

ROBERTO PEREIRA

**FOLKAUTHORITY: A APLICAÇÃO DO CONCEITO DE
AUTORIDADE COGNITIVA POR MEIO DE FOLKSONOMIA**

MARINGÁ

2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

ROBERTO PEREIRA

**FOLKAUTHORITY: A APLICAÇÃO DO CONCEITO DE
AUTORIDADE COGNITIVA POR MEIO DE FOLKSONOMIA**

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

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

MARINGÁ

2008

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá – PR., Brasil)

P436f Pereira, Roberto
Folkauthority : a aplicação do conceito de autoridade
cognitiva por meio de folksonomia / Roberto Pereira. --
Maringá : [s.n.], 2008.
171 f. : il., figs.

Orientador : Prof. Dr. Sérgio Roberto Pereira da Silva.
Dissertação (mestrado) - Universidade Estadual de
Maringá, Programa de Pós-Graduação em Ciência da
Computação, 2008.

1. Folksonomia. 2. Autoridade cognitiva. 3. Sobrecarga
de informação. 4. Tags. 5. Organização da informação na
internet. I. Universidade Estadual de Maringá. Programa de
Pós-Graduação em Ciência da Computação. II. Título.

CDD 22.ed. 025.04

ROBERTO PEREIRA

FOLKAUTHORITY: A APLICAÇÃO DO CONCEITO DE AUTORIDADE COGNITIVA POR MEIO DE FOLKSONOMIA

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

Aprovado em 23/06/2008.

BANCA EXAMINADORA

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

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

Profa. Dra. Junia Coutinho Anacleto
Universidade Federal de São Carlos – DC/UFSCar

AGRADECIMENTOS

Agradeço e dedico este trabalho aos meus pais, Luiz e Bernadete, e aos meus irmãos, Edson (*in memoriam*) e Angelita, e à Palmira B. Bonetti, por todo incentivo, atenção e apoio.

Ao Ser Superior, Deus, fonte de toda força e razão suprema de nossa existência.

Ao professor Dr. Sérgio Roberto P. da Silva, meu orientador, pela orientação, auxílio e confiança. Agradeço também pelo incentivo, atenção e apoio durante todo o período do desenvolvimento desse trabalho.

Aos colegas do GSII – Grupo de Sistemas Interativos Inteligentes, pelas discussões, idéias, incentivos e amizade, em especial: Josiane M. P. Ferreira, Valéria Delisandra Feltrim, Raqueline R. M. Penteado e Edgard Mota, a quem agradeço por me apresentar o assunto que viria a se tornar minha dissertação de mestrado.

Aos grandes amigos, mestrandos: José Valderlei da Silva (Jesus), Daniela F. G. Trindade, Késsia R. C. Marchi, Cristiane Y. H. Castro, Carlos A. M. Basso e Cheila M. Bergamini.

Às professoras do mestrado: Tania F. C. Tait, Elisa H. M. Huzita e Maria M. Dias; aos professores da graduação: Giovane Motter, Edivane Bellé, Fábio A. Taffe, Luciani Pizato e Victor F. A. Santander; e a todos os professores que durante minha vida escolar apoiaram, incentivaram e contribuíram para minha formação.

À Maria Inês Davanço, secretária do mestrado, agradeço a atenção, carinho e apoio. Sua competência e dedicação são grandes exemplos e vão muito além de suas atribuições.

Finalmente, agradeço ao apoio financeiro da CAPES, ao Programa de Pós-Graduação em Ciência da Computação (PCC-UEM) e ao Departamento de Informática (DIN-UEM).

“If we cannot end now our differences, at least, we can help make the world safe for diversity”. (John F. Kennedy)

RESUMO

O excesso de informação sendo produzida e publicada na Internet sem nenhum controle de qualidade faz com que ocorra uma sobrecarga de informação cuja qualidade não pode ser facilmente verificada, tornando o processo de recuperação de informação nesse ambiente cada vez mais complexo. Os sistemas baseados em folksonomia surgiram como uma iniciativa para auxiliar no processo de atribuição de significado e organização da informação na Internet, concedendo aos usuários um papel mais ativo nesse processo. Neste trabalho, propõe-se o emprego do reconhecimento da autoridade cognitiva das fontes de informação em sistemas baseados em folksonomia para a criação de uma cadeia de autoridades, a qual formará uma rede social que representará o reconhecimento das competências destas entidades. Essa cadeia foi empregada para melhorar a qualidade dos resultados obtidos em um processo de recuperação de informação neste tipo de sistemas. Para validar essa hipótese, foram realizados experimentos por meio da simulação dos processos de categorização de informação, de concessão de autoridade cognitiva e de recuperação de informação. Foram Desenvolvidos algoritmos que permitiram priorizar as categorizações de informação efetuadas por autoridades no assunto e demonstrou-se que a utilização da cadeia de autoridades melhora a precisão e a qualidade dos resultados recuperados. Além disso, demonstrou-se também alguns benefícios que são obtidos por meio da exploração da cadeia de autoridades, como a identificação de autoridades em um determinado assunto, o reconhecimento de competências, até a análise da propagação e da retenção de autoridade cognitiva. Estes resultados podem ser muito úteis na emergência de conhecimento em softwares sociais.

Palavras-Chave: Folksonomia, Autoridade Cognitiva, Sobrecarga de Informação, *Tags*.

ABSTRACT

The excess of information being produced and published on the Internet without any quality control causes an information overload whose quality cannot be easily verified, making the information recovery process in this environment more and more complex. Folksonomy-based systems appeared as an initiative to aid in the process of meaning attribution and information organization on the Internet, granting to users a more active role in this process. In this work, we propose the employment of the cognitive authority recognition of the information sources in folksonomy-based systems for the creation of a chain of authorities, which will form a social network that will represent the recognition of the competences of these entities. This chain was used to improve the quality of the results obtained in a information recovery process in this type of systems. To validate our hypothesis, we accomplished experiments through the simulation of the processes of information categorization, of cognitive authority ascription and of information recovery. We developed some algorithms that allow us to prioritize information categorizations made by authorities in the subject and we demonstrated that the use of the chain of authorities improves the precision and the quality of the recovered results. Besides, we demonstrated also some benefits that are obtained through the chain of authorities' exploration, as the identification of authorities in a certain subject, the recognition of competences, until the analysis of cognitive authority propagation and retention. These results can be very useful in the knowledge emergency in social softwares.

Keywords: Folksonomy, Cognitive Authority, Information Overload, Tags.

SUMÁRIO

LISTA DE FIGURAS.....	11
LISTA DE QUADROS E TABELAS	13
LISTA DE ABREVIATURAS E SIGLAS	14
1 Introdução	15
2 Folksonomia	22
2.1 A Sobrecarga de Informação na Internet	23
2.2 A Técnica de Folksonomia	26
2.2.1 Aspectos Gerais sobre a Folksonomia.....	29
2.2.2 Sistemas Baseados em Folksonomia	35
2.2.3 Dificuldades e Desafios na Folksonomia	37
2.3 Perspectivas na Folksonomia e o Problema da Sobrecarga de Informação.....	42
3 Autoridade Cognitiva	46
3.1 A Questão da Autoridade.....	47
3.2 A Teoria da Autoridade Cognitiva.....	49
3.2.1 Conhecimento de Primeira e de Segunda-Mão	50
3.2.2 Autoridade, Influência, Credibilidade	51
3.2.3 Autoridade Cognitiva e Qualidade	55
3.3 Recuperação de Informação, Autoridade Cognitiva e Folksonomia	59
4 Folkauthority	62
4.1 Autoridade Cognitiva e Folksonomia	63
4.1.1 A Origem da Proposta	63
4.1.2 A Folksonomia na Concessão de Autoridade Cognitiva.....	65
4.2 <i>CAW – Cognitive Authority on the Web</i>	67
4.2.1 O Processo de Concessão de Autoridade Cognitiva.....	72
4.2.2 A Cadeia de Autoridades	79
4.2.3 Extração de Informação na Cadeia de Autoridades.....	84
4.2.4 Análises Sobre a Cadeia de Autoridades.....	86
4.3 Reflexões sobre o Processo de Concessão de Autoridade	87
4.3.1 A Concessão de Autoridade Cognitiva.....	88
4.3.2 Considerações Sobre a Interação dos Usuários com a Ferramenta <i>CAW</i>	91
4.3.3 A Cadeia de Autoridades	94
4.4 Trabalhos Relacionados	98
5 Simulações: Concessão de Autoridade Cognitiva e Recuperação de Informação.....	103
5.1 Metodologia	104
5.1.1 A Definição dos Pivôs da Folksonomia	105

5.1.2	A Categorização de Objetos	107
5.1.3	A Concessão de Autoridade Cognitiva.....	110
5.2	A Recuperação de Informação.....	112
5.2.1	Objetivos da Análise.....	113
5.2.2	Análise sobre o posicionamento dos objetos.....	114
5.2.3	Análises e Resultados	115
5.3	Cadeia de Autoridades	122
5.3.1	Distribuição das autoridades na cadeia: graus de separação	126
5.3.2	Distribuição dos objetos de interesse das entidades	127
5.3.3	Consultas: extração de informação na cadeia.....	129
6	Conclusão	136
7	Referências	144
	Anexo I – Processo de Simulação	150
	Anexo II - Princípio de Pareto, <i>Long Tail</i> das <i>Tags</i>.....	157

LISTA DE FIGURAS

Figura 1 - Os três pivôs da folksonomia.....	28
Figura 2 - Folksonomia Vs. organização por especialistas.	31
Figura 3 – Curva de <i>power law</i> , distribuição de Pareto, <i>long tail</i> (WAL, 2005a).	33
Figura 4 - Página de pesquisa do sistema <i>Delicious</i>	36
Figura 5 - Sistemas controlados Vs. sistemas baseados em folksonomia.	40
Figura 6 - Limite dos resultados considerados pelos usuários	44
Figura 7 - Características da qualidade e da autoridade cognitiva.	58
Figura 8 - Meta-categorização: folksonomia em nível duplo.....	65
Figura 9 - CAW: mapa conceitual.....	69
Figura 10 - Arquitetura do sistema CAW.	70
Figura 11 - Modelo conceitual simplificado do banco de dados.	71
Figura 12 - Conjunto de entidades-fontes que categorizam informações.	74
Figura 13 - Entidade “A” concedendo autoridade cognitiva para entidade “B”.	74
Figura 14 - Fluxograma do processo de concessão de autoridade cognitiva.....	75
Figura 15 - Atribuição de peso à autoridade concedida	77
Figura 16 - Redes de autoridades formando uma cadeia de autoridades.....	80
Figura 17 – Relacionamentos: amizade (bidirecional) Vs. autoridade (unidirecional).....	93
Figura 18 - Propagação e retenção de autoridade cognitiva.....	98
Figura 19 - Processo de simulação.	105
Figura 20 - Seleção de termos para a categorização de objetos.	108
Figura 21 - <i>Long tail</i> formado pela freqüência da utilização das <i>tags</i>	110
Figura 22 - Simulação do processo de concessão de autoridade cognitiva.....	112
Figura 23 - Processo de recuperação de objetos.....	114
Figura 24 - Posicionamento dos objetos na recuperação de informação.....	116
Figura 25 - Posicionamento dos objetos na recuperação convencional.	117
Figura 26 - Posicionamento dos objetos priorizando as autoridades (gráfico com linhas). ...	119
Figura 27 - Recuperação normal Vs. recuperação considerando autoridades.....	119
Figura 28 - Cadeia de autoridades formada pelas redes de autoridades das entidades.	122
Figura 29 - Cadeia de autoridades – “entidade 3”.....	124
Figura 30 - Identificação do menor caminho entre duas entidades.	125
Figura 31 - Distância entre entidades.	125
Figura 32 - Distribuição das autoridades por graus na cadeia.....	126

Figura 33 - Distribuição dos objetos de interesse das entidades na cadeia de autoridades.	128
Figura 34 - Rede de autoridades da “entidade 27”.	133
Figura 35 - Fãs da “entidade 27”.	133
Figura 36 - Cadeia de autoridades da “entidade 27”.	135

LISTA DE QUADROS E TABELAS

Quadro 1 - Identificação da popularidade e do peso geral da autoridade cognitiva.....	78
Quadro 2 - Pseudo-algoritmo para a recuperação de informação	78
Quadro 3 - Construção da cadeia de autoridades	82
Quadro 4 - Busca por autoridades	82
Quadro 5 - Reconhecimento de competências	83
Tabela 1 - Apresentação de resultados.	79
Tabela 2 - Frequência de utilização das <i>tags</i>	109
Tabela 3 - Posicionamento dos Objetos na Recuperação Convencional.....	116
Tabela 4 - Posicionamento dos objetos na recuperação priorizando autoridades.	118
Tabela 6 - Autoridades encontradas pela cadeia de autoridades.	131
Tabela 7 - Autoridade recebida pela “entidade 27”.....	131

LISTA DE ABREVIATURAS E SIGLAS

URL	<i>Uniform Resource Locator</i>	15
API	<i>Application Programming Interface</i>	67

Capítulo I

Introdução

Há muito tempo discute-se sobre o crescimento das informações e dos conteúdos produzidos e publicados pelo mundo. A quantidade de informação disponível às pessoas tem apresentado um aumento em grandes proporções (HIMMA, 2007) (AIRES, 2003) (LYMAN, 2000; 2003), e a disseminação do uso da Internet, juntamente com a facilidade de produção e publicação de conteúdos informacionais acarretaram em um aumento da quantidade de materiais disponíveis em taxas nunca antes observadas. Lyman (2000), em suas pesquisas para tentar estimar a quantidade de informação produzida no mundo desde a década de 80, concluiu, no ano 2000, que a quantidade de informação disponível na Internet praticamente duplicava a cada ano; em 2004, Smith *et al.* fundamentaram-se nas evoluções dos estudos de Lyman (2000; 2003) para fornecer uma estimativa conservadora da existência de 10 bilhões de documentos; Shaw (2005) estimou o número de *Websites* existentes em aproximadamente 13 bilhões e Himma (2007), para ilustrar essa quantidade, mencionou que cada pessoa existente no planeta poderia ser, literalmente, soterrada pelo montante de informação existente na *Web*.

O termo **sobrecarga de informação** (*information overload*) é amplamente utilizado para definir a quantidade excessiva de informações disponibilizadas na Internet que, de tão elevada, chega a ser prejudicial — embora alguns autores como Himma (2007) e Levy (2005) sugerem que “sobrecarga de conteúdos” seria um termo mais adequado. De acordo com Lyman (2000; 2003), Levy (2005) e Himma (2007), a sobrecarga de informação é causada, principalmente, pela facilidade de produção e de publicação de conteúdos, juntamente com a ausência de mecanismos de controle de qualidade de tudo o que é produzido e publicado. Há uma quantidade de dados, conteúdos e informações muito superior a capacidade humana de absorver, interpretar e gerenciar, apresentando desafios tanto do ponto de vista técnico quanto social, originados das limitações existentes frente a essa sobrecarga.

Em um ambiente como a Internet não é plausível considerar a possibilidade da existência de profissionais controlando e verificando todos os conteúdos publicados. A complexidade e o custo são inviáveis e incompatíveis com a própria natureza aberta da *Web*. Entretanto, como qualquer indivíduo que utilize a Internet é um consumidor e um produtor de informação em potencial, é relevante considerar também esses indivíduos como organizadores da informação produzida, publicada e encontrada (WAL, 2005b). Os sistemas baseados em folksonomia (*folksonomy-based systems*) representam uma iniciativa para auxiliar no processo de organização e de atribuição de significado aos conteúdos disponíveis na Internet. Esses sistemas permitem que seus usuários categorizem (MATHES, 2004), por meio da atribuição de um conjunto de termos, ou rótulos (*tags*), livremente selecionados, qualquer objeto acessível na *Web* (*URLs*¹, fotos, vídeos), e são denominados como sistemas baseados em folksonomia por fazerem do uso das *tags* pelos usuários sua estrutura central de funcionamento.

¹ *Uniform Resource Locator*

O desenvolvimento de sistemas baseados em folksonomia, como *Flickr* e *Delicious*, foi impulsionado pelo advento da *Web 2.0*. Nesses sistemas a atribuição de significados e a organização de informação não são restritas aos autores dos conteúdos ou aos editores profissionais (RIDDLE, 2005), os próprios usuários organizam, categorizam e atribuem significado às informações, rotulando-as por meio de *tags*, da forma que julgarem mais conveniente e sem nenhum tipo de controle. Nesse contexto, as informações encontradas em um sistema baseado em folksonomia são informações que já foram categorizadas e analisadas por outras pessoas, e não por algoritmos de indexação como, por exemplo, o *Google Pagerank* (LANGVILLE *et al.*, 2004). No entanto, na medida em que cresce o número de usuários desses sistemas e, conseqüentemente, das informações sendo categorizadas por eles, o problema de sobrecarga de informação também se torna evidente. A grande quantidade de informação livremente categorizada implica que, no momento da recuperação, o usuário possa confiar apenas na “opinião coletiva” da maioria dos usuários que realizam categorizações ou em sua própria opinião, porque aspectos relacionados à autoridade e confiança dos categorizadores e, também, à qualidade das categorizações por eles realizadas não são considerados.

O ponto crítico da utilização da técnica de folksonomia está no fato de que a total liberdade dos usuários é considerada tanto a razão do seu sucesso quanto a causa dos seus problemas — a folksonomia funciona por causa dos usuários ao mesmo tempo em que suas falhas também são ocasionadas por eles. Esse paradoxo leva a visões extremas relacionadas desde a proposta de uma folksonomia mais controlada (OHMUKAI, 2005), até a argumentação de que a liberdade deve ser mantida e, o que é realmente necessário, é o desenvolvimento de recursos para minimizar os seus efeitos colaterais e para oferecer maior apoio aos usuários (WU, 2006). Independentemente dessas discussões sobre a liberdade dos usuários e sobre qual é a visão mais adequada, não se pode dizer que os usuários estão errados

ou que não estejam categorizando a informação corretamente. O que os usuários categorizam é adequado para eles mesmos ou para um pequeno grupo, porém, nem sempre será adequado ou interessante para o coletivo, para o total de usuários de um sistema baseado em folksonomia. Existem áreas nas quais os usuários são mais (ou menos) apropriados como fontes de informação e isso depende do conhecimento que os mesmos possuem sobre o assunto que estão categorizando, depende das habilidades e das competências desses usuários — da autoridade cognitiva que eles detém.

Russell (2005), ao discutir sobre a técnica de folksonomia, considera os trabalhos de Rieh e Belkin (2005; 2000) (RIEH, 1998), sobre as relações da **autoridade cognitiva** com a qualidade e com a relevância das informações na Internet. Russel sugere a utilização de folksonomia para conceder autoridade e, assim, identificar a autoridade das entidades e reconhecer suas competências dentro de uma comunidade de usuários na Internet. A autoridade cognitiva é uma teoria da epistemologia social proposta por Wilson (1983) e que discute uma autoridade baseada nas competências, habilidades e capacidades das pessoas ou entidades. Rieh e Belkin realizaram experimentos que demonstraram a relação da autoridade cognitiva com a qualidade e com a relevância das informações e Russell (2005) inferiu que a folksonomia seria um meio adequado de conceder e representar a autoridade cognitiva na Internet. Deste modo, Russel propôs um sistema baseado em folksonomia em que os próprios usuários seriam categorizados, mas, em vez dessa categorização representar o conteúdo ou o significado de uma informação, ela representaria a autoridade cognitiva dos usuários — as *tags* atribuídas ao usuário descreveriam suas competências ou habilidades.

Considerando que nos sistemas baseados em folksonomia as fontes responsáveis pelas categorizações de informação são os próprios usuários, o reconhecimento da autoridade cognitiva que esses usuários possuem permite identificar as informações que são categorizadas por quem é autoridade no assunto. Wilson (1983) argumenta que autoridades

tendem não apenas a organizar melhor suas informações, mas também a possuírem conteúdos de qualidade e a manterem contato com outras pessoas que entendam ou que tenham interesse no mesmo assunto. Dessa forma, se a concessão de autoridade cognitiva for realizada sobre os usuários de um sistema baseado em folksonomia, torna-se possível, além de reconhecer competências e de identificar autoridades em determinado assunto, utilizar essa autoridade para melhorar o processo de recuperação de informação. Como a autoridade é subjetiva (WILSON, 1983), dependendo de quem a concede, da mesma forma que a qualidade e a relevância dos resultados também o são (RIEH & BELKIN, 1998; 2000) (RIEH, 2005), obter melhores ou piores resultados no processo de recuperação de informação em um sistema baseado em folksonomia é uma consequência mais relacionada aos usuários-fontes, que categorizaram as informações, do que à forma como as categorizações foram realizadas.

