da informação como uma nova abordagem em Organização do conhecimento. A criação do *Science Citation Index*, nos anos 1960, trouxe mais uma abordagem, a abordagem bibliométrica. Já nos anos 1970 e 1980, a ênfase dada ao usuário gerou o desenvolvimento de abordagens cognitivas e baseadas nos usuários. Os anos 1990 foram marcados pela influência das novas tecnologias, como a possibilidade de buscas em texto completo e do modelo de Web Semântica. Essa fase é considerada por Hjørland (2007a) como uma continuação da abordagem em recuperação da informação. Ainda nos anos 1990, cresce o interesse em abordagens sociais, como a análise de domínio. Vale destacar ainda as tendências atuais em Organização do conhecimento, que encontram bases na Arquitetura da Informação.

Entre as diversas abordagens, propõe-se agora explorar algumas específicas, nas quais, acredita-se, a folksonomia pode encontrar bases para ser analisada como um esquema de representação do conhecimento.

5.1.1 Abordagem baseada no usuário

A abordagem baseada no usuário surge nos anos 1970, auge dos estudos de uso e usuários. Segundo Hjørland (2007g), os esquemas de representação do conhecimento e seu processo de elaboração podem envolver, de uma forma ou de outra, dados fornecidos pelos usuários ou sobre os usuários. É importante fazer uma distinção básica entre abordagens baseadas em usuários e abordagens amigáveis (*user-friendly*): enquanto na primeira o esquema de representação é gerado a partir de dados fornecidos pelos usuários ou sobre eles, a segunda está relacionada com facilidade de uso, interatividade e capacidade de intuição do usuário no sistema. Nesse sentido, considera-se, por exemplo, que os sistemas de classificação bibliográfica como a CDU não são abordagens amigáveis, já que as notações que representam os assuntos não são intuitivas e podem confundir os usuários.

Restringindo-se a abordagem baseada no usuário apenas a “abordagens em Organização do conhecimento que são principalmente baseadas nos dados obtidos dos usuários [...]” (HJØRLAND, 2007g, tradução nossa), pode-se encontrar uma clara relação com a folksonomia, já que o processo de etiquetagem, que gera os termos do sistema, é totalmente realizado pelo usuário.

Outro aspecto considerado básico por Hjørland (2007g) nas abordagens orientadas ao usuário é que estas devem ter como foco a garantia do usuário em detrimento da garantia literária. Deve-se, então privilegiar, para a inclusão de termos no sistema, a terminologia do usuário e não a dos documentos/objetos informacionais. Neste caso, a folksonomia também pode ser considerada: o próprio usuário adiciona o termo, garantindo seu uso posterior na recuperação da informação. Pode-se questionar o processo de escolha dos termos pelo usuário, já que este pode utilizar termos do próprio documento, como o título, por exemplo. Mas supõe-se que o usuário só fará desta forma se o termo do documento coincide com aquele de seu uso habitual.

Desta forma, entende-se que a folksonomia encontra bases nesta abordagem para ser considerada um esquema de representação do conhecimento, já que possui duas das principais características da abordagem baseada no usuário.

5.1.2 Abordagem cognitiva

A abordagem cognitiva em Biblioteconomia e Ciência da Informação surgiu como paradigma para a Organização do conhecimento a partir de 1992, quando a Segunda Conferência Internacional da International Society for Knowledge Organization (ISKO) teve essa abordagem como tema. Hjørland (2007b) faz uma crítica ao editorial escrito por Dalhberg sobre a visão

cognitiva em Organização do conhecimento, pois o ensaio traz a visão cognitiva como uma tautologia: todas as abordagens em Organização do conhecimento devem se preocupar, de uma forma ou de outra, com questões conceituais e cognitivas. Segundo Hjørland (2007b), muito pouco é realmente dito sobre a visão cognitiva em Organização do conhecimento nesse artigo. O autor afirma ainda que, apesar de algumas discussões sobre o assunto, não há uma avaliação sobre o que a visão cognitiva pode trazer para o campo da Organização do conhecimento. Aparentemente, não há um consenso entre os autores da área.

A visão cognitiva é também relacionada ao comportamento de busca do usuário: “Paradigmas cognitivos indicam o comportamento de busca de conhecimento de indivíduos e grupos de indivíduos.” (NEELAMEGHAN *et al.* 1992 *apud* HJØRLAND, 2007b, tradução nossa). É sugerido ainda que a análise de tal comportamento de busca possa fornecer diretrizes para a organização da informação em base de dados e ambientes similares.

Uma conclusão muito citada sobre o significado da visão cognitiva é a de De Mai (1980, *apud* HJØRLAND, 2007b, tradução nossa): “um sistema de recuperação de informação deve refletir em suas operações, de uma forma ou de outra, o mundo cognitivo do usuário.” Esse *mundo cognitivo do usuário* não deve ser melhor representado do que se gerado pelo *próprio* usuário. De forma semelhante, Foskett (1980) afirma que “o serviço de informação mais eficaz é aquele que é projetado especificamente para cada usuário, baseado em suas necessidades conhecidas.” Levando-se em conta esses autores, tem-se então uma forte base para a folksonomia – acredita-se que nenhuma outra forma de representação do conhecimento está tão diretamente ligada ao mundo cognitivo e às necessidades do usuário.

Sobre esse aspecto, Hjørland (2007b) afirma ainda que a representação do conhecimento não é somente tentar criar representações tão objetivas quanto possíveis, mas otimizá-las em relação a certas tarefas e valores. Neste caso, a folksonomia também pode ser citada: o usuário descreve o objeto de informação de forma a facilitar uma tarefa futura, além de fazer um julgamento de valor, o que não necessariamente corresponderia à forma mais objetiva.

Em artigo, Sinha (2005) faz uma análise cognitiva de processo de etiquetagem e como seu baixo custo o torna popular. A autora descreve como se dá tal processo, além de comparar a etiquetagem com a categorização, em termos cognitivos. A figura 7 mostra como ocorre o processo de etiquetagem no nível cognitivo. O estágio 0 corresponde à seleção de um item (objeto de informação) que vale a pena ser recuperado num momento posterior. O próximo estágio é um processo de comparação de similaridade entre o item e os conceitos candidatos a virarem etiquetas. Nessa fase, conceitos semânticos relacionados são ativados: tanto conceitos gerais quanto os mais pessoais a que o objeto remete e ainda características físicas podem ser lembradas, segundo estudos da psicologia cognitiva. A partir do momento que esses conceitos são ativados, externalizá-los, escrevendo-os na forma de etiquetas é bastante simples: não há um processo de filtragem ou julgamento nesse estágio, pode-se usar tantas associações quanto

necessário. Sinha (2005) afirma que é dessa forma que o processo de etiquetagem funciona do ponto de vista cognitivo.

Processo cognitivo por trás da etiquetagem



Figura 7 – Processo cognitivo por trás da etiquetagem (adaptado de SINHA, 2005)

Segundo a mesma autora, já o processo de categorização, ilustrado pela figura 8, inclui uma nova fase, que é o processo de decisão. As categorias candidatas devem ser julgadas, e decide-se então qual é a correta para descrever dado item. Trata-se de um processo cognitivo básico: categorizamos as coisas o tempo todo, de forma quase automática.

Processo cognitivo por trás da categorização



Figura 8 – Processo cognitivo por trás da categorização (adaptado de SINHA, 2005)

O estágio que dificulta o processo de categorização, segundo a autora, é por ela denominado “post-activation analysis paralysis”, algo como uma *parada para análise*, que, na figura 8, corresponde ao estágio 2. Uma das dificuldades se relaciona com a questão cultural: falta consenso cultural no ambiente digital, e a categorização é com frequência baseada em conhecimento cultural. Além disso, a autora afirma que, no ambiente digital, a etiquetagem não só fornece a categorização de um objeto, como também aumenta sua recuperabilidade (*findability*). Dessa

forma, deve-se considerar não só a categoria mais adequada, mas também aquela com maior probabilidade de se encontrar o item no momento da busca. Essas duas questões levam a respostas conflitantes e complicam o processo de categorização, pois se deve considerar ainda o esquema de categorização maior em que se insere a informação. (SINHA, 2005).

O estágio de *analysis-paralysis*, segundo a autora, corresponde ao temor de se tomar uma decisão errada ao se categorizar um objeto, o que implicaria na perda do item em termos de recuperabilidade. É justamente essa parte da decisão (escolha da categoria correta) que é eliminada no processo de etiquetagem. Em suma: o custo cognitivo do processo de etiquetagem é bem mais baixo.

É importante lembrar que para a organização do conhecimento o item teórico mais essencial é o fato de que a organização do conhecimento deve ser baseada em unidades do conhecimento – que são os conceitos (DAHLBERG, 1993). Nas folksonomias, as etiquetas (*tags*) representam termos, mas o que é ativado no processo cognitivo visto nas figuras 7 e 8 são os conceitos, ou seja, a estrutura é construída a partir dos conceitos fornecidos pelos usuários.

A partir desta análise do processo cognitivo, conclui-se que a etiquetagem é relativamente simples sob esse ponto de vista. Tal fato deve contribuir para a expansão de folksonomias no ambiente digital. A abordagem cognitiva é, portanto, relevante para a análise da folksonomia.

5.1.3 Organização social do conhecimento

Segundo Hjørland, a organização social do conhecimento se contrapõe à organização intelectual do conhecimento. A primeira se refere à organização do conhecimento – categorização das disciplinas, por exemplo – feita de acordo com interesses institucionais e acadêmicos, enquanto a segunda se refere à organização do conhecimento com bases mais inerentes ao próprio conhecimento, ou seja, à área do saber em questão. Hjørland (2007e) afirma que esta distinção é fundamental para a teoria da organização do conhecimento, e que não está relacionada apenas à organização das ciências, mas do conhecimento em geral.

A organização social não chega a ser uma abordagem de Organização do conhecimento, mas uma vertente em que se podem apoiar várias abordagens, como, por exemplo, a bibliométrica. Nesse sentido, a discussão aqui levará em conta a organização social do conhecimento como aquela feita pelo homem, com base em seus princípios e diretrizes (e gerada por uma determinada comunidade), e não com base no próprio conhecimento do domínio. É, portanto, uma forma ainda mais arbitrária de representação do conhecimento – já que todas são em algum grau.

Posto isso, tem-se então que a folksonomia como representação do conhecimento gerada pelo usuário, ou por comunidades afins, pode-se ser dita como uma forma de organização social

do conhecimento: arbitrária, baseada nos princípios dos próprios usuários e compartilhada num meio social determinado. Tal visão pode ser analisada a partir do artigo de Sinha (2006), que após uma análise cognitiva do processo de etiquetagem, fez uma análise social do mesmo.

Sinha (2006) afirma que a formação social básica permitida pela etiquetagem está mais para “multidão” que para um verdadeiro grupo: os usuários não se conhecem e não têm nenhum tipo de contato uns com os outros. O processo de etiquetagem levaria então a uma “sabedoria das multidões” – conceito defendido por James Surowieck. Para Sinha (2006), o processo de etiquetagem estaria livre dos aspectos negativos de comportamentos de “multidão”, e os quatro princípios da “sabedoria das multidões” estão presentes: diversidade de opiniões, independência dos membros, descentralização e método de agregar opiniões. A figura 9 ilustra a interação social e a transmissão conceitual com etiquetas.

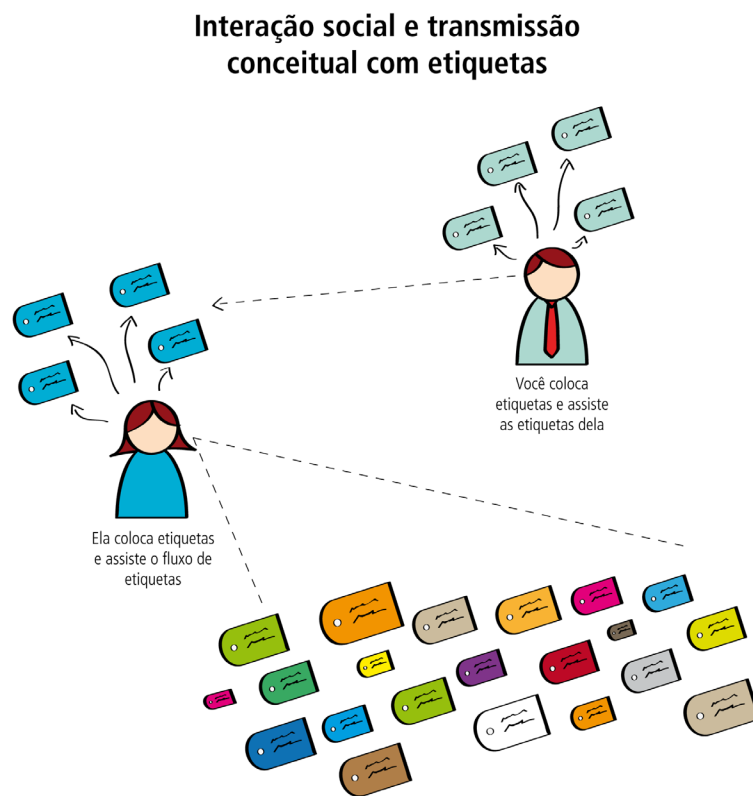


Figura 9 – Interação social e transmissão conceitual de etiquetas (adaptado de SINHA, 2006)

Pode-se ressaltar ainda que na própria denominação do processo que gera a folksonomia está inserido o conceito “social”: classificação social, etiquetagem social, categorização social. Conclui-se então que a folksonomia, se considerada como um esquema de representação do conhecimento, estará inserida nessa vertente da organização social do conhecimento.

5.1.4 Abordagem da recuperação da informação

Entre as abordagens da Organização do conhecimento discutidas por Hjørland (2007d), encontra-se a abordagem da Recuperação da informação (RI). Inicialmente o autor questiona se a RI pode ser caracterizada também como uma abordagem de Organização do conhecimento, já que normalmente ela é considerada como uma sub-disciplina da Biblioteconomia e Ciência da Informação.

A recuperação da informação é baseada no pressuposto fundamental da correspondência entre a questão do usuário e a representação do documento – que é o princípio dos mecanismos de busca. Há algumas fraquezas em relação a esse conceito. Pode-se, por exemplo, considerar relevante identificar co-citações em artigos, independentemente dos documentos serem ou não similares. Neste caso, uma simples “transformação da pergunta” não resolveria (HJØRLAND, 2007d).

Sendo assim, de acordo com Hjørland (2007d), deve-se questionar os resultados obtidos por essa abordagem. Caso contrário, não seria mais necessário o campo de pesquisa em Organização do conhecimento, pois este poderia ser substituído por Recuperação da informação. Por esta razão, o autor considera importante incluir a RI entre as abordagens em Organização do conhecimento, para que se possam identificar assim seus pontos fortes e fracos.

Considerando-se então a Recuperação da informação como uma abordagem em Organização do conhecimento, pode-se identificar como a folksonomia se encaixaria nessa abordagem.

O que se pode considerar da folksonomia para a abordagem de RI é a chamada *etiquetagem implícita*. Trata-se de usar as próprias questões de busca (*queries*) como etiquetas para os documentos recuperados. Morisson (2007) faz uma análise das razões que levam os usuários a etiquetarem. Entre as encontradas pelo autor está a etiquetagem implícita: “os usuários colocam etiquetas incidentalmente enquanto executam outras tarefas de recuperação da informação.” (MORISSON, 2007, p. 14), ou seja, os usuários colocam etiquetas sem nem saber que estão fazendo isso. Tal processo ocorre quando um sistema de informação utiliza o texto da busca e o associa ao documento recuperado escolhido pelo usuário. O problema encontrado nesse método é quando um documento não foi descrito por um termo que é usado pelo usuário para buscá-lo. Assim, a interseção entre pergunta e resultado não ocorrerá, e a etiqueta usada nunca será aplicada a determinado documento. A solução encontrada pelo autor para este caso é adicionar ele mesmo a etiqueta manualmente no sistema.

Como a abordagem da RI se baseia na relação entre a pergunta do usuário e a resposta do sistema, a etiquetagem implícita seria utilizada da mesma forma: a pergunta do usuário seria aproveitada como etiqueta para um objeto informacional.

Pode-se dizer então que, quando desta forma, a etiquetagem – processo que gera a folksonomia, guarda relações estreitas com abordagem de RI. Neste processo, o usuário participa

da classificação dos recursos indiretamente e sem saber que está o fazendo, o que, acredita-se, afastaria a folksonomia da visão cognitiva.

5.2 Folksonomia como esquema de representação

Após a análise das características da folksonomia comuns às abordagens da organização do conhecimento, pode-se agora analisar seus aspectos positivos e negativos para, então, chegar-se a uma definição da questão proposta inicialmente. A figura 10 resume tais características:

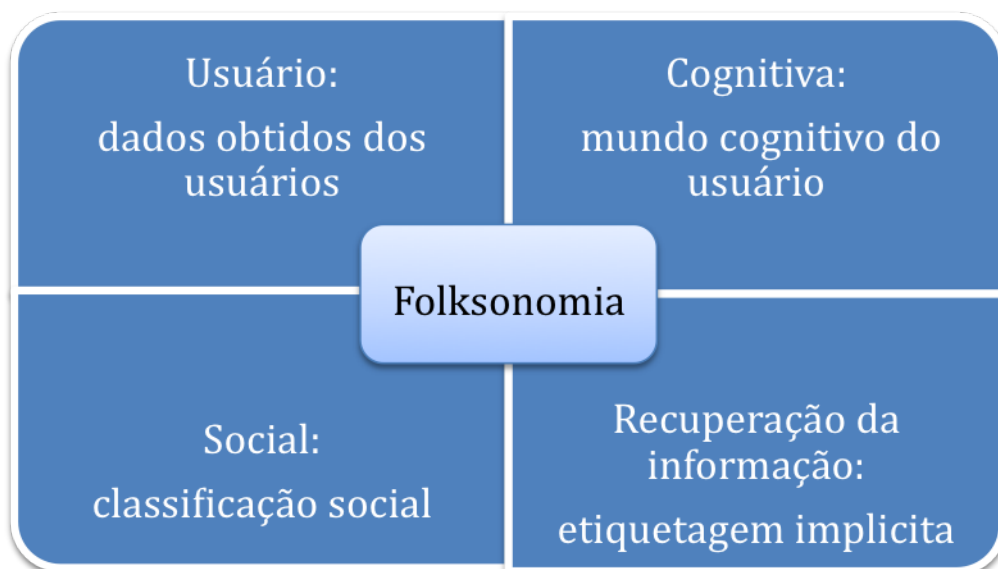


Figura 10 – Abordagens em organização do conhecimento

5.2.1 Pontos positivos e negativos

Dentro das características que permitem incluir a folksonomia nas abordagens em Organização do conhecimento, podem ser encontrados pontos positivos e negativos. O fato de a folksonomia ser construída a partir de dados obtidos dos próprios usuários é algo positivo no sentido da garantia do uso, ou seja, o termo usado para representar o documento será o mesmo usado para recuperá-lo posteriormente, por determinado usuário. Neste caso, a garantia de uso do termo na literatura (garantia literária), parece não importar muito, já que a folksonomia é construída a posteriori – não há uma etapa de análise dos documentos do domínio para então

se coletar termos e criar posteriormente a base para o sistema de representação, como no caso dos tesouros, por exemplo.

Dentro das abordagens orientadas aos usuários, é interessante a discussão feita por Broughton *et al.* sobre o que é organizado em cada abordagem, ou seja, o que é considerada como unidade a ser organizada. Para esses autores, nas abordagens cognitivas e orientadas a usuários, a unidade organizável é a estrutura cognitiva, o individual. Nessa perspectiva, a folksonomia poderia certamente ser considerada uma estrutura de representação do conhecimento.

Apesar disso, o uso da estrutura cognitiva como unidade organizável pode ser questionada. Na abordagem cognitiva, a mesma característica pode ser considerada positiva e negativa: o mundo cognitivo do usuário. Representar o mundo cognitivo do usuário é algo que nenhum outro tipo de representação pode fazer tão bem quanto a folksonomia – ninguém melhor que o próprio usuário para representar seu mundo cognitivo. É sabido que um sistema de representação do conhecimento tão mais preciso será quanto mais restrito e específico for o domínio representado. Levando-se essa especificação ao extremo, deve-se chegar ao usuário, a uma única pessoa – seu mundo cognitivo é certamente o extremo de uma especificação de conhecimento. Sendo assim, a representação do conhecimento gerada numa folksonomia teria uma grande vantagem. Porém, para apenas um usuário:

Pessoas colocando etiquetas para seu próprio uso (ou reuso) selecionam palavras que usam e acreditam que entendem. Essa perspectiva significa que as etiquetas estão corretas para pelo menos uma pessoa e nem sempre para uma comunidade ou disciplina a que pertence. (WAL, 2005).

A representação do conhecimento fica, desta forma, atomizada, o que suscita a questão da utilidade dessa representação. Além disso, a grande vantagem do mundo cognitivo do usuário é diluída à medida que mais e mais usuários colaboram com suas etiquetas em um serviço que vai gerar uma folksonomia.

De qualquer forma, acredita-se que a representação gerada em cada serviço deverá trazer, em algum grau, algo válido para determinada comunidade de usuários que a gerou. Os usuários de determinados serviços de informação na web podem ter interesses comuns e até mesmo formar essas comunidades dentro dos serviços: “a web se auto-organiza de tal forma que a estrutura de *links* permite uma identificação eficiente de comunidades.” (FLAKE *et al.* 2002) Acredita-se que a etiquetagem é um processo que contribui nessa auto-organização da web.

Em todo caso, haverá sempre um *trade off* nos sistemas de representação do conhecimento: quanto mais específicos e próximos da cognição, menor sua utilidade para um grande número de usuários.

A característica de classificação social encontrada no modelo de organização social do conhecimento traz os mesmos problemas dos outros sistemas gerados nesta mesma vertente: a arbitrariedade nas decisões. Apesar disso, a folksonomia tem a vantagem de ser um esquema *botton-up*, em que os termos são gerados a partir do mais específico, não passando pela fase da

categorização, como visto anteriormente. Além disso, apesar dos termos serem gerados pelos próprios usuários, estes o fazem de uma forma um pouco mais independente que um determinado grupo social escolhido para organizar um domínio, como no exemplo de Hjørland (2007e) da categorização de disciplinas em uma universidade. Os usuários que geram as folksonomias podem ser influenciados pelas etiquetas de outros usuários, mas acredita-se que o grau de parcialidade e as questões políticas são menores que num grupo que de fato se conhece e se reúne para elaborar uma categorização de um domínio.

Por fim, a etiquetagem implícita encontrada na abordagem da Recuperação da informação apresenta como vantagem o uso das próprias perguntas de busca como etiquetas para os objetos informacionais, sem que o usuário mesmo saiba. Isso por um lado é bom, pois muitas vezes o usuário de um sistema de informação que disponibiliza a ferramenta de etiquetagem não usa o recurso, ou seja, não contribui para gerar a folksonomia naquele sistema. Já na etiquetagem implícita, o usuário executa uma tarefa comum de busca da qual necessita e nem fica sabendo que está colaborando para o sistema de etiquetagem – dispensa-se assim a “boa vontade” do usuário. A parte falha é que nem todo objeto de informação poderá ser etiquetado dessa forma, pois alguns tipos de rótulos usados pelos usuários não serão encontrados no recurso.

5.2.2 Como esse esquema reflete ou falha em refletir conhecimento

A partir do exposto anteriormente, tem-se então que o processo de etiquetagem gera uma folksonomia, e esta pode ser considerada uma representação do conhecimento, no sentido de, no mínimo, refletir o conhecimento de determinada comunidade. Assim, a representação gráfica usual das folksonomias – a nuvem de etiquetas (figura 11), reflete os assuntos mais recorrentes em determinado momento e em determinado sistema. Dessa forma, pode-se dizer que a folksonomia pode ser usada como forma de descoberta de conhecimento: a partir da análise das etiquetas mais freqüentes é possível determinar tendências e interesses da comunidade analisada.

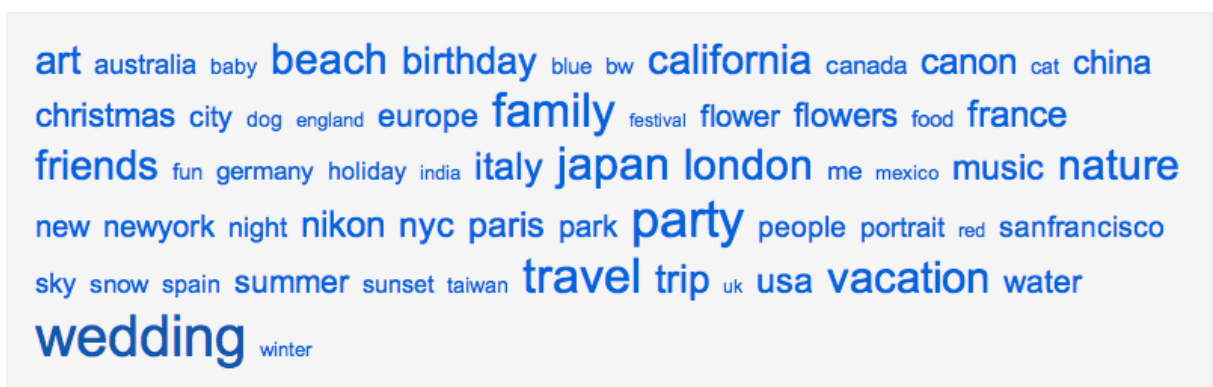


Figura 11 – Exemplo de nuvem de etiquetas: Flickr em 15/06/08

As falhas mais citadas do processo de etiquetagem são as de falta de padronização: na maioria dos casos, não há regras para a aplicação de etiquetas. Desta forma, o vocabulário é totalmente livre, gerando os conhecidos problemas que costumam ser resolvidos por um vocabulário controlado.

Entende-se que a folksonomia, comparada aos outros esquemas de representação do conhecimento, é gerada de forma inversa: primeiro se classificam os objetos informacionais, e, posteriormente, surge uma folksonomia, representada visualmente pela nuvem de etiquetas. Já nos outros sistemas como as taxonomias e os tesouros, os objetos informacionais são classificados somente quando o sistema já existe: um tesouro, por exemplo, é usado para a escolha dos termos que irão descrever o conteúdo de um documento.

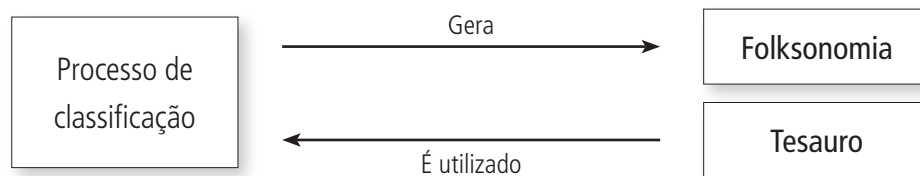


Figura 12 – Formação de folksonomia

Ou seja, a folksonomia é construída *a posteriori* em relação aos outros sistemas. Isso traz algumas falhas para a representação do conhecimento, como observado por Sinha (2006), o surgimento de termos sem conceito, já que não há uma formalização ou um consenso explícito da comunidade envolvida.

Assim, pode-se entender que a folksonomia estaria inserida nas duas vertentes de Organização do conhecimento propostas por Broughton *et al.* (2005): a dos processos de organização e a dos sistemas de organização do conhecimento. Sobre esse aspecto, alguns autores como Trant (2006) fazem uma clara distinção entre folksonomia e etiquetagem, que também é adotada nesta pesquisa. Neste caso, a etiquetagem estaria na vertente dos processos de organização (como forma de indexação colaborativa) e a folksonomia na dos sistemas de organização do conhecimento.

A proposta deste capítulo foi analisar a folksonomia em cada uma das abordagens e seus pontos fortes e fracos para a representação do conhecimento. Feito isso, tem-se então uma base que permite enquadrar a folksonomia nos esquemas de representação do conhecimento, ao menos em certo sentido. Assim, supõe-se que as características da folksonomia encontradas nas abordagens propostas possam indicar que esta estrutura tenha sua utilidade para a organização da informação na web. Acredita-se ainda que esta análise pode servir como contribuição para o entendimento das folksonomias e sua contextualização na teoria da Organização do conhecimento.

Capítulo 6

6 Estudo de usuários de folksonomia: motivações para a etiquetagem

Este capítulo buscou investigar o que leva o usuário a executar tarefas de etiquetagem dos recursos da web. Pretendeu-se identificar quais as motivações dos usuários para a realização de tal tarefa. Para isso, foi realizado um estudo de usuários, com base em uma abordagem alternativa, ou seja, pela óptica do usuário da informação - o usuário de folksonomia.

Acredita-se que o levantamento das possíveis motivações dos usuários poderá oferecer subsídios para a análise da folksonomia em relação à sua utilidade para a organização e recuperação da informação na web.

A pesquisa foi aplicada em dois dos mais populares sítios que permitem a etiquetagem: Delicious e Flickr. Por serem os pioneiros, sugere-se que os usuários de tais serviços já estejam bastante familiarizados com o processo de etiquetagem, e, portanto, devem executá-lo com maior frequência, fornecendo melhores resultados (maior número de respostas) para a pesquisa. Além disso, a análise dos resultados em sítios com diferentes tipos de objetos informacionais – fotos e sítios favoritos – permitirá uma análise comparativa das motivações em diferentes ambientes de compartilhamento.

6.1 Contexto

De forma ampla, o contexto para a realização desse estudo de usuários, considerando contexto como um *quadro de referência para práticas informacionais* (COURTRIGHT, 2007), é a própria Internet, no âmbito das aplicações de Web 2.0. De forma mais restrita e para análise dos resultados individualmente, pode-se delimitar o contexto aos serviços de informação onde se aplicou a pesquisa (sítios). Pode-se estabelecer aqui um paralelo com as Organizações – que costumam ser utilizadas nos estudos de usuários para delimitar contextos.

6.2 Sobre os sítios pesquisados

O Delicious é um serviço de armazenamento e gerenciamento de sítios favoritos em que o usuário tem a possibilidade de adicionar etiquetas aos sítios colecionados. Segundo Golder e

Huberman (2006, p. 201), o Delicious é um sistema de etiquetagem colaborativa e seu criador Joshua Schachter o denomina como “gerenciador social de sítios favoritos.”

A grande vantagem de se utilizar o Delicious, ou outro serviço de armazenamento de sítios favoritos, é que o usuário pode acessar seus favoritos de qualquer computador e não somente de seu navegador, que já possui essa função. Outra característica do serviço é a possibilidade do usuário acessar os sítios favoritos adicionados por outros usuários, além de permitir buscas a partir das etiquetas. Por esse motivo ele é considerado “social.” Apesar disso, os usuários não têm como se comunicar diretamente uns com os outros, o que enfraquece o conceito de comunidade para esse sistema.

O Flickr é “um gerenciador de fotos e aplicativo de compartilhamento *online*.” (FLICKR, 2008, tradução nossa). O sistema permite ao usuário cadastrado armazenar, organizar (por meio de etiquetagem) e compartilhar suas fotos. Segundo Marlow *et al.* (2006, tradução nossa), “as etiquetas são parte importante deste ambiente, em que atuam como ferramenta navegacional primária para se encontrar recursos similares e pessoas.” Há vários tipos de controle de privacidade e configurações, assim o usuário pode restringir ou ampliar o acesso a seu conteúdo, bem como a adição de comentários e etiquetas a suas fotos.

Os usuários também podem usar as etiquetas para buscar fotos em sua própria coleção, na coleção de outro usuário ou em todo o rol de fotos públicas armazenadas no sítio (AMES; NAAMAN, 2007, tradução nossa). No caso de imagens, a etiquetagem é particularmente importante para a recuperação da informação, visto que não há outra forma eficiente de se buscar um recurso não-textual. “Uma recuperação eficiente é amplamente baseada em anotações apropriadas para cada imagem, já que formas automáticas de análise de conteúdo ainda são limitadas.” (SHNEIDERMAN; BEDERSON; DRUCKER, 2006, tradução nossa).

O sistema também oferece outras funcionalidades que aumentam a interatividade, como o FlickrMail, um sistema interno de comunicação direta com qualquer usuário do sistema, tornando o sistema mais “social.” Apesar de ser mais conhecido por armazenar fotos, recentemente o Flickr ampliou seu leque possibilitando também o armazenamento de vídeos. Os serviços citados serão descritos com mais detalhes no capítulo 7.

A popularização da prática de etiquetagem na web está sendo vista como uma importante mudança em sua forma de organização. Para se analisar melhor tal fenômeno, deve-se compreender o que leva os usuários a adotarem tal comportamento, ou seja, suas motivações. Tal comportamento interfere diretamente nas práticas informacionais dos usuários da web. Será abordado a seguir o tema da motivação no comportamento humano.

6.3 Sobre motivação

De acordo com a psicologia, “motivo ou motivação, refere-se a um estado interno que pode resultar de uma necessidade. É descrito como ativador ou despertador, de comportamento geralmente dirigido para a satisfação da necessidade instigadora” (DAVIDOFF, 2005, p. 326). Essa autora cita dois modelos básicos de motivação: o modelo homeostático e o de incentivo. O primeiro está mais relacionado com necessidades fisiológicas básicas e o segundo com questões psicossociais.

Já os autores Huffman, Vernoy e Vernoy (2003, p. 429) separam em três categorias as teorias que explicam a motivação: teorias biológicas, teorias psicossociais e interacionismo, este último com base na hierarquia das necessidades de Maslow.

Segundo Amaral (1996 p. 220):

Os estudos sobre motivação podem responder às perguntas do tipo “o que faz as pessoas usarem, freqüentarem ou não as unidades de informação?” Entretanto, as pessoas são complexas e as respostas podem não ser tão simples, pois dependem de pesquisas mais detalhadas sobre motivação.

Estudos sobre motivação também podem responder perguntas do tipo “o que leva o usuário a utilizar ou não um recurso de um sistema ou serviço de informação?.” Essa foi a pesquisa aplicada neste trabalho.

A motivação estudada nessa pesquisa está inserida na categoria das psicossociais, que inclui as teorias do incentivo e a cognitiva. Acredita-se que o modelo mais relacionado com os fins desse trabalho para se explicar motivação seria o do incentivo, que

[...] diz que experiências e incentivos freqüentemente alteram cognições e emoções, levando à motivação. A motivação aciona o comportamento, o qual pode novamente alterar cognições e emoções, aumentando ou diminuindo o nível de motivação. (DAVIDOFF, 2005, p. 326-327).

Ou seja, neste caso, a motivação estaria diretamente relacionada a uma nova ação ou comportamento do individuo, como explicado no artigo de Marlow *et al.* (2006).

Os autores elaboraram duas taxonomias organizacionais para sistemas de etiquetagem social. Uma delas descreve o *design* e atributos do sistema e a outra os incentivos do usuário. Para os autores, o comportamento do usuário é amplamente ditado pelas formas de contribuição permitidas e pelas motivações sociais e pessoais para adicionar dados ao sistema:

Incentivos e motivações para os usuários também exercem um papel significativo afetando as etiquetas que emergem de sistemas sociais de etiquetagem. Usuários são motivados por interesses sociais e pessoais. As motivações de alguns usuários vêm de um propósito prescrito, enquanto outros usuários conscientemente mudam para um outro propósito os sistemas disponíveis para satisfazer suas próprias necessidades e desejos, e outros ainda buscam contribuir para o progresso coletivo. Uma grande parte das motivações e influências dos

usuários de sistemas de etiquetagem é determinada pelo desenho do sistema e pelo método pelo qual eles são apresentados para práticas de etiquetagem inerentes. (MARLOW *et al.*, 2006, tradução nossa).

Os autores categorizaram as motivações em dois níveis: organizacional e social. O nível organizacional se refere ao uso de etiquetas como uma estrutura de arquivamento, e a motivação para esses usuários deve estar relacionada com padrões pessoais e uso de etiquetas comuns criadas por outros usuários. Já a vertente social da etiquetagem, segundo os autores, estaria relacionada com a natureza comunicativa da etiquetagem, onde os usuários se expressam, opinam e julgam o recurso etiquetado (MARLOW *et al.*, 2006).

Além dessa categorização mais ampla, os autores elencam um rol de motivações potenciais que influenciam o comportamento de etiquetagem. São elas:

- Recuperação futura
- Contribuição e compartilhamento
- Atrair atenção
- Jogar e competir
- Auto-apresentação
- Expressar opinião

Na seção seguinte desta revisão, serão citados outros trabalhos relacionados com a motivação dos usuários para a etiquetagem de objetos informacionais na web.

6.4 Trabalhos relacionados

Em maio de 2007, foi apresentado por Alla Zollers, na 16a. Conferência Internacional da *World Wide Web*, um trabalho sobre motivações para etiquetagem, que identificou três motivações sociais para a tarefa: expressar opinião, performance e ativismo. A autora realizou uma pesquisa a partir da análise das próprias etiquetas atribuídas aos recursos informacionais de dois sítios: o Amazon, uma loja virtual e o Last.fm, uma rede social voltada para o tema de música. A hipótese de pesquisa da autora é de que o poder de organizar a informação pessoal em um contexto público motiva as pessoas a utilizarem etiquetas com um propósito social, além da organização da informação e geração folksonomias. Uma hipótese adicional é de que o uso de etiquetas em sistemas que não foram originalmente desenhados para a organização da informação irá mostrar grandes sinais de socialização.

A pesquisa começou o estudo exploratório com a observação das nuvens de etiqueta de cada sítio, e então se iniciou uma investigação de etiquetas específicas. Para explorar as

motivações para etiquetagem no Amazon, foram selecionados aleatoriamente 25 livros e 25 produtos musicais. Já no Last.fm, o sistema não permite que álbuns, artistas ou faixas musicais seja selecionadas aleatoriamente, mas permite que usuários o sejam. Neste caso, a autora selecionou aleatoriamente 50 usuários e a partir das listas que possuíam, selecionou 25 artistas e 25 faixas musicais aleatórias.

De acordo com os resultados encontrados, a motivação referente a expressar opinião está emergindo no Last.fm, e prevalece no Amazon. A motivação “performance” refere-se à atividade de participantes, em uma determinada ocasião, que de certa forma influencia outros participantes. A última motivação citada é utilizar a etiquetagem como forma de ativismo, ou seja, utilizar uma etiqueta para fazer campanha de alguma coisa. A autora conclui que apesar dessas três motivações terem sido encontradas em ambos os sítios pesquisados, elas não se encontram na mesma proporção nos dois.

O LibraryThing, sítio que permite que os usuários cadastrem sua biblioteca pessoal num ambiente de compartilhamento e também a etiquetagem dos livros adicionados, fez uma pesquisa que indicou uma quantidade muito maior de etiquetas em seu sítio do que no Amazon, que é proporcionalmente muito maior em termos de visibilidade na web. Tim Spalding, autor da pesquisa postada no *blog* do LibraryThing, acredita que o principal fator que leva os usuários a etiquetarem mais recursos do LibraryThing que os do Amazon é que no LibraryThing o usuário é “dono” dos recursos etiquetados, ou seja, está fazendo um serviço para ele mesmo. Já no Amazon, o usuário estaria fazendo algo para terceiros, colocando etiquetas em produtos comerciais de um sítio de compras. Além disso, o autor indica que o Amazon não incentiva muito essa prática, colocando esse recurso de forma não muito acessível no sítio.

Outro estudo, iniciado por pesquisadores do Yahoo! e publicado em 2007, identificou motivações para etiquetagem por meio de um estudo de usuários. Os autores responsáveis pela publicação final foram Morgam Ames, da Universidade de Stanford, e Mor Naaman, do Yahoo! Research Berkeley.

O estudo foi feito com usuários de um aplicativo para celulares com câmera, o ZoneTag, que permite que as fotos recém tiradas recebam etiquetas e seja feito o *upload* imediato para o Flickr do usuário. Além disso, o sistema tem a funcionalidade de sugestão de etiquetas (*tags suggestion*). A combinação do Flickr com o ZoneTag os permitiu estudar essa nova classe de organização, anotação e compartilhamento de fotos. Segundo Ames e Naaman (2007, tradução nossa), “as motivações e incentivos para participação e etiquetagem nesses sistemas ainda não foram estudadas com profundidade.”

A parte principal do estudo envolveu entrevistas semi-estruturadas com 13 participantes, incluindo alguns dos usuários do ZoneTag que já tiraram muitas fotos. As entrevistas consistiram de conversas casuais sobre o uso do Flickr e do ZoneTag e as motivações para captura de imagens e etiquetagem. As motivações encontradas foram divididas em sociais e funcionais, e estruturadas numa forma de taxonomia, ilustrada na figura 13.

		Função	
		Organização	Comunicação
Sociabilidade	Individual	<ul style="list-style-type: none"> • Recuperação, Diretório • Busca 	<ul style="list-style-type: none"> • Contexto para si mesmo • Memória
	Social	<ul style="list-style-type: none"> • Contribuição, chamar atenção • Associação ad hoc de fotos 	<ul style="list-style-type: none"> • Descrição de conteúdo • Sinalização Social

Figura 13 – Taxonomia das motivações para etiquetagem (adaptado de AMES; NAAMAN, 2007)

Concluiu-se que é possível motivar os usuários a adicionarem anotações aos conteúdos, encorajando-os a utilizarem as etiquetas. Destaca-se, particularmente, o incentivo social como importante motivação para os usuários colocarem etiquetas em suas fotos. Os autores acreditam que as pessoas são mais inclinadas a etiquetarem seus conteúdos quando são dados os incentivos corretos e boas condições para anotação (AMES; NAAMAN, 2007).

Baseados na pesquisa anterior, que sugere que a motivação predominante ocorre na categoria social, Nov e Ye (2007) realizaram uma nova pesquisa com os usuários do Flickr e desenvolveram uma escala de motivações para a etiquetagem, distinguindo-as entre os que colocam etiquetas para si mesmos, os que colocam para seus amigos e familiares e os que colocam para o público em geral.

Os autores esperavam encontrar uma relação positiva entre o número de grupos a que determinado usuário pertence no Flickr, há quanto tempo usa o sítio e a quantidade de etiquetas que ele usa, o que foi confirmado posteriormente. A metodologia utilizada foi aplicação de questionários via web usando uma combinação de dados gerados pelo usuário e dados gerados pelo próprio sistema, utilizando-se um *Application Programming Interface* (API³) do Flickr. A pesquisa ainda está em curso, mas as conclusões preliminares sugerem que tanto motivações sociais quanto individuais afetam o nível de etiquetagem do usuário, com exceção da motivação para amigos e familiares. Recomenda-se que o desenvolvimento de sistemas colaborativos leve em conta a presença social para que se tenha um grande uso das etiquetas.

Em outro artigo publicado no Boletim da ASIS&T, Jason Morrison analisa a função das folksonomias e faz um levantamento das motivações dos usuários para a etiquetagem. O autor não faz nenhum tipo de pesquisa direta com usuários nem um estudo empírico estrito. O levantamento das possíveis motivações que os levam a adicionar etiquetas a conteúdos foi feito

³ Interface de Programação de Aplicativos: é um conjunto de rotinas e padrões estabelecidos por um *software* para a utilização das suas funcionalidades por programas aplicativos, isto é: programas que não querem envolver-se em detalhes da implementação do *software*, mas apenas usar seus serviços. (Fonte: Wikipedia, <<http://pt.wikipedia.org/>>)

com base na observação de vários sistemas de etiquetagem, buscando analisar por que os sítios empregam etiquetagem e por que os usuários participam.

O autor identificou as seguintes motivações: encontrar depois (recuperação), ganhar exposição (divulgação), expressar opiniões, ganhar pontos em um jogo (ESP Game) e encontrar itens relacionados. Além disso, o autor encontrou outra possibilidade de etiquetagem pelo usuário: quando realiza buscas e o buscador utiliza a expressão utilizada para classificar o conteúdo recuperado e acessado pelo usuário. Neste caso, não há motivação para a etiquetagem explicitamente, pois o usuário nem sabe que está realizando tal tarefa.

Outra recente encontrada nessa temática, publicada em maio de 2008, testa duas possíveis explicações para a motivação dos usuários: a etiquetagem para uso próprio e a etiquetagem para outros usuários, como uma forma de contribuição voluntária para o sistema:

Esses *resultados* sugerem que há dois conjuntos distintos de motivos que guiam os que contribuem com recursos etiquetados: motivos orientados si mesmo (para recursos etiquetados para uso pessoal) e motivos orientado aos outros (para recursos etiquetados para outros usuários). (BENBUNAN-FICH ; KOUFARIS, 2008).

Para tal estudo, os autores propõem um modelo de pesquisa e desenvolvimento de hipóteses. A pesquisa foi realizada em dois populares sítios de *social bookmarking* (gerenciador de sítios favoritos). Os sítios estudados não são identificados, pois não houve permissão dos mesmos. Foram aplicados questionários a 381 usuários que aceitaram participar da pesquisa, mas somente 94 completaram o questionário. Os autores confirmam algumas de suas hipóteses e afirmam que pesquisas futuras devem explorar os antecedentes das motivações individuais dos usuários.

É interessante apontar que em todos os serviços que possibilitam a etiquetagem aqui citados, o uso das etiquetas é opcional. Não há obrigatoriedade de se adicionar etiquetas, e, segundo Porter (2006, tradução nossa), é justamente por esse motivo que os usuários colaboram:

Então, por que as pessoas colocam etiquetas? Bem, parte da razão deve ser por não haver uma penalidade por não colocar. Diga-me que eu tenho que fazer algo e eu não farei. Dê uma escolha, e talvez algum reforço positivo, e eu posso vir a fazer isso. Essa é a forma como os *softwares* devem funcionar. Usuário primeiro. Sistema depois.

Vários dos estudos aqui citados reforçam a importância de novas pesquisas sobre a motivação dos usuários para a etiquetagem. Na seção seguinte, será feita a análise dos dados do estudo de usuários proposto neste trabalho para uma nova investigação deste tema.

6.5 Análise dos dados

Ao final da coleta de dados, foram obtidas 56 respostas, sendo 23 no Delicious e 33 no Flickr.

6.5.1 Dados demográficos

Os dados demográficos da pesquisa, resultados referentes a sexo, idade, localização, escolaridade e tempo de uso do sistema dos respondentes de cada sistema e, posteriormente, no geral, somando-se os resultados dos dois serviços, por não apresentarem influência na análise da questão-foco da pesquisa, serão apresentados no apêndice B ao final do trabalho.

6.6 Motivações para etiquetagem

A pergunta-chave do estudo foi sobre a motivação do usuário para a etiquetagem dos recursos dos sistemas analisados, no caso do Delicious, sítios favoritos e do Flickr, fotos. Foi feita pergunta aberta sobre o tema, deixando o usuário livre para escrever sua resposta.

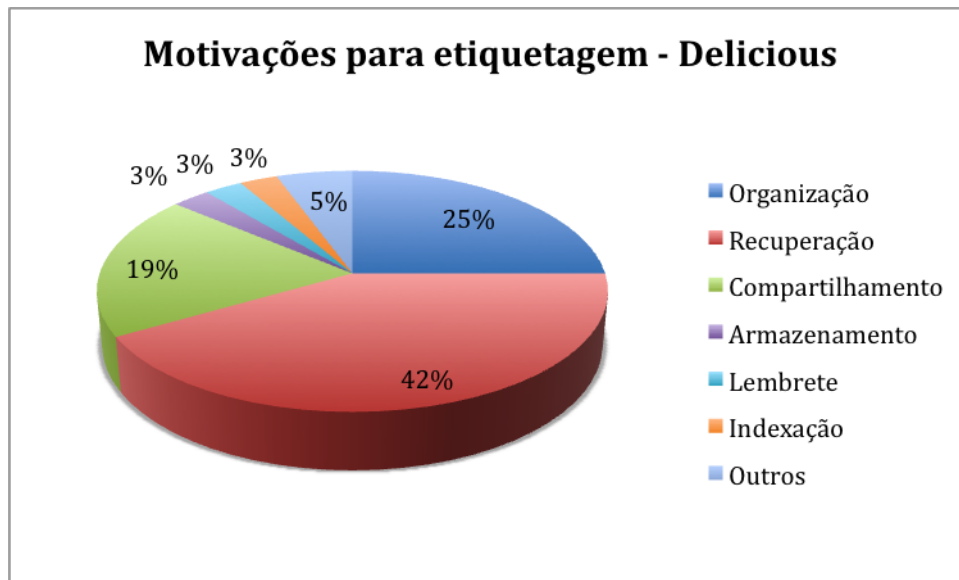
Serão apresentados a seguir os resultados encontrados nos serviços separadamente, e, após, o resultado geral condensado e a análise comparativa.

6.6.1 Delicious

A partir da análise do conteúdo das respostas, pode-se obter um rol de motivações citadas pelos usuários para a etiquetagem de recursos, encontrando-se assim as seguintes categorias:

- Organização
- Recuperação
- Compartilhamento
- Lembrete
- Armazenamento
- Indexação
- Outros

Buscou-se preservar ao máximo a linguagem das respostas do usuário para a terminologia das categorias, incluindo as especificidades como a citação de "Indexação" como motivação para etiquetagem. Além disso, observou-se que os fatores "facilidade" e "rapidez" foram citados por alguns usuários. Três usuários comentaram a rapidez e seis a facilidade do uso de etiquetas



no Delicious. Para a obtenção dos gráficos com os resultados percentuais, utilizou-se os 23 questionários respondidos no Delicious como equivalente a 100%, ou seja, 100% das respostas = 23 respostas. A partir daí, tem-se os percentuais por categoria de motivação.

Quantitativamente, os resultados obtidos foram os seguintes, representados no gráfico 1:

Gráfico 1 - Motivações para etiquetagem no Delicious

Tem-se então que as motivações mais citadas são a recuperação, seguida da organização e o compartilhamento. Na categoria compartilhamento, foram consideradas as funções sociais e de disseminação da informação.