A hipótese deste trabalho é de que a aplicação do conceito de autoridade cognitiva em sistemas baseados em folksonomia eleva a relevância e a qualidade dos resultados, e ameniza os problemas da sobrecarga de informação. Assim, consideramos a proposta de Russell (2005) para a concessão de autoridade por meio de folksonomia e propomos que essa autoridade seja utilizada no processo de recuperação de informação. Como consequência do reconhecimento e da identificação da autoridade dos categorizadores, torna-se possível priorizar as informações categorizadas por quem entende do assunto e, assim, apresentar conteúdos de maior relevância e qualidade para aqueles que concedem ou que concordam com a autoridade das fontes de categorização. Para validar nossa hipótese, efetuamos simulações sobre o processo de concessão de autoridade cognitiva por meio de folksonomia em um sistema baseado em folksonomia e consideramos essa autoridade no processo de recuperação de informação. Por meio dos resultados obtidos, verificamos que priorizar as categorizações realizadas por autoridades proporciona melhores resultados do que a recuperação tradicional, em que os resultados são apresentados em ordem cronológica de

categorização e em ordem de popularidade. Observamos também que a concessão de autoridade por parte dos usuários cria uma cadeia de autoridades que possibilita a extração de informações que vão muito além da simples identificação de autoridades e do reconhecimento de competências, tais como: as autoridades mais populares, a formação de grupos, a análise da distribuição das entidades na cadeia de autoridades, etc. Essa cadeia de autoridades representa uma rede social baseada na confiança, no mérito e na autoridade de seus componentes, sendo de grande importância na validação do conhecimento coletivo estruturado pelo sistema de folksonomia.

Os experimentos realizados demonstraram que a hipótese deste trabalho é válida e indicam a viabilidade de aplicar o conceito de autoridade cognitiva em um sistema baseado em folksonomia com usuários reais. Entretanto, devemos salientar que por se tratar de uma simulação, detalhes sobre os aspectos sociais foram ignorados e questões de interação humano-computador não foram consideradas, ao contrário do que acontece em um sistema com usuários reais que exigirá atenção e que envolverá o desenvolvimento de pesquisas adicionais devido ao papel crítico que esses fatores desempenham para o êxito de um sistema baseado em folksonomia.

Esta dissertação está organizada da seguinte forma. No capítulo II, apresentamos a técnica de folksonomia como uma alternativa para a organização de conteúdos na Internet; discutimos os seus pontos favoráveis e desfavoráveis, assim como suas particularidades; apresentamos alguns sistemas baseados em folksonomia e abordamos o problema da sobrecarga de informação nesses sistemas. No capítulo III, expomos a teoria da autoridade cognitiva; discutimos sobre autoridade e sobre sua relação com a qualidade; apresentamos e justificamos a autoridade cognitiva como uma alternativa para reduzir os problemas de sobrecarga de informação nos sistemas baseados em folksonomia e, da mesma forma, argumentamos por que concordamos em utilizar a folksonomia para concessão de autoridade

cognitiva. No capítulo IV, descrevemos como chegamos à nossa proposta de conceder autoridade cognitiva por meio de folksonomia e de utilizá-la para a recuperação de informação; exemplificamos o processo de concessão de autoridade e apresentamos a arquitetura de um sistema para suportá-lo; levantamos discussões sobre aspectos sociais e relacionados à interação humano-computador e outras questões pertinentes; e finalizamos apresentando os principais trabalhos relacionados. No capítulo V, expomos o processo de simulação utilizado, a execução da simulação para a concessão de autoridade cognitiva e para a recuperação de informação e, então, as análises sobre os resultados obtidos nesse processo. Finalmente, no capítulo VI, apresentamos nossas conclusões sobre o trabalho desenvolvido, discutimos suas limitações e sugerimos pesquisas futuras que podem ser realizadas para complementá-lo ou que podem ser desenvolvidas a partir dele.

Capítulo II

Folksonomia

A Internet é um meio de disseminação de informação abrangente, rápido e de baixo custo (LYMAN, 2000; 2003). A facilidade de acesso à Internet juntamente com a facilidade de produção e de publicação de conteúdos, são fatores que contribuem para a existência de uma quantidade, nunca antes vista, de conteúdos disponíveis e de fácil obtenção cuja qualidade, infelizmente, não pode ser facilmente verificada. Neste capítulo, aborda-se o conceito de sobrecarga de informação na Internet, apresenta-se a técnica de folksonomia aplicada para auxiliar na organização de informação e expõem-se discussões teóricas e práticas relacionadas a essa técnica. Após essas discussões, realiza-se uma explanação sobre a existência da sobrecarga de informação nos sistemas que aplicam a técnica de folksonomia, sobre as implicações da sobrecarga nesses sistemas e, também, sobre medidas que visam obter melhores contribuições para a organização de informação em sistemas baseados em folksonomia.

2.1 A Sobrecarga de Informação na Internet

A informação, de todo e qualquer tipo, está se proliferando a uma taxa enorme (HIMMA, 2007) (LYMAN, 2003). As discussões sobre o crescimento dos conteúdos produzidos e publicados no mundo não é recente. Em 1987 Warren Thorngate já discutia sobre o considerável aumento das publicações acadêmicas e não-acadêmicas (HIMMA, 2007), porém, a Internet é a responsável por fazer com que esse crescimento torne-se cada vez mais elevado e mais constante. No ano 2000, Lyman desenvolveu um estudo concluindo que a quantidade de conteúdos disponíveis na Internet duplicava anualmente e, segundo suas considerações, essa quantidade continuaria evoluindo a altas taxas, pois a Internet ainda não havia alcançado seu período máximo de expansão. Lyman estimou a existência de dois bilhões e meio de documentos e de mais de dois bilhões e cem mil páginas disponíveis na Internet no início do ano 2000 crescendo a uma taxa de sete milhões e trezentas mil páginas por dia. Em 2003, os estudos de Lyman apresentam uma estimativa atualizada de 10 bilhões de páginas. Já em 2005, Shaw estima esse número em aproximadamente treze bilhões demonstrando a evolução e o considerável aumento na quantidade de conteúdos disponibilizados pela *Web* em um curto intervalo de tempo.

Segundo Lyman (2000, 2003), uma das vantagens da Internet, assim como dos meios de comunicação como o rádio e a televisão, é o fluxo da informação. Na mídia impressa um exemplar normalmente é lido por apenas uma ou duas pessoas, enquanto que na Internet uma única unidade de conteúdo pode ser acessada por um número muito maior de pessoas. Com a disseminação do uso da Internet, e com o surgimento de tecnologias para a facilitação da produção e da publicação de conteúdos por qualquer pessoa, impulsionou-se ainda mais a quantidade de páginas, de documentos e de recursos disponibilizados diariamente (HIMMA, 2007) (LEVY, 2006). Entretanto, uma vez que a quantidade de conteúdos produzida e publicada é alta, não é plausível considerar a existência de profissionais especializados

controlando e analisando todos esses conteúdos (as dificuldades técnicas e os custos de execução desse processo seriam muito altos). Assim, com a ausência de mecanismos, ou de medidas para assegurar a qualidade das informações na Internet e para organizá-las, rapidamente percebeu-se o problema da existência de um volume expressivo de informação, sem controle, sem padrões ou regulamentações e, principalmente, sem garantia de qualidade. Nesse contexto é que o termo **sobrecarga de informação** (*information overload*) tornou-se popular.

Himma (2007) se propõe a analisar conceitualmente o termo “sobrecarga de informação” e resume sua análise ao afirmar que esse termo define: “*Exposure to or provision of too much information; a problematic situation or state of mental stress arising from this*”². É importante salientar que o conceito de informação utilizado pelo autor refere-se a todo e qualquer conteúdo disponibilizado e acessível na Internet. Himma afirma que os éticos, normalmente, julgam a informação como um bem primário e, assim sendo, o aumento da informação seria sempre uma boa nova, pois existiria um excesso de algo que é bom. Entretanto, dificilmente alguém reclamaria por ter algo bom em grande quantidade. Por essa razão Himma (2007), assim como Levy (2006), chega a sugerir que o termo “sobrecarga de conteúdos” seria mais adequado, pois além de nem todo conteúdo ser considerado informação, o problema existente não está somente no excesso de quantidade, mas também, na falta de qualidade dos conteúdos que são encontrados na Internet.

Definindo o conceito de informação como todo e qualquer conteúdo acessível na Internet, Himma (2007) passa a abordar o termo “sobrecarga”. O autor formaliza o conceito de sobrecarga dizendo que: P tem um excesso de e , se e somente se, P tem muito de e . Mas para dizer que P tem muito de e , é necessária a existência de algum padrão ou norma que defina o que realmente é muito. Além disso, o autor apresenta o excesso como uma condição

² Tradução do autor: “Exposição ou provisão de muita informação; uma situação problemática ou estado de tensão mental que surge em função disto”.

necessária, mas não suficiente, para causar a sobrecarga de informação. O excesso é verificado de acordo com alguma norma e corresponde a ter mais que o necessário, mas a sobrecarga corresponde a um excesso indesejado e que causa algum efeito prejudicial.

Para justificar o excesso de informação, Himma (2007) conclui sobre a complexidade em se estabelecer uma medida ou em se definir um valor que represente a quantidade de informação (ou de conteúdos) que seria ideal. Nessas conclusões, o autor prefere definir “excesso” como o fato de ter-se acesso, ou de estar exposto, a uma quantidade de informação útil maior do que um indivíduo seria capaz de processar produtivamente. Para interpretar esse excesso como sobrecarga, Himma (2007) cita Levy (2005) e afirma que a quantidade de informação disponível na Internet é, de algum modo, prejudicial às pessoas — a informação não pode ser processada sem que se preste atenção à mesma (LEVY, 2005), e existe muito mais informação disponível do que alguém conseguiria prestar atenção durante toda a sua vida. Levy (2005, 2006) argumenta em seus trabalhos que o problema de ter que lidar com muita informação é que se perde um tempo que seria melhor empregado pensando, contemplando e raciocinando. O autor menciona que:

*“...today the library problem encompasses more. It is not simply about the amount of information, or about organizing and finding the most relevant information, although these are surely important concerns. There is an additional dimension to the problem: time — to think, to reflect, to absorb, to muse...”*³ (LEVY, 2005).

Himma (2007) finaliza suas argumentações em torno do termo “sobrecarga de informação” deixando claro que o excesso de conteúdos disponíveis na Internet só é considerado como sobrecarga devido à sua falta de qualidade e ao impacto negativo que esse excesso causa para as pessoas que precisam lidar com ele. O autor conclui afirmando que, se distinguir conteúdo útil de inútil é um desafio, distinguir conteúdo verdadeiro de falso é um

³Tradução do autor: “hoje o problema de biblioteca abrange mais. Não se trata simplesmente da quantidade de informação, ou sobre organizar e encontrar a informação mais relevante, embora estas preocupações sejam certamente importantes. Existe uma dimensão adicional ao problema: tempo - para pensar, refletir, absorver, meditar”.

desafio ainda maior. O problema está no fato de que ninguém é infalível e não existe um algoritmo capaz de distinguir um conteúdo que seja verdadeiro de um conteúdo que seja falso. Assim, é necessário que se trabalhe cuidadosamente com todos os conteúdos, pensando tão profundamente e cuidadosamente quanto os materiais permitirem.

O conceito de sobrecarga de informação utilizado nesse trabalho corresponde às exposições realizadas acima. Entretanto, nos sistemas baseados em folksonomia, compreende-se como informação todo o tipo de objeto categorizado pelos usuários desses sistemas; e compreende-se como sobrecarga o excesso de informação que dificulta a utilização de um sistema baseado em folksonomia. Na seqüência, realiza-se uma explanação sobre a técnica de folksonomia e sobre seu papel dentro do contexto de organização e de recuperação de informação na Internet.

2.2 A Técnica de Folksonomia

Folksonomia é um neologismo criado por Thomas Vander Wal (MATHES, 2004) em 2004 por meio da junção das palavras “*folks*” (em inglês, pessoas) e “*taxonomy*” (taxonomia) resultando em *folksonomy* — para a língua portuguesa, folksonomy foi adaptado principalmente como “folksonomia”, mas também podem ser encontradas sugestões menos utilizadas como “gentonomia”, “classificação popular” ou mesmo “folcsonomia”. O termo folksonomia designa a “classificação realizada pelas pessoas”, ou seja, as pessoas atribuindo sentido, realizando organizações e descrições dos objetos (recursos) presentes na *web*.

Na prática, a folksonomia não representa nada de novo e nem de inovador. A partir de meados de 2005, rótulos de textos livres ou “*tags*” cresceram muito em popularidade e muitos *websites* passaram a permitir que seus usuários realizassem anotações ou utilizassem simples palavras (*tags*), ou frases curtas, na forma de um texto não-estruturado (RUSSELL, 2006). Essas anotações atribuem significados e criam uma espécie de organização dos conteúdos por seus usuários. Nesse sentido, o termo folksonomia passou a ser utilizado para representar a

técnica em que as pessoas utilizam *tags* para a organização de informação na *Web* categorizando URLs, fotos, postagens de *blogs* ou qualquer outro objeto passível de ser referenciado. Nesse mesmo foco, os sistemas que suportam o processo de categorização de objetos por meio de *tags* e que fazem do uso dessas *tags* sua estrutura fundamental de organização, passaram a ser denominados de **sistemas baseados em folksonomia**.

A utilidade das *tags* é muito ampla (RIDDLE, 2005). Quando aplicadas para a categorização de informação na Internet, essas *tags* podem auxiliar na recuperação de informações pessoais, descrevendo-as, organizando-as e estruturando-as, atribuindo ordem e significado. Uma das características principais em torno do uso de *tags* é a liberdade quanto a quantidade e ao tipo de palavras-chaves que podem ser utilizadas para realizar uma categorização. A categorização é totalmente livre, o que proporciona poder e liberdade de trabalho aos usuários — não existe nenhum controle, nenhum esquema a ser seguido ou respeitado, os usuários atribuem as *tags* que quiserem e da forma que julgarem mais adequada. Outra característica principal da folksonomia e que é consequência da liberdade oferecida pela mesma (LEW, 2006), é que seu sucesso está totalmente ligado e dependente dos usuários: se houver vários usuários efetivamente categorizando informações, então ocorrerá uma construção de “sabedoria” por uma multidão — o trabalho individual de organização, de categorização e de atribuição de sentido aos objetos por cada usuário colabora para o coletivo, para o global. Por outro lado, se não houver usuários não haverá sabedoria e nenhum outro tipo de informação a ser analisada. Compreende-se, então, que o sucesso e a obtenção de benefícios em sistemas baseados em folksonomia dependem fortemente do modo como os usuários se sentem ao utilizar esses sistemas: os usuários precisam sentir-se seguros e guiados, pois da mesma maneira que a liberdade é atrativa uma liberdade total poderá deixar esses usuários sem saber o que fazer ou como proceder.

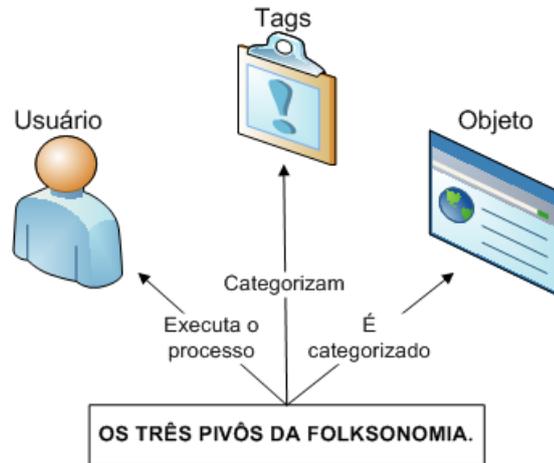


Figura 1 - Os três pivôs da folksonomia.

O processo de categorização de informação fundamenta-se basicamente nos três pivôs da folksonomia, ilustrados pela Figura 1: o **usuário** – que realiza a categorização; o **objeto** – que é categorizado; e as **tags** – que categorizam este objeto (MATHES, 2004) (SHEN *et al.*, 2005). Considerando um exemplo em que o objeto é uma imagem, o usuário atribuirá um conjunto de *tags* selecionadas livremente de acordo com sua preferência para descrever, classificar ou organizar a imagem em questão. Na seção anterior, a impossibilidade da existência de profissionais organizando e avaliando as informações existentes na Internet foi mencionada. A folksonomia representa justamente uma alternativa para essa ausência de controle e de organização, pois em vez de limitar os processos de organização, de avaliação e de atribuição de significado às informações aos autores dos conteúdos e aos editores profissionais, na folksonomia todo e qualquer usuário pode realizar a sua própria organização e da forma que preferir. Os sistemas baseados em folksonomia aplicam a filosofia de que, como os usuários atuam como produtores e publicadores de conteúdos, também é válido considerar que esses usuários possam atuar como organizadores e avaliadores do que é produzido.

2.2.1 Aspectos Gerais sobre a Folksonomia

Existem algumas considerações relevantes sobre a técnica de folksonomia, suas aplicações e implicações que devem ser mencionadas. A primeira particularidade está no consenso em se utilizar a palavra “categorização” no lugar da “classificação” (WU, 2006) (LEW, 2006) (KOME, 2005) (STURTZ, 2004) (MATHES, 2004). A palavra “categorizar” sugere um esquema menos rígido de se organizar algo. Na **categorização** um objeto pode ter vários termos a ele relacionados e pode ser adequado para várias categorias ou para várias formas de organização. Ao contrário, em uma **classificação**, geralmente se aplica ao objeto um único termo hierarquicamente organizado e com relacionamentos bem definidos. Nesse contexto, um artigo científico sobre a economia brasileira poderia se encaixar em várias categorias: economia, estatística, política, investimentos, história entre outros. Isso demonstra que nem sempre um único termo é capaz de classificar um objeto com precisão.

Um exemplo de classificação é a taxonomia de Lineu (ROSSI, 2000) que organiza os seres vivos em hierarquia: Reinos, Filos, Classes, Ordens, Famílias, Gêneros e Espécies. Cada ser vivo existente no planeta é incluído em uma, e apenas uma, dessas classificações. Quando a posição taxonômica de um ser não pode ser determinada, esse ser é chamado de *Incertae sedis*, isto é, tem posição incerta na classificação. Este esquema de organização dos seres vivo é clássico da biologia e não se discute suas contribuições. Entretanto, existem casos em que um esquema mais flexível de organização seria mais adequado: o morcego e o golfinho são classificados como mamíferos. O morcego voa e o golfinho nada, mas nem por isso o morcego é classificado como uma ave ou o golfinho como um peixe, embora ambos também apresentem semelhanças com seres dessas classificações (aves e peixes).

Na folksonomia não existe hierarquias pré-definidas formalmente para o vocabulário de termos utilizado (MATHES, 2004). Não há nenhuma especificação direta de relacionamento de descendência ou de irmandade entre os termos utilizados para a

categorização. As *tags* geradas são relacionadas automaticamente e se agrupam em *tags* baseadas em objetos comuns. Isso é muito diferente da taxonomia formal e de outros esquemas de classificação nos quais existem múltiplos tipos de relacionamentos explícitos entre os termos. Entretanto, o uso de um vocabulário não-controlado é um dos pontos de conflitos na folksonomia (STURTZ, 2004), e discute-se que essa característica tão peculiar seja a fonte de muitos benefícios e de muitos aspectos considerados desvantajosos na utilização dessa técnica (SHEN *et al.*, 2005). Um vocabulário controlado pode melhorar e auxiliar na recuperação de informação (*findability*), mas a folksonomia permite que sejam realizadas grandes descobertas não intencionadas por meio de relacionamentos, categorizações, termos utilizados, etc.

Russel (2005) argumenta que os sistemas baseados em folksonomia permitem que sejam encontradas informações às quais os usuários provavelmente nunca seriam apresentados. Mathes (2004) e Sturtz (2004) caracterizam isso como **serendipismo**⁴, e afirmam que a folksonomia favorece a descoberta de conhecimento e de informações úteis e interessantes, tornando possível a análise de vários aspectos interessantes, tais como: tendências no uso de termos, evolução do conhecimento e do vocabulário das pessoas, autoridades, e uma série de itens que podem ser obtidos por meio do estudo das categorizações realizadas pelos usuários.

Ao considerar as categorizações para organização, descrição e atribuição de significado aos objetos torna-se possível estabelecer uma ligação com **metadados** (dados para descrever dados). A diferença é que o significado construído na folksonomia existe na superfície (STURTZ, 2004), em alto-nível, visível e útil, ao contrário dos *metadados* utilizados que normalmente ficam escondidos dos usuários. A técnica de folksonomia permite a criação distribuída, dinâmica e descentralizada de *metadados* pelos usuários (ver Figura 2),

⁴ Do inglês, *serendipity*, designa a arte de fazer descobertas felizes e não-intencionadas (FILHO, 2002).

agregando valor e realizando boa parte do trabalho que seria executado somente por especialistas (RUSSELL, 2005). Entre as vantagens em permitir que o próprio usuário realize o processo de categorização, colocando ordem e sentido nos objetos por meio da atribuição de *tags*, está a redução de custos de tempo e de investimentos em serviços especializados (MATHES, 2004). O custo existente é o esforço dos usuários em executar o processo de categorização, um custo em termos de esforço e de tempo aplicado para a categorização, mas que se torna pequeno quando comparado ao investimento necessário para manter uma organização realizada por profissionais ou especialistas.