A partir das respostas, observou-se que alguns usuários colocaram suas motivações para a utilização do serviço Delicious como um todo, e não somente das etiquetas. Um exemplo desse tipo de resposta foi com o respondente #13, que citou como motivação "ser alertado (por meio do RSS) de novos sítios em assuntos específicos." Tal alerta é obtido a partir da assinatura de uma etiqueta, em que o serviço mostra todos os novos sítios adicionados e etiquetados daquela forma. Neste caso, o usuário estaria se beneficiando da etiquetagem de outros usuários, e não da sua própria.

Outro caso desse tipo de resposta é em que o usuário cita como motivação a descoberta de sítios novos, como o respondente #13 "Para encontrar novos recursos em assuntos muito específicos (raros)", e o respondente #21.." e procurar por algumas URLs com o mesmo interesse." Em ambos os casos, os usuários estariam tendo usufruto da etiquetagem de outros usuários.

Uma observação interessante do respondente #15: “para evitar hierarquia nos sítios favoritos (onde você fica imaginando em qual pasta você deve adicionar seu sítio favorito) que está presente em todos os navegadores hoje em dia.” A partir desse comentário, percebe-se uma vantagem identificada pelo respondente na folksonomia, que é a dispensa de uma estrutura hierárquica rígida como a das classificações e taxonomias, conforme visto na revisão de literatura. A flexibilidade das etiquetas permite que o usuário classifique um recurso em mais de uma categoria, evitando a necessidade de uma decisão estrita.

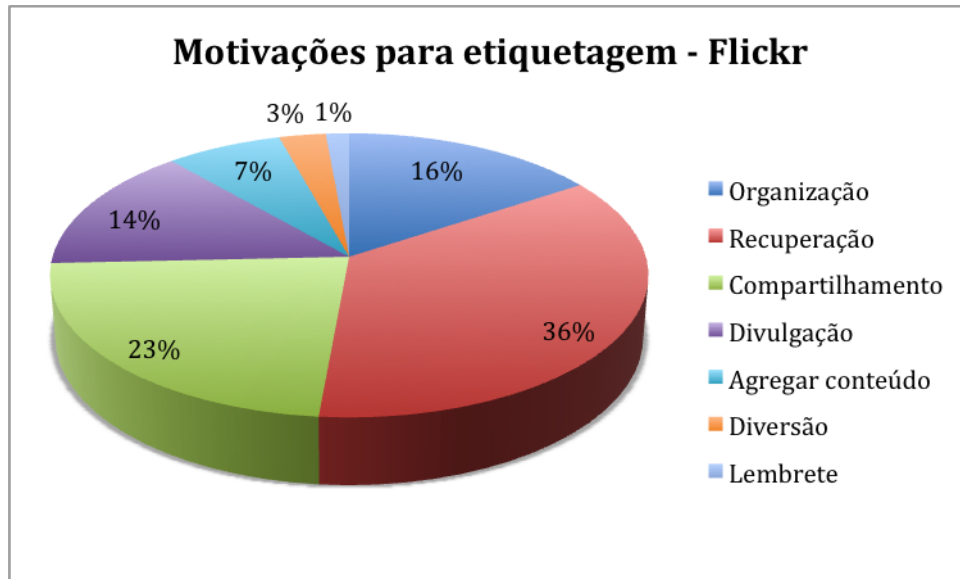
Entre as respostas, encontram-se duas que destoam bastante do tipo de respostas no geral: “requerimento do cliente” e “nada em especial.” Tais respostas foram enquadradas na categoria “outros.”

6.6.2 Flickr

Para o Flickr os resultados foram bem parecidos, acompanhando a análise anterior. A novidade foi o surgimento de novas motivações, incluído uma muito citada: divulgação. As motivações encontradas no Flickr foram:

- Organização
- Recuperação
- Compartilhamento
- Divulgação
- Agregar conteúdo informativo
- Diversão
- Lembrete

Outra motivação que surgiu nas respostas para esse sistema foi a de diversão. Dois usuários responderam que colocam etiquetas nas fotos por que acham uma tarefa divertida. Essa motivação pode explicar também o jogo ESP Game, em que o usuário ganha pontos por colocar etiquetas nas imagens como citado por Morrison (2007). Recentemente, o Google Imagem começou a permitir o mesmo: o usuário pode adicionar etiquetas nas imagens para ajudar na recuperação posterior, e o sistema funciona na forma de um jogo em que o usuário ganha pontos. No estudo de Marlow *et al.*, essa categoria também é citada: jogar e competir, o que também estaria relacionado com a motivação da diversão.



Assim como no Delicious, os fatores rapidez e facilidade foram citados. A facilidade foi citada com referência tanto ao sistema de etiquetagem quanto ao uso de etiquetas para facilitar as buscas posteriormente.

Segue o quantitativo das motivações, com os cálculos de porcentagem realizados da mesma forma anterior, com 33 questionários respondidos correspondendo a 100% e calculando a partir daí os valores das motivações por categoria, conforme apresentado no gráfico 2.

Gráfico 2 - Motivações para etiquetagem no Flickr

As motivações mais citadas foram organização (16%), recuperação (36%) e compartilhamento (23%). Destaque aqui para a motivação da divulgação: o tipo de recurso etiquetado no Flickr (fotos) explica o surgimento dessa motivação. Os usuários geralmente colocam suas próprias fotos no Flickr, e muitos buscam visibilidade a partir do serviço. Várias respostas ilustram esse fato:

- “Quero ter o maior número de visitas que eu puder”
- “Ter o máximo de pessoas possíveis para ver as fotos que eu posto”
- “Etiquetagem também ajuda as minhas fotos a serem vistas por um grande número de pessoas na Internet”
- “Para que possam ser encontradas por outros usuários do Flickr”

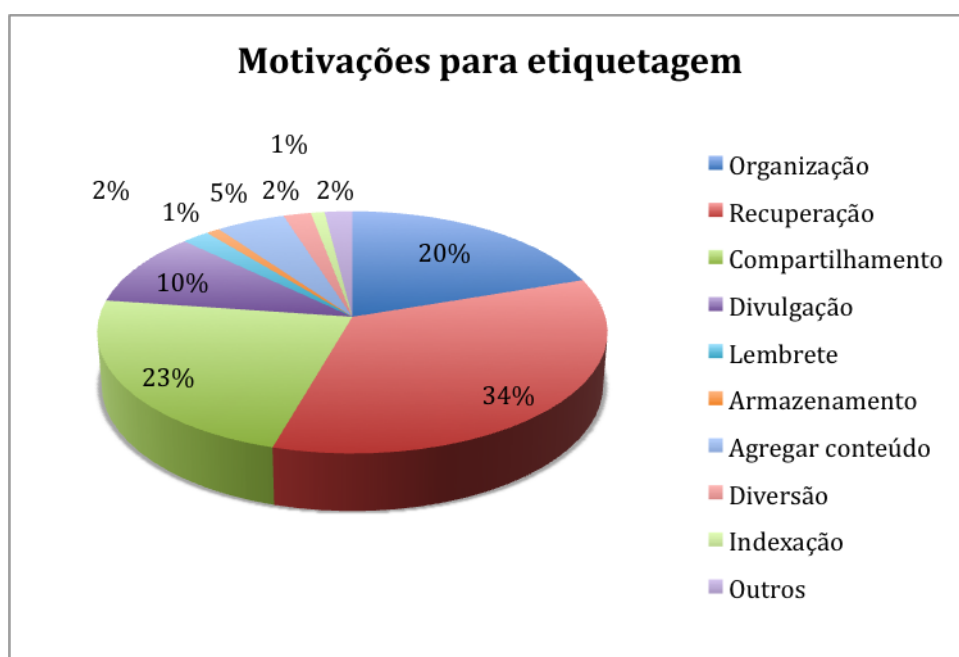
O uso de etiquetas facilita a recuperação das imagens, pois estas não possuem outra forma de serem recuperadas por buscas. Assim, as etiquetas funcionam como elementos descritivos do conteúdo da imagem. Essa mesma razão explica a alta porcentagem de respostas que consideram a motivação da recuperação. Alguns usuários já percebem essa questão, como

o respondente #09, que fala que usa etiquetas “para melhorar a informação fornecida”, e o #15 “para adicionar informação sobre a foto.” Além disso, o respondente #13 comenta que coloca etiquetas “para ter palavras que representam seu trabalho”, ou seja, descrição de conteúdo da imagem como forma de representação daquela informação.

6.6.3 Resultado geral e comparação nos serviços

Considerando-se as respostas recebidas nos dois sistemas, ao todo foram encontradas as seguintes motivações:

- Organização
- Recuperação
- Compartilhamento
- Divulgação
- Agregar conteúdo informativo
- Diversão
- Indexação
- Lembrete
- Armazenamento



- Outros

Para a obtenção deste resultado, foram somados os questionários recebidos nos dois serviços, 23 do Delicious e 33 do Flickr, tendo-se assim 56 questionários respondidos considerados equivalentes a 100%. A partir daí, calculou-se as porcentagens de acordo com a categoria das motivações encontradas, somando-se as respostas da mesma categoria nos dois serviços. Segue o gráfico 3 com as porcentagens:

Gráfico 3 - Motivações para etiquetagem

As categorias propostas por Marlow *et al.* (2006) de motivações sociais e organizacionais, estão presentes nos resultados dos sistemas analisados. Além disso, os resultados encontrados se enquadram em duas das motivações encontradas no rol proposto por pelos mesmos autores, citadas anteriormente: *recuperação futura e contribuição e compartilhamento*. Já em relação ao trabalho de Zollers (2007), que sugere três categorias sociais para a etiquetagem (expressar opinião, performance e ativismo), a presente pesquisa não encontrou relações. Sugere-se que isso se deve aos tipos de recursos e serviços analisados, que são diferentes nos dois estudos.

Em relação ao estudo de Ames e Naaman (2007), as motivações encontradas aqui podem ser enquadradas na taxonomia proposta pelos autores (ver Figura 13), com as funções sociais e individuais, de organização e comunicação. Observou-se na pesquisa que alguns usuários têm realmente uma preocupação social e de colaboração, como o respondente #8, do Delicious, que afirma coloca etiquetas “para contribuir com a comunidade de usuários (é por isso que coloco etiquetas em inglês).” Já outros usuários se preocupam em etiquetar somente para sua própria recuperação, como o respondente #16, do Delicious, que fala “eu não poderia me importar se ajuda outras pessoas a encontrar ou não, o Delicious não é social, independente do que os criadores afirmem.”

Tal resultado também corrobora a pesquisa recente de Benbunan-Fich R. e Koufaris, M. (2008), que supõem que há dois tipos de etiquetagem: para uso próprio e para outros usuários. No Flickr também é possível encontrar essas duas vertentes, mas com destaque para a etiquetagem voltada para outros usuários, que se relaciona com a motivação da divulgação, como visto antes. Exemplos de respostas que ilustram a colaboração: “..compartilhar com a família, amigos e público em geral”, “é também uma boa forma de rede de trabalho (*network*) porque encontro pessoas/fotógrafos com gostos similares, posso ver o trabalho deles e me inspirar”, “.. fornece uma forma fácil de compartilhar informação com outras pessoas.” Em relação a etiquetagem para si mesmo, identificam-se os seguintes exemplos de respostas: “Eu coloco etiquetas nas fotos para que eu possa buscar e encontrá-las depois”, “Etiquetagem é uma forma fácil para mim de catalogar e procurar minhas fotos, já que eu tenho uma grande quantidade delas” e “Para me ajudar a encontrar minhas próprias fotos.”

6.7 Análise dos resultados

Desde a antiguidade, o homem possui uma necessidade de classificar os objetos ao seu redor e os fenômenos da natureza. Conforme Araújo (2006, p. 118):

Todos os seres humanos classificam a todo momento as outras pessoas e os objetos do mundo, com quem e com os quais se relacionam, incessantemente. Todas as nossas ações no mundo são envolvidas por atos classificatórios [...].

Os sistemas de informação na web que permitem a etiquetagem de recursos estão em voga nos dias atuais e em constante crescimento. Acredita-se que as folksonomias têm um valor considerável para organização e recuperação da informação nesse ambiente, conforme sugerido pela presente pesquisa: 54% das respostas consideram como motivação para a etiquetagem os fatores organização e recuperação dos recursos etiquetados. A recuperação da informação pressupõe sua prévia organização, portanto, as duas motivações foram agrupadas, somando-se os 54%.

Entende-se então que essa alta motivação para organizar e recuperar relaciona-se com o instinto humano de categorizar e classificar as coisas (organização) para que sejam encontradas posteriormente (recuperação).

Outra motivação com alta porcentagem de resposta foi o compartilhamento, ou seja, a função social da etiquetagem como forma de colaborar com o sistema e com a comunidade de usuários. Essa motivação encontra base no próprio contexto da pesquisa: os sítios pesquisados fazem parte do conceito de Web 2.0, em que as principais características são a interação do usuário, o compartilhamento de recursos e conteúdos e a colaboração com o sistema e com os outros usuários.

Dessa forma, acredita-se que a etiquetagem possui um grande potencial para a organização e recuperação dos recursos da web, além de ser um elemento de socialização e colaboração entre os usuários.

Capítulo 7

7 Uso de folksonomia na web: diferentes aplicações e contextos

O último elemento de validação definido para etiquetagem e folksonomia é verificar sua aplicabilidade na arquitetura dos sítios, analisando seu potencial para organização e recuperação da informação. Desta forma, este capítulo faz um estudo descritivo, trazendo um panorama de alguns serviços de informação que utilizam etiquetagem e geram folksonomias, identificando seus principais elementos, características e abordagens.

O critério utilizado para a seleção dos sítios incluídos nesse capítulo é seu objeto informacional principal. Marlow at al. (2006) criaram uma taxonomia definindo alguns atributos pelos quais os sítios que utilizam etiquetagem podem ser classificados e entre as dimensões propostas está o tipo de objeto. Segundo os autores:

O tipo de objeto etiquetado é uma consideração importante [...] Na realidade, qualquer objeto que possa ser representado virtualmente pode ser etiquetado ou utilizado em um sistema de etiquetagem [...] As implicações para a natureza das etiquetas resultantes são inúmeras; um exemplo trivial é que suspeitamos que etiquetas dadas a recursos textuais devem diferir das etiquetas atribuídas a recursos/objetos que não contenham representação textual, como imagem ou áudio, apesar disso não ter sido testado empiricamente. (MARLOW *et al.*, 2006, tradução nossa).

Assim, buscou-se contemplar sítios com diferentes objetivos e tipos de objeto informacional para que se tenha uma visão ampla da utilização de etiquetagem e folksonomia na organização e recuperação da informação na web. Não se pretendeu uma seleção exaustiva de sítios e sim uma visão geral dos serviços de informação por tipo de objeto. Da mesma forma, a descrição dos sítios não será exaustiva e não pretende destacar todos os elementos contidos nos serviços de informação, mas focar no uso da etiquetagem e geração de folksonomias e na organização e recuperação da informação no sítio. Além dos 7 tipos de objeto selecionados, incluiu-se um objeto repetido que é o sítio favorito, pois há um serviço de informação com foco nesse objeto que incorpora o controle de vocabulário nas etiquetas. Por esse motivo o sítio foi incluído, já que pode se tratar de uma evolução dos sistemas de etiquetagem. Chegou-se então a 8 sítios para serem descritos, conforme quadro 1 a seguir:

Sítio	Objeto informacional principal	URL
1. Delicious	Sítios favoritos	http://delicious.com/
2. Flickr	Fotos	http://www.flickr.com/
3. YouTube	Vídeos	http://www.youtube.com/
4. Library Thing	Livros	http://www.librarything.com/

Sítio	Objeto informacional principal	URL
5. Amazon	Produtos comerciais	http://www.Amazon.com/
6. Last.fm	Música	http://www.last.fm/
7. Steve	Objetos de arte	http://www.steve.museum/
8. Zigtag	Sítios favoritos - controle de vocabulário	http://www.Zigtag.com/home.jsp

Quadro 1 – Sítios escolhidos

7.1 Delicious

O Delicious é um sítio de *social bookmarking*, ou seja, um gerenciador de sítios favoritos em que o usuário se cadastra e adiciona seus sítios favoritos da web. A função de armazenamento de sítios favoritos já existe nos navegadores praticamente desde que estes surgiram, porém, a grande vantagem dos sítios de *social bookmarking* é poder acessar os favoritos de qualquer computador e não somente do seu navegador. A função social se deve ao recurso de rede social, é possível adicionar pessoas que também estão cadastradas no sítio e se comunicar com elas por meio do envio de URLs, assim como explorar as etiquetas e sítios favoritos de outros usuários.

Segundo o próprio sítio, o Delicious “é um serviço de *social bookmarking* que permite que os usuários usem etiquetas, gerenciem e compartilhem páginas da web de uma fonte centralizada.” O serviço é gratuito e surgiu em 2003. Relata-se na literatura que o Delicious foi um dos primeiros a utilizar etiquetagem em sua arquitetura. O serviço define etiqueta como “uma simples palavra que você usa para descrever um sítio favorito. Diferente das pastas, você cria as etiquetas quando precisa delas e pode criar quantas quiser. O resultado é um forma melhor de organizar seus sítios favoritos e uma ótima forma de descobrir coisas interessantes na Web”, e nuvem de etiquetas como “uma lista de etiquetas em que o tamanho reflete a popularidade.” (DELICIOUS, 2009, tradução nossa). O sítio possui versão somente em inglês até o momento e pode ser acessado sem problemas em qualquer navegador. Sua página inicial pode ser visualizada na figura 14 a seguir.

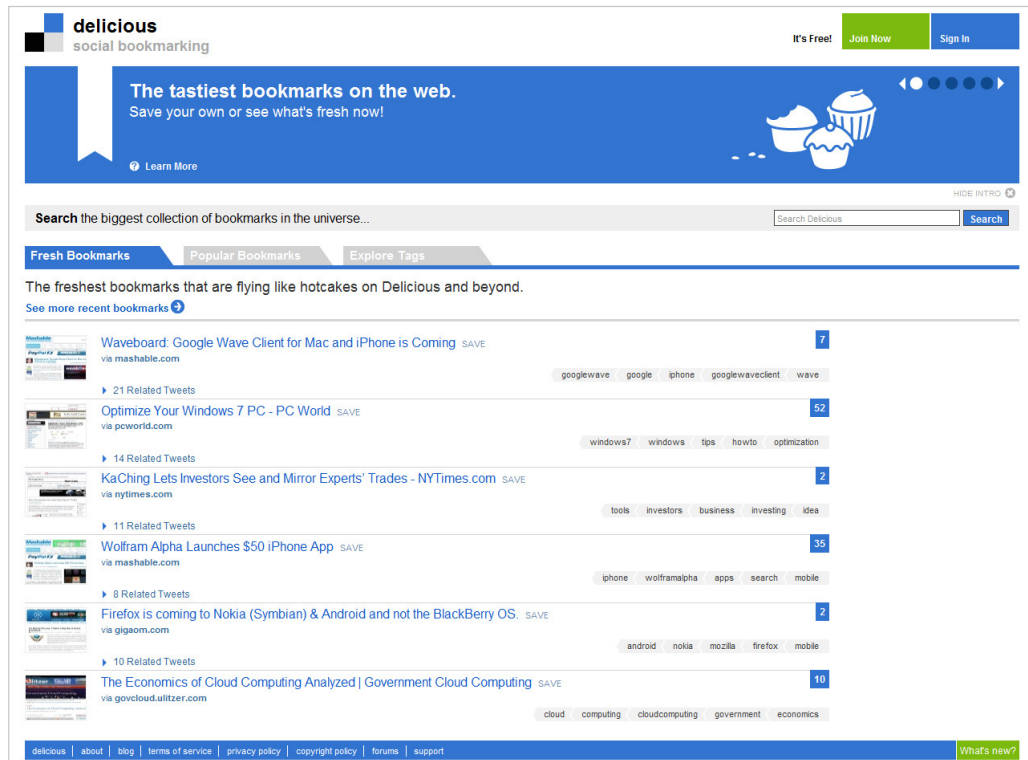


Figura 14 – Página inicial do Delicious

Seu uso é bastante simples, o usuário faz um cadastro e a partir daí pode adicionar URLs de seus sites favoritos da web. Há duas maneiras de adicionar um site. A primeira é por meio da opção “Save a new bookmark”, que remete a uma nova tela para preenchimento da URL a ser adicionada. Após, aparecem as opções de descrição do item adicionado: título, notas, etiquetas e enviar. Quando a URL possui um título, esse é preenchido automaticamente. As demais informações devem ser preenchidas pelo usuário. O preenchimento do título é obrigatório, os outros campos são opcionais. O campo de notas é um campo livre para o usuário fazer anotações, comentários ou adicionar qualquer outro tipo de informação. O último campo a ser preenchido é o “enviar”, função que permite enviar a URL adicionada para outros usuários do Delicious, bem como para usuários da rede Twitter ou por *e-mail*. Essa é uma função que permite a disseminação das informações adicionadas. No campo de adicionar etiquetas aparecem como opções etiquetas recomendadas e etiquetas populares relacionadas quando a URL já foi adicionada nos favoritos de outros usuários. Há também a opção de ver todas as etiquetas já utilizadas pelo usuário e selecionar desta lista as etiquetas que serão utilizadas para o novo site favorito adicionado. Por fim, é possível tornar o site adicionado privado, de forma que os outros usuários não tenham acesso às etiquetas utilizadas nem ao site adicionado, pois este fica oculto na lista pública. A outra forma de se adicionar um site é por meio de uma ferramenta que se instala ao navegador e funciona como um botão chamado “post to delicious.” Com esta ferramenta instalada, basta clicar neste botão para que o site aberto no navegador seja adicionado ao Delicious, abrindo na tela de descrição do recurso adicionado, com a URL já preenchida.

O Delicious permite também que o usuário “assine” uma etiqueta para que seja avisado de toda nova URL, adicionada por qualquer outro usuário do Delicious, que receba a etiqueta escolhida.

7.1.1 Organização e recuperação da informação

O principal elemento de organização da informação utilizado são as etiquetas. Os outros campos que aparecem na descrição do sítio adicionado poderiam ser utilizados para a organização da informação, porém, não há como ordenar por esses critérios na visualização da lista de favoritos. Esta fica disponível para o usuário em ordem cronológica do mais recente para o mais antigo, sendo possível reverter esta ordem. No campo de etiquetas da descrição do sítio, as etiquetas são separadas por um espaço. Dessa forma, quando ocorrem etiquetas formadas por mais de uma palavra, estas devem ser escritas justapostas (sem espaço) ou com separação por hífen, ponto, underline ou outro sinal. Caso o usuário escreva as palavras separadas o sistema entende que são etiquetas diferentes. Há um limite de 128 caracteres por etiqueta.

Na página do usuário aparecem, na lateral direita, duas listas de etiquetas como padrão: as 10 mais usadas e a lista completa com todas as etiquetas. A visualização das etiquetas pode ser em forma de lista ou de nuvem, e podem ser ordenadas alfabeticamente ou por frequência de uso. Há ainda uma opção chamada *tag bundle*, que permite a criação de categorias maiores para agrupar etiquetas semelhantes de acordo com algum critério definido pelo usuário. Caso o usuário crie esses pacotes de etiquetas, eles aparecem na visualização na lateral direita também, bem como as etiquetas soltas (*unbundled tags*). As etiquetas podem pertencer a mais de um pacote. Dessa forma elas podem ser organizadas em grupos maiores, o que facilita na visualização quando se tem muitas etiquetas, conforme pode ser visto na figura 15.



Figura 15 – *Tag bundles* com visualização em forma de nuvem

A recuperação da informação no sítio também baseia-se fundamentalmente em etiquetas. A página inicial do Delicious mostra os últimos sítios adicionados e possui uma aba para os sítios mais populares adicionados e outra para explorar as etiquetas. A principal forma de recuperação é por motor de busca (figura 16), não há diretórios com taxonomias ou classificações dos assuntos.



Figura 16 – Campo de busca principal

O campo de busca principal recupera informações do título, das notas e das etiquetas e é possível restringir a busca apenas em *bookmarks* selecionados (o do próprio usuário, o da rede de trabalho do usuário ou todo os de todos os usuários). Há também uma opção de refinamento da busca na tela de resultados, e é exibido um gráfico com a data em que foram inseridos os sítios adicionados recuperados. Na tela de resultados aparecem também as etiquetas relacionadas à etiqueta buscada. Na aba “explore tags” há um outro campo de busca que procura somente nas etiquetas, ilustrado na figura 17.



Figura 17 – Campo de busca por etiquetas

Além da busca, outra forma de encontrar informações é explorar as próprias etiquetas pela aba “explore tags”, que mostra várias etiquetas populares e, ao clicar em alguma delas, os sítios adicionados marcados com tal etiqueta são recuperados. Pode-se escolher entre a visualização dos mais recentes ou dos mais populares. A tela de resultados mostra quantas pessoas adicionaram cada sítio como favorito no Delicious, e ao clicar no número é possível saber quais usuários utilizaram aquela etiqueta. Há ainda, na tela de resultados (figura 18), uma coluna do lado direito com as etiquetas relacionadas ao assunto da etiqueta pesquisada.

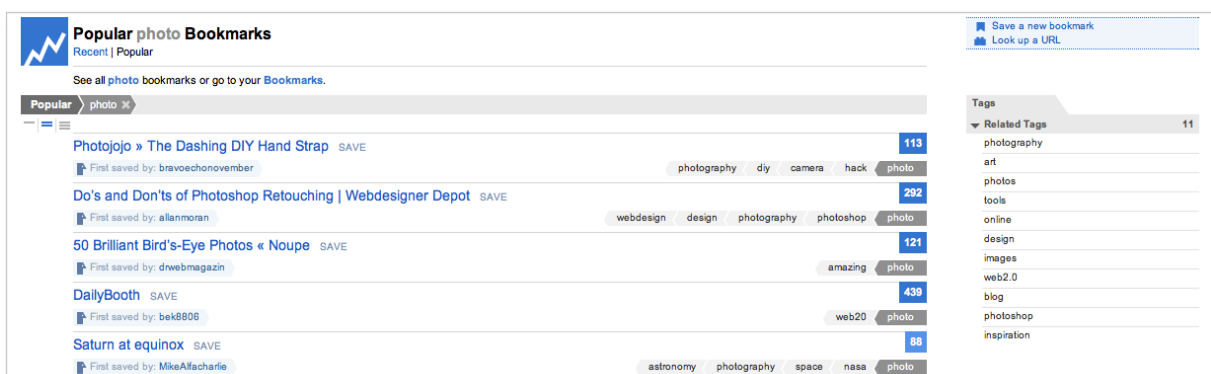


Figura 18 – Resultados da busca pela etiqueta ‘photo’

A outra possibilidade de busca é pela visualização da nuvem de etiquetas principal do Delicious (figura 19), com as mais populares, escolhendo-se o assunto a ser pesquisado a partir de então. O resultado aparece como na aba de explorar as etiquetas e também pode ser visualizado por popularidade ou pelos mais recentes.

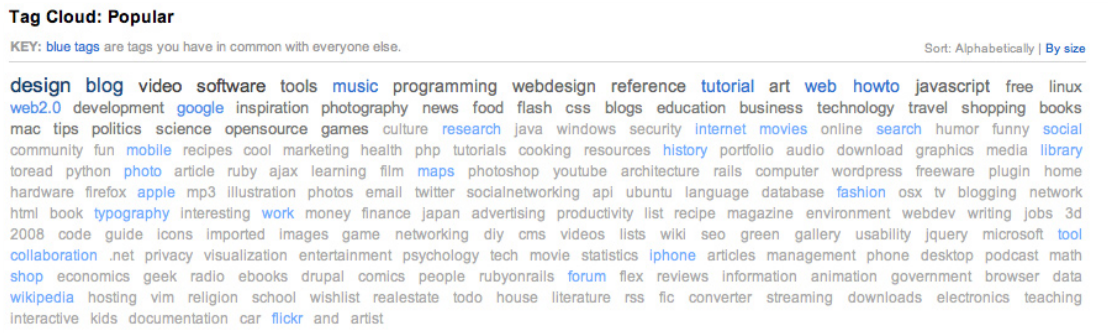


Figura 19 – Nuvem de etiquetas do Delicious

A folksonomia aqui formada se encaixa no tipo de folksonomia aberta, pois qualquer usuário pode colocar etiqueta nos objetos informacionais. O Delicious possui a maior coleção de *bookmarks* do mundo, segundo o próprio sítio.

7.2 Flickr

O Flickr é um sítio voltado para a organização, gerenciamento e disseminação de fotos e, recentemente, vídeos. Pela proposta dessa análise, o objeto informacional focado nesta seção será somente a foto. O sítio se define como um aplicativo de gerenciamento e compartilhamento e tem como objetivos principais ajudar as pessoas a disponibilizar conteúdos e permitir novas maneiras de organizar fotos e vídeos. O Flickr surgiu em 2003 e foi um dos primeiros, junto com o Delicious, a usar etiquetas para organização de seu conteúdo. O sítio está disponível em 8 idiomas, incluindo português, e não apresenta problemas de acessibilidade em diferentes navegadores. A página inicial do Flickr pode ser vista na figura 20 a seguir.

flickr

[Entrar](#)

[Criar sua conta](#)

É rápido e fácil com a sua ID Yahoo!

**Compartilhe suas fotos.
Explore o mundo.**

[BUSCAR](#)

5.831 uploads no [último minuto](#) · 601 coisas com a tag [linhas](#) · 2,6 milhão de coisas [georeferenciadas](#) este mês · [Fazer um tour](#)

Compartilhe **Organize**

Cortar, corrigir, editar **Explore...**

[Fazer um tour](#) Explore o [Blog do Flickr](#), o [Mapa-Múndi](#), [Localizador de Câmeras](#) ou uploads interessantes dos [últimos 7 dias](#).

Escolha um idioma: [繁體中文](#) · [Deutsch](#) · [English](#) · [Español](#) · [Français](#) · [한글](#) · [Italiano](#) · [Português](#)

[Blog do Flickr](#) · [Sobre o Flickr](#) · [Termos do Serviço](#) · [Sua privacidade](#) · [Política de Direitos Autorais](#) · [Regras da Comunidade](#) · [Denuncie aqui](#)

Copyright © 2009 Yahoo! Inc. Todos os direitos reservados.

Figura 20 – Página inicial do Flickr

O usuário pode se cadastrar no sítio gratuitamente ou adquirir uma conta paga com algumas vantagens como maior espaço para incluir arquivos. Há várias formas de enviar os arquivos para o sítio, como aplicativos próprios, envio pela página web do sítio, por *e-mail*, por aplicativos de celulares e *smartphones*, entre outros. Após enviar a foto, esta pode ser editada com uma ferramenta do sítio, caso o usuário deseje.

Em relação à função de rede social, o Flickr possibilita que se crie uma rede de relacionamentos, adicionando pessoas que podem se comunicar, enviar comentários às fotos e adicionar etiquetas, caso o usuário permita. A comunicação pode ser direta entre os usuários usando a ferramenta FlickrMail, que funciona como um email interno do Flickr e pode ser enviado para qualquer usuário. Na parte de compartilhamento, há opções de permitir que a foto seja visível por qualquer usuário, por grupos de usuários ou por usuários específicos autorizados. Os grupos do Flickr são temáticos e podem ser públicos ou particulares, com níveis de privacidade, licença de uso, tipo de conteúdo e nível de segurança definidos pelo usuário. Há também a possibilidade de convidar pessoas que não são cadastradas no Flickr para visualizar algum álbum específico.

7.2.1 Organização e recuperação da informação

Há algumas maneiras de se organizar as fotos no Flickr. Após enviar a foto para o sítio, há a opção de descrição, com os campos título, descrição e etiquetas. Nenhum deles é obrigatório. Essas informações podem ser incluídas ou alteradas a qualquer momento posteriormente. Podem também ser criados álbuns (grupos de fotos) e coleções (agrupamentos de álbuns ou outras coleções) e há a ferramenta “Organizr”, onde os álbuns e coleções são criados e que permite fazer modificações na descrição de lotes de fotos, como por exemplo trocar etiquetas, data e hora. Há ainda a opção de organizar as fotos geograficamente, também pelo Organizr, arrastando-as para o mapa da localidade onde a foto foi tirada. A informação da localidade, chamada *geotag*, aparece na página da foto.

Os álbuns e coleções citados como forma de organizar as fotos adicionadas no Flickr funcionam como super classes e têm uma abrangência temática maior. Para uma descrição específica de cada foto, o mais adequado é o uso de etiquetas. Segundo o Flickr (2009), “As tags são como palavras-chave ou títulos que você adiciona a uma foto para facilitar encontrá-la posteriormente.” As etiquetas adicionadas aos recursos podem ser editadas ou excluídas na página de gerenciamento de etiquetas, que apresenta as etiquetas mais utilizadas pelo usuário (com um limite de 150 etiquetas) em ordem alfabética e permite esse gerenciamento.

Os campos de título e descrição também podem ser entendidos como elementos de organização. As etiquetas são separadas por espaço, porém, há como criar uma só etiqueta com 2 palavras, usando aspas. Há um limite de 75 etiquetas por foto. O sistema captura automaticamente metadados como data e modelo da câmera utilizada quando se envia a foto, e esses dados aparecem na página da foto como informações adicionais. A figura 21 mostra os dados que aparecem após a descrição da foto.

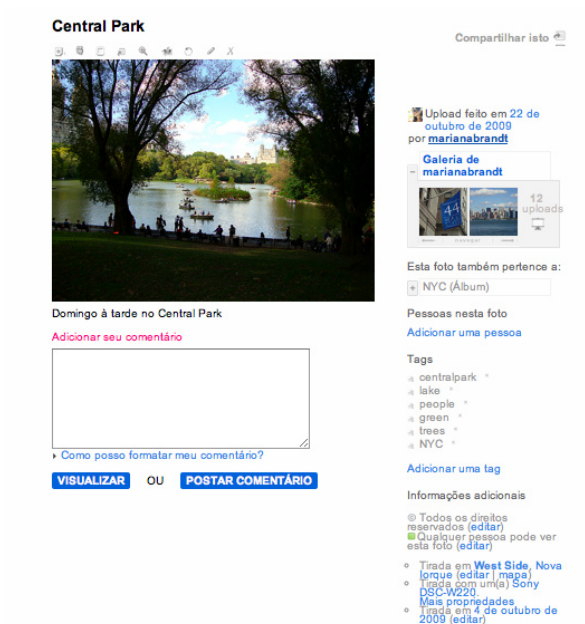


Figura 21 – Página da foto no Flickr

O usuário pode ainda criar uma exposição escolhendo 18 fotos para serem incluídas e mostradas ao público em um único lugar. É uma forma de se destacar algumas fotos da coleção do usuário. Recentemente o Flickr implementou mais uma forma de descrição dos recursos, que é chamado de “pessoas na foto” e permite identificar membros do Flickr que aparecem nas fotos e vídeos.

Para a recuperação das fotos no Flickr, a forma mais destacada é o motor de busca, assim como no Delicious. A recuperação no Flickr pode ser dividida em dois grandes grupos: fotos próprias e fotos de outros usuários. Para buscar nas próprias fotos, pode-se usar o Organizr. Dentro do Organizr, é possível buscar em suas próprias fotos pelas etiquetas ou por texto completo, somente nas fotos, somente nos vídeos ou em fotos e vídeos, e delimitar a busca por data de inclusão, conteúdo sem etiqueta, conteúdo com ou sem *geotag* e conteúdo fora de um álbum e ainda por filtros de privacidade.

Para buscar fotos de outros usuários, há como opções o motor de busca que aparece em várias páginas, ou a busca pelas etiquetas. O buscador da página inicial e os que aparecem em outras páginas têm como padrão a busca nos títulos e nas etiquetas. Há possibilidade de limitar a busca em partes específicas do sítio:

- *Uploads* de todos
- dos seus contatos
- dos seus amigos
- da sua galeria
- por grupos
- por membros do Flickr
- por localização

Há também a possibilidade de configurar o motor de busca para procurar em campos determinados, clicando-se em ‘buscar’ mas com a caixa de texto em branco. Abrirá uma página de busca mais detalhada (figura 22) onde é possível definir se a busca será em texto completo, somente nas etiquetas e ainda uma opção para busca avançada.



Figura 22 – Busca detalhada no Flickr

Esta última opção traz uma tela com vários critérios e formas de ordenar os argumentos

de busca que podem ser combinados para uma busca mais específica. Inclui as seguintes opções para os argumentos: todas as palavras, frase exata, qualquer palavra e nenhuma dessas palavras, e para os critérios: busca em texto completo, apenas nas etiquetas, por filtros de privacidade, por coleção (incluindo fotos licenciadas pelo Creative Commons), por conteúdo (captura de telas, arte, foto, vídeo, animação, etc), por mídias (foto, vídeo e ambos) e por data de inclusão no Flickr ou de quando a foto foi tirada. Em vários dos critérios citados há ainda sub-critérios que podem ser definidos. Ou seja, há uma ampla gama de possibilidades para a busca de objetos no Flickr. O motor de busca do Flickr trabalha com os operadores booleanos AND, OR e NOT, sendo padrão o AND.

Outra forma de recuperação de imagens no Flickr é pelo menu Explorar, onde é possível encontrar fotos por data, hora, exposições, fotos analógicas, etiquetas, lugares, câmera usada e o chamado “interestingness.” Este último é um recurso funciona como uma compilação das fotos mais interessantes do Flickr, com base em alguns critérios, como explicado no sítio:

Existem diversos elementos que tornam algo ‘interessante ‘ (ou não) no Flickr. De onde vêm as taxas de cliques; quem faz comentários e quando; quem favorita; suas tags e muitas outras coisas que mudam constantemente. O interestingness muda com o tempo, à medida que conteúdos e histórias mais fantásticas são adicionados ao Flickr. Adicionamos algumas páginas (e alteramos outras existentes) para ajudá-lo a explorar o conteúdo mais interessante do Flickr. (FLICKR, 2009).

Para a navegação nas etiquetas, há a opção “explorar tags”, que mostra uma nuvem de etiquetas mais populares (figura 23). Ao clicar na etiqueta desejada, todas as fotos marcadas com aquela etiqueta são recuperadas, sendo mostradas das fotos mais recentes para as mais antigas, por data de inclusão no sítio. Esta página mostra também as etiquetas mais usadas nas últimas 24 horas e na última semana.

The screenshot shows the 'Explorar / Tags' section of the Flickr website. It features two columns of 'Tags quentes' (Hot Tags) for the last 24 hours and the last week. Below these is a search bar with the text 'Saltar para:' and a search button labeled 'IR'. The main part of the image is a large, dense cloud of tags, where each tag is a word in a different color and size, representing its popularity. The most prominent tags include 'art', 'beach', 'japan', 'wedding', 'nature', 'city', 'family', 'festival', 'paris', 'travel', 'vacation', 'water', 'white', 'winter', 'york', and 'zoo'.

Figura 23 – Explorando as etiquetas do Flickr

O sítio possui ainda clusters de etiquetas, que são conjuntos de fotos com etiquetas semanticamente próximas, e permite que a busca seja feita nesse arranjo. A folksonomia formada no Flickr é restrita, pois nem todas as fotos podem receber etiquetas de qualquer usuário.

7.3 YouTube

O YouTube é um sítio para o armazenamento e compartilhamento de vídeos *online*. Foi lançado em fevereiro de 2005 e permite que o usuário disponibilize seus vídeos pela Internet por meio de *blogs*, sítios, dispositivos móveis como celulares e *smartphones* e por *e-mail*. Qualquer pessoa pode assistir aos vídeos no YouTube e não precisa ser cadastrado para isso. Segundo o próprio sítio, quanto mais as pessoas colocarem seus vídeos lá, mais poder o YouTube ganha para ser a grande ferramenta de transmissão do futuro. Pouco mais de um ano depois de ter sido lançado, o YouTube foi comprado pela Google. O sítio tem parcerias com várias empresas de produção de som e imagem como a BBC e a Sony Music. Está disponível em 19 línguas e não apresenta problemas de acessibilidade em diferentes navegadores. O sítio afirma ser a maior comunidade mundial de vídeos na Internet e sua página inicial é ilustrada na figura 24 a seguir.

The image shows the YouTube homepage interface. At the top left is the YouTube logo with the tagline "Broadcast Yourself™". To its right is a search bar with a "Pesquisar" button. Further right are links for "Criar conta" and "Fazer login". Below the search bar are navigation links: "Página inicial", "Vídeos", and "Canais". On the right side, there are links for "Inscrições", "Histórico", and an "Enviar" button.

The main content area is divided into two sections:

- Spotlight: Get Animated (exibir todas)**: A section titled "MovieSpotlights" celebrating International Animation Day. It features four video thumbnails:
 - SLEEPING BETTY (ENTIRE)**: 1 ano atrás, 352 exibições, rating 5 stars, channel nfb.
 - Sita Sings the Blues**: 4 dias atrás, 20325 exibições, rating 5 stars, channel questioncopyright.
 - Fantaisie in Bubblewrap**: 4 semanas atrás, 113434 exibições, rating 5 stars, channel metcallovesyou.
 - Archon Defender**: 1 mês atrás, 25640 exibições, rating 5 stars, channel dave4096.
- Recomendado para você (exibir todas) Saiba mais**: A grid of recommended videos:
 - Kim Kardashian inspired Make-up**: 7 meses atrás, 300273 exibições, rating 5 stars, channel pixiwoo.
 - U2 3D**: 1 ano atrás, 1179471 exibições, rating 5 stars, channel NationalGeograph.
 - Eyeliner Tutorial - Pencil. Crea...**: 2 anos atrás, 1701239 exibições, rating 5 stars, channel MichellePhan.
 - **SNOWFLAKE NAILS****: 10 meses atrás, 460645 exibições, rating 5 stars, channel julieg713.
 - U2 - Stuck In A Moment You**: 2 anos atrás, 458089 exibições, rating 5 stars, channel universalmusicgro.
 - U2 - New Years Dav**: 2 anos atrás, 3267504 exibições, rating 5 stars, channel universalmusicgro.
 - MAC Sugar Sweet & Mv doos**: 7 meses atrás, 28762 exibições, rating 5 stars, channel TheCurrentCustom.
 - Heavy Gothic Look**: 1 ano atrás, 487762 exibições, rating 5 stars, channel kuupo1207.

On the right side of the page, there is a large video player with a play button and the text "Watch In Widescreen". Below the player is a "Publicidade" section with a button that says "Clique e saiba mais.". At the bottom right, there is a yellow box with the text "O que há de novo" and a link to "Leia mais no nosso blog".

Figura 24 – Página inicial do YouTube

Para adicionar vídeos ao sítio, o usuário deve se cadastrar e criar uma conta no YouTube, o que pode ser feito gratuitamente. “Ao se registrar, os usuários podem enviar e compartilhar vídeos, salvar favoritos, criar listas de reprodução e fazer comentários sobre vídeos.” (YOUTUBE, 2009). Depois de cadastrado o usuário passa a possuir um canal no YouTube e pode também assinar outros canais dos quais deseja receber informações de atualizações, além de encontrar amigos que tem canal no YouTube, customizar seu perfil e, entre outras opções de personalização, comunicação e compartilhamento, adicionar seus próprios vídeos.

Ao adicionar um vídeo, uma tela de descrição de conteúdo é apresentada ao usuário, que deve preencher as seguintes informações: título, descrição, etiquetas, categoria (escolhida em uma lista pré-definida) e opções de compartilhamento. As etiquetas podem conter mais de uma palavra e são separadas por vírgula na entrada dos dados. Além disso, há opções de descrição de data e local, que pode ser escolhido a partir de um mapa do Google maps, com pesquisa de localização. É possível deixar o vídeo visível por qualquer pessoa que acessar o YouTube, o que é recomendado pelo sítio, ou restringir a um grupo de até 25 pessoas definidas pelo usuário. A página de descrição do vídeo pode ser conferida na figura 25 a seguir.

The screenshot shows the YouTube video description page for "Unknown Caller - U2 360 Tour". At the top, there are navigation links: "Minha conta" (with a dropdown arrow), "Detalhes do vídeo: Unknown Caller - U2 360 Tour", and a "Retornar à página anterior" button. Below these are links for "Informações & configurações", "AudioSwap", "Anotações", and "Legendas".

The main content area is divided into several sections:

- Salvar alterações** ou cancelar
- Informações do vídeo**:
 - Título:** Unknown Caller - U2 360 Tour
 - Descrição:** Unknown Caller at U2 360 Tour - Giants Stadium - NY (24 sept. 2009)
 - Palavras-chave:** U2 "360 tour" NY "Giants Stadium" "Unknown caller" 2009
 - Categoria:** Música
- Miniatura do vídeo**:
 - Three thumbnail images are shown.
 - Text: "Escolha uma imagem estática acima para representar seu vídeo nos resultados de pesquisa e em outras exibições. É possível escolher uma imagem estática diferente clicando nela. Observação: pode demorar até seis horas para que sua imagem seja atualizada."
- Opções de divulgação e compartilhamento**:
 - Privacidade:**
 - Compartilhar seu vídeo com o mundo (Recomendado)
 - Privado (pode ser visível para até 25 pessoas)
 - Comentários**

On the right side of the page:

- Video Player:** Shows the video title "Unknown Caller - U2 360 Tour" and a play button. The progress bar indicates 0:00 / 0:34.
- Data e mapa:**
 - Data:** 24 Setembro 2009. Buttons for "Hoje" and "Limpar".
 - Local do mapa:** Giants Stadium. Buttons for "Pesquisa" and "Limpar".
 - Altitude (em pés):** Buttons for "Definir" and "Limpar".
 - Map:** A map of the New York City area with a red pin at Giants Stadium. A speech bubble says "Este local foi gravado." Buttons for "Mapa", "Satélite", and "Híbrido".

At the bottom of the map, it says "Dados do mapa © 2009 Google; Tele Atlas - Termos de utilização".

Figura 25 – Página de descrição do vídeo

É possível ainda, mudar o áudio e adicionar legendas e anotações ao vídeo, bastando navegar pelas abas da tela ilustrada na figura 25.

7.3.1 Organização e recuperação da informação

Conforme visto na seção anterior, a organização do conteúdo dos vídeos do YouTube é feita com base em metadados gerados pelos usuários. O YouTube possui três opções principais de navegação: página inicial, vídeos e canais. Na página inicial são apresentados os vídeos em destaque, os recomendados para o usuário (com base nas últimas buscas feitas) os que estão sendo assistidos agora e os mais populares por categoria. Já na página “Vídeos”, há um menu com as categorias dos vídeos, conforme mostra a figura 26 a seguir.

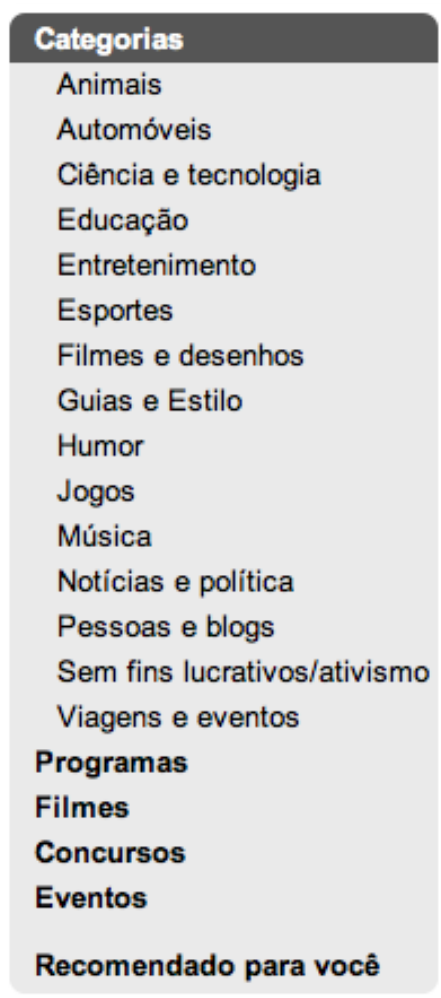


Figura 26 – Categorias de vídeos no YouTube

As categorias “Programas” e “Filmes” possuem ainda subcategorias mais específicas de gêneros como comédia, drama, ação, terror, etc. Nas páginas específicas de cada categoria são

mostrados vídeos mais vistos, mais populares, mais comentados, mais recentes ou outras opções que o usuário escolher. Essas opções podem mudar conforme a categoria escolhida, por exemplo, na categoria “Filmes”, há opção de ordenação alfabética.

A terceira página de navegação encontrada é a de “Canais.” Esta página também possui um menu de categorias, diferentes das categorias de vídeo, e possui opções de visualização por canais mais vistos ou com mais usuários inscritos. As três páginas mostram no topo o buscador principal do YouTube, que faz busca em qualquer das categorias, vídeos e canais, e o usuário pode escolher a visualização dos resultados de acordo com vários critérios, como pode ser visto na figura 27 a seguir.