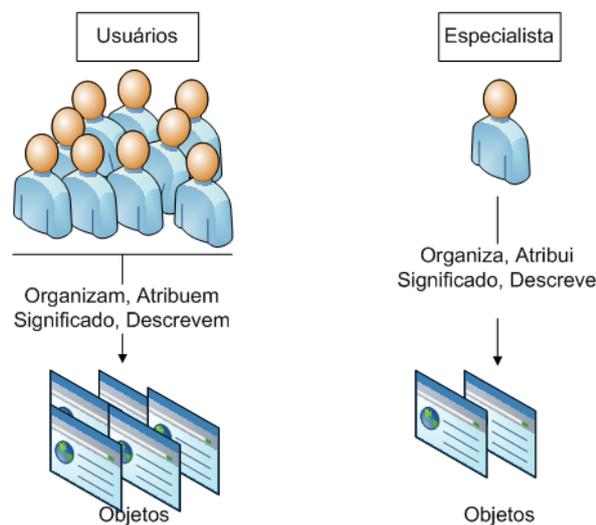


Figura 2 - Folksonomia Vs. organização por especialistas.

Os mecanismos de busca tradicionais, a exemplo do Google (LANGVILLE *et al.*, 2004), possuem algoritmos complexos para a indexação das páginas existentes na Internet de modo a apresentar melhores resultados no processo de recuperação de informação. Os sistemas baseados em folksonomia possuem a vantagem de representar exatamente a opinião dos usuários — são as pessoas que estão realizando categorizações e organizando os objetos que poderão ser analisados e encontrados por outras pessoas. Não é um algoritmo que está atribuindo as *tags* aos objetos, são usuários com opiniões e visões distintas, em contextos diferentes e com objetivos diferentes. De acordo com Mathes (2004) e com Sturtz (2004), talvez o benefício mais importante da folksonomia seja a sua capacidade de refletir o

vocabulário dos usuários. Os autores argumentam que num sistema de recuperação de informação existem vários vocabulários envolvidos, entre eles os vocabulários: do usuário do sistema, do projetista, do autor do material, do responsável pela definição das classificações, e assim por diante. Porém, esses vocabulários nem sempre são adequados ou correspondem ao vocabulário dos usuários, comprometendo, assim, a utilidade e a eficiência do sistema. Diferentemente, os sistemas baseados em folksonomia refletem diretamente as ações realizadas pelos usuários, suas escolhas, terminologias e precisões — os usuários constroem e escolhem o vocabulário que será utilizado. Mathes (2004) afirma que, com uma natureza não-controlada e com um crescimento orgânico, a folksonomia tem a capacidade de se adaptar muito rapidamente às mudanças no vocabulário e às necessidades dos usuários, o que não acontece em nenhum outro tipo de sistema.

A folksonomia utilizada para a categorização de objetos é dividida, segundo Wal (2005a), em dois tipos: folksonomia larga (*broad folksonomy*) e folksonomia estreita (*narrow folksonomy*). Wal (2005a) explana que na **folksonomia larga** existem muitas pessoas categorizando o mesmo objeto e cada pessoa pode categorizar o objeto com os termos de seu próprio vocabulário. No sentido prático, as pessoas criam conteúdos (objetos) e os disponibilizam aos outros. Outras pessoas categorizam os objetos com termos, ou *tags*, que escolheram ou que julgaram adequados e, também, encontram informações baseadas nas *tags* que foram utilizadas para a categorização — o *Delicious*, descrito na próxima seção, é um sistema que aplica o conceito de folksonomia larga. A **folksonomia estreita** pode ser então compreendida como uma única pessoa, ou um pequeno grupo de pessoas, categorizando um objeto para recuperar ou reencontrar novamente determinada informação. Normalmente, o criador do conteúdo atribui um conjunto de *tags* ao objeto para organizá-lo e encontrá-lo novamente com mais facilidade — o *Flickr* é um exemplo de aplicação do conceito de folksonomia estreita. Deste modo, pode-se dizer que, enquanto na folksonomia estreita o

usuário categoriza apenas os próprios conteúdos, na folksonomia larga os usuários categorizam além dos seus conteúdos, os conteúdos produzidos pelos outros usuários.

Tanto a folksonomia larga quanto a folksonomia estreita apresentam a vantagem de permitir que os usuários organizem, analisem e descrevam as informações, (RIDDLE, 2005) (RUSSELL, 2005) (SHAW, 2005) (WAL, 2005) (MATHES, 2004) (STURTZ, 2004). Entretanto, na folksonomia larga é que se observa o aspecto mais interessante da aplicação da técnica de folksonomia: o fenômeno da coletividade que é uma característica dos softwares sociais. Quando várias pessoas categorizam o mesmo objeto, a utilização das *tags* para realizar essas categorizações forma um *long tail* (ANDERSON, 2006) — ou seja, forma-se uma curva que respeita uma lei de potência (*power law*) (BRYNJOLFSSON *et al.*, 2007) (HALPIN *et al.*, 2007), demonstrando o modo como uma grande quantidade de pessoas vê esse objeto.

A curva de lei de potência é uma representação particular (RECUERO, 2005), uma curva logarítmica que decresce abruptamente a níveis próximos de um mínimo e mantém-se assim, construindo uma cauda longa (o *long tail*). A distribuição de lei de potência é relacionada ao Princípio de Pareto (NEWMAN, 2006) que determina que grande parte dos efeitos são causados por uma pequena parcela das causas, e que grande parte das causas acarretam numa pequena parcela dos efeitos. No contexto da utilização de *tags* na folksonomia, isso implica que uma abundância de *tags* é utilizada poucas vezes para a categorização de objetos, e que uma pequena, mas significativa minoria de *tags* é aplicada para a maioria das categorizações — a Figura 3 ilustra essa distribuição.

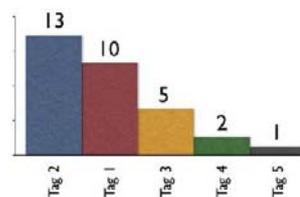


Figura 3 – Curva de *power law*, distribuição de Pareto, *long tail* (WAL, 2005a).

O gráfico criado por Wal (2005a), ilustrado na Figura 3, é a primeira representação utilizada para demonstrar o *long tail* criado pelas *tags*. Wal explica que as *tags* “1” e “2” são as mais utilizadas e demonstram com facilidade a “concordância” dos categorizadores sobre o objeto. Essas *tags* poderiam ser utilizadas para a extração de um vocabulário controlado ou, pelo menos, para identificar como o objeto é rotulado ou organizado pela maioria das pessoas. Por outro lado, também é possível identificar as *tags* situadas no lado direito da curva que formam o *long tail*. Nesta parte da curva é que se encontra a minoria de pessoas que categorizam o objeto por um termo específico, mas essas pessoas quando classificam os objetos deste modo mais diferenciado, permitem que outras pessoas com um vocabulário mental similar encontrem esse objeto, mesmo quando não são utilizados os termos empregados pelas massas ao realizar as categorizações. Os benefícios da curva de lei de potência e do *long tail* são obtidos com a folksonomia larga, na qual muitas pessoas categorizam o mesmo objeto de forma compartilhada e aberta.

Steven Johnson discute em seu livro sobre emergência (JOHNSON, 2003) que os sistemas emergentes dispensam o controle centralizado e se organizam de baixo para cima (*bottom-up*). A organização dos objetos na folksonomia é citada e compreendida como o resultado de uma taxonomia criada seguindo essa lógica (OHMUKAI, 2005) (STURTZ, 2004), sem controle, sem restrições, emergindo do trabalho coletivo realizado para a categorização de informações individuais. Existem questões interessantes e importantes em torno da organização de informação por folksonomia que podem ser relacionadas e estudadas sob o ângulo da emergência de informação. Identificar e utilizar o conhecimento que pode ser gerado em sistemas baseados em folksonomia (WU, 2006) é um grande desafio para a área de emergência e para os próprios estudos em torno da folksonomia. Neste contexto, Wu (2006) argumenta que na folksonomia estão contidas as estruturas de conhecimento dos indivíduos

sobre os objetos, que definem como os conceitos em um domínio estão relacionados, e que podem trazer benefícios para a sociedade como um todo.

2.2.2 Sistemas Baseados em Folksonomia

Os sistemas baseados em folksonomia representam uma iniciativa importante para auxiliar na construção de significados e na organização de informação na *Web*. Esses sistemas permitem que os usuários categorizem os conteúdos disponíveis na Internet anotando-os com um conjunto de palavras-chaves livremente selecionadas (SHAW, 2005). Com o advento da *Web 2.0* (O'REILLY, 2005) (STURTZ, 2004), sistemas baseados em folksonomia começaram a ser desenvolvidos para permitir que os usuários atuassem ativamente no processo de organização de informação. Dentre os sistemas baseados em folksonomia existentes, os sistemas *Delicious*⁵ e *Flickr*⁶ são considerados os mais maduros e são os principais exemplos citados (STURTZ, 2004) (SHEN, *et al.*, 2005) (RIDDLE, 2005) (RUSSELL, 2006).

O sistema *Delicious* (YAHOO, 2003) pode ser definido como um gerenciador de favoritos (*bookmarks*) social que permite que URLs de qualquer página da Internet sejam adicionadas, classificadas com *tags*, acessadas em qualquer máquina e, principalmente, compartilhadas com outros usuários (Figura 4). Ele é considerado como o exemplo mais completo da aplicação da técnica de folksonomia atual. O *Delicious* implementa o conceito de folksonomia larga (*broad folksonomy*) na qual os usuários categorizam seus próprios conteúdos e os conteúdos produzidos por qualquer outra pessoa, entidade ou organização. Apesar de não ser o único sistema a oferecer esse tipo de serviço, em uma pesquisa realizada pela KDnuggets (KDNUGGETS, 2007) o *Delicious* aparece como o gerenciador de *bookmarks* mais utilizado na Internet (25%), seguido pelo *Digg*⁷ e pelo *Slashdot*⁸ (12.9% e 9.5% respectivamente). Segundo divulgado pelos administradores do sistema, em 2006 a

⁵ <http://del.icio.us/>

⁶ <http://www.flickr.com/>

⁷ <http://www.digg.com/>

⁸ <http://slashdot.org/>

marca de um milhão de usuários já havia sido ultrapassada (DELICIOUS, 2006). A principal diferença entre o *Delicious* e os demais gerenciadores de *bookmarks* é justamente a ênfase no usuário adicionando as *tags* como uma construção organizacional fundamental, o que o caracteriza como um sistema baseado em folksonomia.

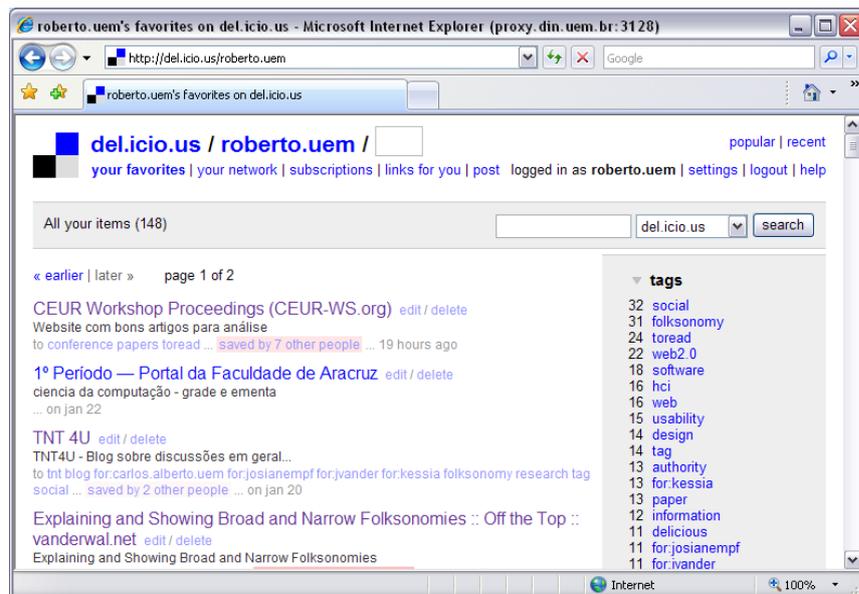


Figura 4 - Página de pesquisa do sistema *Delicious*.

Na outra perspectiva da utilização da técnica de folksonomia, aplicando o conceito de folksonomia estreita (*narrow folksonomy*), está o sistema *Flickr* (YAHOO, 2004). O *Flickr* é um compartilhador de fotos pessoais que permite aos seus usuários a inserção de *tags* para categorizar e organizar suas imagens. Inicialmente o *Flickr* permitia apenas ao proprietário da imagem a atribuição de *tags* para categorizá-la, mas em seguida desenvolveu um recurso em que o proprietário da imagem pode definir um ou mais usuários que terão permissão para categorizá-la também. A anotação de imagens por meio de *tags* permite que mecanismos de busca e outras ferramentas possam trabalhar de forma mais eficaz essas imagens e seus significados — comparando com outros tipos de conteúdos, as imagens, normalmente, ficam restritas à poucas informações de texto relacionadas às mesmas, sendo mais difícil de se realizar um processamento ou uma análise com essas imagens. Os administradores do *Flickr* pretendem permitir que qualquer usuário categorize qualquer foto, o que deixará de

caracterizar o sistema como de folksonomia estreita para caracterizá-lo como de folksonomia larga.

O *Youtube*⁹ (GOOGLE, 2005) também pode ser considerado um sistema baseado em folksonomia que implementa o conceito de folksonomia estreita, no qual os usuários atribuem *tags* para a categorização de vídeos. Mais recentemente, o *Gmail*¹⁰ e o *Orkut*¹¹ também passaram a trabalhar com *tags*, para a organização de mensagens eletrônicas no caso do *Gmail*, e para a descrição de perfis virtuais e para a realização de pesquisas no caso do *Orkut*, mas isso não os caracteriza como sistemas baseados em folksonomia por que eles não fazem do uso de *tags* a sua estrutura de funcionamento central. Os sistemas citados são alguns exemplos que demonstram a tendência da disponibilização de recursos que incentivem a utilização de *tags*, a participação dos usuários e a utilização dos benefícios obtidos das informações geradas.

2.2.3 Dificuldades e Desafios na Folksonomia

Ao mesmo tempo em que a aplicação da folksonomia proporciona benefícios para a organização de informação, também identifica-se alguns problemas que podem comprometer a sua utilidade (SHAW, 2005). Em um sistema como o *Delicious*, na medida em que aumenta o número de usuários os mesmos problemas relacionados à sobrecarga de informação são encontrados: existe uma grande quantidade de informação sendo categorizada por um grande grupo de usuários, e isso acarreta na dificuldade de encontrar a informação desejada e de avaliá-la.

A natureza não-controlada da folksonomia é fundamentalmente caótica e sofre com problemas de falta de organização e de ambigüidade que o desenvolvimento de vocabulários controlados e de esquemas de hierarquia efetivamente melhora (MATHES, 2004). A ausência de controle sobre as categorizações e sobre as *tags* utilizadas implica em uma tendência para a

⁹ <http://br.youtube.com/>

¹⁰ <http://www.gmail.com/>

criação de informações inúteis do ponto de vista coletivo (WU, 2006), para a idiossincrasia das *tags*, para a geração de *metadados* ruidosos e, conseqüentemente, para a redução da utilidade dos sistemas baseados em folksonomia.

O fato das categorizações serem fundamentadas em *tags* também apresenta algumas limitações, pois *tags* não funcionam bem para medidas de precisão e de recuperação (RIDDLE, 2005). Nessa situação, como o usuário pode aplicar diferentes *tags* para um mesmo objeto, ou uma única *tag* para vários objetos distintos, o resultado de uma pesquisa pode estar tanto cheio de ruídos quanto incompleto. Devido à total liberdade dos usuários no processo de categorização em sistemas baseados em folksonomia, não há um mecanismo de controle de sinônimos e, assim, problemas de polissemia e de ambigüidade também se tornam claramente visíveis (MATHES, 2004). A *tag* “janela”, por exemplo, pode ser aplicada em diferentes contextos e assim gerar ambigüidades nos resultados: janela para uma casa — construção civil; janela de um programa — desenvolvimento de software; janela de oportunidades — linguagem figurada.

O impacto causado pela ausência de controle nos termos utilizados para a categorização torna-se maior devido à heterogeneidade dos usuários (idioma, cultura, nível de instrução). Palavras em formatos diferentes, siglas, abreviações, gênero e número também causam problemas. O uso de um vocabulário controlado é empregado em muitas configurações justamente por causa de problemas dessa natureza. Ohmukai (2005) expõe discussões sobre uma folksonomia controlada. No entanto, como Mathes (2004) e Riddle (2005) discutem, um vocabulário controlado parece ser praticamente inviável no caso de aplicativos como o *Delicious* e *Flickr*. Por outro lado, Shen (2005), Sturtz (2004) e Mathes (2004) defendem a pesquisa e o desenvolvimento de recursos que permitam usufruir melhor dos benefícios proporcionados pela folksonomia. Estes autores argumentam que a principal

¹¹ <http://www.orkut.com>

causa do sucesso de sistemas como o *Delicious* e o *Flickr* é o *feedback* imediato que eles proporcionam aos usuários. Além disso, esses sistemas deixam visíveis os aspectos individuais e coletivos relacionados ao comportamento dos usuários na organização de informação, algo que não pode ser obtido com tamanha riqueza de detalhes em um sistema com vocabulário controlado. Os autores defendem a necessidade de investir em recursos adicionais que melhorem os sistemas baseados em folksonomia, pois tentar controlá-la é o mesmo que descaracterizá-la e passar a adotar uma taxonomia mais liberal.

As *tags* também não são boas para a geração de ontologias, no sentido de estabelecer relacionamentos entre conceitos precisamente definidos, uma vez que elas são criadas sem a presença de qualquer tipo de controle (RIDDLE, 2005). Existem pesquisas, como em (GRUBER, 2007) (ECHARTE *et al.*, 2007) e (WU *et al.*, 2006), sendo desenvolvidas para tentar gerar ontologias a partir dos termos utilizados para a categorização de objetos em sistemas baseados em folksonomia. Além da geração de ontologias, a recomendação de documentos, de autoridades e de conteúdos baseada nas informações emergidas, assim como a análise quantitativa e qualitativa das *tags*, também são citadas como desafios (WU, 2006). Wu (2006) considera que o “aproveitamento do conhecimento social” é o grande benefício da folksonomia, e o autor também concorda que é necessário o desenvolvimento de pesquisas e de ferramentas que permitam o aproveitamento das informações geradas por meio dessa técnica.

O fato de a folksonomia ser uma técnica relativamente recente levantou discussões sobre a viabilidade de sua aplicação. Mathes (2004) realizou um dos primeiros estudos sobre essa viabilidade e procurou compreender por que a folksonomia funcionava. O autor levantou questões sobre a inexistência de métricas para medir o sucesso de um sistema baseado em folksonomia. Existem fatores como a quantidade de usuários e o montante de categorizações sendo realizadas que podem ser considerados, mas não há nenhuma métrica formal que

permita identificar se um sistema realmente está dando certo. Também não é possível saber se vale a pena investir no desenvolvimento desses sistemas, da mesma forma que os usuários não podem prever se vale a pena empregar esforços categorizando informações. No entanto, desde a exposição de suas argumentações, a popularidade da técnica de folksonomia aumentou tanto entre os usuários quanto no meio acadêmico. Alguns trabalhos posteriores a Mathes (2004), como (RUSSELL, 2006), (WU, 2006) e (HALPIN *et al.*, 2007), já não discutem a aceitação ou a viabilidade da utilização da folksonomia. A principal questão está no desenvolvimento de recursos que tornem os sistemas baseados em folksonomia mais úteis, que ofereçam apoio aos usuários e que permitam a emergência ou o aproveitamento do conhecimento social gerado nesses sistemas. Segundo esses autores, a folksonomia tem se apresentado bem ajustada a campos em desenvolvimento como a Internet, em contraste com esquemas de hierarquias (ex: diretórios do Yahoo®) descritos como “excelentes” para a representação de conhecimento em domínios maduros e nos quais as naturezas das entidades das relações significantes são bem conhecidas (ver Figura 5).