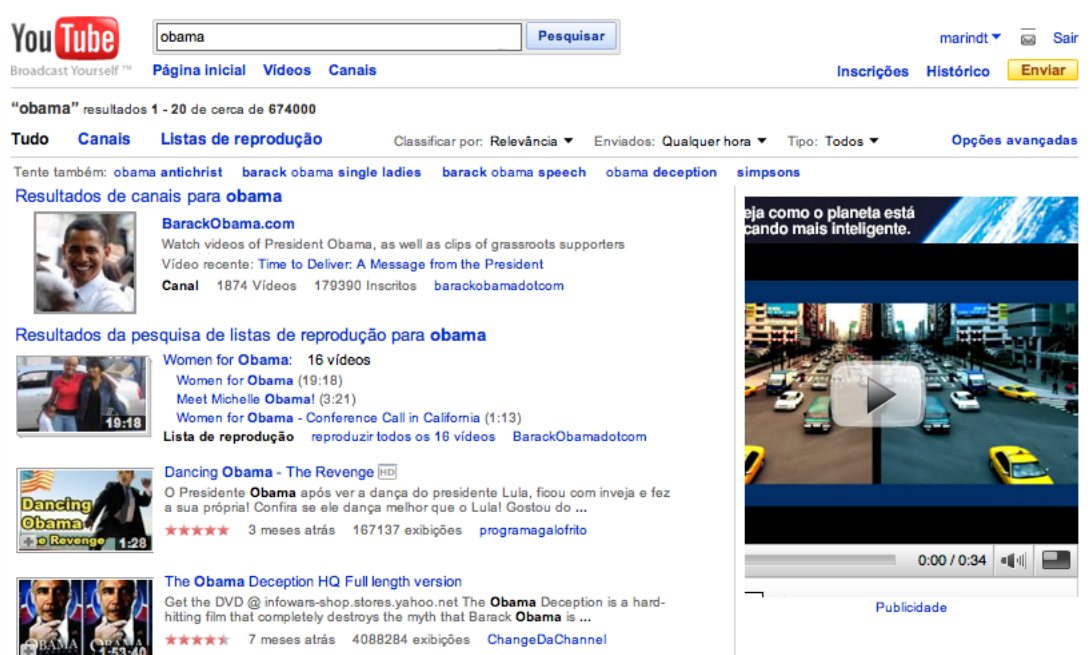


Figura 27 – Resultados de busca no YouTube

O YouTube tem também uma ferramenta de busca avançada em que o usuário pode pesquisar combinando vários critérios, como ilustra a figura 28.

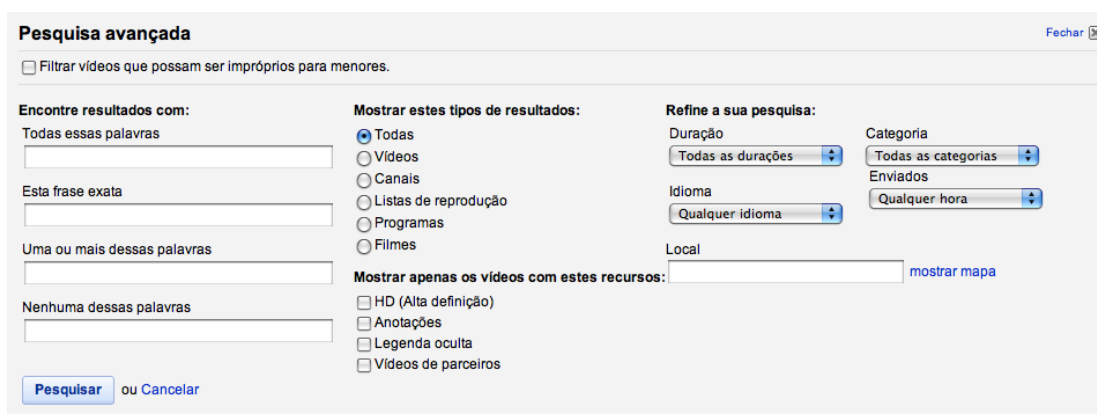


Figura 28 – Busca avançada no YouTube

As buscas no YouTube são feitas pela indexação de todos os campos de descrição do vídeo preenchidos pelo usuário. Na página dos vídeos, as etiquetas adicionadas são *links* que levam a outros vídeos que receberam a mesma etiqueta, assim como a categoria escolhida. Não há, mesmo na busca avançada, opção de busca somente nas etiquetas ou em qualquer outro campo específico. O sítio utiliza o operador AND como padrão em suas buscas, porém, os outros operadores booleanos não são utilizados. A folksonomia aqui é restrita já que apenas o usuário que postou o vídeo pode colocar etiquetas e o YouTube não mostra a representação gráfica da nuvem de etiquetas.

7.4 LibraryThing

O LibraryThing é um sítio para a organização de livros pessoais ou acervos de pequeno porte, formando-se um catálogo. Além disso, o sítio permite o contato com outros usuários cadastrados, formando-se grupos em torno de assuntos de interesse ou por afinidade literária. Assim, as pessoas que estão lendo a mesma coisa podem se comunicar, trocar idéia, discutir, etc. O sítio se define como um serviço *online* que ajuda as pessoas a catalogarem seus livros facilmente. O serviço é gratuito para a inclusão de até 200 livros. Há opções de pagamento de anuidade ou de uma taxa só para a vida toda. Nesses casos, não há limite de inclusão de livros. Há ainda uma opção para contas institucionais. A figura 29 mostra a página inicial do LibraryThing.

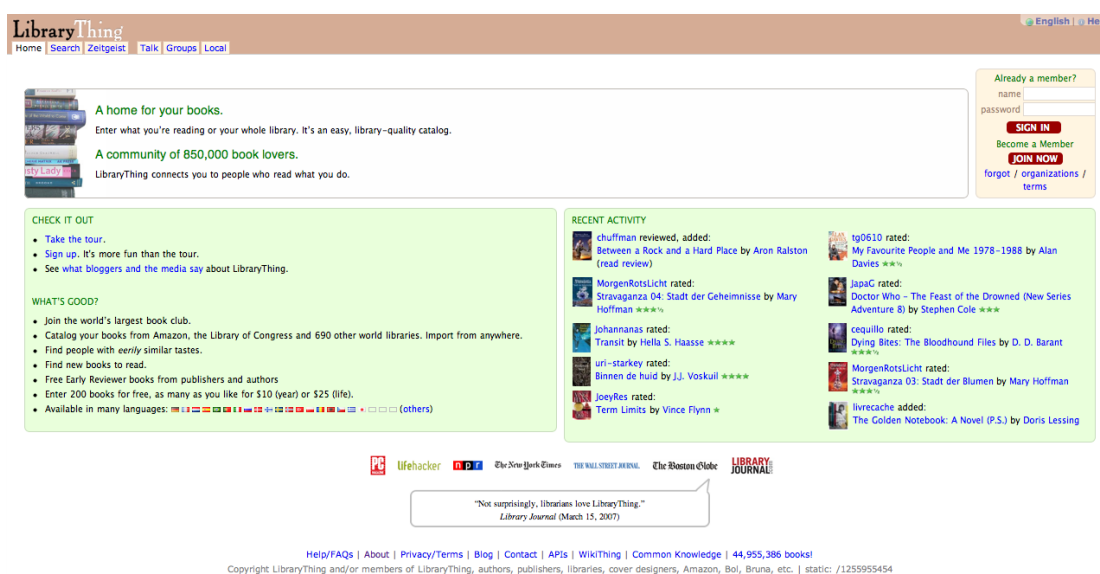


Figura 29 – Página inicial do LibraryThing

Segundo o LibraryThing, ele é o maior clube de leitura do mundo. O sítio começou em 2006 e está disponível somente em inglês. Existe uma função de migração de dados, usando o

protocolo z39.50, de catálogos de mais de 500 bibliotecas em todo o mundo, incluindo a Library of Congress, e do sítio de comércio eletrônico Amazon. Segundo o próprio sítio, o LibraryThing é um espaço social incrível e costuma ser definido como o “Facebook dos livros.” O serviço faz ainda recomendações de leitura aos usuários, baseado nos catálogos de outros usuários com itens semelhantes.

O cadastro é bem simples, requer apenas um nome de usuário e uma senha. Há opção de privacidade em que os livros adicionados não podem ser vistos pelos outros usuários. Para adicionar os livros a sua conta, o usuário deve procurar por título, autor, International Standard Book Number (ISBN) ou outro campo e o sítio recupera as informações em catálogos de biblioteca ou no Amazon, a critério do usuário, e gera uma lista de resultados. A partir daí, o usuário escolhe o livro que quer adicionar e pode editar as informações do item incluído.

7.4.1 Organização e recuperação da informação

Ao adicionar um livro em seu catálogo pessoal, os metadados de onde o registro foi migrado já são adicionados automaticamente, e podem ser editados posteriormente pelo usuário. O sistema divide a descrição dos livros por obra e por item, utilizando para obra o mesmo conceito⁴ do modelo Functional Requirements for Bibliographic Records (FRBR)⁵ da International Federation of Library Associations and Institutions (IFLA). Os campos disponíveis para a obra são título, autor, quantas pessoas possuem, resenhas, número de classificação da Library of Congress, número de classificação de Dewey, assuntos e outras línguas, enquanto os do item são: título, autor, avaliação, resenha, outros autores, etiquetas, coleções, editora, data de publicação, ISBN, número de classificação da Library of Congress, número de classificação de Dewey, língua primária, língua secundária, língua original, comentários, comentários privados, resumo, número de exemplares, Book Crossing Identification (BCID), data de aquisição, período de leitura, membro, data de entrada do registro e fonte do registro. Apenas os campos do item podem ser editados. A página com as informações do registro contém ainda várias outras informações, como outras capas, resenhas de outros usuários, recomendações, e outras, conforme pode ser visto na figura 30.

⁴ Obra: noção abstrata de uma criação intelectual ou artística.

⁵ <<http://www.ifla.org/en/frbr-rg>>

Extremely Loud and Incredibly Close: A Novel
by Jonathan Safran Foer

Members	Reviews	Popularity	Average rating	Conversations
6,169	138	270	★★★★ (4.19)	93

Book Information: Edit book | Review | Delete book

Extremely Loud and Incredibly Close: A Novel
Jonathan Safran Foer
Mariner Books (2006), Paperback, 368 pages

Member: nanaaa Rating: ★★★★★
Collections: Your library Tags: NYC, WTC, 11/9 (edit)

Members: all members
Tags: numbers | show all

2007 21st century **9/11** America **American** American literature book club childhood coming of age contemporary contemporary fiction **death** Dresden family favorite **fiction** grief literature loss **New York** **New York City** novel NYC own read Roman TBR unread USA

Recommendations: view history | edit this data

LibraryThing members' description

Book descriptions

Open Shelves Classification

Buy, borrow or swap: Abebooks, Alibris, Amazon.com (top edition), Barnes & Noble, BookFinder.com, Indiebound, WorldCat. Swap this book (13/255)

Member Giveaway: Give this away

Popular covers: (see all 46 covers)

Rating: Average: ★★★★★ (4.19)
0.5: 7
1: 14
1.5: 10
2: 57
2.5: 25
3: 233
3.5: 100
4: 620
4.5: 123
5: 848

Figura 30 – Registro de livro no LibraryThing

Como os campos de descrição bibliográficas são preenchidos automaticamente ao se adicionar um item, o usuário fica responsável apenas por completar ou alterar alguma informação que desejar e escolher as etiquetas que vai adicionar. Conforme já visto anteriormente, é interessante que a escolha de termos de assunto para obras de ficção seja feita pelo usuário, devido ao caráter subjetivo. No LibraryThing as etiquetas são separadas por vírgula e são limitadas a 30 caracteres. É possível adicionar ou alterar etiquetas de vários livros de uma vez. Há ainda a opção de combinar etiquetas, que é uma forma de controle de vocabulário pois permite agrupar etiquetas com o mesmo significado e que possuem o mesmo uso.

A recuperação dos registros pode ser feita de várias maneiras. Há opções de busca somente na biblioteca do usuário ou no sítio todo. A busca somente na biblioteca do usuário percorre, por padrão, os campos de título, autor, etiqueta, ISBN, data, fonte, número de classificação Library of Congress e Dewey. Pode-se também incluir os campos resenha, comentário e assunto. Já a busca no sítio todo recupera por obra (título, autor e ISBN), autores, etiquetas, grupos, revendedores e membros por localidade, conforme aparece na figura 31 a seguir.

Search your library

Your library: Search

Examples:
greek history
"greek history"
greek -history
greek hist*
tag:greek
title:Iliad source:Library of Congress
review:Excellent

Fields: all, most, title, author, tag, isbn, date, icon, dewey, source, subject.
By default it searches "most" (all, minus review, comment and subject).

Site search

Works (Title/Author/ISBN): Search

Authors: Search

Tags: Search
(Separate with commas)

Talk messages: Search

Groups: Search

Venues by location: Search

Venues: Mobile, AL | 5 miles | address or postal code

Members and locations: Search

"Also On": Search

On: Search

by location | by name

Figura 31 – Campos de busca no LibraryThing

O mecanismo de busca do LibraryThing opera por definição com o operador AND e os outros operadores booleanos não estão disponíveis. A informação pode ser recuperada também pela navegação nas nuvens formadas, acessíveis pelo *link* “clouds.” Há opção de visualização das nuvens somente das informações do usuário ou das informações de todo o sítio. Na opção de todo o sítio podem ser visualizadas as nuvens de etiquetas de autores e do que o sítio denomina “conhecimento comum”, que engloba as 200 séries mais populares, os 200 lugares mais mencionados nas obras, os 200 nomes de personagens mais citados, as 100 premiações literárias mais citadas, os 50 editores mais citados, entre várias outras listas. Já para as nuvens apenas do usuário, há opção de nuvem de etiquetas, nuvem de autores e espelho de nuvens, que são as etiquetas que outros usuários deram para os livros desta coleção.

Ao se clicar em uma etiqueta da nuvem, são recuperadas informações da etiqueta como a quantidade de vezes que ela foi utilizada, quantos usuários já utilizaram, as obras que mais comumente recebem esta etiqueta, os usuários que mais utilizam esta etiqueta, as obras que mais recentemente receberam esta etiqueta, as etiquetas relacionadas, os assuntos relacionados e as chamadas “tagmashes”, que são etiquetas formadas por 2 assuntos, uma espécie de cabeçalho pré-coordenado. Por meio da função de combinar etiquetas, o sítio faz uma tipo de controle ortográfico, pois nas informações da etiqueta aparece a indicação de que aquela etiqueta inclui outras com grafias semelhantes, como pode ser observado na figura 32.

Tag info: Comics

Includes: Comics, COMIC BOOK, comic, comic books, comic_book, comic_books, comicbook, comicbooks, comix

Figura 32 – Informações da etiqueta

O LibraryThing informa em sua página que possui atualmente 59 milhões de etiquetas. Estas geram uma folksonomia aberta.

7.5 Amazon

O Amazon é um sítio de comércio eletrônico que vende vários tipos de produtos e, diferente dos anteriores, não é, por definição, uma rede social. Apesar disso, o usuário cadastrado pode criar um perfil e adicionar outros usuários em sua lista de contatos. A página inicial do Amazon é ilustrada na figura 33.



Figura 33 – Página inicial do Amazon.com

Em 2005 o Amazon implementou um sistema de etiquetagem que possibilita aos consumidores colaborarem para uma das formas de organização do sítio, colocando etiquetas nos produtos. Para isso o usuário deve ser cadastrado como cliente no Amazon. Depois de se cadastrar, basta procurar um produto e, na página de informações, há um espaço para as etiquetas em que o usuário pode adicionar e ver as já existentes para o produto buscado (figura 34).



Figura 34 – Etiquetas no Amazon.com

Há também a opção de editar as etiquetas ou inseri-las acessando uma nova janela “Tag this product”, conforme ilustrado na figura 35.



Figura 35 – Etiquetas no produtos

Cada usuário pode adicionar até 15 etiquetas por produto e é possível ver todos os produtos em que ele colocou etiquetas. As etiquetas são separadas por vírgula e é possível criar termos compostos. Forma-se uma folksonomia aberta, já que qualquer usuário pode colocar etiquetas nos produtos do Amazon.

7.5.1 Organização e recuperação da informação

O Amazon possui várias formas de organização da informação em seu sítio. Os produtos estão separados por departamentos, como em uma loja real. Há 12 grandes departamentos, como livros; filmes, músicas e jogos; brinquedos, crianças e bebês; eletrônicos; casa e jardim, etc. Esses departamentos são divididos em sub-departamentos, com categorias mais específicas, como pode ser visto na figura 36.

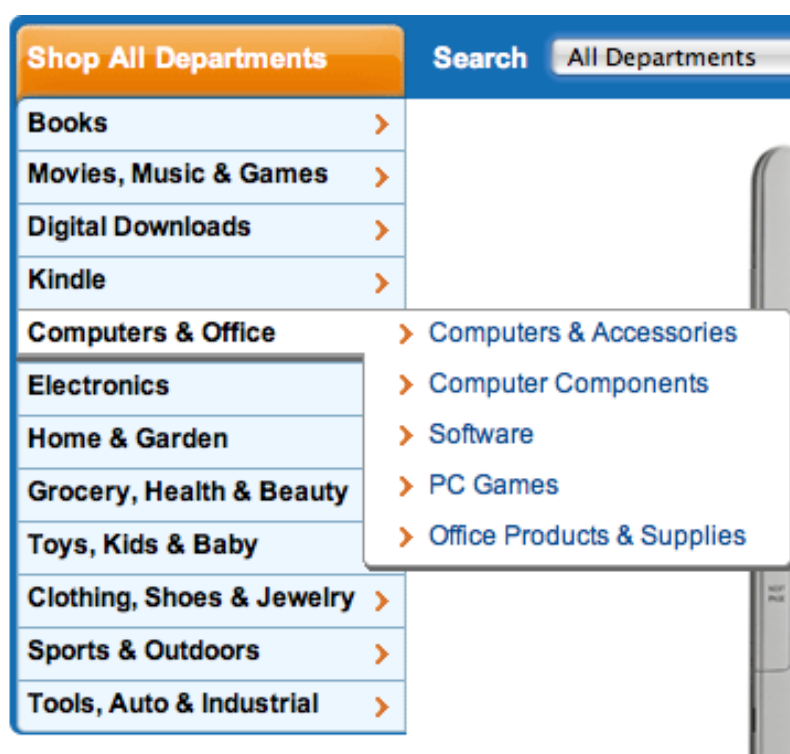


Figura 36 – Taxonomia dos departamentos no Amazon.com

Em cada departamento há ainda listas mais específicas de tipos de produtos, listas dos mais vendidos, das tendências, das marcas vendidas, produtos masculinos, femininos e para crianças, entre outras, que variam de acordo com o departamento. Nas páginas dos produtos há várias informações específicas como descrição do item, especificações técnicas, recomendações, resenhas, avaliação, comentários de outros usuários, lista de assuntos, lista de etiquetas, e outras informações que variam conforme o produto.

Para encontrar produtos no sítio, além das várias taxonomias dos departamentos, é possível recuperar informação pelo mecanismo de busca, que permite escolher em qual departamento buscar os produtos, ou buscar em todos (figura 37). O padrão de busca é recuperar todos os termos do argumento utilizado e os operadores booleanos não são utilizados.



Figura 37 – Mecanismo de busca no Amazon

A navegação pelas taxonomias permite que se encontrem coisas mais específicas, pois funcionam como filtros de busca restringindo os produtos recuperados, como pode ser visto na figura 38 a seguir.



Figura 38 – Filtros de busca no Amazon

Como pode ser observado, o Amazon possui a informação do sítio organizada de diferentes maneiras e isso possibilita vários tipos de busca. Por esse motivo, a etiquetagem não se destaca como forma de recuperação de informação no sítio. Apesar disso, é possível usar as etiquetas para encontrar produtos, e há inclusive uma busca somente pelas etiquetas e formação de uma folksonomia representada pela nuvem de etiquetas mais populares. (figuras 39 e 40)



Figura 39 – Busca somente nas etiquetas



Figura 40 – Nuvem de etiquetas do Amazon

7.6 Last.fm

O Last.fm é um serviço de recomendação de músicas baseado no gosto do usuário. O sítio funciona também como uma rede social pois há interação entre os usuários, é possível entrar em fóruns de discussão, além de outras formas de comunicação como deixar comentários em páginas de bandas, adicionar conhecidos em seu perfil e anunciar eventos musicais. O serviço começou a ser pensado no final dos anos 1990, lançou seu sítio em 2002 e foi comprado pela empresa de mídia CBS em 2007. As gravadoras Warner e Sony BMG autorizaram que o Last.fm adicionasse seus catálogos de músicas ao sítio. O serviço é gratuito, está disponível em 12 línguas e pode ser acessado de qualquer navegador. Há opção paga de assinatura da rádio Last.fm. Sua página inicial é ilustrada a seguir (figura 41).

The image shows the homepage of Last.fm. At the top, there is a red navigation bar with the logo 'lost.fm' and menu items: 'Músicas', 'Rádio', 'Eventos', 'Vídeos', 'Tabelas', and 'Comunidade'. On the right side of the bar, there are links for 'Inscreva-se' and 'Login', and a search bar with the text 'Pesquisar música'. Below the navigation bar, there is a main content area. On the left, there is a large dark blue banner with the text 'A Last.fm recomenda músicas, vídeos e shows com base no que você ouve.' and a form to search for an artist. On the right, there is a section for 'Rádio visual personalizada' with a small image of a person's face. Below these, there is a horizontal menu of music genres: 'Todas', 'Rock', 'Indie', 'Electronic', 'Pop', 'Metal', 'Hip-hop', 'Jazz', 'Classical', and 'Country'. The 'Eventos de Shows na sua área' section lists several events with photos and details, including Nina Becker, The Exploited, and Mariana Aydar. There are also promotional banners for 'FEI VESTIBULAR 2010' and 'Downloads gratuitos de músicas de Faixas gratuitas populares'.

Figura 41 – Página inicial do Last.fm

Após se cadastrar, o usuário passa a possuir um perfil que deve ser personalizado. Depois deve instalar uma ferramenta do sítio que adiciona automaticamente as músicas executadas do computador ou tocador de mp3 do usuário a seu perfil. Com base nessas informações o sistema recomenda músicas, bandas e pessoas ao usuário. Segundo o sítio, “Quanto mais a Last.fm aprende sobre o seu gosto musical, melhor ela fica. Receba recomendações de músicas, vídeos, MP3s gratuitos e listas de shows — tudo de acordo com seu gosto musical.” (LAST.FM, 2009).

7.6.1 Organização e recuperação da informação

A informação é organizada no Last.fm de acordo com vários critérios. Há guias no topo da página que levam para as seguintes opções: música, rádio, eventos, vídeos, tabelas e comunidade, conforme ilustra a figura 42.



Figura 42 – Guias de navegação do Last.fm

Em cada guia, há categorias de acordo com o melhor critério para cada tipo de informação. Por exemplo, na guia “Música”, o menu de categorias se refere aos gêneros musicais. Há ainda outras formas de organização que podem ser mostradas ao usuário, como músicas mais populares, músicas do momento e músicas populares no país do usuário. As bandas são exibidas em ordem de popularidade seguindo esses critérios, e, para cada uma, há recomendações de artistas similares. Esta página mostra, ainda, recomendações de artistas com base na biblioteca do usuário. Em cada uma das guias, a informação é organizada seguindo critérios que melhor se aplicam. Na guia “Eventos”, há opções de organização por data, local, gênero musical, entre outras. A guia “Tabelas” corresponde a estatísticas das músicas mais tocadas, artistas principais, faixas preferidas, etiquetas mais utilizadas, entre outras opções. Os dados são relativos à última semana. Com exceção das guias “Tabela” e “Vídeos”, há opções de procurar por mecanismos de busca da própria página.

Nas páginas dos artistas, os usuários podem colaborar tanto com o conteúdo, adicionando fotos do artista e escrevendo descrições da banda e das músicas e biografias, em formato wiki, quanto com a organização da informação, adicionando etiquetas aos artistas, músicas (faixas), e álbuns (figura 43).



Figura 43 – Etiquetagem no Last.fm

É possível ainda criar descrições para as etiquetas, o que seria um tipo de nota explicativa. Na parte de fotos, o sistema se integra ao Flickr, também por meio das etiquetas. Basta que o usuário coloque uma etiqueta de máquina determinada em sua foto do Flickr que ela vai automaticamente para a galeria do Last.fm.

A recuperação da informação no sítio pode ser feita por meio de navegação nas guias, categorias, etiquetas e outras opções, além de mecanismos de busca espalhados por vários lugares do sítio. O buscador principal do sítio apresenta opções de busca por músicas, artistas, álbuns, faixas, etiquetas e gravadoras, como pode ser visto na figura 44.

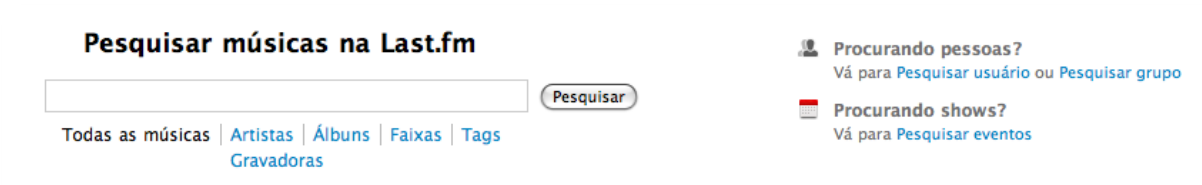


Figura 44 – Busca no Last.fm

O mecanismo de busca do Last.fm permite que os operadores booleanos AND, OR e NOT sejam utilizados. O padrão da busca mostra resultados com a maioria dos termos pedidos, mas não usa o operador AND. Por exemplo, uma busca por 5 termos sem operadores entre eles pode trazer no resultado registros encontrados com 4 dos termos pedidos. Há opção de busca somente nas etiquetas, mostrando no resultado a definição do termo, etiquetas relacionadas, artistas relacionados mais populares, eventos, entre outras informações (figura 45).