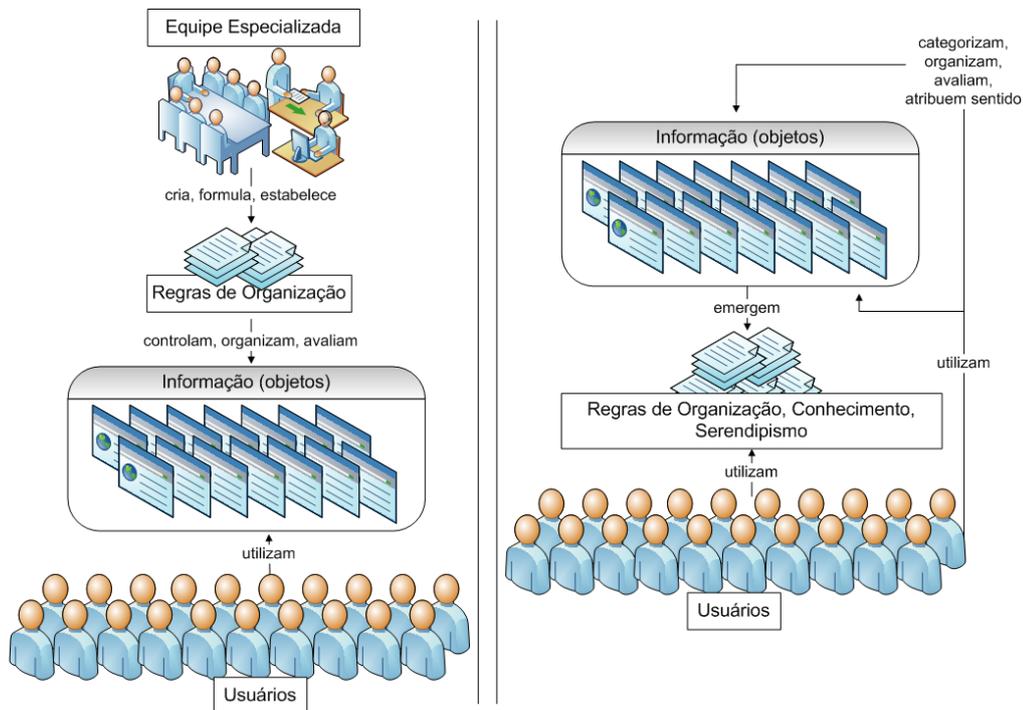


Figura 5 - Sistemas controlados Vs. sistemas baseados em folksonomia.

A Figura 5 ilustra a diferença entre a estrutura de organização de sistemas controlados e de sistemas baseados em folksonomia. Os sistemas que utilizam um esquema de hierarquia ou uma estrutura rígida (por exemplo: diretórios do Yahoo![®]) possuem suas regras de organização definidas por especialistas, ou por um pequeno grupo de pessoas. Essas regras normalmente são estáticas, bem definidas e estruturadas. Por fim, os usuários utilizam esses sistemas e precisam se adaptar às suas regras, terminologias, vocabulários, entre outros. Por outro lado, nos sistemas baseados em folksonomia não existe nenhum tipo de regra ou de controle. Os usuários trabalham individualmente para a categorização e organização de informação e, desse trabalho individual, emerge uma estrutura coletiva de organização. Os usuários podem então utilizar essa estrutura para auxiliar na realização de futuras categorizações e para a utilização do sistema de uma forma geral. Nesses casos, o sistema reflete diretamente o vocabulário, o conhecimento, os interesses e os relacionamentos entre seus usuários, além de vários aspectos sociais que podem ser inferidos por meio dessas informações.

Considerando as vantagens e as desvantagens dos sistemas baseados em folksonomia, é importante compreender que a técnica de folksonomia não pretende substituir os esquemas de organização tradicionais e controlados e que não há dúvidas sobre as contribuições geradas por esses esquemas (SHEN *et al.*, 2005) (KOME, 2005) (SHAW, 2005) (STURTZ, 2004). A folksonomia é uma alternativa para ambientes em que não existe ou não é possível (ou viável) manter um esquema de organização rígido. Ela tem se mostrado adequada para ambientes mutáveis, com grande quantidade de usuários e de informações, permitindo uma organização, mesmo que superficial dessas informações. Em ambientes com essas características, é preferível adotar um esquema, mesmo que com características anárquicas como a folksonomia, do que não adotar nada.

2.3 Perspectivas na Folksonomia e o Problema da Sobrecarga de Informação

De acordo com discussões sobre a técnica de folksonomia que foram realizadas neste capítulo, é possível afirmar que os sistemas baseados em folksonomia são uma alternativa interessante e têm muito a colaborar com a organização de informação na Internet. Considerando os conflitos existentes entre os pontos favoráveis e desfavoráveis que envolvem esse tipo de organização, Stewart Butterfield, um dos criadores do *Flickr*, argumenta que a diferença na complexidade entre folksonomia e esquemas de classificação rígidos é muito importante (SHEN *et al.*, 2005). Segundo Stewart, é justamente a falta de hierarquia, de controle de sinônimos e de precisão semântica que constróem a folksonomia, e não é viável tentar controlar tudo isso, pois esse controle resultaria em muito trabalho cognitivo para o usuário, custo em tempo e em esforço, o que teria pouco valor na prática. É nesse sentido que se encontra a necessidade de desenvolvimento de recursos para oferecer suporte aos usuários de sistemas baseados em folksonomia sem que a liberdade de trabalho, a facilidade de uso e a praticidade sejam comprometidas.

Russell (2005) menciona que as pessoas, normalmente, possuem compulsão por compartilhar, por emitir uma opinião, por mostrar o que estão fazendo ou pensando. O êxito dos *blogs*, *fotologs* e do *Youtube* são exemplos claros disso. Entretanto, é necessário que essas ferramentas disponibilizadas aos usuários atentem efetivamente para as questões relacionadas à interação humano-computador e a aspectos éticos e sociais (TAIT *et al.*, 2003). Os sistemas baseados em folksonomia devem ser projetados tendo o usuário em mente. É essencial que a interface desses sistemas tenha controles com operações e efeitos óbvios, claros, e que proporcionem um *feedback* imediato e útil aos seus usuários (ROCHA & BARANAUSKAS, 2003). Se o custo em utilizar um sistema for muito alto frente ao benefício obtido com essa utilização, o usuário não hesitará em procurar um recurso alternativo.

Os sistemas *Delicious* e *Flickr* já permitem a obtenção de algumas informações geradas pelos seus usuários. Existem *plug-ins* e interfaces de comunicação sendo desenvolvidos para trabalhar com esses sistemas. A análise das informações geradas pelos usuários de sistemas baseados em folksonomia, suas categorizações, seus vocabulários, interações, etc., podem trazer contribuições para diversas áreas de pesquisa e, também, possibilitar o desenvolvimento de estudos que vão desde assuntos da computação até outras áreas como sociologia, lingüística e psicologia. Como mencionado por Wu (2006), o aproveitamento de conhecimento social é o maior benefício (e também desafio) que pode ser obtido dos sistemas baseados em folksonomia.

No que se relaciona a manutenção e a eficiência, a ausência de controle nas categorizações e nos vocabulários dos usuários impacta diretamente na qualidade das informações recuperadas em sistemas baseados em folksonomia. Nesse sentido, um sistema como o *Delicious* tem a utilidade comprometida proporcionalmente ao aumento da quantidade de usuários e de informação produzida e armazenada. A questão central novamente aparece: “os sistemas baseados em folksonomia funcionam por causa de seus usuários ao mesmo tempo em que suas falhas também aparecem por causa deles”. O que acontece nesses sistemas é que a grande quantidade de informação sendo categorizada sem controle de organização e de qualidade faz com que os usuários, ao recuperarem informação, possam confiar apenas no “senso comum” da maioria dos usuários ou em sua própria opinião, uma vez que aspectos relacionados à autoridade e à confiança de cada categorizador não são considerados.

Tomando como exemplo o sistema *Delicious*, ao solicitar a recuperação de *bookmarks* categorizados com a *tag* “janela”, o usuário receberá 1094 resultados¹² distribuídos em aproximadamente 110 páginas¹³. Segundo dados divulgados pela iProspect (2007), cerca de

¹² Recuperação efetuada em 24 de Março de 2008.

¹³ Considerando a utilização do padrão do *website* que exibe 10 *bookmarks* por página de resultados.

81.7% dos usuários de ferramentas de busca não consideram os resultados apresentados além da terceira página de informações recuperadas. A pesquisa detalha ainda que 22.6% dos usuários analisam somente os primeiros resultados, 18.6% consideram toda a primeira página, 25.8% analisam até a segunda página e apenas 14.7% consideram a terceira página de resultados. Caso as informações encontradas não forem satisfatórias, novos termos são utilizados e a busca é reiniciada (ver Figura 6).

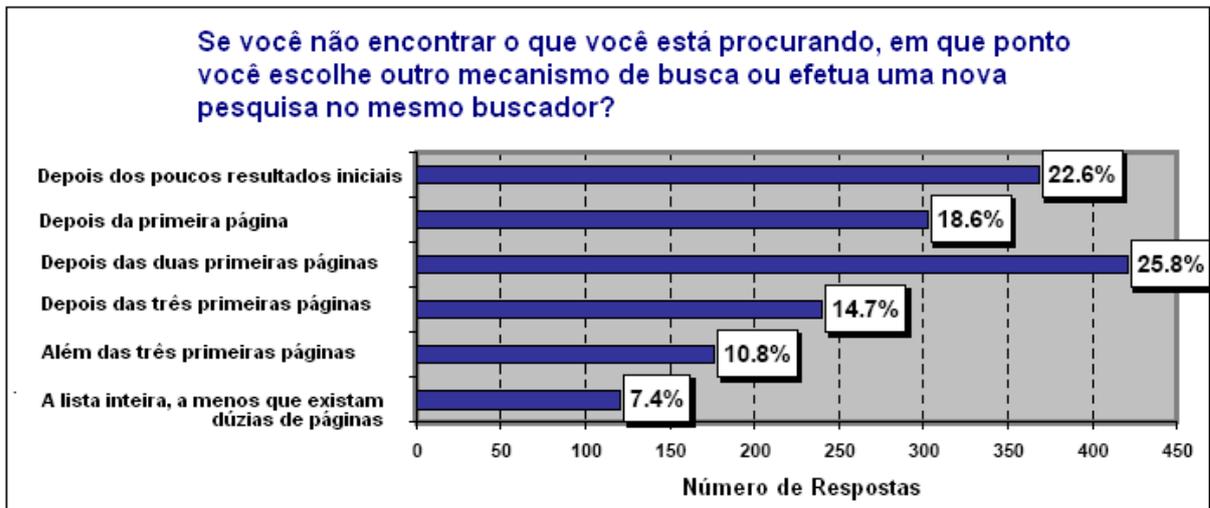


Figura 6 - Limite dos resultados considerados pelos usuários, adaptado (IPROSPECT, 2007).

Ainda considerando o exemplo anterior de recuperação de informação com a *tag* “janela”, são retornados *bookmarks* referentes a *blogs* de poesia, *websites* de economia e oportunidade de negócios, construção e arquitetura, técnicas de programação, *design* de *software*, publicidade, entre outros (em uma recuperação utilizando a *tag* “ant”, os resultados são ainda mais diversificados). Isto se deve ao fato de que o *Delicious*, com mais de um milhão de usuários, considera todas as categorizações realizadas por todos os usuários. O resultado é o retorno de uma grande quantidade de informação categorizada por diferentes usuários, em diferentes contextos, com diferentes culturas e conhecimentos. Como não existe nenhum aspecto relacionado à qualidade que possa ser considerado, a sobrecarga de informação também começa a se fazer presente. Entretanto, as diferenças nas categorizações não significam que os usuários estejam errados ou que eles não estejam categorizando

corretamente. O que os usuários estão categorizando, normalmente, é apropriado para eles mesmos ou para um pequeno grupo, porém, nem sempre suas categorizações trarão benefícios para o coletivo ou para todos os demais usuários do sistema. Wal (2005) expõe essa questão demonstrando o *long tail* formado pelas *tags* aplicadas nas categorizações de objetos. Deve-se levar em conta que existem áreas nas quais os usuários são mais, ou menos, apropriados como fontes de informação, mas isso depende da capacidade, da base e da competência que eles possuem nos assuntos sobre os quais estão categorizando — a autoridade cognitiva desses usuários.

O algoritmo *PageRank* (LANGVILLE *et al.*, 2004), utilizado pelo buscador *Google*¹⁴ para a indexação de páginas, aplica um esquema de pesos que influencia nos resultados de uma pesquisa. O peso de um objeto é determinado em um cálculo que considera o peso dos outros objetos que apontam para ele: assim um *website* possuirá maior peso e ocupará uma boa posição na recuperação de informação se ele for apontado por outros *websites* que também possuem um peso alto. Como nos sistemas baseados em folksonomia são os próprios usuários que realizam as categorizações de informação, obter melhores ou piores resultados num processo de recuperação é mais relacionado aos usuários categorizadores que são considerados no processo, do que à forma como as categorizações foram realizadas.

No próximo capítulo, apresenta-se a teoria da autoridade cognitiva e discute-se a sua aplicação como uma alternativa para a melhoria da qualidade e da relevância e para a redução dos problemas causados pela sobrecarga de informação nos sistemas baseados em folksonomia.

¹⁴ <http://www.google.com.br>

Capítulo III

Autoridade Cognitiva

Como uma iniciativa para auxiliar nos processos de organização e de construção de significado das informações na Internet, os sistemas baseados em folksonomia inovaram ao permitir que os próprios usuários fossem os organizadores dessas informações. No capítulo II, juntamente com a apresentação da técnica de folksonomia, foram expostas discussões sobre o fato de que a heterogeneidade do conjunto de usuários, aliada à total liberdade de trabalho que os usuários possuem ao categorizar informações, colabora para que o problema da sobrecarga de informação também seja encontrado nos sistemas baseados em folksonomia. Considerando que aspectos relacionados à qualidade e à relevância das informações categorizadas são diretamente dependentes e relacionados aos usuários que realizaram essas categorizações, percebe-se que reconhecer e utilizar a autoridade cognitiva dos usuários (entidades categorizadoras de informação) para auxiliar no processo de recuperação de informação é uma alternativa interessante. Neste capítulo, apresenta-se a teoria da autoridade cognitiva, discute-se a relação entre autoridade e

qualidade, e sugere-se a utilização das autoridades para auxiliar no processo de recuperação de informação em sistemas baseados em folksonomia.

3.1 A Questão da Autoridade

Em disciplinas como Filosofia, Educação, Ciência Política, Religião e Ciência da Informação são encontradas diferentes definições para o termo “autoridade”. Wilson (1983) definiu o termo “autoridade cognitiva” (*cognitive authority*) em seu livro “*Second-hand Knowledge: An Inquiry into Cognitive Authority*”, para explicar o tipo de autoridade que influencia na forma como as pessoas pensam, acreditam ou compreendem as coisas. De acordo com Wilson, a autoridade cognitiva é um tipo particular e diferenciado de autoridade que se distingue, principalmente, do que se denomina de “autoridade administrativa” (*administrative authority*) — uma autoridade estabelecida em uma posição hierárquica, por imposição ou por situação.

Rieh & Belkin (1998; 2000) (RIEH, 2005) e Russell (2005) citam que R. T. De George, autor do livro “*The Nature and the Function of Epistemic Authority*”, apresenta considerações semelhantes às de Wilson, expondo, porém, a relação de um indivíduo com a autoridade como **sendo uma autoridade** (*epistemic*), ou **estando em autoridade** (*deontic*). Ser uma autoridade para De George equivale ao conceito de autoridade cognitiva de Wilson (1983), representando a influência que a autoridade pode causar na forma de pensar de um indivíduo, pois este a julga própria, digna de crédito e de confiança — uma autoridade que, segundo Wilson, define “*who knows what about what*” (quem sabe o quê sobre o quê). Ainda, estar em autoridade na visão de De George corresponde à autoridade administrativa de Wilson. Uma autoridade fundamentada em princípios hierárquicos na qual um indivíduo tem o poder para dizer aos outros o que eles devem fazer.

No livro “*Ensaio de Sociologia*”, Max Weber (1963) levanta questões relacionadas a “quando” e “por quê” os homens obedecem e respeitam autoridades e, também, questões

relacionadas às justificativas e aos meios exteriores que suportam esse domínio. Weber aborda a questão de autoridade classificando-a de acordo com sua origem e influência em: autoridade tradicional; autoridade carismática; e autoridade legal. Em um foco distinto Davenport (1998), abordando o gerenciamento de conhecimento nas organizações, indica a necessidade de uma mudança nos modelos de organização para que se passe a considerar o conhecimento pessoal dos indivíduos envolvidos — a autoridade cognitiva. Essa mudança é apresentada por Vasconcelos (2002) em sua abordagem sobre a questão da racionalidade, da autoridade e da burocracia, servindo como bases para a definição de um tipo organizacional pós-burocrático. Vasconcelos fundamenta-se nas definições de autoridade de Max Weber e menciona que, em uma organização, a autoridade deve ser vista como resultado não apenas de propriedades formais do cargo, mas também das competências pessoais de um indivíduo e de sua capacidade de liderança pessoal. Assim, Vasconcelos cita a ascensão de uma definição de autoridade que não deriva de posições formais na hierarquia, mas de relações pessoais e grupais centradas na capacidade pessoal. Pode-se, assim, compreender que a autoridade de que Vasconcelos fala deriva da definição de autoridade carismática de Max Weber e fundamenta-se nos mesmos conceitos da autoridade epistêmica de De George e da autoridade cognitiva de Wilson.

Ao justificar a escolha da teoria da autoridade cognitiva para o reconhecimento de competências dos indivíduos, Russel (2005) argumenta que as pessoas são criaturas sociais e que as interações no mundo envolvem, pelo menos até certo nível, outras pessoas. Essas interações, segundo o autor, estão constantemente mudando a visão sobre o mundo e sobre as pessoas que existem nele, pois todos passam por mudanças semelhantes. A reputação de uma pessoa é definida pela opinião que as outras pessoas possuem dela, assim como a identidade social e o lugar que ela ocupa no mundo também o são. Entretanto, Russel afirma que essas questões de reputação, opinião e identidade social não podem ser facilmente medidas e

identificadas e, então, propõe a concessão de autoridade cognitiva por meio de folksonomia para o reconhecimento das habilidades das pessoas e para a identificação de autoridades.

A autoridade cognitiva no contexto deste trabalho refere-se a uma autoridade relacionada à forma como as pessoas confiam nas opiniões, conselhos e pensamentos das outras pessoas (WILSON, 1983). Não se pretende discutir aqui a autoridade administrativa e as relações entre chefes e subordinados e nem as razões sobre as quais uma pessoa concede autoridade a outra, essas discussões podem ser encontradas em (WEBER, 1963). O foco principal está nas mesmas questões levantadas por Russell (2005) para o reconhecimento e para a identificação de autoridades e, também, nas discussões realizadas por Rieh & Belkin (1998; 2000) (RIEH, 2005) sobre a forma como as pessoas julgam a qualidade das informações e a autoridade cognitiva na Internet. Nas próximas seções apresentam-se as bases teóricas e filosóficas da teoria da autoridade cognitiva e situa-se a sua aplicação no contexto deste trabalho.

3.2 A Teoria da Autoridade Cognitiva

A teoria da autoridade cognitiva é uma teoria da epistemologia social sobre o modo como as pessoas identificam, reconhecem e atribuem autoridade umas às outras (WILSON, 1983). Uma autoridade fundamentada na competência e nas capacidades intelectuais de quem a recebe, e cuja concessão é compreendida como o reconhecimento e como um mérito por essas capacidades — uma autoridade que define “quem sabe o quê sobre o quê” (*who knows what about what*). Wilson inicia suas explanações sobre autoridade cognitiva mencionando:

*“All I know of the world beyond the narrow range of my own personal experience is what others have told me. It is all hearsay. But I do not count all hearsay as equally reliable. Some people know what they are talking about, other do not. Those who do are my cognitive authorities.”*¹⁵

(WILSON, 1983. p.13)

¹⁵ Tradução do autor: “Tudo o que eu sei além dos limites de minha experiência pessoal é o que outros me contaram. Tudo é boato. Mas eu não considero todos os boatos como igualmente confiáveis. Algumas pessoas sabem sobre o que estão falando, outras não. Aquelas que sabem são as minhas autoridades cognitivas.”

3.2.1 Conhecimento de Primeira e de Segunda-Mão

A teoria da autoridade cognitiva de Wilson (1983) explica que as pessoas constroem seu conhecimento por meio de duas formas distintas: a primeira é baseada em suas próprias experiências (conhecimento de primeira mão, em inglês: *first-hand knowledge*); a segunda é aquela em que as pessoas aprendem por meio de outras pessoas (conhecimento de segunda-mão, ou *second-hand knowledge*). Deste modo, o que as pessoas aprendem em primeira-mão depende do grupo de idéias que essas pessoas possuem e que elas tentam interpretar de acordo com o que elas vivenciaram no mundo real — as suas experiências diretas. Por outro lado, também é válido dizer que muito do que essas pessoas pensam sobre o mundo deve-se às idéias adquiridas por conhecimento de segunda-mão — idéias vindas de outras pessoas. Isso possibilita que os indivíduos tenham idéias e informações além dos limites de suas experiências diretas.