Figura 45 – Página da etiqueta “rock” no Last.fm

O sítio apresenta também nuvens de etiquetas de artistas, formadas por uma folksonomia aberta, como no exemplo da figura 46 a seguir.

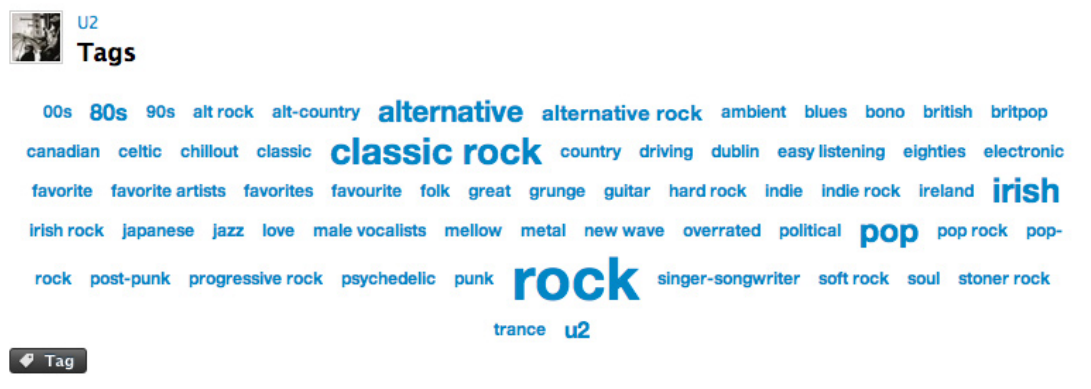


Figura 46 – Nuvem de etiquetas de artista no Last.fm

Além de ser um elemento para a organização e recuperação da informação, outro benefício da etiquetagem para o Last.fm é permitir que os usuários ouçam uma rádio escolhendo uma etiqueta. Dessa forma, todas as músicas que receberam a etiqueta serão tocadas naquela rádio. É possível combinar até três etiquetas diferentes para a escolha da rádio, pela opção “rádio multi-tags”, permitindo assim uma especificidade maior.

7.7 Steve

O Steve é um projeto relacionado com etiquetagem para objetos de museus e obras de arte. Inclui tanto a prática da etiquetagem quanto o estudo dessa forma de descrição dos objetos, além do desenvolvimento de ferramentas que permitam esta prática. Os envolvidos no projeto acreditam que “a etiquetagem social pode fornecer novas formas de descrever e acessar coleções de heranças culturais e encorajar o visitante a um comprometimento com os objetos da coleção.” (STEVE, 2009, tradução nossa). O projeto foi fundado em 2005, devido à preocupação com o acesso as coleções *online* dos museus de arte. O sítio do projeto está disponível em inglês, como pode ser visto na figura 47.

HELLO MY NAME IS Steve

STEVE: THE MUSEUM SOCIAL TAGGING PROJECT

search... search

Home Research Software Steve in Action FAQ Links & Resources Forum Steve Tagger

Welcome to the Steve Project

Steve is a collaboration of museum professionals and others who believe that social tagging may provide profound new ways to describe and access cultural heritage collections and encourage visitor engagement with collection objects. Our activities include researching social tagging and museum collections; developing open source software tools for tagging collections and managing tags; and engaging in discussion and outreach with members of the community who are interested in implementing social tagging for their own collections.

Steve as Research

The project team is engaged in systematic research into how social tagging can best serve the museum community and its visitors. Our current [research project](#), "T3: Text, Tags, Trust," a partnership with the University of Maryland's School, is funded, in part, by a National Leadership Grant for Research from the U.S. Institute of Museum and Library Services. [Results](#) of a recently-completed IMLS research project, "Researching Social Tagging and Folksonomy in the Art Museum," are presented in the Research section of this website. Data collected at <http://tagger.steve.museum> provides a testbed for researching our hypotheses about social tagging.

Steve as Software

To support our research, the steve team has developed a suite of open source tagging tools. The current version of the steve tagger tool is available at the project's [SourceForge](#) site, and members of the community are encouraged to download and install the tool. At its core, the tagger is a php web application that can publish a basic online collection and record tags. The project has issued a rich API that allows the integration of the tagger tool with existing tools and supports new development efforts. By itself, tagging is not technically difficult, but it is our intention to provide a easily-deployed common platform to allow users to share the costs of future tool development.

tranquil
interior
water jug
woman
shadow

Figura 47 – Página inicial do Steve

O relatório de umas das pesquisas do projeto está disponível no sítio e aponta o estudo que foi feito entre outubro de 2006 e dezembro de 2008. A autora explica o problema do acesso a coleções *online* e investiga a possível contribuição da etiquetagem nesse contexto. O estudo foi aplicado em vários grandes museus norte americanos como o Metropolitan Museum of Art, o Boston Museum of Fine Arts, o San Francisco Museum of Modern Art, entre outros.

Para colocar etiquetas nas obras de arte disponíveis no sítio do projeto Steve, o usuário não precisa se cadastrar. O cadastro no sítio é opcional e gera dados demográficos para a pesquisa. Caso o usuário não queira se cadastrar, basta clicar em “start tagging” e escolher o conjunto de obras em que deseja colocar etiquetas (figura 48).

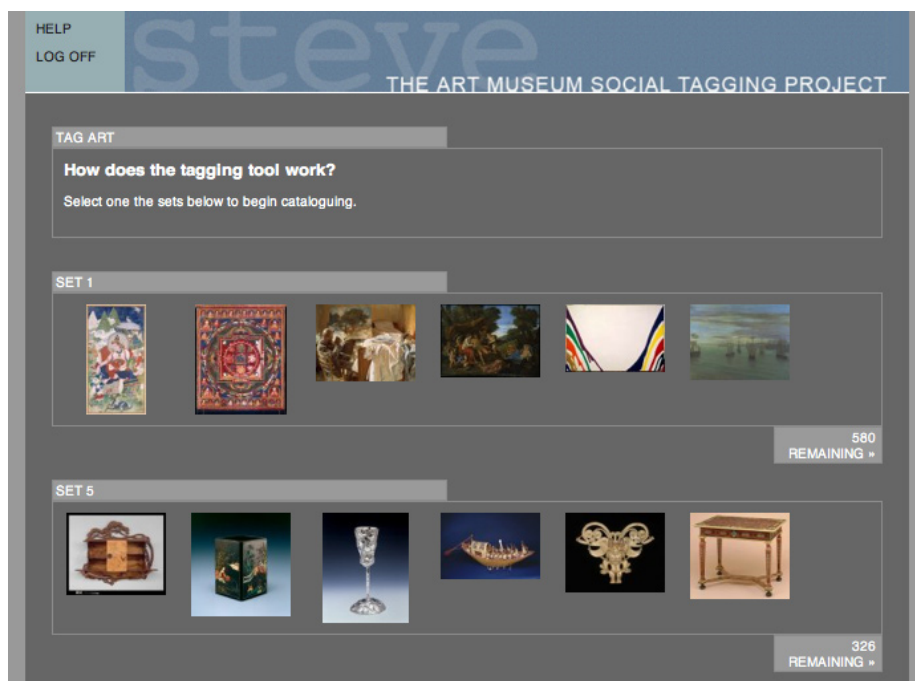


Figura 48 – Etiquetagem de obras de arte no projeto steve.museum

Ao escolher um conjunto de obras, o usuário é levado a uma tela para adicionar as etiquetas às obras daquele conjunto. É possível ampliar a imagem clicando nela. A tela de etiquetagem pode ser vista na figura 49 a seguir.

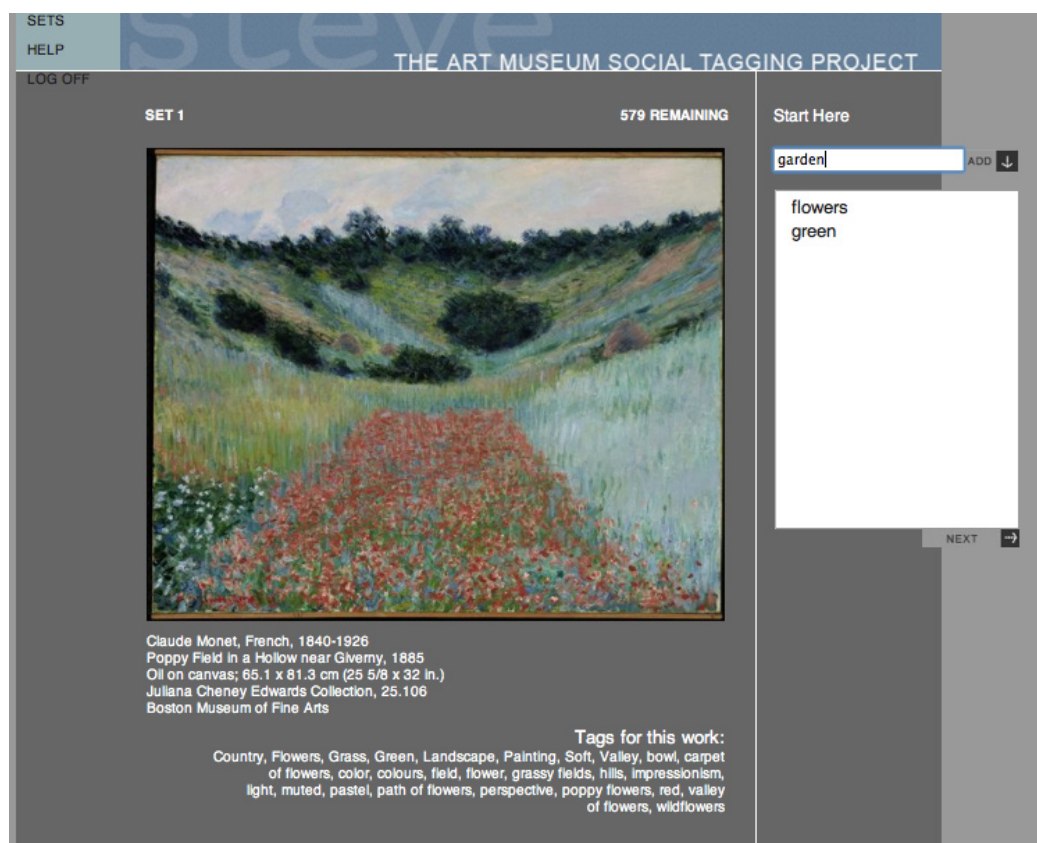


Figura 49 – Etiquetagem no steve.museum

Após adicionar as etiquetas, ainda é possível editá-las ou excluí-las antes de seguir para a próxima obra.

7.7.1 Organização e recuperação da informação

O projeto Steve visa gerar etiquetas adicionadas pelos usuários a objetos de vários museus, e o sítio *steve.tagger* é apenas uma ferramenta para coletar essas informações. Assim, a informação etiquetada não está disponível na página do projeto, que funciona apenas como um meio de acesso ao usuário que deseja obter informações e colaborar para a etiquetagem das coleções.

A equipe do projeto informa que, pelo fato do projeto original ter sido orientado a pesquisa, não há um local público onde as etiquetas e imagens possam ser vistas. Nos trabalhos recentes, as etiquetas coletadas em cenários reais podem ou não ser disponibilizadas ao público, e esta é uma decisão que cabe aos museus que fizeram parte do projeto. O projeto Steve por si só não planeja fazer isso, seu objetivo é auxiliar os museus a que desejam adotar a prática e as ferramentas para etiquetagem, individualmente ou de forma colaborativa. A folksonomia que pode ser gerada nesse projeto seria do tipo aberta.

7.8 Zigtag

O último sítio escolhido para esta análise é o Zigtag, que tem a mesma função do Delicious, de armazenar sítios favoritos (*bookmarks*), porém, com uma grande inovação que é o controle de vocabulário nas etiquetas. O sítio se define como uma ferramenta de *social bookmarking* inteligente. O aspecto inteligente se deve às etiquetas definidas, o que permite agregar valor semântico aos termos e proporciona recuperação de informação mais relevante. É também uma rede social, os usuários podem adicionar amigos a seu perfil e comunicar-se com eles por meio dos grupos e fóruns. Com base em afinidade de etiquetas utilizadas, o sítio recomenda usuários para a lista de amigos. O sítio foi lançado em 2008 e está disponível somente em inglês. É possível acessá-lo em diferentes navegadores, porém, para ter acesso a todas as funcionalidades do sítio é recomendado o Firefox. A figura 50 a seguir mostra a página inicial do Zigtag.

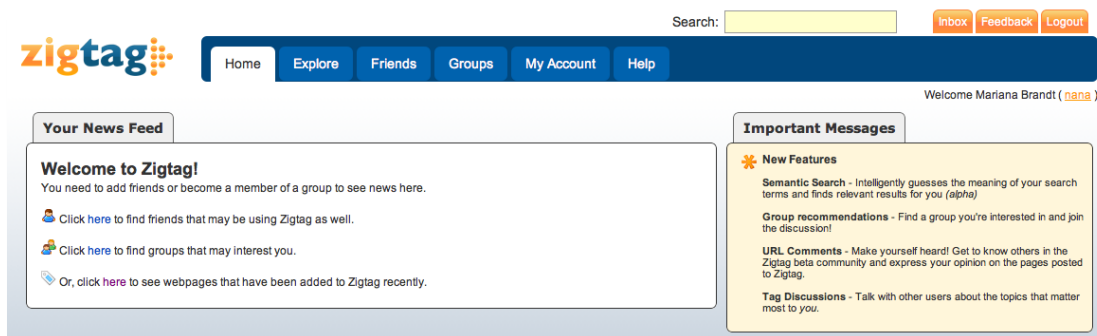


Figura 50 – Página inicial do Zigtag

O usuário deve se cadastrar no sítio e então criar um perfil. Para adicionar sítios favoritos, há uma opção de colocar o botão “Zigtag – tag” no navegador, e então basta entrar no sítio desejado e clicar neste botão. Uma janela é aberta para o usuário preencher informações relativas ao sítio, com os campos título, URL e etiquetas. O campo título já vem preenchido se a URL possuir um, e o campo URL vem sempre preenchido. Nessa janela de preenchimento dos campos, aparecem ainda as etiquetas sugeridas para aquele sítio. A figura 51 a seguir mostra como é feita a descrição dos sítios favoritos no Zigtag.

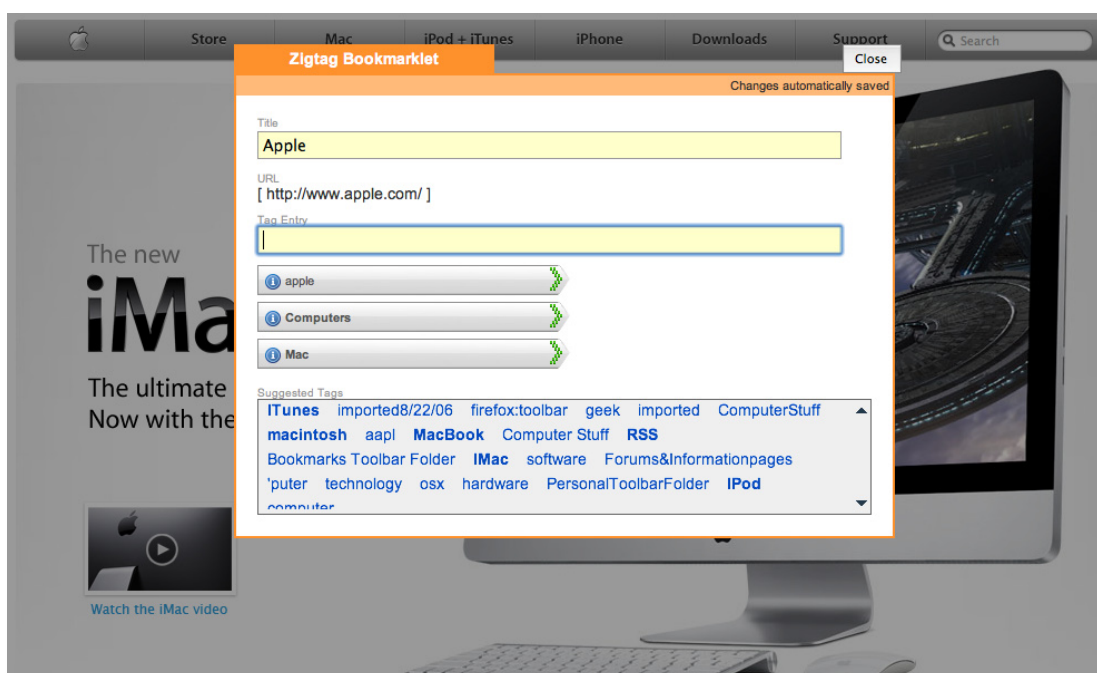


Figura 51 – Etiquetagem no Zigtag

As etiquetas adicionadas podem ser públicas, privadas ou restritas a um grupo de pessoas escolhidas pelo usuário. Caso o usuário utilize o navegador Firefox, há opção de instalar uma ferramenta com vários recursos do Zigtag direto no navegador, conforme pode ser visto na figura 52 a seguir.

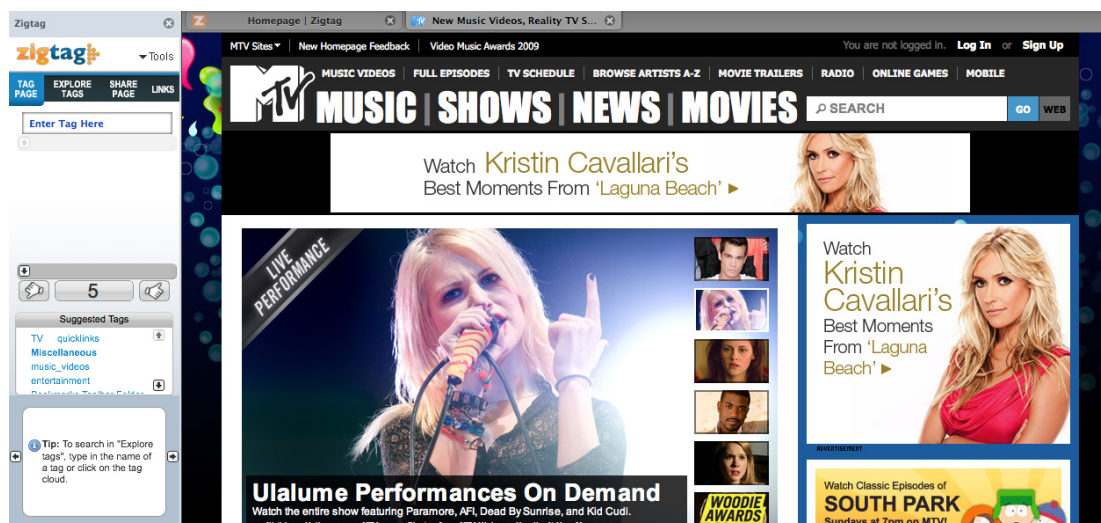


Figura 52 – Ferramenta do Zigtag no Firefox

É possível também importar bookmarks da ferramenta de favoritos do navegador ou do Delicious para o Zigtag, e exportar os bookmarks do Zigtag em forma de arquivo. Outra forma de adicionar bookmarks é no menu “my account”, opção “add tag.” Abrirá uma tela para o usuários colocar a URL do sítio que deseja adicionar e então aparecerá a janela de preenchimento das etiquetas.

7.8.1 Organização e recuperação da informação

A organização da informação no Zigtag é baseada nas informações fornecidas pelo usuário, principalmente as etiquetas. No campo de descrição “tags”, o usuário adiciona suas etiquetas, uma a uma. Não há separação por vírgula, o usuário deve entrar com uma etiqueta de cada vez. Caso coloque várias palavras, mesmo que separadas por vírgula ou outro sinal, o sistema registra como uma etiqueta só.

A página inicial do sítio mostra os últimos bookmarks adicionados e outros podem ser encontrados pela guia de navegação “explore” e também pelo mecanismo de busca. A guia de navegação “explore” oferece as opções: todas as etiquetas, minhas etiquetas, etiquetas dos amigos, *blog* do Zigtag e fórum do Zigtag, como visto na figura 53 a seguir.

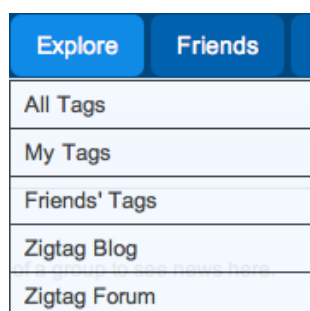


Figura 53 – Guia “explore” do Zigtag

As opções referentes a etiquetas mostram as nuvens de etiquetas correspondentes, e, ao clicar em na etiqueta desejada, todos os sítios que a receberam são mostrados. A etiqueta escolhida é destacada, e, para as etiquetas que possuem definição, esta é mostrada no lado esquerdo da tela, bem como o *link* para a discussão da etiqueta, e ainda as etiquetas relacionadas, como pode ser visto na figura 54.

Figura 54 – Página de resultados de busca por uma etiqueta

Além da navegação na nuvem de etiquetas, a recuperação de informação no sítio é possível pelo mecanismo de busca, que pode ser encontrado em todas as páginas. O motor de busca do Zigtag não trabalha com operadores booleanos.

A inovação apresentada pelo Zigtag está na possibilidade de gerenciar as etiquetas, definindo-as e resolvendo problemas de ambigüidades, sinonímia e diferenças de ortografia. Por meio da opção “manage my tags” do menu “my account”, o usuário é levado a uma tela com todas as etiquetas já utilizadas por ele, em ordem decrescente de frequência de uso (figura 55).

Tag (# times used)	Defined Tag Suggestions <i>Allows you to assign a definition to undefined tags</i>	Rename Tag <i>Use this to change one tag to a completely different tag</i>	Delete Tag <i>Delete all instances of this tag</i>
folksonomy (21)	Folksonomy	Select Tag Here [dropdown] [Rename]	[Delete]
tagging (18)	Tagging (Semantic Tagging) Tagging (Tag (metadata))	Select Tag Here [dropdown] [Rename]	[Delete]
motivation (14)	Motivation (Motivation (song)) Motivation Motivation (Motivation (episode))	Select Tag Here [dropdown] [Rename]	[Delete]
folksonomia (13)		Select Tag Here [dropdown] [Rename]	[Delete]
web2.0 (9)	WEB2.0 (Web 2.0)	Select Tag Here [dropdown] [Rename]	[Delete]
flickr (8)	Flickr	Select Tag Here [dropdown] [Rename]	[Delete]

Figura 55 – Gerenciador de etiquetas do Zigtag

É possível também buscar por uma etiqueta determinada para então gerenciá-la. Entre as opções de gerenciamento, o usuário pode renomear uma etiqueta, definir seu significado entre uma lista de definições ou excluí-la. Para definir uma etiqueta, basta clicar no significado que deseja atribuir, escolhendo na lista de sinônimos com qualificadores que aparecem na coluna ao lado, conforme pode ser visto na figura 56 a seguir.

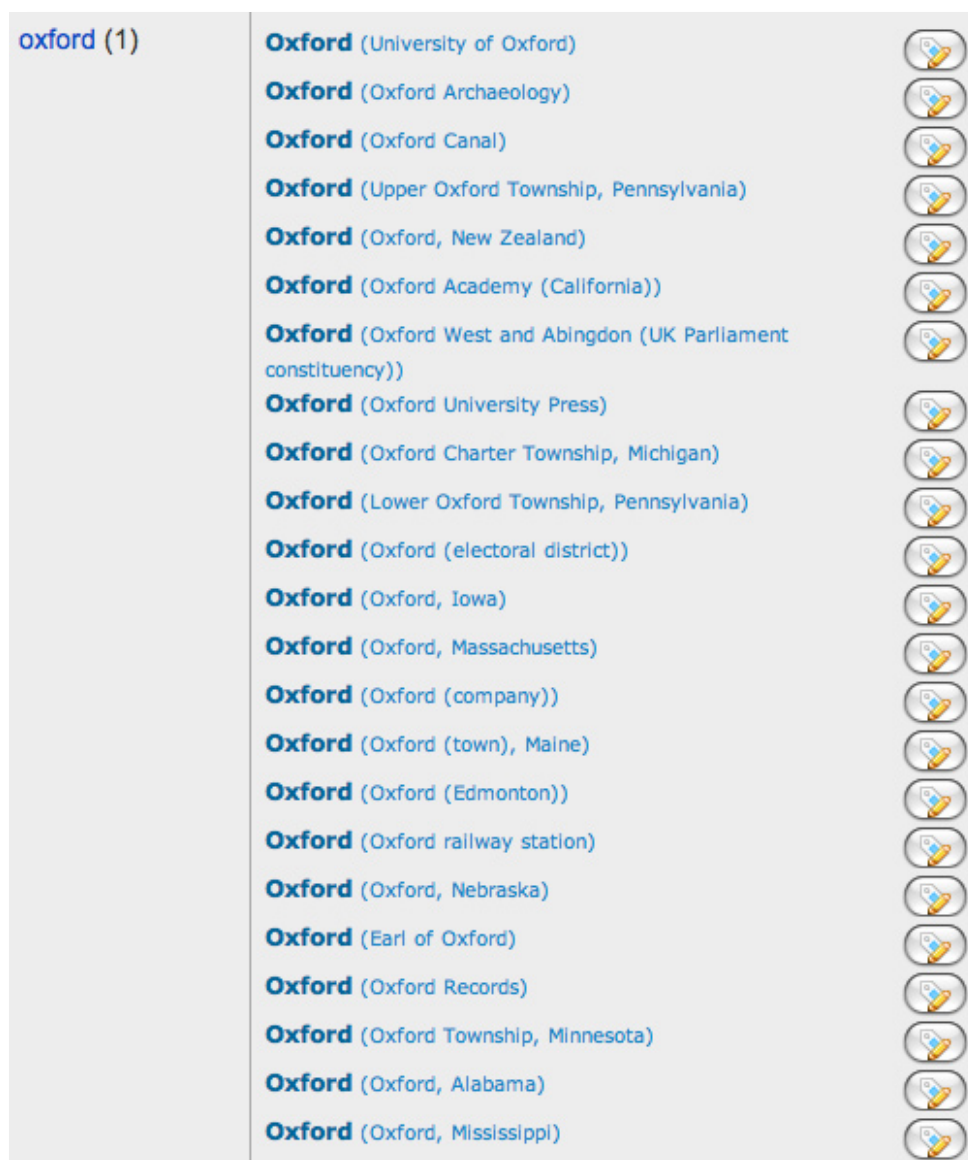


Figura 56 – Qualificadores de etiquetas no Zigtag

Pode-se ainda ver a definição da etiqueta, basta deixar o cursor do mouse parado em cima da etiqueta desejada (figura 57).