Durante toda a vida as pessoas ocupam um número muito limitado de posições sociais, cargos, funções, etc. (WILSON, 1983). Se tudo o que um indivíduo pensar depender exclusivamente de suas experiências diretas, das idéias que ele traz e interpreta de acordo com o mundo real, então esse indivíduo criará um sentido muito pequeno do mundo. As pessoas dependem principalmente dos outros para ter idéias, bem como para obter informações sobre assuntos que estão fora de suas experiências diretas, que se encontram além de sua capacidade de observação e de suas perspectivas.

Obviamente, não faz sentido dizer que o conhecimento de segunda-mão é mais importante que o conhecimento de primeira-mão — ouvir alguém falar sobre uma música não é a mesma coisa que escutá-la. Entretanto, além da possibilidade das pessoas aprenderem e conhecerem mais do que suas experiências diretas permitem, existem casos, como o de experiências ruins ou difíceis, nos quais o conhecimento de segunda-mão é mais desejado. A teoria da autoridade cognitiva levanta duas questões importantes: 1 – “O que leva as pessoas a

procurar o conhecimento de segunda-mão?” e 2 – “A quem elas recorrem?”. A resposta para a primeira questão pode ser resumida em uma palavra: “necessidade” e para a segunda da seguinte forma: “as pessoas recorrem àqueles que elas julgam saberem algo que elas mesmas não sabem”. Wilson enfatiza que a questão não é estudar, ou entender, “o que é conhecer sobre o mundo”, mas sim, o que as pessoas pensam sobre conhecimento — como as pessoas decidem “quem sabe o quê sobre o quê” — essa é a questão sobre autoridade cognitiva.

3.2.2 Autoridade, Influência, Credibilidade

Wilson (1983) levanta algumas questões particulares da autoridade cognitiva. A primeira delas é que a autoridade **envolve pelo menos duas pessoas**: ninguém pode ser uma autoridade apenas para si. Isso é diferente de ser um especialista que, mesmo sendo a última pessoa da terra, continuaria sendo um especialista em determinado assunto. A autoridade cognitiva é **unidirecional**: o fato de “A” ser uma autoridade para “B” não implica em “B” ser uma autoridade para “A”, mesmo isso sendo perfeitamente possível — logo, autoridade é diferente de um relacionamento como amizade, que é bidirecional.

As autoridades cognitivas de um indivíduo são pessoas que, na opinião desse indivíduo, entendem e sabem do que estão falando (WILSON, 1983). É uma questão de valor (**peso**, de acordo com Wilson). Se o que “A” fala sobre determinado assunto possui valor e relevância para “B”, então, “A” tem autoridade nesse assunto aos olhos de “B”. Autoridade também é uma questão de **grau** — pode-se possuir muito ou pouco dela. Em certa área “A” pode ser uma grande autoridade e suas palavras terem muito valor, enquanto que em outra, “A” pode falar sem autoridade alguma. Todavia, Wilson afirma que as pessoas, normalmente, expressam a autoridade cognitiva por meio de bases indiretas e que nem sempre é possível demonstrar essa autoridade em um valor ou escala mensurável. Além disso, a autoridade varia de acordo com quem a está recebendo e com a **esfera de interesse** (área ou assunto que está sendo julgado). Isso implica que duas pessoas podem receber autoridades em níveis diferentes

para um mesmo assunto e, também, que uma mesma pessoa pode receber níveis distintos de autoridade dependendo do assunto, área ou item analisado. Na prática, raramente existe o caso de autoridades absolutas, uma pessoa, normalmente, possui várias autoridades em assuntos distintos, ou em um mesmo assunto, porém com pesos diferentes.

A autoridade cognitiva é claramente relacionada com a **credibilidade**, pois ela representa a influência ocasionada no modo de pensar de uma pessoa, uma influência que essa pessoa conscientemente julga própria e adequada. Se “A” não possui credibilidade para com “B”, então, não existem razões para que “B” considere “A” uma autoridade cognitiva. “B” também pode considerar “C” uma pessoa digna de crédito, embora “C” não exerça influência sobre o modo de pensar e não seja uma autoridade para “B”. Essa relação entre a autoridade cognitiva e a credibilidade é facilmente percebida nas discussões de Wilson (1983). Segundo o autor, uma autoridade é aquela pessoa a quem se procura para obter informações, mas pode ser também aquela pessoa a quem se recorre para obter um conselho, uma opinião, mesmo que não exista conhecimento para ser empregado no assunto. As autoridades não são avaliadas apenas por seus “estoques de conhecimento”, ou o que Wilson chama de “habilidade para responder questões fechadas”. As autoridades também são avaliadas por suas opiniões, seus conselhos sobre atitudes adequadas, sobre exemplos, discussões, respostas propostas, etc. — a “habilidade para responder questões abertas”. Essa autoridade pode se estender por diversas áreas (moral, política, filosófica, técnica, científica) e, embora a autoridade cognitiva não possa dizer a uma pessoa “o que se deve pensar”, ela pode influenciar em qualquer crença ou atitude.

É importante diferenciar autoridade cognitiva de especialidade e de reputação. Ser um especialista não é o mesmo que ser uma autoridade cognitiva em determinado assunto. A autoridade é dependente do reconhecimento e do valor da **especialidade**. Wilson (1983) utiliza como exemplo a astrologia: uma pessoa que tenha se dedicado muito aos estudos da

astrologia pode ser considerada uma especialista na área, no entanto, para que essa pessoa seja considerada uma autoridade, sua especialidade precisa ser de interesse e de importância para outras pessoas — o que, no caso do autor, não acontece com o assunto astrologia. As autoridades normalmente são especialistas na área e possuem essa reputação, mas o inverso não é necessariamente verdade. Ser uma autoridade depende do reconhecimento por uma ou por um conjunto de pessoas que vai além dos limites da especialização.

A autoridade cognitiva também é diferente de **reputação**. Wilson (1983) menciona a reputação como uma base para justificar a autoridade concedida a alguém. Segundo Wilson, todas as pessoas possuem uma reputação, seja ela boa ou ruim, mas nem sempre uma pessoa possui uma boa reputação, e nem sempre uma boa reputação é suficiente para justificar a concessão da autoridade cognitiva. “A” pode possuir a reputação de ser considerada uma autoridade para o grupo “X”, mas “B” também conhece a reputação de “A” em manipular os resultados de suas pesquisas, o que é um argumento suficientemente forte para que “B” não considere “A” como sua autoridade cognitiva. Os dois conceitos estão fortemente relacionados, mas assim como ocorre entre autoridade e especialidade, autoridade e reputação não significam a mesma coisa — a autoridade cognitiva vai além dos limites e das justificativas da reputação.

Na ciência da computação existem vários trabalhos relacionados à reputação como, por exemplo, (CRUZ *et al.*, 2007) (GOLBECK & HENDLER, 2006) e (MATSUO *et al.*, 2004). No entanto, Russell (2005) fundamenta-se em Muller (2003) para afirmar que, normalmente, os algoritmos desenvolvidos na ciência da computação e na teoria das redes sociais que têm sido utilizados para destilar a reputação e a confiança em um valor calculável estão procurando por uma “opinião agregada” sobre todas as áreas ou tópicos. Russell (2006) argumenta que nesse processo existe muita informação relacionada a detalhes da reputação sendo perdida, a qual não será recuperada depois. Demonstrar a reputação por meio de um

valor auxilia na tomada de decisões, mas não permite nada além disso, pois não há como verificar aspectos relacionados ao contexto da reputação e da confiança identificada.

Além da reputação, Wilson (1983) também cita a experiência, o treinamento, a formação educacional, o desempenho, a opinião geral e a plausibilidade intrínseca como bases para justificar a concessão de autoridade. A reputação depende do grupo, do critério que está sendo avaliado e do que está sendo considerado. O papel da reputação é adequado para o reconhecimento de grandes especialistas, mas é preciso lembrar que um especialista pode não ser uma autoridade cognitiva. A **experiência**, a **formação educacional** e o **desempenho** também são bases válidas, pois estão diretamente ligadas às competências que podem justificar uma concessão de autoridade. Entretanto, Wilson afirma que não existem meios fáceis e infalíveis de medir essas competências. A **opinião geral** é ligada, de certa forma, à reputação — se muitas pessoas que “B” conhece consideram “A” uma autoridade, isso é uma razão considerável para que ele também pense a respeito; se as autoridades de “B” consideram “A” uma autoridade, “B” tem uma justificativa ainda mais forte para também considerar a autoridade de “A”. Segundo Wilson, nós usamos essas análises que envolvem reputação quando não temos bases próprias suficientes para avaliar e justificar a autoridade em questão.

Wilson (1983) afirma que existem dois tipos de bases para justificar a concessão de autoridade: diretas e indiretas. A base direta, e crucial, está relacionada ao fato de que se uma pessoa (ou fonte) constantemente, e repetidamente, fornece informações esclarecedoras para um indivíduo, esse indivíduo pode vir a esperar mais, a contar mais com essa pessoa\fonte, a falar de outros para ela e dela para outros — ela adquiriu autoridade sobre esse indivíduo. O oposto pode vir a acontecer caso a autoridade não continue com as mesmas contribuições significativas, ou mesmo se ela falhar em algum momento. Do outro lado, as bases relacionadas à experiência, reputação, treinamento e desempenho são consideradas como

bases indiretas para a concessão de autoridade. Entretanto, mesmo que todas essas bases estejam presentes para uma pessoa\fonte, um indivíduo pode não conceder autoridade porque o que essa fonte fala ou faz, não passa no teste intrínseco de plausibilidade. Ela simplesmente não pode ser levada a sério. Como em grande parte dos casos o indivíduo não possui conhecimentos suficientes para chegar a essa conclusão, então, o julgamento de plausibilidade estará sujeito aos conhecimentos que o indivíduo já possui ou nas coisas em que ele já crê.

Existe um tipo diferenciado de autoridade que é adquirido e fundamentado simplesmente na confiança ou na crença em uma pessoa\fonte. Esses casos particulares são fortemente relacionados com a autoridade carismática de Max Weber (1963) e se caracterizam por estarem fundamentados pela base mais forte: a da confiança — não se pode provar nada, nem a favor e nem contra. Wilson (1983) menciona que todas as bases são questionáveis, inconclusivas, apresentando-se como sinais ou índices da autoridade cognitiva, mas que, mesmo estas bases sendo insuficientes para estabelecer a autoridade cognitiva além de qualquer desafio, a maioria das pessoas, pelo menos uma vez na vida, recorreu ou recorrerá a estas bases para fazer algum tipo de julgamento.

3.2.3 Autoridade Cognitiva e Qualidade

Antes de iniciar as discussões sobre qualidade é preciso deixar claro que não são apenas indivíduos que recebem autoridade cognitiva (WILSON, 1983). É possível reconhecer a autoridade também em livros, instituições, instrumentos e organizações. Os dicionários geralmente são considerados autoridades absolutas com relação à pronúncia, ortografia e significado — ao consultar um dicionário, as informações encontradas, normalmente, são consideradas suficientemente seguras para sanar a dúvida. Os textos religiosos são exemplos incontestáveis e infalíveis relacionados à moral e à crença dos seus seguidores. No caso dos instrumentos, um termômetro, barômetro ou relógio, fornecem informações que na maioria dos casos são confiáveis e não são contestadas. Também existe a autoridade intrínseca do

cargo, ocupação ou profissão (exemplo: papa, médicos, professores) que por si só já são basicamente competentes e, mesmo que um indivíduo não a mereça, o grupo todo representado pelo cargo ou profissão merece e propaga a autoridade. Pela autoridade não ser restrita a pessoas, a partir desse momento nos referenciaremos a indivíduos, objetos, organizações e instituições simplesmente como “entidades”.

Pode-se afirmar que a propagação da autoridade é natural e praticamente inerente: um livro produzido (escrito ou publicado) por uma grande autoridade herda o seu respeito e o seu reconhecimento, e o inverso também pode ser verdadeiro — um autor pode ser reconhecido e respeitado por causa de um trabalho desenvolvido e considerado de grande importância. Um indivíduo também pode ser considerado autoridade por ocupar determinada posição em uma instituição, da mesma forma que essa instituição pode se tornar respeitada devido às autoridades que trabalham nela. A propagação da autoridade é comum entre indivíduos: se “A” considera “B” uma autoridade e “B” considera “C”, a autoridade concedida por “B” à “C” pode também ser respeitada e considerada por “A” devido a influência exercida por “B”.

Wilson (1983) afirma que a autoridade possui ligações fortes com a qualidade. O autor justifica que a qualidade dos trabalhos produzidos, das opiniões emitidas, das recomendações realizadas, está inserida nas bases que justificam a concessão de autoridade a uma entidade. Wilson argumenta que as autoridades tendem a possuir, produzir ou indicar conteúdos de qualidade e, também, a manter contato com outras entidades que se interessam ou que também entendem do assunto. Nessa mesma linha, Wu (2006) menciona que indivíduos especialistas em determinada área — e que, portanto, são ou podem vir a ser autoridades — não apenas usam documentos de qualidade como também desenvolvem esquemas mais organizados de classificação e de relacionamento entre conceitos. As discussões teóricas de Wilson (1983) expõem que a autoridade cognitiva possui impactos na qualidade, mas é nos

trabalhos de Rieh & Belkin (1998; 2000) que esses impactos são estudados com mais profundidade.

Os trabalhos desenvolvidos por Rieh & Belkin (1998; 2000) aplicaram a teoria da autoridade cognitiva de Wilson (1983) para examinar o conceito de autoridade e de qualidade na *Web* da perspectiva dos indivíduos buscadores de informação. Os autores mencionam que “*people understand authority on the basis of personal experience with particular sources or reputations.*”¹⁶ (RIEH & BELKIN, 1998) concluindo que a avaliação da autoridade cognitiva é subjetiva, relativa e situacional em vez de objetiva, absoluta e universalmente reconhecível — essas conclusões vêm de encontro com as afirmações de que o julgamento da autoridade cognitiva e suas bases não são compreensíveis a um nível de expressões verbais, mas a um nível cognitivo bem mais profundo (MCKENZIE, 2003). No que diz respeito à qualidade, no livro “*Non-Functional Requirements in Software Engineering*” (CHUNG *et al.*, 1999) a qualidade é mencionada como o requisito não-funcional mais complexo. Chung *et al.* (1999), assim como Sommerville (2003), determinam que os requisitos não-funcionais possuem três características-chaves: subjetividade, relatividade e interdependência. Segundo essas características, pode-se considerar que a autoridade cognitiva se adéqua às definições de Chung sobre requisitos não-funcionais e que, da mesma forma, o requisito não-funcional qualidade se adéqua às conclusões de Rieh & Belkin sobre a autoridade cognitiva. Assim, é possível estabelecer que a autoridade cognitiva e a qualidade possuem bases comuns por serem, ambas: subjetivas, relativas, situacionais e interdependentes (ver Figura 7).

¹⁶ Tradução do autor: As pessoas compreendem autoridade com base em suas experiências pessoais com certas fontes ou reputações.

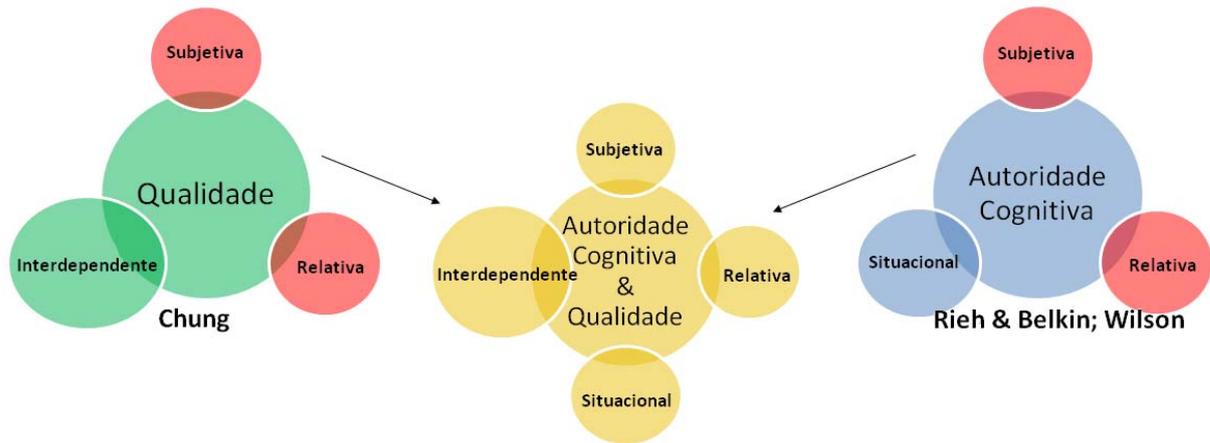


Figura 7 - Características da qualidade e da autoridade cognitiva.

A relação que se pretende deixar clara nessas discussões é a relação existente entre a autoridade cognitiva e a qualidade das informações no contexto de entidades produtoras e organizadoras de informação. Todavia, os argumentos demonstram-se válidos para o julgamento da autoridade e da qualidade separadamente e independentemente do domínio ou aplicação. Deste modo, dizer que tanto a autoridade quanto a qualidade possuem a característica de serem **subjetivas**, é dizer que elas possuem características fortemente pessoais e que variam de acordo com quem faz o julgamento: o que é qualidade para “A” pode não ser para “B”, da mesma forma que uma autoridade de “B” pode não exercer nenhuma influência sobre “A” — (ALDE *et al.*, 2005) e (LOPES, 2005) apresentam discussões relacionadas ao assunto. Também é possível reconhecer as características de serem relativas e situacionais: **relativas** porque uma informação ou uma autoridade pode ser adequada para um objetivo específico (estudos iniciais ou compreensão geral sobre um assunto), mas completamente inadequadas para outro (referência em um artigo científico); **situacionais** porque tanto a autoridade quanto a qualidade sofrem alterações ao decorrer do tempo, da evolução do conhecimento das pessoas, de suas necessidades, experiências, etc. Por fim, dizer que a autoridade cognitiva e a qualidade são **interdependentes** refere-se ao fato de ambas sofrerem influências, tanto das bases utilizadas para julgá-las (experiência, reputação,

plausibilidade intrínseca) quanto de outros requisitos não-funcionais (tempo, segurança, custo).

Chung *et al.* (1999) refletem sobre qualidade e dizem que as pessoas não precisam que alguém lhes diga o que é ou não é bom. As pessoas possuem seus próprios fundamentos, diretos ou indiretos para concluir isso e nem sempre esses fundamentos são explicáveis — esses são os argumentos de Wilson (1983) sobre a autoridade cognitiva. Russell (2005) fundamentado principalmente nas explanações de Rieh (2005), de Rieh & Belkin (1998; 2000) e na teoria da autoridade cognitiva de Wilson (1983), afirma que ninguém pode dizer quem é autoridade para quem, da mesma forma que ninguém pode definir o que é qualidade para outra pessoa. Conseqüentemente, cabe apenas às pessoas fazerem esse julgamento, de forma livre e do modo que acharem mais adequado.

3.3 Recuperação de Informação, Autoridade Cognitiva e Folksonomia

Wilson considera a autoridade cognitiva como um dos componentes do controle de qualidade na recuperação de informação. Segundo ele “*a text usually has one chance to capture our attention; reading a few words of it may be enough to discourage us from continuing on to reading the whole thing*”¹⁷ (WILSON, 1983. p.169).

Os resultados obtidos por Rieh & Belkin (1998; 2000) para validar a teoria de Wilson demonstraram que, ao procurar por informações na *Web*, os pesquisadores realizam seus julgamentos de autoridade e de qualidade primeiramente fundamentados pelos seus conhecimentos do domínio e do sistema. Como complemento, utiliza-se características das fontes tais como: domínio da URL, reputação, credenciais do autor\criador e, também, características do objeto de informação: tipo, título, conteúdo, organização, apresentação, funcionalidades. Rieh & Belkin (2000) desenvolveram um estudo com usuários reais e concluíram que freqüentemente esses usuários selecionavam páginas da Internet nas quais

¹⁷ Tradução do autor: “Um texto normalmente tem uma chance de capturar nossa atenção; lendo umas poucas palavras dele pode ser suficiente para nos desencorajar a continuar a leitura”.

existia alguma indicação sobre a autoridade da fonte, seja por experiência própria, por recomendação de outros ou por algo que eles já haviam ouvido. Os autores perceberam que os usuários, normalmente, indicavam suas autoridades cognitivas como sendo amigos, colegas, acadêmicos e até mesmo como jornais ou artigos científicos.