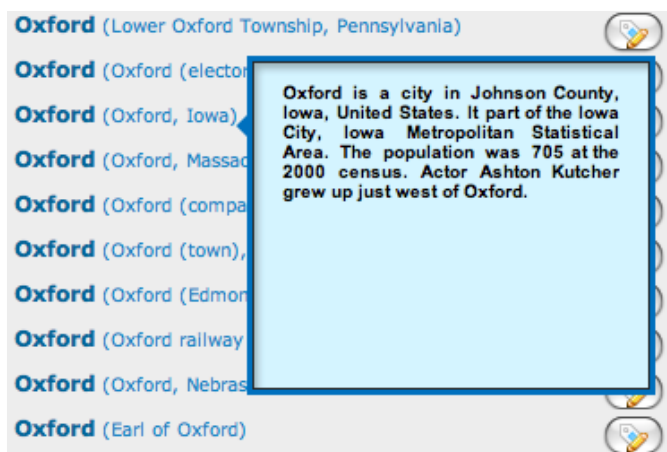


Figura 57 – Definição de etiqueta no Zigtag

As folksonomias formadas no Zigtag são abertas pois qualquer usuário pode colocar etiquetas nos sítios favoritos adicionados ao sistema.

7.9 Observações sobre os serviços descritos

Após a observação dos serviços é possível perceber a utilidade e a utilização da etiquetagem e da folksonomia em diversas formas, para diferentes objetos informacionais e com diferentes abordagens. Fica claro que para alguns dos sites analisados, como o Delicious e o Zigtag, a etiquetagem é elemento chave para a organização da informação e a recuperação é basicamente baseada em etiquetas. Já em outros, como o Amazon, o processo é apenas mais uma ferramenta que permite a busca de produtos, pois o sítio possui taxonomias diversas e mecanismos de busca como formas primárias de organização e recuperação de informação. No LibraryThing a abordagem é diferente, mesclando elementos tradicionais da descrição bibliográfica com a etiquetagem feita pelo usuário. Nesse serviço são geradas folksonomias diversas que permitem analisar tendências, descobrir novos livros e conhecer pessoas com interesses comuns.

Para os recursos visuais descritos no Flickr e YouTube, as etiquetas e os outros dados da descrição são a única forma de se encontrar um objeto informacional de acordo com seu conteúdo. Já para as músicas no Last.fm, a etiquetagem e folksonomias formadas não são essenciais para a organização do sítio, porém, permitem a descoberta de novas músicas, bandas, eventos e álbuns com base nas etiquetas adicionadas. O projeto Steve, apesar de não disponibilizar uma base de dados com recurso de busca das obras etiquetadas, traz uma grande inovação na descrição desse tipo de recurso e permite uma aproximação do usuário com o mundo da arte.

Além da etiquetagem e folksonomia, outros elementos influenciam no sucesso da recuperação da informação nos sistemas, como os operadores booleanos. Conforme citado, nem

todos os serviços utilizam esse recurso e quando utilizam as formas são variadas, não há um padrão e isso pode confundir o usuário. Outra coisa que poderia ser padronizada é a forma de entrada das etiquetas. Em alguns sistemas, conforme visto, as etiquetas são separadas por um espaço, em outras, por vírgula. Alguns permitem o uso de etiquetas compostas, outros entendem palavras diferentes separadas como etiquetas diferentes, e há ainda os que permitem etiquetas compostas, porém, com o uso de aspas na entrada dos dados. Isso confunde o usuário e tende a gerar mais problemas terminológicos nas etiquetas geradas.

A questão do controle do vocabulário também é algo que pode ser minimizado com recursos de gerenciamento de etiquetas, que nem todos os serviços possuem. Questões relativas ao próprio desenho das páginas, disposição dos elementos e infra-estrutura tecnológica também são fatores importantes nesse contexto, possibilitando experiências de uso melhores ou piores ao usuário que busca informação.

De qualquer forma, foi possível verificar a aplicabilidade de sistemas de etiquetagem e folksonomia em diferentes contextos e para diferentes objetos informacionais. A implementação desses sistemas nos sítios deve ser estudada e considerada como forma de organização e recuperação em sistemas de informação em que o usuário interage diretamente com os objetos informacionais.

Capítulo 8

8 Considerações finais e sugestões para pesquisa futuras

O problema da recuperação da informação em grande escala e de forma precisa é algo que está longe de ser solucionado. Observam-se vários esforços nesse sentido com o desenvolvimento de sistemas especializados em áreas específicas do conhecimento, mas para o grande repositório mundial que é a web, com suas características tão variadas, heterogêneas e seu volume infinito, os mecanismos de busca tradicionais, como o Google, continuam sendo o que há de mais eficiente para a recuperação da informação. Apesar disso, encontram-se problemas relacionados à precisão da informação recuperada dessa forma, principalmente quando a questão de busca é algo menos trivial que uma simples consulta de localização de fato.

A Web 2.0 mostra que a participação do usuário é fundamental para a sobrevivência dos serviços de informação *online*. Os processos colaborativos fazem com que o usuário se torne parte integrante e necessária da web e essa tendência está em pleno crescimento. Observam-se inclusive esforços de empresas públicas e privadas para entrarem nesse universo, numa tentativa de se aproximar de seus clientes. É natural que essa tendência também seja utilizada para a organização e recuperação de informação, o que acontece por meio de etiquetagem e folksonomia. A etiquetagem vem sendo utilizada para organização de conteúdos na web, gerando folksonomias que permitem a recuperação de informações. Essa forma de indexação coletiva, democrática, colaborativa e em linguagem natural, livre de regras, categorizações estritas e hierarquias, permite que qualquer pessoa participe do processo.

Após a análise desses processos de acordo com as quatro etapas propostas, foi possível identificar uma base teórica em que a etiquetagem e a folksonomia podem apoiar-se, no âmbito da Ciência da Informação. Com relação ao estudo da etiquetagem paralelamente à indexação de assuntos, foram encontradas várias características comuns às duas práticas. Ambos são processos cognitivos, com foco no usuário e permitem a recuperação da informação. Em áreas mais específicas como a indexação de imagens e de obras de ficção, fica clara a utilidade da etiquetagem – mesmo os processos tradicionais de indexação desses recursos apontam a necessidade da presença do usuário para sua realização. Algo que chama atenção é a proposta de indexação democrática feita nos anos 1990 por autores da Ciência da Informação. O proposto naquela época está muito próximo da etiquetagem que vem ocorrendo hoje na web. Por tudo isso acredita-se que o elemento “indexação” está presente e possui uma ligação estreita com o processo de etiquetagem.

Para o elemento seguinte, da análise das abordagens em Organização do Conhecimento, encontram-se, em maior ou menor grau, características da folksonomia em quatro abordagens: cognitiva, social, baseada no usuário e da recuperação da informação. É possível ainda considerar a folksonomia como um esquema de representação do conhecimento caso entenda-se que a unidade organizável depende da abordagem. Segundo propõem alguns autores, a unidade

organizável para a abordagem cognitiva é a estrutura cognitiva individual, e neste caso a folksonomia cumpre precisamente esse papel. Apesar de não possuir formalidades em sua estruturação, que é realizada automaticamente após a etiquetagem, é possível entender a folksonomia como uma estrutura de representação do conhecimento dos domínios em que está inserida.

Como todo esse processo depende fundamentalmente do usuário, este foi considerado o terceiro elemento de análise no percurso metodológico da pesquisa. A investigação da motivação para a etiquetagem mostrou que mais da metade dos usuários que responderam a pesquisa tem objetivos de organização e recuperação da informação ao realizar a tarefa. Apesar do estudo de usuário ter sido realizado em ambientes limitados e ter uma pequena quantidade de participantes, ele trouxe indícios de que as pessoas participam do processo colaborativo de etiquetagem com objetivos de organização e recuperação da informação etiquetada. Dessa forma, pode-se indicar que a peça-chave do sistema de etiquetagem, que é o usuário, está motivada de acordo com objetivos válidos para o proposto nesta pesquisa, sendo assim um elemento também presente na validação do modelo etiquetagem e folksonomia para a organização e recuperação da informação na web.

A aplicabilidade desse sistema em diferentes contextos e para diferentes tipos de objeto informacional é ampla e as possibilidades são muitas. Isso já era previsto pois até mesmo pela definição de Wal para folksonomia tem-se que os objetos etiquetados podem ser “qualquer coisa com URL.” Pela análise das diferentes aplicações verificou-se que a etiquetagem e a folksonomia são incluídas de forma a destacarem-se mais ou menos dependendo do sítio e do objeto informacional, o que se justifica pelo objetivo de cada sítio. Vale destacar aqui a necessidade de metadados para recursos visuais como fotos e vídeos. Nestes casos, as etiquetas cumprem bem este papel de descrição da imagem. Além disso, nos sistemas analisados para esses conteúdos, Flickr e YouTube, o usuário é responsável não só pela etiquetagem como também pelo fornecimento de todas as informações referentes aos recursos (título, categoria, etc). Ou seja, são sistemas que dependem do usuário para a organização de sua informação.

Tem-se então que, apesar de algumas limitações, a etiquetagem e a folksonomia possuem características suficientes para que sejam consideradas como modelo válido para a organização e recuperação da informação na web, com maior ou menor potencial, dependendo da natureza do sítio em que se insere e do objeto informacional com que se relaciona. Além de possuir elementos que devem estar presentes para a organização e recuperação da informação, a folksonomia deve ser considerada válida também por ser um sistema de simples utilização, que coloca o usuário à frente do sistema, e, além disso, é uma solução de baixo custo em sua implementação, manutenção, e atualização, se comparado aos sistemas tradicionais como ontologias e taxonomias.

Aqui cabe uma observação importante. Acredita-se que esse tipo de sistema só tem boas possibilidades de funcionar em sítios onde a participação do usuário de forma colaborativa é grande e há motivação do usuário para a etiquetagem. Ou seja, a implementação desse tipo de sistema pressupõe que o serviço de informação conheça seu usuário e tenha consciência de seu potencial para a realização da tarefa de etiquetagem.

Acredita-se ainda que o grande potencial da folksonomia para organização e recuperação de informação ainda não esteja sendo utilizado pelos sistemas de informação e que a recuperação de informação relevante pode melhorar muito com alguma tecnologia que permita tratamento dos dados inseridos pelos usuários nos sistemas.

Dessa forma, seria possível o que sugeriu Foskett nos anos 1980, quando o autor fala que o melhor sistema de informação é aquele projetado especificamente para cada usuário. Com o próprio usuário tratando o conteúdo de seu sistema, criando suas categorias pessoais por meio da etiquetagem, tem-se um sistema de gerenciamento pessoal de informação único e específico, gerando folksonomias também pessoais.

A revisão de literatura sobre folksonomia permitiu verificar que os autores ainda não entraram num consenso, tanto terminológico quanto conceitual, para a área. Há muita confusão entre os termos e entre os processos, e parece não haver grande interesse em uma definição mais formal que permita a criação de uma base teórica da área. Acredita-se que esta pesquisa permitiu, por meio de comparação com a base teórica da Ciência da Informação, algo que pode ser considerado o início de um entendimento mais formal dos processos de etiquetagem e folksonomia.

Como sugestões de pesquisas futuras, propõe-se então pesquisas e discussões teóricas sobre o assunto, que permitam a criação de uma base mais sólida para a área. Em relação à área de Organização do conhecimento, sugere-se um estudo que identifique o que faz com que um modelo seja considerado um esquema de representação do conhecimento, ou seja, quais os requisitos formais básicos que permitem tal definição. Sugere-se que se identifiquem esses requisitos nos sistemas tradicionais, como tesauros, taxonomias e ontologias, e, a partir daí, se faça uma comparação com as características da folksonomia.

Outro estudo que poderia ser realizado, em relação aos usuários de folksonomia, seria a ampliação do estudo de usuários sobre motivação para outros serviços, que trabalhem com outros tipos de recursos, como o YouTube, o LibraryThing e o Amazon. Mais um tipo de pesquisa que pode ser feita é a comparação das motivações para a etiquetagem em outros sítios que utilizem os mesmos tipos de recursos, o que permitiria uma análise comparativa entre objetos informacionais similares.

O uso da etiquetagem para indexação de obras de ficção também é algo que vale a pena ser estudado, com um projeto de aplicação em catálogos de bibliotecas já existentes e análise das etiquetas geradas e da efetividade da recuperação da informação por essa forma. Estudos de natureza terminológica a partir das etiquetas utilizadas também são considerados importantes para essa área. Além disso, sugere-se que estudos na área de indexação de imagens também considerem a folksonomia em seu escopo.

Referências

ABBOTT, N. J.; CALVERT, D. Psychology and neuroscience of indexing: what goes on in the mind of the indexer? *The indexer*. v. 25, n. 3, abril 2007.

ALMEIDA, M.B.; BAX, M.P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, v.32, n.3, p.7-20, set./dez. 2003.

AMARAL, S. Análise do consumidor brasileiro do setor de informação: aspectos culturais, sociais, psicológicos e políticos. *Perspectivas em Ciência da Informação*, Belo Horizonte, v.1, n.2, p. 207 - 224, jul./dez. 1996

AMES, M.; NAAMAN, M. *Why we tag*: motivations for annotation in mobile and *online* media. Proceedings of the SIGCHI conference on Human factors in computing systems, April 28-May 03, 2007, San Jose, California, USA. Disponível em: <<http://yahooresearchberkeley.com/blog/wp-content/uploads/2007/02/chi2007-Ames-whyWeTag.pdf>>. Acesso em: 03 out. 2008.

ARAÚJO, C. Fundamentos teóricos da classificação. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, n. 22, 2º sem. 2006

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 12676 – Métodos para análise de documentos – determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro: ABNT, 1992.

AUSTIN, D. W. *Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngues*. Brasília: Inst. Bras. de Inform. em Ciência e Tecnologia, 1993. 86 p.

BAEZA-YATES, R.; RIBEIRO-NETO, B. *Modern Information Retrieval*. New York: Addison Wesley Longman, 1999.

BATES, M. J. Indexing and Access for digital libraries and the Internet: human, database, and domain factors. *Journal of the American Society for Information Science*. v. 49, n. 13, p. 1185-1205, 1998.

BAX, M. Introdução às linguagens de marca. *Ci. Inf.*, Brasília, v. 30, n. 1, p. 32-38, jan./abr. 2001.

BEAUDOIN, J. Flickr image tagging: patterns made visible. *Bulletin of the American Society for Information Science and Technology*. p. 26-29 Oct./Nov. 2007.

BENBUNAN-FICH, R.; KOUFARIS, M. Motivations and contribution behaviour in social bookmarking systems: an empirical investigation. *Electronic Markets*, v. 18, n. 2, p. 150 - 160, maio 2008.

BERINSTEIN, P. do you see what I see? Image indexing principles for the rest of us. *Online Weston, Conn.* v. 23, n. 2, p. 85-86, 1999.

BERNERS-LEE, T. *Information Management: a proposal*. 1990. CERN Disponível em: <<http://www.w3.org/History/1989/proposal.html>>. Acesso em: 31 mar. 2009.

BERNERS-LEE, T. *The World Wide Web: a very short personal history*. May, 1998. Disponível em: <<http://www.w3.org/People/Berners-Lee/ShortHistory.html>>. Acesso em: 02 abr. 2009.

BORKO, H. Information science: what is it? *American Documentation*, v. 19, n. 1, 1968.

BRASCHER, M.; CAFÉ, I. *Organização da informação ou organização do conhecimento?* In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9, 2008, São Paulo. *Anais...* São Paulo: USP, 2008.1 CD-ROM. ISBN: 978-85-60922-01-7.

BROUGHTON, V. *et al. Knowledge Organization*. Report of working group. IN: LIS-education in Europe. Working seminar held in Copenhagen 11-12 August 2005 at the Royal School of Library and Information Science. Disponível em: <http://www.db.dk/bh/lifeboat_ko/HISTORY%20&%20THEORY/LIS%20Education%20in%20EuropeDraft_KO_Submitted.htm>. Acesso em: 08 jul. 2008.

BRUCE, R. *Descriptor and folksonomy concurrence in education related scholarly research*. *Webology*, v. 5, n. 3, artigo 59. 2008. Disponível em: <<http://www.webology.ir/2008/v5n3/a59.html>>. Acesso em: 29 jun. 2009.

CAMPOS, F.; BAX, M. Como os mecanismos de busca da web indexam páginas HTML. In: Congresso Brasileiro de Biblioteconomia e Documentação, 19., 2000, Porto Alegre. *Anais...* Porto Alegre: ARB, 2000.

CAMPOS, M.L.A.; GOMES, H.E. Taxonomia e Classificação: a categorização como princípio. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB),

8, 2007, Salvador. *Anais...* Disponível em: < <http://www.enancib.ppgci.ufba.br/artigos/GT2--101.pdf> >. Acesso em 10 set. 2008.

CATARINO, M. E.; BAPTISTA, A. A.. Folksonomia: um novo conceito para a organização dos recursos digitais na Web. *Datagramazero*. Rio de Janeiro, v. 8, n. 3, p. 1-26, jun. 2007. Disponível em: < http://www.dgz.org.br/jun07/Art_04.htm >. Acesso em: 1 nov. 2007.

COATES, S. Teaching book indexing: cognitive skills and term selection. *The Indexer*. v. 23, n. 1, abr. 2002. p. 15-17

CONWAY, S.; Sligar, C. *Unlocking knowledge assets*. Redmond, Wash.: Microsoft Press, 2002.

COURTRIGHT, C. Context in information behavior research. *Annual Review of Informations Science and Technology*. v. 41, 2007. chapter 6.

CUNHA, M. B. *Dicionário de biblioteconomia e arquivologia*. Brasília: Briquet de Lemos, 2009.

DALHBERG, I. Knowledge organization: its scope and possibilities. *Knowledge Organization*, v. 20, n. 4, 211-222. 1993.

DAVIDOFF, L. L. *Introdução à psicologia*. São Paulo: Pearson Makron Books, 2005. 798 p. ISBN 8534611254.

DCMI. 2009. Disponível em: <<http://dublincore.org/about/>>. Acesso em: 21 maio 2009.

DODEBEI, V. L. D. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.

EDOLS, L. *Taxonomies are what?* 2001. Disponível em: <<http://www.freepint.com/issues/041001.htm#feature>>. Acesso em: jan. 2009.

FARROW, J. All in the mind: concept analysis in indexing. *The Indexer*. v. 19, n. 4, oct., 1995. pp. 243-247.

FARROW, J. A cognitiva process model of document indexing. *Journal of Documentation*. v. 47, n. 2, jun. 1991, pp. 149-166.

FEITOSA, A. *Organização da informação na web: das tags à web semântica*. Brasília: Thesaurus, 2006.

FIDEL, R. User-centered indexing. *Journal of the American Society for Information Science*. v. 45, n. 8 p. 572-576, 1994.

FLAKE, G. W. *et al.* Self-organization and identification of web communities. *IEEE Computer*. v. 35, n. 3, p. 66-71, 2002.

FOSKETT, D. J. Psicologia do usuário. In: *Contribuição da psicologia para o estudo dos usuários da informação técnico-científica*. Rio de Janeiro: Calunga, 1980. p. 11-29.

FROHMANN, b. Rules of mentalism: a critique of mentalism in information retrieval theory. *Journal of Documentation*. v. 46, n. 2, jun. 1990. pp. 81-101.

GOLDER, S.; Huberman, B.A. Usage patterns of collaborative tagging Systems. *Journal of Information Science*, v. 32, n. 2, p. 198-208, 2006.

GOMES, H. E. (coord.). *Manual de elaboração de tesouros monolíngües*. Brasília: Programa nacional de bibliotecas das instituições de ensino superior, 1990.

GOMES, H. E; CAMPOS, M. L. A. Tesouro e normalização terminológica: o termo como base para intercâmbio de informações. *Datagramazero*. v. 5, n. 6, dez. 2004.

GRAEF. *Managing taxonomies strategically*. 2001. <<http://www.montague.com/abstracts/taxonomy3.html>>. Acesso em: 08 jan. 2009.

GRUBER, T. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*. v. 5, n. 2, pp.199-220, 1993.

GRUBER, T. *What is an ontology?* 1996. Disponível em: <<http://www.wksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em jan. 2009.

GUARINO, N.; GIARETTA, P. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In: N. MARS (ed.) *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*. Amsterdam: IOS Press, 1995. pp. 25-32.

HAMMOND, T. *et al.* Social Bookmarking Tools (I): a general review. *D-Lib Magazine*, v.11, n.4, apr. 2005. Disponível em: <<http://www.dlib.org/dlib/april05/hammond/04hammond.html>>. Acesso em: 14 nov. 2007.

HIDDERLEY, R.; RAFFERTY, P. Democratic indexing: an approach to the retrieval of fiction. *Information services & use*, v. 17, 1997, p. 101-109.

HJØRLAND, B. Towards a theory of aboutness, subject, topicality, theme, domain, field, content... and relevance. *Journal of the American Society for Information Science and Technology*. v. 52, n. 9, p. 774-778, 2001.

HJØRLAND, B. *Cognitive views in knowledge organization (KO)*. 2007b. Disponível em: <http://www.db.dk/bh/lifeboat_ko/HISTORY%20&%20THEORY/cognitive_view_in_knowledge_orga.htm> Acesso em: 1 jun. 2008.

HJØRLAND, B. *Information retrieval (IR) as challenger to knowledge organization (KO)*. 2007d. Disponível em: <http://www.db.dk/bh/lifeboat_ko/HISTORY%20&%20THEORY/information_retrieval%20approach.htm> Acesso em: 7 jun. 2008

HJØRLAND, B. “*Traditional approaches*” to knowledge organization (KO). 2007f. Disponível em: <http://www.db.dk/bh/lifeboat_ko/HISTORY%20&%20THEORY/traditional_approaches_to_knowle.htm>. Acesso em: 26 maio 2008.

HJØRLAND, B. *Approaches to knowledge organization. (Theories of KO)*. 2007a. Disponível em: <http://www.db.dk/bh/lifeboat_ko/HISTORY%20&%20THEORY/approaches_to_knowledge_organiza.htm> Acesso em: 26 maio 2008.

HJØRLAND, B. *Discussion of Dahlberg's theory of concepts and knowledge organization (KO)*. 2007c. Disponível em: <http://www.db.dk/bh/lifeboat_ko/CONCEPTS/discussion_of_dahlberg.htm> Acesso em: 24 maio 2008.

HJØRLAND, B. *Indexing*. 2007h. Disponível em: <http://www.db.dk/bh/lifeboat_ko/CONCEPTS/indexing.htm>. Acesso em 30 jan. 2009.

HJØRLAND, B. *Knowledge organization systems*. 2008. Disponível em: <http://www.db.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_systems.htm> Acesso em: 27 maio 2008.

HJØRLAND, B. *Social organization of knowledge*. 2007e. Disponível em: <http://www.db.dk/bh/lifeboat_ko/CONCEPTS/social_organization_of_knowledge.htm>. Acesso em: 26 maio 2008.

HJØRLAND, B. *User and user studies in knowledge organization (KO)*. 2007g. Disponível em: <http://www.db.dk/bh/lifeboat_ko/CONCEPTS/user_and_user_studies.htm>. Acesso em: 31 maio 2008.