Rieh & Belkin, por meio de seus trabalhos (1998; 2000) (RIEH, 2005), confirmaram as afirmações de Wilson sobre a importância da autoridade cognitiva para a qualidade na recuperação de informação. Rieh (2005) cita Fogg (2003) e afirma que assuntos de autoridade estão recebendo maior atenção não apenas da ciência da informação, mas também de outros campos que incluem a interação humano-computador, a educação e a ciência da computação. No entanto, Rieh (2005) enfatiza que as pesquisas sobre qualidade da informação e sobre problemas de credibilidade na *Web* que são desenvolvidas fora do campo da ciência da informação dificilmente mencionam a teoria da autoridade cognitiva de Wilson ou qualquer outra teoria ou modelo de comportamento da informação. Para a autora, é necessário que as teorias sobre o comportamento da informação (*information behavior theories*) sejam discutidas e aplicadas além dos limites do campo da ciência da informação.

Rieh & Belkin (1998; 2000) (RIEH, 2005) também enfatizam em seus trabalhos, que a utilização da teoria da autoridade cognitiva de Wilson (1983) pode trazer contribuições em nível de relevância e de qualidade na recuperação de informação. No capítulo anterior foram discutidos os problemas causados pela sobrecarga de informação nos sistemas baseados em folksonomia. Considerando as discussões realizadas neste capítulo em torno da autoridade cognitiva, pode-se supor que se a autoridade das entidades que categorizam informações nos sistemas baseados em folksonomia for levada em conta, poderão ser obtidos resultados mais relevantes e de maior qualidade na recuperação de informação nesses sistemas. Os usuários poderão identificar as informações categorizadas por autoridades no assunto que, pelo menos, fornecem uma base maior para o julgamento de sua qualidade e de sua relevância.

Russell (2005) apresenta uma proposta para a concessão de autoridade cognitiva por meio de folksonomia. Esta proposta mostra-se adequada por quatro razões: 1) a folksonomia é caracterizada pela liberdade de expressão do usuário, e é disso que o usuário precisa. Ninguém pode dizer quem é uma autoridade cognitiva para outra pessoa, a própria pessoa deve concluir isso; 2) a autoridade possui características peculiares (esfera de interesse, valores, etc.) que as *tags* são capazes de representar adequadamente; 3) nos sistemas baseados em folksonomia as categorizações de informação são realizadas pelos próprios usuários. Deste modo, além de serem mais adequados para avaliar os categorizadores com os quais mantém contato (ou conhecem), os usuários já estão familiarizados com a utilização das *tags* e com o funcionamento da técnica de folksonomia; e 4) a folksonomia permite que se identifique mais do contexto em que a autoridade foi concedida do que qualquer outro esquema de avaliação poderia proporcionar. As autoridades não serão identificadas com base num único peso ou valor estabelecido por um algoritmo com critérios pré-definidos. Elas serão identificadas com base na opinião das pessoas, na imprevisibilidade e na diversidade dessas opiniões, sendo a forma mais próxima de representar e de tornar explícita a autoridade já existente e já concedida no mundo real.

No próximo capítulo realiza-se uma discussão em torno da proposta de se utilizar a concessão de autoridade cognitiva por meio de folksonomia para a redução dos problemas de sobrecarga de informação e para o aumento da qualidade e da relevância das informações recuperadas em sistemas baseados em folksonomia. Serão realizadas discussões sobre a viabilidade e sobre as implicações dessa concessão de autoridade que caracterizará uma meta-categorização — uma folksonomia em nível duplo, na qual no primeiro nível estão as entidades categorizando informações, enquanto que no segundo estarão entidades categorizando outras entidades.

*Capítulo IV**Folkauthority*

Uma das principais características da folksonomia é a participação essencial, e necessária, dos usuários para a categorização de informação. Essa característica representa diretamente a opinião, o entendimento e mesmo os objetivos dos usuários de sistemas baseados em folksonomia, sendo diretamente responsável tanto pelo êxito, quanto por grande parte dos problemas resultantes nesses sistemas. Este capítulo aborda a concessão de autoridade cognitiva em sistemas baseados em folksonomia — uma concessão realizada pelos usuários sobre os próprios usuários categorizadores de informação — com o intuito de recuperar informações mais relevantes e de maior qualidade, e de reduzir o impacto da sobrecarga de informação nesses sistemas. Na seqüência, discute-se sobre aspectos teóricos e práticos da concessão de autoridade cognitiva, e apresenta-se a arquitetura de um sistema para suportar a concessão de autoridade cognitiva por meio de folksonomia e sobre sistemas que sejam baseados em folksonomia.

4.1 Autoridade Cognitiva e Folksonomia

Nos capítulos anteriores foram expostos os problemas relacionados tanto com a sobrecarga quanto com a qualidade das informações disponíveis na Internet. A folksonomia foi apresentada como uma técnica emergente aplicada para a atribuição de significado pelos usuários aos conteúdos disponíveis na *web* e a teoria da autoridade cognitiva foi apresentada como um *framework* a ser considerado na redução dos problemas mencionados. A hipótese levantada neste trabalho é a de que a aplicação do conceito de autoridade cognitiva em sistemas baseados em folksonomia eleva a relevância, a qualidade e a confiabilidade das informações recuperadas e, também, reduz o impacto da sobrecarga de informação nesses sistemas.

Folkauthority (do inglês, *folk* = pessoas + *authority* = autoridade) é um neologismo que usaremos para designar a autoridade concedida às fontes de informação por meio de folksonomia (PEREIRA & DA SILVA, 2007). Aplicar o conceito de *folkauthority* corresponde a permitir que os próprios usuários de um sistema concedam autoridade cognitiva às fontes de informação com as quais o sistema trabalha. Essa concessão de autoridade é realizada por meio da atribuição de *tags* que representem as competências, habilidades ou áreas nas quais as fontes de informação são consideradas referências pelos seus categorizadores (concessores de autoridade). O fato de os sistemas baseados em folksonomia possuírem os usuários como categorizadores de informação torna-os um ambiente propício para a aplicação do conceito de *folkauthority*, pois a concessão de autoridade passa a ocorrer sobre os próprios usuários propagando a autoridade concedida para as categorizações de informação por eles realizadas.

4.1.1 A Origem da Proposta

Nas discussões conceituais sobre a técnica de folksonomia, autores como Russell (2006) e Mathes (2004) sugerem que a aplicação dessa técnica seria mais adequada em

sistemas que reconhecessem, ou tornassem explícitas, as habilidades de seus usuários. Sturtz (2004) indica que os usuários categorizadores de informação são diretamente responsáveis pela qualidade das informações categorizadas e pelo sucesso dos sistemas baseados em folksonomia de uma forma geral. Wu (2006) reforça essa visão e vai além ao mencionar que “*Folksonomy may hold the key to developing a Semantic Web*”¹⁸. Ao contrário do trabalho de Ohmukai (2005) que propõe uma folksonomia controlada, esses trabalhos fornecem bases para que se pense em alternativas que não comprometam a liberdade dos usuários, mas que busquem melhores resultados por meio do desenvolvimento de recursos que apoiem os usuários na utilização de sistemas baseados em folksonomia.

Rieh (2005), separadamente das discussões sobre folksonomia, afirma que a aplicação do conceito de autoridade cognitiva pode auxiliar nos processos de organização e de recuperação de informação. Rieh & Belkin (1998; 2000) utilizam esta teoria para demonstrar como as pessoas julgam a qualidade das informações na *Web* e, então, realizam um estudo que conclui que a autoridade cognitiva é um critério predominante nesse julgamento. Russell (2005), fundamentando-se nas reflexões de Rieh & Belkin sobre autoridade cognitiva, propõe a concessão desta em uma comunidade de usuários por meio de folksonomia com o intuito de reconhecer autoridades e de identificar competências.

Como os problemas motivadores desta pesquisa são a ausência de controle de qualidade nas categorizações e a sobrecarga de informação nos sistemas baseados em folksonomia, os trabalhos de Russell (2005), de Rieh & Belkin (1998; 2000) e de Rieh (2005), todos citando Wilson (1983), formam a base da hipótese de que aplicar o conceito de autoridade cognitiva eleva a qualidade e a relevância das informações recuperadas e ameniza o impacto da sobrecarga de informação. Do trabalho de Russell (2005) segue-se a idéia de utilizar a folksonomia para a concessão de autoridade cognitiva, mas direcionando sua

¹⁸ Tradução do autor: “A folksonomia pode possuir a chave para o desenvolvimento de uma *web*-semântica”

aplicação em sistemas que sejam baseados em folksonomia. Dos estudos realizados por Rieh e por Belkin (1998; 2000) (RIEH, 2005), considera-se as relações entre autoridade, qualidade e relevância, propondo-se que a concessão de autoridade cognitiva seja utilizada para melhorar os resultados do processo de recuperação de informação em sistemas baseados em folksonomia, além de identificar autoridades e reconhecer competências, como proposto por Russell (2005).

4.1.2 A Folksonomia na Concessão de Autoridade Cognitiva

Utilizar a técnica de folksonomia para a concessão de autoridade (*folkauthority*) em sistemas baseados em folksonomia implica no que pode ser chamado de “meta-categorização”. Como discutido anteriormente, no processo de categorização de objetos o usuário-fonte atribui *tags* que descrevem o conteúdo ou o significado de um objeto. Já a concessão de autoridade, proposta em nosso trabalho, representa um processo no qual a categorização é realizada sobre os próprios usuários-fontes de um sistema, descrevendo as competências, as habilidades, os conhecimentos ou as áreas nas quais esses usuários são considerados como referência pelos categorizadores (RUSSELL, 2005). Deste modo, tem-se uma categorização sobre os objetos (informação) num primeiro nível e sobre os usuários-fontes que realizam a categorização de objetos num segundo nível — uma categorização dos categorizadores pelos categorizadores. A Figura 8 ilustra essa meta-categorização.

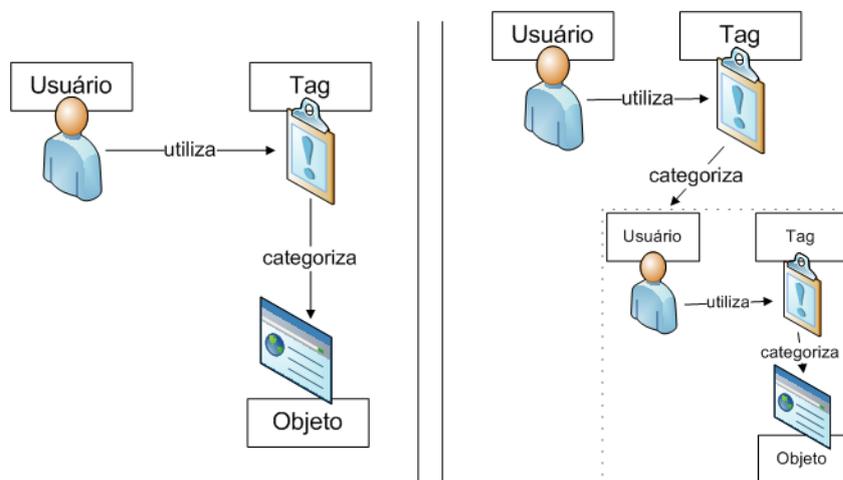


Figura 8 - Meta-categorização: folksonomia em nível duplo.

No capítulo III argumentamos que uma das relações entre a qualidade e a autoridade é a natureza subjetiva de ambas, sendo esta subjetividade justamente o que torna a folksonomia adequada para a concessão de autoridade cognitiva. O conessor de autoridade é livre para categorizar suas fontes de informação da forma que julgar mais adequada, sem restrições ou limitações, representando, assim, a sua opinião sobre aquela fonte.

Ao argumentar sobre a concessão de autoridade, Russell (2005) deixa implícita a viabilidade de utilizar folksonomia para concedê-la e cita a identificação das competências e das habilidades dos usuários-fontes como o principal benefício a ser obtido. Em nosso trabalho, o fato do foco ser direcionado para sistemas baseados em folksonomia permite a adição de outras duas vantagens às já expostas por Russell: 1) a recuperação de informações categorizadas por autoridades, e 2) a exploração da cadeia de autoridades formada pela atribuição de autoridades.

Conforme mencionado no capítulo II, a ausência de mecanismos de controle da qualidade das informações produzidas e publicadas na Internet é um ponto crítico para o processo de recuperação de informação, uma vez que não existe um meio direto de identificar a veracidade das informações e nem a competência de seus produtores (ALDE *et al.*, 2005). Ao reconhecer as competências dos usuários-fontes nos sistemas baseados em folksonomia, as informações categorizadas por esses usuários passam a sofrer influência da autoridade possuída pelos mesmos. Considerando os estudos de Rieh & Belkin (1998; 2000) (RIEH, 2005), pode-se inferir que na recuperação de informação a identificação das informações categorizadas por autoridades proporciona resultados mais relevantes, mais seguros e de melhor qualidade para aqueles que concederam ou que concordam com a autoridade da fonte categorizadora. Isso porque os usuários possuem habilidades e conhecimentos distintos, (RUSSELL, 2005) (RIEH, 2005) o que os torna mais, ou menos, adequados como fontes de informação de acordo com o assunto, área ou tópico em análise. Segundo as discussões de

Wilson (1983) sobre especialistas e generalistas, pode-se considerar que um usuário com um maior conhecimento em determinado assunto tende não só a organizar seus conteúdos de uma melhor forma, como também a possuir melhores informações e a manter contatos com outras pessoas que conhecem tal assunto. Utilizando esses argumentos, é válido considerar que quem concede autoridade a uma fonte também reconhece a qualidade das categorizações por ela realizada.

4.2 CAW – *Cognitive Authority on the Web*

Em nosso trabalho, propomos o desenvolvimento de um sistema baseado em folksonomia para oferecer apoio à concessão de autoridade cognitiva, o qual denominamos de CAW (acrônimo de *Cognitive Authority on the Web*). Existem dois pontos, ambos citados por Russell (2005), que diferenciam um sistema para a concessão de autoridade cognitiva de um sistema baseado em folksonomia convencional. O primeiro é que, em vez dos três pivôs da folksonomia (usuário, objeto, *tags*) têm-se apenas dois: o usuário e as *tags* — o usuário é o próprio objeto. O segundo é que o objeto é categorizado em relação ao que ele “sabe”, no que ele é “bom” ou no que ele é considerado “referência”, diferentemente da categorização convencional que rotula o objeto com o intuito de descrever o seu conteúdo ou significado.

Além dos pontos acima citados, identificamos duas outras diferenças relacionadas à quantidade de objetos categorizados e à natureza e consequência das categorizações.

No que diz respeito à quantidade de objetos categorizados, haverá uma menor carga cognitiva para o usuário no momento de categorizar seus usuários fontes. Isto ocorre porque, por maior que seja a rede social de um indivíduo, e a quantidade de entidades que possam figurar como autoridades nessa rede, esse número será inferior ao total de recursos *on-line* que o indivíduo pode categorizar, pois a quantidade de recursos *on-line* (páginas, documento, fotos, vídeos) há muito já ultrapassou a quantidade de pessoas existentes no mundo (LEVY, 2006) (HIMMA, 2007).

Com relação à natureza e às conseqüências das categorizações, diferentemente do que ocorre na categorização convencional, os usuários terão um maior conhecimento sobre o objeto que está sendo categorizado e, conseqüentemente, já terão uma melhor noção dos termos que podem ser utilizados para categorizá-lo. Os próprios termos utilizados para a concessão de autoridade poderão ser diferentes dos termos aplicados para a categorização de objetos, pois a concessão de autoridade deverá ser efetuada com termos que o usuário que a recebe utiliza para categorizar objetos, e esses termos nem sempre estarão presentes no vocabulário do usuário que está concedendo a autoridade cognitiva. Além disso, os aspectos sociais existentes na folksonomia são maximizados e tornam-se críticos na concessão de autoridade cognitiva, pois de acordo com Russell (2005), questões relacionadas à privacidade, ao relacionamento entre os indivíduos, e à interação humano-computador devem ser levadas em conta para que um sistema que apóie esse processo de concessão de autoridade obtenha êxito.

Quanto a questões técnicas, a principal é a integração do sistema proposto com outros sistemas baseados em folksonomia nos quais os usuários já possuem informações. Deste modo, a recuperação das informações poderá ser efetuada sobre as categorizações que as autoridades já realizaram em outros sistemas. Os sistemas *Delicious* e *Flickr*, por exemplo, já disponibilizam APIs¹⁹ que possibilitam uma comunicação que permite recuperar e atualizar seus dados. Assim, o sistema *CAW* também deve estar apto tanto a trabalhar e a se comunicar com esses sistemas quanto a permitir que outros sistemas se comuniquem com ele.

O mapa conceitual da Figura 9 demonstra as idéias que fundamentam o desenvolvimento do sistema *CAW*. Neste mapa, o conceito **usuário** representa uma “pessoa, entidade ou organização que pode acessar a ferramenta e utilizar seus recursos” — esse conceito é referenciado simplesmente como **entidade**. Uma entidade utiliza *tags* tanto para

¹⁹ *Application Programming Interface*

recuperar objetos previamente categorizados e armazenados no(s) sistema(s) sobre o(s) qual(is) o sistema CAW opera, quanto para categorizar outras entidades atribuindo autoridade às mesmas. A relação entre as *tags* aplicadas às entidades com as *tags* atribuídas aos objetos possibilita que em um processo de recuperação de informação, sejam priorizados os objetos que foram categorizados por autoridades naquelas *tags*. Os conceitos em laranja referem-se aos pivôs de um sistema baseado em folksonomia, enquanto que os conceitos em verde, azul e lilás representam diferentes especializações desses pivôs.

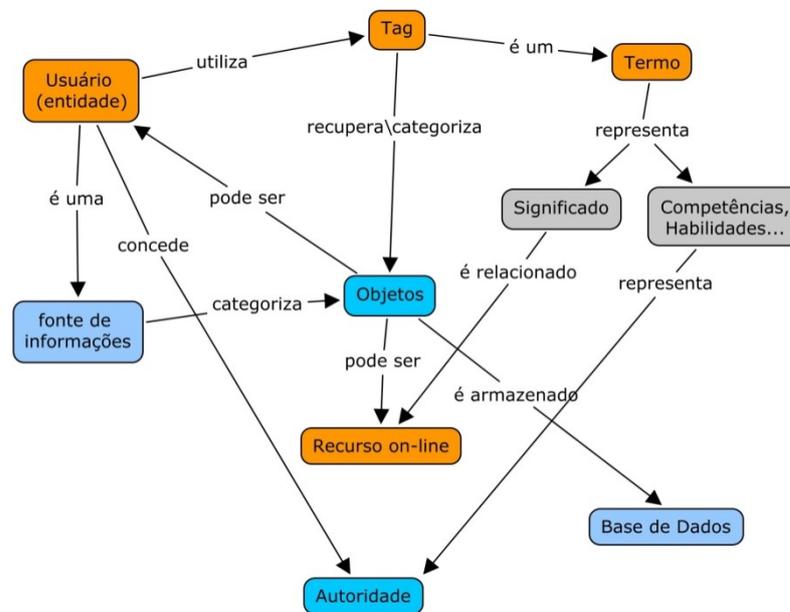


Figura 9 - CAW: mapa conceitual.

A Figura 10 ilustra a arquitetura proposta para o sistema CAW, a qual é composta por três módulos principais (Interface do Sistema; Processamento; e Comunicação) e por um Banco de Dados.

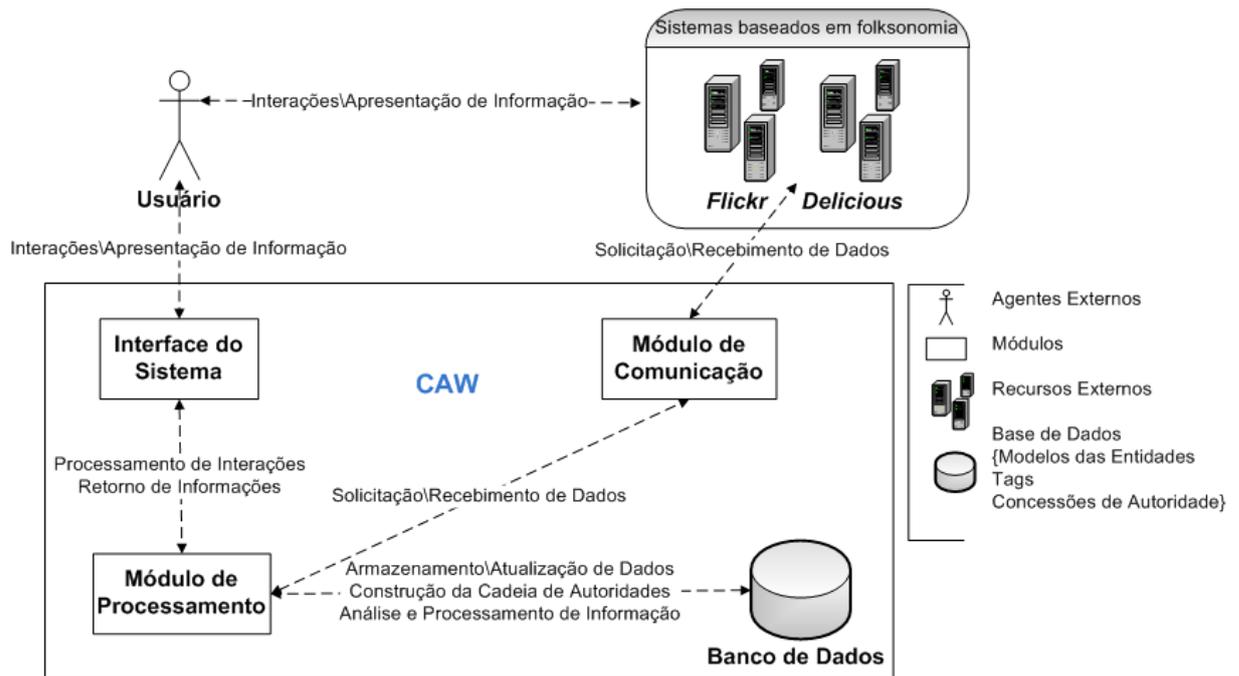


Figura 10 - Arquitetura do sistema CAW.

O **módulo de interface** do sistema foi especificado por meio de protótipos devendo fornecer suporte às ações dos usuários e apresentar os resultados dos processamentos realizados. É este módulo que deve oferecer suporte as interações do usuário e possibilitar a execução de operações, tais como: concessão de autoridade, busca por autoridades, reconhecimento de competências e, também, a recuperação de informação considerando as categorizações realizadas por autoridades em outros sistemas.

O **módulo de processamento** é o responsável por processar as requisições realizadas pelos usuários, por armazenar e atualizar as informações no banco de dados, por solicitar informações de sistemas externos e por realizar qualquer outro processamento para o armazenamento ou para a apresentação de informações. No módulo de processamento foram desenvolvidos os algoritmos para inserção, atualização e pesquisa de informações no banco de dados do sistema CAW. Implementamos os algoritmos para construção da cadeia de autoridades, para o reconhecimento de competências das entidades, para a identificação de autoridades, e criamos a estrutura para recuperação de informação e reordenação de resultados

recuperados. Os principais algoritmos serão apresentados na medida em que seus recursos forem discutidos neste capítulo.

O **módulo de comunicação** deve receber do módulo de processamento as instruções sobre as informações que precisam ser solicitadas aos sistemas externos. Esse módulo solicita as informações de acordo com as instruções recebidas e retorna os resultados ao módulo de processamento. Em nosso trabalho, não desenvolvemos o módulo de comunicação com outros sistemas baseados em folksonomia. Por essa razão, o **banco de dados** foi ampliado para apoiar a categorização de qualquer objeto que possa ser referenciado (*URLs*, vídeos, fotos, documentos), além das informações necessárias para o funcionamento do sistema (são elas: o registro de usuários, a concessões de autoridade e os termos utilizados — ver Figura 11). Assim, a arquitetura implementada para o sistema *CAW* oferece apoio para a categorização de objetos, porém, a arquitetura proposta prevê também a comunicação e o aproveitamento de categorizações já realizadas pelas entidades em outros sistemas baseados em folksonomia.

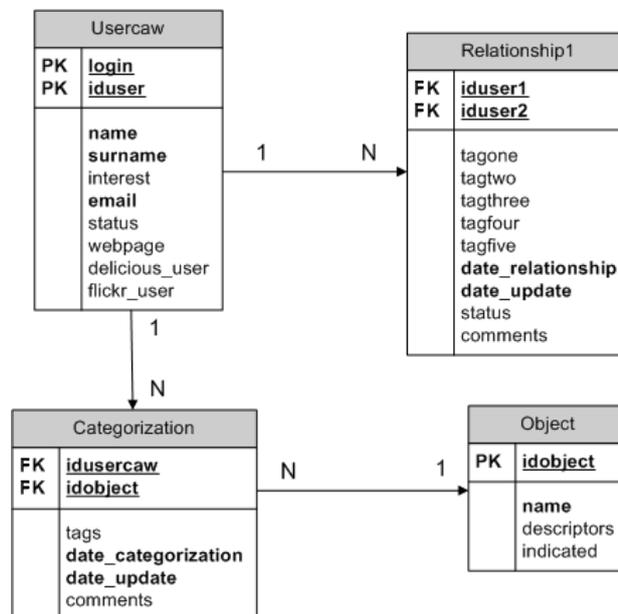


Figura 11 - Modelo conceitual simplificado do banco de dados.

A Figura 11 apresenta o modelo conceitual simplificado do banco de dados com as principais tabelas e seus relacionamentos. A tabela **UserCAW** é a responsável pelo

armazenamento de informações relacionadas aos usuários, ou entidades como também os chamamos. Na tabela **Relationship** são armazenados os relacionamentos de concessão de autoridade entre os usuários (entidades). A tabela **Object** é responsável por armazenar os objetos que serão categorizados pelas entidades, enquanto que a tabela **Categorization** armazena a categorização de objetos propriamente dita. Essas quatro tabelas suportam a concessão de autoridade, a categorização de informação e, conseqüentemente, a recuperação de informação priorizando as categorizações realizadas por autoridades. Foram desenvolvidas tabelas adicionais para armazenar informações estatísticas, tais como: *tags* mais utilizadas, distribuição das *tags* e dos objetos por entidade, ligação entre as entidades na cadeia de autoridades, etc. No próximo capítulo serão apresentados maiores detalhes sobre as informações geradas e armazenadas no sistema CAW.

4.2.1 O Processo de Concessão de Autoridade Cognitiva

Para melhor exemplificar o processo de concessão de autoridade cognitiva por meio de folksonomia, optamos por apresentar uma descrição desse processo, considerando a concessão de autoridade cognitiva sobre os usuários do sistema de *bookmarking* social *Delicious*.

No sistema *Delicious*, os usuários categorizam seus objetos (que nesse caso são *bookmarks* — *URLs*) com as *tags* que eles escolheram para rotulá-los, descrevê-los e organizá-los. Essas categorizações, normalmente, ficam publicamente disponíveis e podem ser visualizadas por qualquer pessoa que acesse a página do usuário categorizador. O sistema permite, também, que sejam efetuadas pesquisas por uma *tag* específica, recuperando os *bookmarks* categorizados por qualquer usuário do sistema com esta *tag*. Por exemplo, uma recuperação de informação pela *tag* “java” apresentará todos os *bookmarks* categorizados com essa *tag*, independentemente de quem foram os categorizadores.

Um ponto interessante a ser observado no processo de recuperação de informação do *Delicious* é que os resultados obtidos são apresentados em ordem cronológica inversa de categorização, ou seja, os mais recentes primeiro e os mais antigos por último. Entretanto, essa ordenação condena muitos objetos a aparecerem nas últimas posições e a não serem encontrados pelos usuários que, em sua maioria, não consideram os resultados obtidos além da primeira página retornada (IPROSPECT, 2007). Além disso, a heterogeneidade do conjunto de usuários e a ausência de controle nas categorizações são fatores que colaboram para a existência de problemas relacionados à polissemia dos termos, podendo prejudicar ainda mais o posicionamento dos objetos que seriam de interesse de um usuário ao recuperar informação. Uma alternativa à forma cronológica de apresentação é a ordenação com base nas categorizações efetuadas pelas autoridades. Esta ordenação consiste em priorizar os *bookmarks* categorizados por autoridades no assunto, de modo que eles sejam apresentados antes dos *bookmarks* que foram categorizados por “usuários-fontes comuns”.

As autoridades em determinada *tag* (a qual representa um assunto, uma área do conhecimento ou uma competência) compõem um subconjunto dentro do conjunto total de entidades-fontes disponíveis, conforme ilustrado pela Figura 12. Considerando as relações existentes entre autoridade cognitiva, qualidade e relevância das informações, mencionadas por Rieh e Belkin (1998; 2000), podemos afirmar que priorizar os objetos categorizados por esse subconjunto de usuários equivale a priorizar os objetos que foram categorizados por quem “entende do assunto”. Deste modo, mesmo que a categorização de um objeto seja “antiga”, esse objeto poderá ser apresentado em posições mais favoráveis caso essa categorização tenha sido efetuada por uma autoridade.

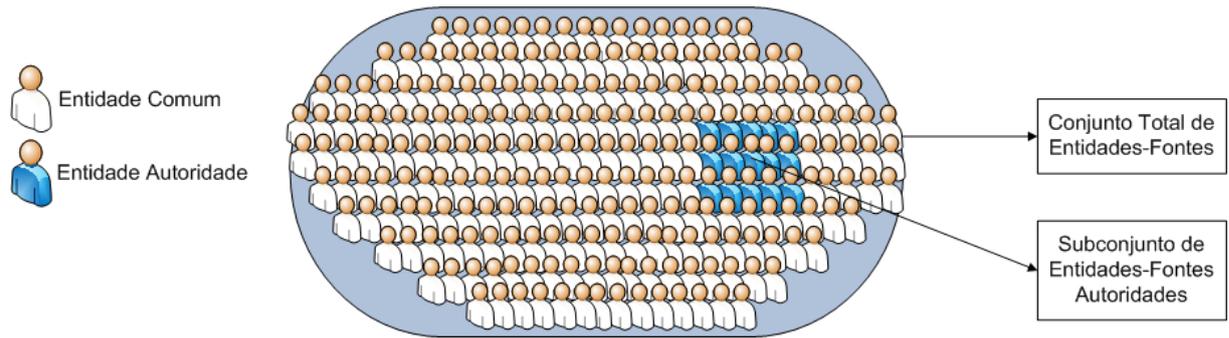


Figura 12 - Conjunto de entidades-fontes que categorizam informações.

Para que o subconjunto ilustrado na Figura 12 seja formado, é necessário que ocorra a concessão de autoridade cognitiva entre as entidades. Essa concessão acontece quando uma entidade atribui *tags* à outra entidade, tornando explícita a autoridade merecida pela segunda. Deste modo, à medida que as concessões vão ocorrendo torna-se possível identificar as autoridades em um determinado tópico, as competências das entidades categorizadoras de informação, e os objetos que são categorizados por autoridades.

A Figura 13 ilustra o processo em que a entidade “A” concede autoridade com as *tags* “usabilidade”, “literatura” e “design”, para a entidade “B”, ou seja, segundo a opinião da entidade “A”, dentre as entidades-fontes existentes (e entre as quais “A” está ciente da existência) “B” é qualificada, domina, ou é referência nesses tópicos e, portanto, as categorizações que “B” realizar com essas *tags* interessam a “A” — podem existir várias entidades consideradas autoridades em um mesmo tópico e em níveis distintos, isso depende exclusivamente da entidade que concede a autoridade.

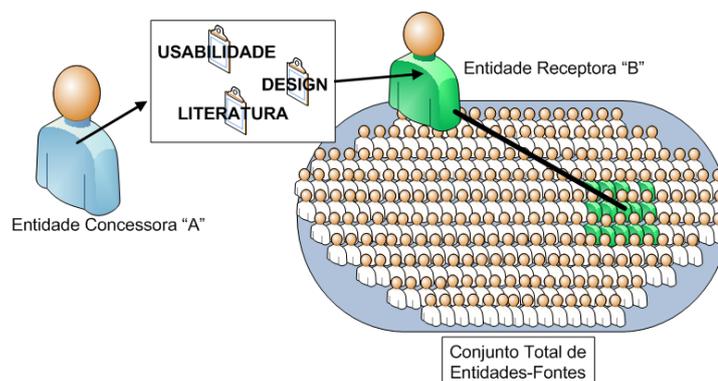


Figura 13 - Entidade “A” concedendo autoridade cognitiva para entidade “B”.

Na prática, o processo de concessão de autoridade cognitiva pode ser executado de duas formas distintas: i) por entidade: atribuindo uma ou mais *tags* para uma entidade, ou ii) por *tag*: relacionando uma ou mais entidades a uma determinada *tag*. O fluxograma da Figura 14 demonstra as diferenças existentes entre os processos de concessão de autoridade cognitiva focado na entidade e o focado na *tag*.

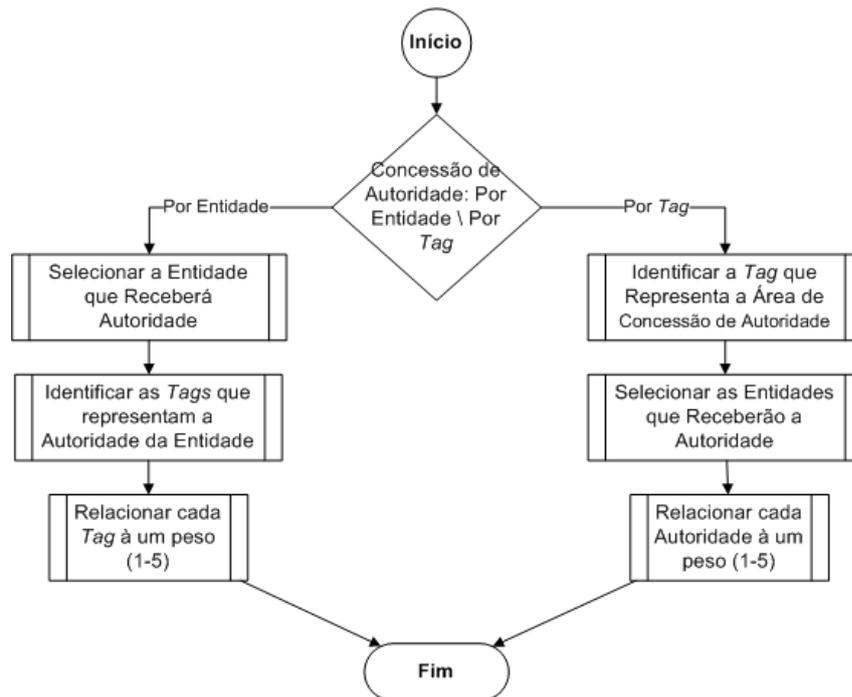


Figura 14 - Fluxograma do processo de concessão de autoridade cognitiva.

O processo focado na entidade ocorre, principalmente, quando o objetivo do usuário é reconhecer as competências de uma entidade específica — o usuário já tem conhecimento da entidade e, então, deseja tornar explícita a autoridade merecida por essa entidade. Já o processo focado na *tag* ocorre nos casos em que o usuário pretende encontrar entidades para serem categorizadas como autoridades em uma determinada *tag*. Por exemplo, considere que a entidade “A” queira encontrar autoridades em “cardiologia” para adicioná-las à sua rede; este é um processo diferente daquele em que a entidade “A” deseja reconhecer a autoridade da entidade “B” nos assuntos que ela a considera referência.

4.2.1.1 A Popularidade e o Peso de uma Autoridade

O processo de concessão de autoridade cognitiva possibilita a identificação das entidades que são consideradas autoridades em determinado assunto. Além dessa identificação, é possível identificar duas informações relacionadas à autoridade recebida por uma entidade: a popularidade e o peso. A importância, a utilidade e a forma de obter essas informações serão descritas a seguir.

A noção de **popularidade** é amplamente utilizada em softwares sociais: i) no *YouTube* (GOOGLE, 2005) ela é determinada pelo número de vezes que um vídeo foi assistido ou adotado como favorito; ii) no *Digg* (DIGG, 2004) pela quantidade de votos que uma notícia recebeu; iii) no *Orkut* (GOOGLE, 2004) a noção de popularidade é relacionada ao número de amigos ou de fãs que um usuário possui; iv) e nos sistemas *Flickr* (YAHOO, 2004) e *Delicious* (YAHOO, 2003) pela quantidade de vezes que uma *tag* foi utilizada ou que um objeto foi categorizado. No caso do sistema *CAW*, a popularidade corresponde à quantidade de vezes que uma entidade recebeu autoridade em determinada *tag* — quanto maior o número de vezes que a entidade recebeu autoridade de outras entidades em uma mesma *tag* maior é a sua popularidade. Da mesma forma, a popularidade de um objeto refere-se a quantidade de categorizações que ele sofreu. Com relação ao conceito de **peso** (ou valor), Wilson (1983) expõe que nenhuma entidade é considerada autoridade absoluta em todas as áreas e, também, que essa autoridade difere em peso ou importância de área para área na própria entidade, ou mesmo entre entidades distintas. Deste modo, todo termo utilizado para a concessão de autoridade está associado a um peso (de 1 a 5) que distingue as áreas em que uma mesma entidade é considerada autoridade e, também, que diferencia as autoridades das entidades umas das outras (ver Figura 15).

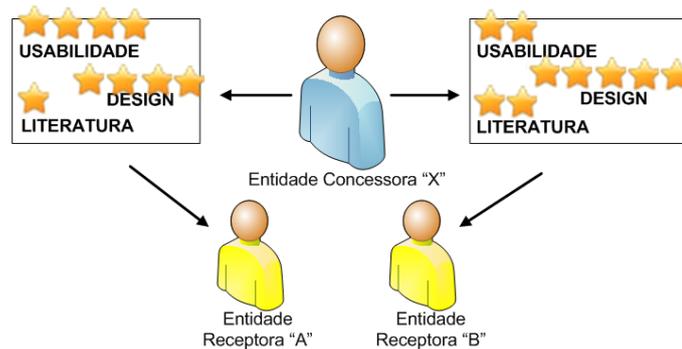


Figura 15 - Atribuição de peso à autoridade concedida

A Figura 15 representa a diferenciação com relação ao peso, ou valor. A entidade concessora atribui pesos distintos às autoridades concedidas para uma mesma entidade: pode-se dizer que a entidade “A” é considerada autoridade com grande peso tanto em “usabilidade” quanto em “design”, enquanto que com um peso mínimo em “literatura”, embora seja considerada uma autoridade nesse assunto. Observando a entidade “B”, é possível dizer que “design” é a sua maior especialidade (do ponto de vista da entidade “X”), seguido por “usabilidade” e “literatura”. Além de diferenciar as áreas em que a própria entidade é autoridade, o peso também serve como base para diferenciar a autoridade entre duas ou mais entidades. No mesmo exemplo da Figura 16, percebe-se que a entidade “A” ostenta maior autoridade em “usabilidade” que a entidade “B”, porém, “B” prevalece em “design” e em “literatura”.

As noções de popularidade dos objetos e de peso da autoridade das entidades são as bases para a recuperação de informação exposta nesse trabalho. Uma recuperação em que os objetos mais populares categorizados pelas autoridades de maior peso são priorizados para ocupar as primeiras posições entre os resultados obtidos.

4.2.1.2 Cálculo da Popularidade e do Peso de uma Autoridade

Para identificar a popularidade de uma autoridade em determinada *tag* é necessário identificar todas as concessões que ela obteve com esta *tag*. O processo de identificação da popularidade e do peso da autoridade que a entidade “X” possui no assunto “T” é demonstrado pelo algoritmo do quadro abaixo:

Quadro 1 – Algoritmo para a determinação da popularidade e do peso geral da autoridade.

<p>Início</p> <p>Selecionar TODAS as concessões de autoridade que a entidade “X” sofreu com a <i>tag</i> “T”</p> <p>Para CADA concessão de autoridade</p> <p style="padding-left: 40px;">Incrementar 1 (um) na popularidade da entidade “X” no assunto “T”</p> <p style="padding-left: 40px;">Contabilizar o peso da autoridade recebida — mínimo 1, máximo 5.</p> <p>Popularidade = total de concessões sofridas pela entidade “X” com a <i>tag</i> “T”</p> <p>Se a entidade “X” está na rede de autoridades da entidade que recupera a informação</p> <p style="padding-left: 40px;">Peso = peso definido pela entidade concessora da autoridade</p> <p>Senão</p> <p style="padding-left: 40px;">Peso = soma de todos os pesos / Popularidade</p> <p>Fim.</p>
--

No processo de recuperação de informação, identifica-se o peso das autoridades na *tag* solicitada e utiliza-se esse peso como base para determinar a ordem na qual os resultados do processo de recuperação de informação são apresentados. Os objetos que foram categorizados pela maior quantidade de autoridades (popularidade do objeto) e por autoridades de maior peso, são priorizados e apresentados antes de outros objetos que foram categorizados apenas por entidades comuns. O quadro 2 apresenta o algoritmo para a recuperação de informação priorizando os resultados que foram categorizados por autoridades.

Quadro 2 - Algoritmo para a recuperação de informação.

<p>Início</p> <p>Selecionar os objetos categorizados com a <i>tag</i> “T”</p> <p>Para CADA objeto</p> <p style="padding-left: 40px;">Recuperar as autoridades que categorizaram o objeto com a <i>tag</i> “T”</p> <p>Para CADA objeto</p> <p style="padding-left: 40px;">Determinar sua ordem de apresentação de acordo com:</p> <p style="padding-left: 80px;">1º - Quantidade de autoridades que o categorizaram</p> <p style="padding-left: 80px;">2º - Peso médio das autoridades categorizadoras</p> <p>Apresentar os objetos na ordem estabelecida</p> <p>Fim.</p>
--

O algoritmo do Quadro 2 prioriza os objetos que foram categorizados pelo maior número de autoridades, e dentre esses, os objetos cujo peso médio das autoridades categorizadoras seja o mais elevado. A Tabela 1 apresenta um exemplo dessa priorização.