HODGE, G. *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Washington, DC: the Council on Library and Information Resources, 2000. Disponível em: <<http://www.clir.org/pubs/reports/pub91/contents.html>>. Acesso em: 15 jul. 2008.

HUFFMAN, K.; VERNON, M.; VERNON, J. *Psicologia*. São Paulo: Atlas, 2003. 814 p. ISBN 8522434727.

KWASNIK, B. H. The role of classification in knowledge representation and discovery. *Library Trends*. v. 48, n.1, p. 22-47, 1999.

LANCASTER, F. W. *Indexação e resumos: teoria e prática*. 2. ed. Brasília: Briquet de Lemos, 2004. 452 p. ISBN 8585637242.

LAYNE, S. S. Some issues in the indexing of images. *Journal of the American Society for Information Science*. v. 45, n. 8, pp. 583-588, 1994.

LE COADIC, Y. *A ciência da informação*. Brasília: Briquet de Lemos, 2004. 124 p. ISBN 8585637234.

MARLOW *et al.* HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, ToRead. *Proceedings of Hypertext'06*, pp. 31-40, 2006.

MATHES, A. Folksonomies: cooperative classification and communication through shared metadata. *Computer Mediated Communication*. Urbana: University of Illinois, 2004. Disponível em: <<http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>>. Acesso em: 31 out. 2007.

MÉNARD, E. Image index: how can I find a nice pair of italian shoes? *Bulletin of the American Society for Information Science and Technology*. p. 21-25, Oct./Nov. 2007.

MORRISON, P. J. Why are they tagging, and why do we want them to? *Bulletin of the American Society for Information Science and Technology*. p. 12-15 Oct./Nov. 2007. Disponível em: <<http://www.asis.org/Bulletin/Oct-07/morrison.html>> Acesso em: 22 set. 2008.

NEAL, D. Folksonomies and image tagging: seeing the future? *Bulletin of the American Society for Information Science and Technology*. p. 7-11, Oct./Nov. 2007

NORUZI, A. Folksonomies: Why do we need controlled vocabulary? *Webology*, v. 4, n. 2, 2007. Disponível em: <<http://www.webology.ir/2007/v4n2/editorial12.html>>. Acesso em: 27 out. 2007.

NOV, O.; YE, C. Why do People Tag? Motivations for Collaborative Content Tagging. *Proceedings for the Sixth Annual Workshop on HCI Research in MIS*. Montreal, Canada. Dez. 2007. Disponível em: < http://sigs.aisnet.org/SIGHCI/Research/ICIS2007/SIGHCI_2007_Proceedings_paper_9.pdf> Acesso em: 08 out. 2008.

O'REILLY, T. *What is Web 2.0*. 2005. Disponível em: <<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=1>> Acesso em: 16 set. 2008.

PEJTERSEN, A. M; AUSTIN, J. Fiction retrieval: experimental *design* and evaluation of a search system based on user's value criteria (part 1). *Journal of documentation*, v. 39, n. 4, dec. 1983, pp. 230-246.

PEJTERSEN, A. M; AUSTIN, J. Fiction retrieval: experimental *design* and evaluation of a search system based on user's value criteria (part 2). *Journal of documentation*, v. 40, n. 1, mar. 1984, pp. 25-35.

PINHEIRO, L. V. R. *Usuários – informação: o contexto da ciência e da tecnologia*. Rio de Janeiro: LCT: IBICT, 1982. 66 p.

PORTER. J. *Why do people tag?* 2006. Disponível em: <<http://bokardo.com/archives/why-do-people-tag/>>. Acesso em: 24 out. 2008.

QUINTARELLI, E. Folksonomies: power to the people. In: INCONTRO ISKO ITALIA - UNIMIB, Milão, 2005. *Papers...* Milan: Università di Milano, 2005. Disponível em: <<http://www.iskoi.org/doc/folksonomies.htm>>. Acesso em: 11 jun. 2009.

RAFFERTY, P.; HEDDERLEY, R. Flickr and democratic indexing: dialogic approaches to indexing. *Aslib Proceedings: new information perspectives*, v. 59, n. 4/5, pp. 397-410, 2007.

ROBREDO, J. *Documentação de hoje e de amanhã: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivistas e museológica*. 4. ed. Brasília: Ed. do autor, 2005. 409 p. ISBN 8590592014

SAARTI, J. Fiction indexing and the development of fiction thesauri. *Journal of Librarianship and Information Science*. v. 31, n. 2, pp. 85-92, June 1999.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. *Perspec. Ci. Inf.*, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SHNEIDERMAN, B.; BEDERSON, B.; DRUCKER S. Find That Photo! Interface Strategies to Annotate, Browse, and Share. *Communications of the ACM*, v. 49, n. 4, p. 69-71, 2006.

SINHA, R. *A cognitive analysis of tagging (or how the lower cognitive cost of tagging makes it popular)*. 2005. Disponível em: <<http://rashmishinha.com/2005/09/27/a-cognitive-analysis-of-tagging/#more-112>> Acesso em: 2 jun. 2008.

SINHA, R. *A social analysis of tagging (or how tagging transforms the solitary browsing experience into a social one)*. 2006. Disponível em: <http://web.archive.org/web/20060130101032/http://www.rashmishinha.com/archives/06_01/social-tagging.html> Acesso em: 06 jun. 2008.

SLAVIC, A. Knowledge organization systems, network standards and semantic Web. In: *Informacijske znanosti u procesu promjena*. Zavod za informacijske studije, Sveuciliste u Zagrebu. 2005. pp. 5-22.

SONNENREICH, W. *A history of search engines*. Wiley, c1997. Disponível em: <http://www.wiley.com/legacy/compbooks/sonnenreich/history.html>. Acesso em: 04 maio 2009.

SOWA, J. F. *Knowledge representation: logical, philosophical, and computational Foundations*. Pacific Grove, CA: Brooks Cole Publishing, 2000. <<http://www.jfsowa.com/krbook/index.htm>> Acesso em: 15 jul. 2008.

SPALDING, T. *When tags work and when they don't*: Amazon and LibraryThing. 2007. *ACM*, v. 49, n. 4, p. 69-71, 2006. Disponível em: < <http://www.librarything.com/thingology/2007/02/when-tags-works-and-when-they-dont.php> > Acesso em: 25 out. 2009.

TAPSCOTT, D.; Williams, A. D. *Wikinomics: como a colaboração em massa pode mudar seus negócios*. Rio de Janeiro: Nova Fronteira, 2007.

TAYLOR, A. *The organization of information*. 2. ed. Library and Information Science Text Series. 2003.

TERRA, J.C.C. *et al.* *Taxonomia: elemento fundamental para a gestão do conhecimento*. 2005. Disponível em: <<http://www.terraforum.com.br>>. Acesso em: 07 jan. 2009.

TRANT, J. Exploring the potential for social tagging and folksonomy in art museums: proof of concept. *New Review of Hypermedia and Multimedia*. v. 12, n. 1, jun. 2006. Disponível em: <www.archimuse.com/papers/steve-nrhm-0605preprint.pdf> Acesso em: 17 jul. 2008.

UDDIN, M. N.; JANECEK, P. Faceted classification in web information architecture: a framework for using semantic web tools. *The electronic library*, v. 25, n. 2, p. 210-233, 2007.

VICKERY, B. C. Knowledge representation. *Journal of documentation*. v. 42, n. 3, p. 145-159. 1986.

VOSS, J. *Tagging, folksonomy and co.: renaissance of manual indexing?* 10th international Symposium for Information Science Retrieved January 31, 2008. Disponível em: <http://arxiv.org/PS_cache/cs/pdf/0701/0701072.pdf>. Acesso em: 01 nov. 2007.

WAL, T. V. *A stale state of tagging*. 2007. Disponível em: <<http://www.vanderwal.net/random/category.php?cat=153>>. Acesso em: 28 jun. 2009.

WAL, T. V. *Explaining and Showing Broad and Narrow Folksonomies*. 2005. Disponível em: <<http://www.vanderwal.net/random/entrysel.php?blog=1635>>. Acesso em: 22 jun. 2009.

WAL, T. V. *Folksonomy*. 2007. Disponível em: <<http://www.vanderwal.net/folksonomy.html>> Acesso em: 16 set. 2008.

WAL, T. V. *From tags to the future*. 2005a. Disponível em: <<http://www.vanderwal.net/random/category.php?cat=153>>. Acesso em: 22 jun. 2009.

WAL, T. V. *Getting to know collective and collaborative*. 2008. Disponível em: <<http://www.vanderwal.net/random/category.php?cat=153>>. Acesso em: 23 jun. 2009.

WAL, T. V. *Tagging for fun and finding*. 2005b. Disponível em: <<http://www.ok-cancel.com/archives/article/2005/07/tagging-for-fun-and-finding.html>>. Acesso em: 28 jun. 2009.

WALL, A. *History of Search Engines: From 1945 to Google* 2007. 2007? Disponível em: <<http://www.searchenginehistory.com/>> Acesso em: 12 maio 2009.

WEAVER, C. The indexer as a consultant: collaborative indexing of community newspaper. *Key Word*. V. 14, n. 1, Jan./Mar. 2006.

WIKIPEDIA. *Diretórios de sites*. Disponível em: <http://pt.wikipedia.org/wiki/Diretórios_de_sites>. Acesso em: 27 abr. 2009.

ZOLLERS, A. Emerging motivations for tagging: Expression, performance, and activism. 2007. 16. International *World Wide Web* Conference. Banff, Canada. Disponível em: <http://www2007.org/workshops/paper_55.pdf> Acesso em: 04 out. 2008

APÊNDICE A - Questionários de coleta de dados

Academic Research - Tagging at Delicious

*** Required**

1. Sex *

male

female

Other:

2. Age *

3. Where are you from? *
(Country)

4. Highest educational level *

5. For how long are you using Delicious? *

less than a year

1-2 years

2-3 years

over 3 years

6. Why do you tag? *
Tell me your motivations to add tags to your bookmarks

Academic Research - Tagging at Flickr

* Required

1. Sex *

- male
 female

2. Age *

under 15 ↓

3. Where are you from? *

(Country)

4. Highest educational level: *

Elementary School ↓

5. For how long are you using Flickr? *

- less than a year
 1-2 years
 2-3 years
 more than 3 years

6. Why do you tag? *

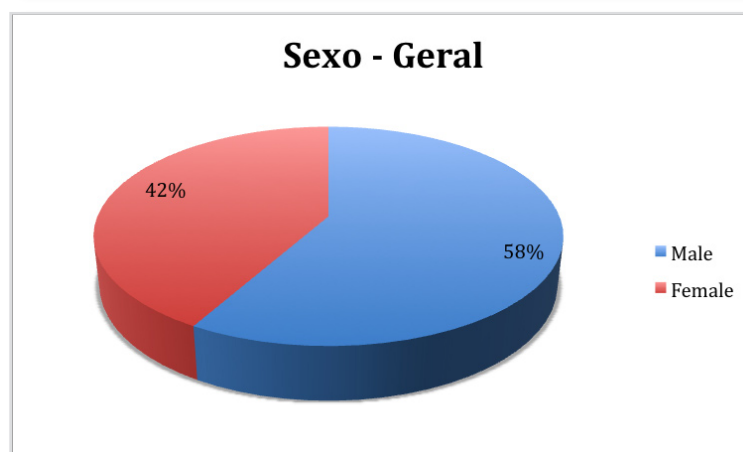
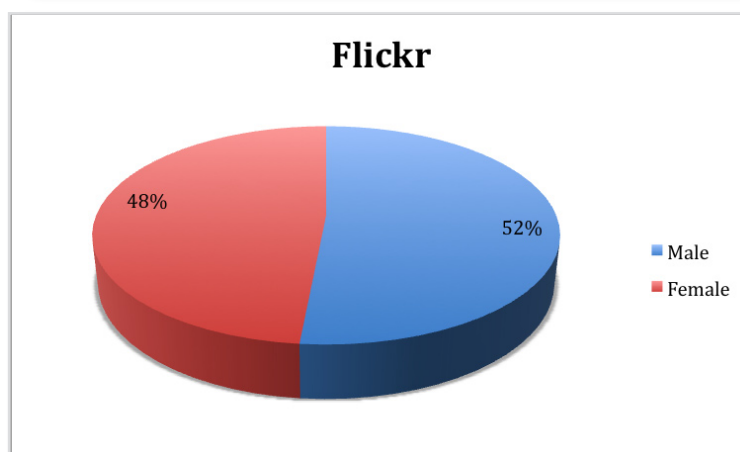
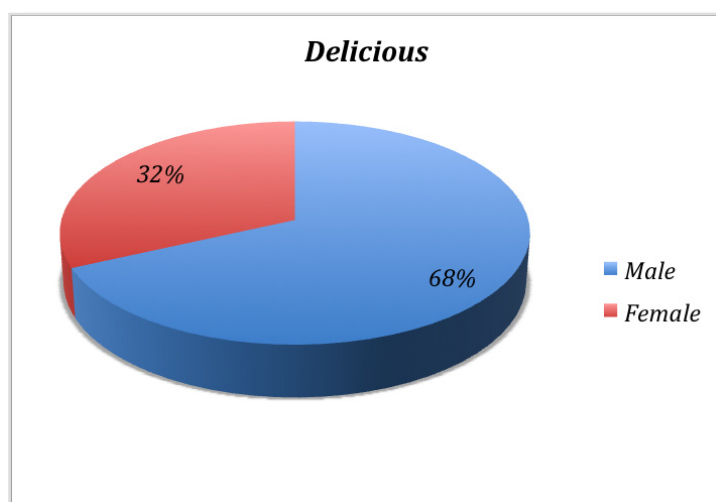
Tell me your motivations to add tags to your photos

Submit

APÊNDICE B - Dados demográficos

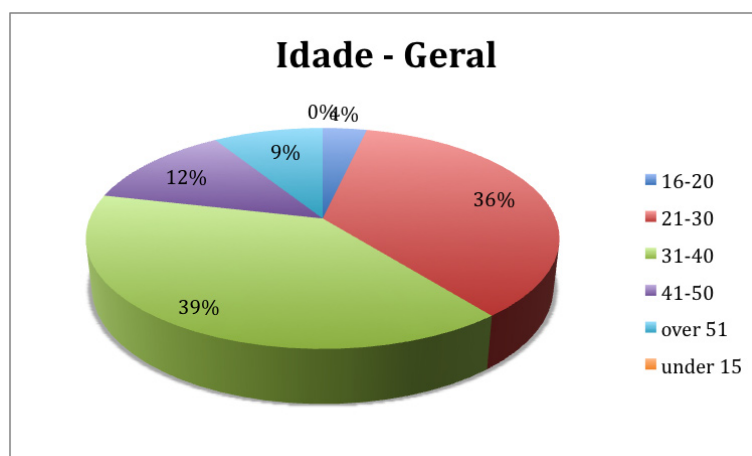
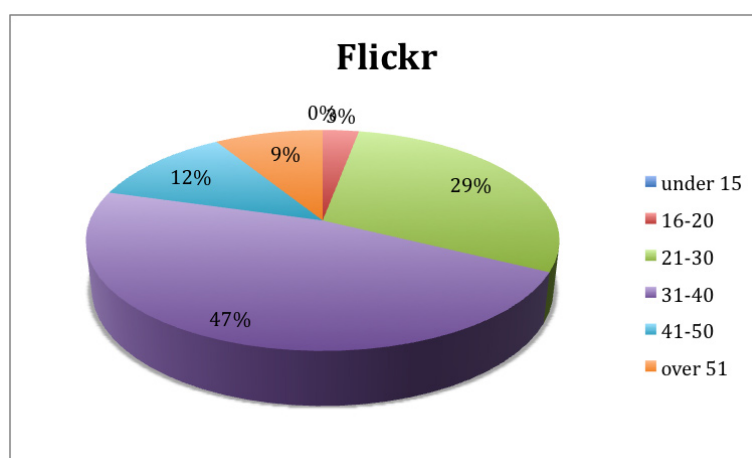
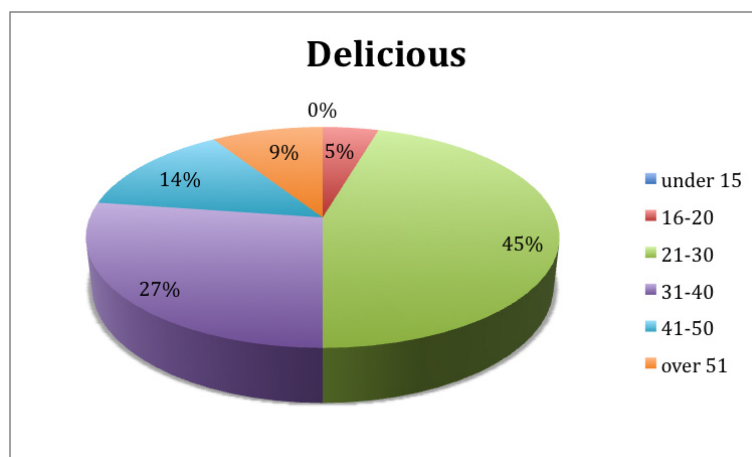
1 Sexo

Em ambos os sistemas pesquisados, houve predominância de respondentes do sexo masculino, com uma predominância acentuada no Delicious.



2 Idade

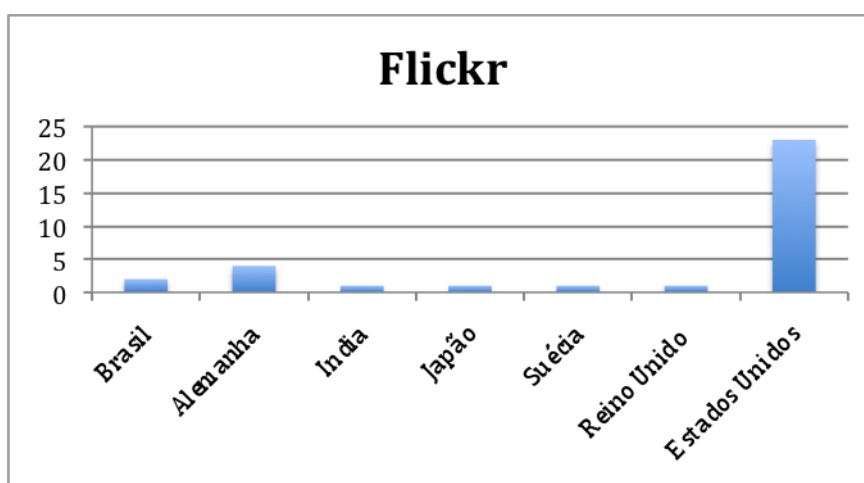
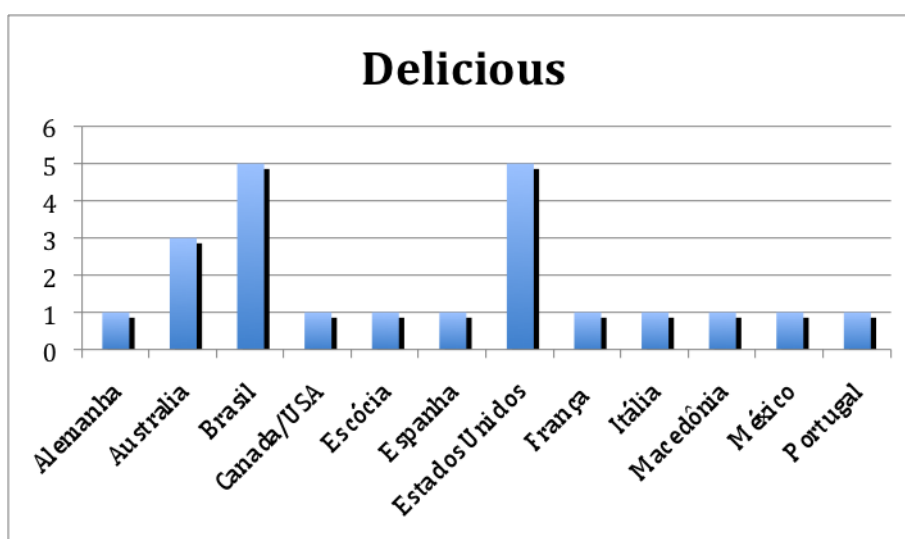
A maioria dos respondentes no Delicious encontram-se na faixa dos 21 aos 30 anos, enquanto no Flickr houve uma predominância de respondentes um pouco mais velhos: 31 a 40 anos. Em ambos os sistemas não houve respondentes com menos de 15 anos.



3 País

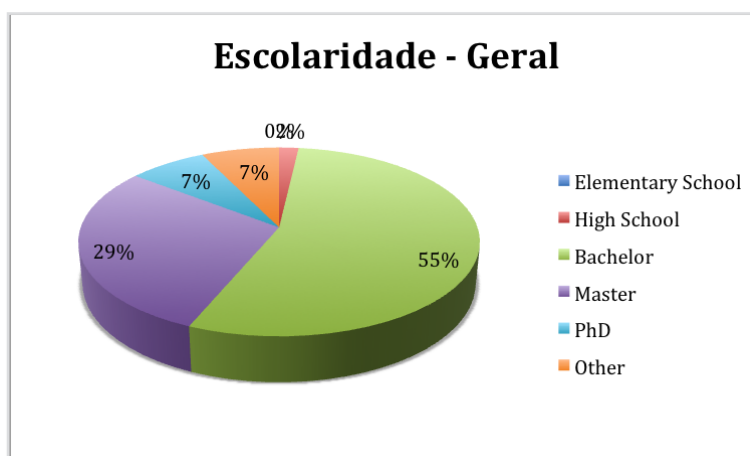
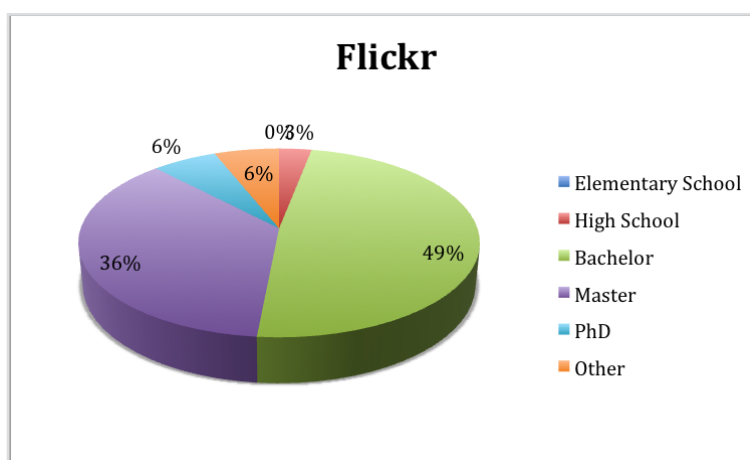
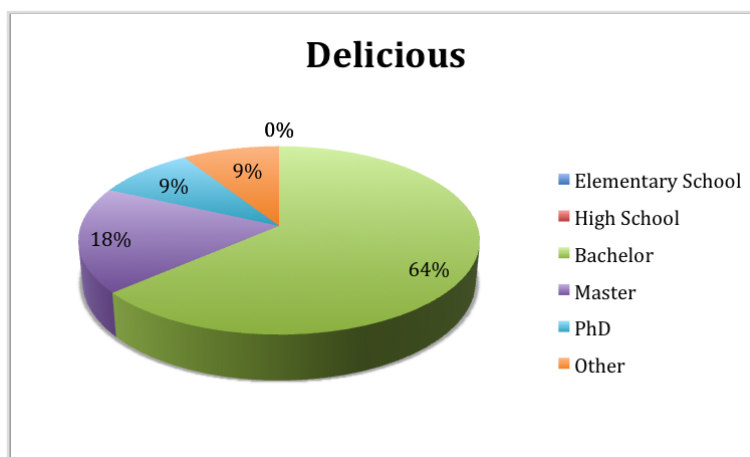
A localização geográfica dos respondentes do Delicious ficou bem variada, com predominância dos Estados Unidos e Brasil. Em relação ao Brasil, o grande número de respostas deu-se devido a uma tentativa de se obter mais respostas, enviando o questionário diretamente a pessoas conhecidas que usam o sistema. Como o fator geográfico não influirá no resultado principal da pesquisa, acredita-se que não há problemas na adoção de tal prática.

Já para o Flickr, houve uma predominância evidente de respostas dos Estados Unidos, e uma gama menor de países respondentes:



4 Escolaridade

Em ambos os sistemas a predominância foi de respondentes graduados, com uma participação significativa de mestres respondendo ao questionário do Flickr.



5 Tempo de uso do sistema

A maioria dos usuários que responderam no Delicious já utilizam o sistema há mais de três anos. Acredita-se que essa taxa deve-se a forma de coleta de dados nesse sistema, por meio do Fórum de discussão e da lista de *e-mails*. Geralmente, quem utiliza tais ferramentas são usuários mais antigos dos sistemas, e não os novatos.

Já para o Flickr, a maioria utiliza há 2-3 anos, e houve uma distribuição mais equilibrada do tempo de uso entre as faixas do questionário. Pode-se supor que isso ocorreu também devido ao método de coleta de dados, que foi feito de forma bastante aleatória nesses sistema, não privilegiando nenhum tipo de usuário especificamente.