Tabela 1 - Apresentação de resultados.

Ordem	Objeto	Quantidade de Autoridades	Peso Médio das Autoridades
1	Objeto3	7	4
2	Objeto1	7	3
3	Objeto4	5	3,5
4	Objeto2	1	4

Os objetos que não foram categorizados por nenhuma autoridade são ordenados pela quantidade de categorizações que obtiveram. Entretanto, como não há um peso para fazer a distinção entre os objetos que foram categorizados o mesmo número de vezes, os objetos categorizados mais recentemente possuem vantagem sobre os mais antigos.

É importante deixar clara a diferença entre a popularidade das autoridades e a popularidade dos objetos. Os algoritmos implementados consideram a popularidade das autoridades para o cálculo do peso que essas autoridades possuem, determinando-a pela quantidade de entidades que concederam a autoridade para a entidade em questão. Já a popularidade dos objetos é determinada pela quantidade de autoridades que os categorizaram com determinado termo, e é utilizada como critério para definir a posição que os objetos ocuparão na recuperação de informação.

4.2.2 A Cadeia de Autoridades

O processo de concessão de autoridade cognitiva por meio de folksonomia formará uma rede de contato social entre a entidade e suas entidades-fontes, aqui denominada de **rede de autoridades**. A rede de autoridades de uma entidade é composta por todas as entidades às quais ela concedeu autoridade por meio da atribuição de *tags*.

A interligação entre as redes de autoridades de cada entidade cria naturalmente uma **cadeia de autoridades**. No exemplo ilustrado pela Figura 16, a entidade “A” localizada no centro da figura estabeleceu que os componentes de sua rede são as entidades “B”, “C” e “D”, e atribuiu a cada uma a autoridade que lhes julgava pertinente. Essas entidades por sua vez, também definiram os componentes de suas redes e concederam-lhes autoridades; esses componentes também construíram suas próprias redes e assim sucessivamente.

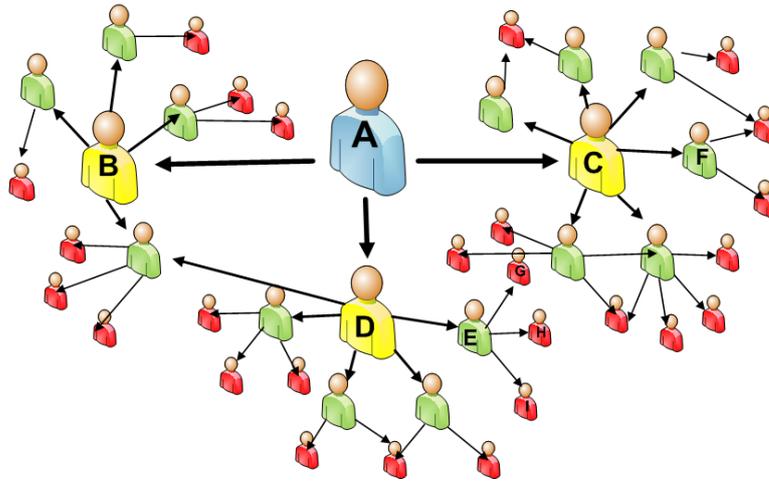


Figura 16 - Redes de autoridades formando uma cadeia de autoridades.

Dentro de uma cadeia de autoridades existem autoridades nas mais diversas áreas e com os mais distintos pesos. No momento da recuperação de informação em um sistema baseado em folksonomia, por meio da cadeia de autoridades formada pelas entidades categorizadoras de informação nesses sistemas, torna-se possível priorizar os objetos que foram categorizados por autoridades no assunto. Como exemplo, considere que a entidade “A” da Figura 16 realiza uma recuperação de informação com a *tag* “cinema”. A entidade “C” foi categorizada por “A” como uma autoridade em “cinema”, portanto, as informações categorizadas com “cinema” por “C” terão prioridades sobre as categorizações que “D” (que não é autoridade em “cinema”) realizou. Se ao percorrer a cadeia houver outras autoridades em “cinema”, as suas categorizações também serão priorizadas frente àquelas realizadas por não-autoridades, e é válido lembrar que essa priorização obedece a popularidade do objeto

categorizado (quantidade de autoridades que o categorizaram) e o peso das autoridades categorizadoras. Presumindo a legitimidade da autoridade das entidades em “cinema”, pode-se dizer que as categorizações realizadas pelas autoridades é de maior relevância, confiança e qualidade do que as demais categorizações realizadas por qualquer outra entidade.

Uma das razões para a construção das redes e, conseqüentemente, de uma cadeia de autoridades, é a sua utilização no processo de recuperação de informação. No entanto, a descoberta de autoridades e o reconhecimento das competências das entidades são dois benefícios importantes mencionados por Russell (2005) e que também são proporcionados pela cadeia exposta na Figura 16.

A **descoberta de autoridades** pode ser compreendida como uma resposta para a pergunta: “Quem são as autoridades no assunto X?”. A cadeia de autoridades possibilita que uma entidade encontre autoridades com as quais não possui contato direto, que estejam além de sua rede, mas que mantenham uma ligação, mesmo que remota, com ela por meio dessa cadeia de autoridades. Na Figura 16, a entidade “A” pode não conhecer diretamente nenhuma autoridade em “usabilidade”, porém, a entidade “D” que está conectada à “A” possui a entidade “E” em sua rede de autoridades e, “E”, por sua vez, definiu a entidade “H” como uma autoridade em “usabilidade”. Por outro lado, o **reconhecimento de competências** das entidades pode ser interpretado como uma resposta para questão: “Quais os assuntos nos quais a entidade Y é considerada autoridade?”. A resposta é a identificação de todas as *tags* utilizadas para conceder autoridade cognitiva à entidade “Y”.

4.2.2.1 Algoritmos para Operar sobre a Cadeia de Autoridades

Por meio dos processos de concessão de autoridade cognitiva pode-se criar uma cadeia de autoridades e, além de utilizar essa cadeia para a recuperação de informação, é possível efetuar a descoberta de autoridades e o reconhecimento de competências de uma determinada entidade.

O Quadro 3 apresenta um algoritmo para a construção da cadeia de autoridades da entidade “X” até o sexto grau de separação. Como uma entidade pode receber autoridade de diversas outras entidades, considera-se sempre o menor grau de separação entre a entidade foco e a entidade que recebeu autoridade (o menor caminho). A limitação da construção da cadeia até o sexto grau visa manter uma viabilidade de processamento e está fundamentada por estudos previamente realizados e que demonstram a validade dessa limitação (BALANCIERI *et al.*, 2005) (YU *et al.*, 2003). No próximo capítulo a cadeia de autoridades é apresentada com maiores detalhes.

Quadro 3 – Algoritmo para a construção da cadeia de autoridades da entidade “X”.

<p>Início Carregar a rede de autoridades da entidade “X” (1º grau da cadeia de “X”) Do primeiro até o quinto grau de separação na cadeia faça Para cada entidade do grau analisado Carregar a rede de autoridades da entidade Fim.</p>

Analisando a cadeia de autoridades de uma entidade, é possível efetuar uma busca por autoridades e mapear a ligação que essas autoridades possuem com a entidade. Considere que a entidade “A” precisa encontrar uma autoridade em “cardiologia”: por meio da cadeia de autoridades é possível identificar se existe alguma entidade “X” que seja considerada uma autoridade em cardiologia e a que grau de separação ela se encontra da entidade “A”. O Quadro 4 apresenta o algoritmo para encontrar uma autoridade na cadeia de autoridades.

Quadro 4 – Algoritmo para busca por autoridades de uma entidade em um assunto.

<p>Início Carregar a cadeia de autoridades da entidade “X” Percorrer a cadeia numa busca em largura até o nível 6 Se a busca for apenas pela autoridade mais próxima Apresentar a entidade categorizada com a <i>tag</i> “T” Apresentar o caminho entre a entidade “X” e a entidade autoridade em “T”</p>
--

<p>Finalizar a busca</p> <p>Senão</p> <p> Apresentar a entidade categorizada com a <i>tag</i> “T”</p> <p> Apresentar o caminho entre a entidade “X” e a entidade autoridade em “T”</p> <p>Fim.</p>
--

Numa outra perspectiva encontra-se o reconhecimento de competências de uma entidade. Enquanto a cadeia de autoridades é criada fundamentada nas concessões de autoridade realizadas sobre as entidades, esse reconhecimento consiste em identificar todas as áreas nas quais uma entidade recebeu autoridade. O Quadro 5 apresenta um algoritmo para o reconhecimento das competências de uma entidade “X”.

Quadro 5 – Algoritmo para o reconhecimento de competências da entidade “X”.

<p>Início</p> <p>Carregar todas as concessões de autoridade que “X” recebeu e com as quais “X” concordou</p> <p>Para cada concessão de autoridade</p> <p> Identificar as <i>tags</i> utilizadas para conceder autoridade e os níveis da autoridade concedida</p> <p> Se a <i>tag</i> “T” já foi utilizada</p> <p> Contabilizar mais 1 (um)</p> <p> Somar o nível da autoridade</p> <p> Identificar a entidade concessora da autoridade</p> <p> Senão</p> <p> Considerar a <i>tag</i> “T” como uma área na qual “X” é autoridade</p> <p> Contabilizar 1 (um) para a <i>tag</i> “T”</p> <p> Identificar o nível da autoridade concedida para “X”</p> <p> Identificar a autoridade concessora da autoridade</p> <p>Apresentar todas as <i>tags</i> em que “X” recebeu autoridade</p> <p>Apresentar a quantidade de vezes que cada <i>tag</i> foi utilizada</p> <p>Apresentar as entidades que concederam autoridade para “X”</p> <p>Se a quantidade de entidades que concederam autoridade para “X” for maior que 1 (um)</p> <p> Apresentar o nível total da autoridade / quantidade de categorizações que “X” recebeu</p> <p>Fim.</p>
--

O algoritmo apresentado no Quadro 5 não demonstra o nível da autoridade nos casos em que houve apenas uma concessão. Nos demais casos, em que houver duas ou mais concessões de autoridade em determinado assunto para uma entidade, o peso médio dessas

concessões é exibido, assim como a quantidade total de entidades que concederam autoridade para “A” e quem são essas entidades concessoras. Essa condição é uma medida para evitar problemas de privacidade originados pela definição do peso da autoridade concedida, por exemplo: a entidade “X” identificou que a entidade “A” lhe concedeu autoridade em um nível baixo. Entretanto, reconhecemos que o problema da privacidade persiste, pois para que a informação do peso médio seja exibida sem o risco de comprometer a privacidade dos concessores, deveremos identificar qual a quantidade mínima ideal de entidades concessoras a ser considerada para que a informação do peso médio seja apresentada. Este estudo não está inserido no escopo do nosso trabalho.

4.2.3 Extração de Informação na Cadeia de Autoridades

A construção da cadeia de autoridades pode ser utilizada para a extração de outras informações além das mencionadas na subseção anterior, assim como para a realização de análises sobre os dados gerados. Na seqüência são expostas algumas considerações adicionais sobre a identificação de autoridades populares, identificação de *hubs* de autoridades, formação de grupos e, também, sobre a recomendação de autoridades.

Identificação das autoridades mais populares: as autoridades mais populares são as entidades que possuem o maior número de categorizações num determinado assunto. Juntamente com a popularidade, é possível também calcular o peso médio dessa autoridade, considerando-se o peso atribuído em todas as concessões de autoridade recebidas.

Identificação de *hubs* de autoridades: os *hubs* são entidades que concedem autoridades à um grande número de entidades. A identificação de *hubs* pode ser útil para auxiliar a encontrar autoridades e até mesmo para identificar “*fakes*” entre as entidades. Os *hubs* que apontam para várias autoridades populares também podem ser considerados como “identificadores de autoridades”. Wilson (1983) chega a discutir sobre “autoridades em autoridades” e admite a existência de entidades qualificadas em sugerir ou encontrar

autoridades, mas uma verdadeira autoridade precisaria possuir bases suficientes para julgar e reconhecer todo tipo de autoridade nas mais distintas áreas — fato que não é considerado plausível. Wilson afirma que o julgamento da autoridade é algo subjetivo que não segue um procedimento explícito e que, dessa forma, não há como uma entidade julgar e conceder autoridade em nome de outras entidades e muito menos fazer esse julgamento por bases próprias.

Formação de grupos: a concessão de autoridade cognitiva propicia o reconhecimento de competências das entidades. Deste modo, é possível identificar autoridades em assuntos ou áreas em comum, com habilidades e competências relacionadas e/ou complementares e que, assim, possibilitem a formação de grupos. O ponto-chave é que estes grupos são formados por autoridades reconhecidas por outras entidades, e os objetivos da formação desses grupos podem ser muito distintos: podem ser gerados grupos de autoridades para a pesquisa e prevenção de uma doença; autoridades que se mobilizem e auxiliem em casos de catástrofes naturais; autoridades para a formação de um comitê científico para a avaliação de artigos; autoridades para discutir assuntos políticos, sociais ou econômicos; autoridades que possuem conhecimentos e interesses em comum para a discussão de idéias; para oferta e procura de empregos, cargos, indicações, entre outros.

Recomendação de autoridades: a identificação de características comuns das autoridades de uma entidade (ex: vocabulário utilizado, objetos categorizados, relacionamento na cadeia) pode fornecer bases para a recomendação de autoridades “relacionadas” que podem ser de interesse da entidade. Existem trabalhos relacionados à recomendação de conteúdos baseada em filtragem colaborativa (SCHAFER *et al.*, 2007) e, também, à determinação da autoridade de usuários de sistemas de recomendação (MILANI & CAZELLA, 2006) que realizam recomendações com base na identificação de características

comuns entre objetos e que servem de base para um estudo sobre a possibilidade de recomendação autoridades no sistema CAW.

4.2.4 Análises Sobre a Cadeia de Autoridades

A cadeia de autoridades também possibilita o desenvolvimento de análises relacionadas tanto ao processo de concessão de autoridade cognitiva quanto ao impacto causado por esse processo. A área de abrangência da autoridade, o impacto da concessão de autoridade sobre a entidade que a recebe e a evolução da autoridade recebida por uma entidade são exemplos de análises que podem ser desenvolvidas sobre a cadeia de autoridades. Na seqüência, expomos algumas considerações sobre essas análises.

Área de abrangência: uma vez construída, a cadeia de autoridades pode tornar possível o desenvolvimento de estudos que identifiquem a área de abrangência, seja da influência de uma entidade, seja da concordância semântica com relação aos termos utilizados para a categorização de objetos. A influência que uma autoridade exerce pode ser maior ou menor, de acordo com o seu grau de separação (VIRGINIA, 2007) com uma determinada entidade (a distância entre elas). Essa influência pode estar relacionada tanto à indicação de novas autoridades e à categorização de informação, quanto à forma como essas categorizações são realizadas e ao vocabulário utilizado pela entidade. Ainda com relação ao vocabulário de termos utilizados, se duas entidades aplicam um mesmo termo para a categorização de objetos, as chances do significado representado por esse termo ser o mesmo para ambas também deve sofrer influência de acordo com o grau de separação entre as duas entidades. É factível acreditar que entidades que possuem um grau de separação próximo (1 grau, por exemplo), atribuam um significado semelhante para um termo “X”. Entretanto, também é factível crer que essa afirmação não seja pertinente para um grau e separação mais distante (5º ou 6º grau).

Análise do impacto da concessão de autoridade cognitiva: a autoridade concedida e recebida pode exercer ou sofrer influência de uma série de fatores, tais como: para quem se está concedendo autoridade, o intuito da concessão, os termos utilizados, os níveis de autoridade recebidos, entre outros. Se uma entidade recebe autoridade em um assunto no qual ela não possui domínio, essa entidade pode sentir-se motivada a informar-se mais sobre o assunto? Se uma entidade é reconhecida como uma autoridade de alto nível em determinado assunto, esse reconhecimento influenciará no rigor ou na forma como a mesma realiza suas categorizações naquele assunto? O fato de uma entidade receber autoridade pode influenciá-la a retribuir com outra concessão de autoridade que não reflita, necessariamente, sua real opinião? As razões para concessão de autoridade (amizade, medo, influência, necessidade, opinião própria, etc.), podem causar o surgimento de falsas autoridades? Como identificar falsas autoridades? Como a autoridade recebida pode impactar na reputação de uma entidade? E, como a rede de autoridades de uma entidade pode influenciar em sua reputação?

Análise de evolução da autoridade: uma análise da evolução dos termos e de seus pesos atribuídos às autoridades no decorrer do tempo pode descrever mudanças relacionadas à capacidade, influência, reputação e reconhecimento de uma determinada entidade. Wilson (1983) argumenta que a autoridade é um item subjetivo e situacional. Rieh & Belkin (1998; 2000) (RIEH, 2005) confirmam essa afirmação considerando que por estar sujeita à subjetividade a autoridade concedida a uma entidade é situacional e, portanto, pode sofrer alterações de acordo com: a evolução do conhecimento de seu concessor; com os contatos e com novas autoridades que esse concessor obtenha; com o grau de confiança ou de proximidade existente entre ambos; e outras decisões de caráter pessoal.

4.3 Reflexões sobre o Processo de Concessão de Autoridade

Nesta seção são expostas algumas considerações sobre o processo de concessão de autoridade cognitiva e sobre a cadeia de autoridades originada nesse processo. As

considerações realizadas buscam expor reflexões sobre particularidades existentes na proposta de utilizar folksonomia para conceder autoridade e, também, sobre a construção e entendimento da cadeia de autoridades.

4.3.1 A Concessão de Autoridade Cognitiva

A primeira distinção entre a concessão de autoridade cognitiva e um processo de categorização de objetos está no objetivo. Como já citamos anteriormente, Russell (2005) menciona que na concessão de autoridade o usuário\entidade se torna o objeto e que, de acordo com suas palavras, “(...) *users are not tagging objects with ‘aboutness’ or labels. They are not labeling what something is or what it is about. They are labeling what someone knows, what they are good at, what they are an authority on.*”²⁰. Assim, uma *tag* atribuída a uma entidade representa a autoridade que essa entidade possui, na opinião de quem a concedeu.

Wilson (1983) defende que toda pessoa possui uma visão do mundo de acordo com seus conhecimentos, opiniões e objetivos — a lente com a qual se vê o mundo. Deste modo, a autoridade concedida por uma entidade representa a opinião da mesma, pois, de acordo com sua visão de mundo, a entidade que recebe essa autoridade a merece. Max Weber (WEBER, 1963) discute sobre as razões que levam ao reconhecimento da autoridade, mas essas razões não serão discutidas no escopo do nosso trabalho.

O peso relacionado a cada *tag* atribuída a uma entidade tem o intuito de distinguir as áreas nas quais a entidade é autoridade. Avaliar a entidade de uma forma global ocasiona a perda de informações relevantes a seu respeito — podem existir áreas nas quais a entidade é mais reconhecida e áreas nas quais ela possui menos autoridade. O peso geral da autoridade de uma entidade em determinado assunto é obtido pela média dos pesos recebidos: uma

²⁰ Tradução do autor: “os usuários não estão categorizando os objetos com relação ao seu conteúdo ou a sua descrição. Eles não estão descrevendo sobre o que alguma coisa é ou o que ela representa. Eles estão descrevendo o que alguém sabe, no que eles são bons, ou em que eles são uma autoridade.”

entidade que recebeu peso 4 e peso 2 em “programação” possui uma avaliação média igual a 3. A representação geral do peso da autoridade recebida por uma entidade se destina a demonstrar a opinião que as demais entidades mantêm sobre ela. Essa informação serve como base para que entidades que estejam procurando por referências em determinado assunto visualizem o que as demais entidades pensam sobre aquela autoridade. É importante mencionar que na recuperação de informação o peso de uma autoridade, existente na rede de autoridades de certa entidade, respeita o valor determinado no momento da concessão dessa autoridade, o peso geral é utilizado nos casos em que a autoridade não se encontra diretamente conectada na rede da entidade.

Preece *et al.* (2005) falam da importância de considerar aspectos sociais, culturais e emocionais dos usuários ao projetar o esquema de interação de um sistema. A atribuição de pesos às *tags* é uma ação que exige planejamento e cuidados. Na concessão de autoridade não estamos avaliando um recurso ou atribuindo uma nota para um objeto qualquer, estamos julgando as competências de uma entidade (pessoa ou organização) expressando a nossa real opinião sobre a mesma. Em nosso convívio social não costumamos rotular explicitamente nossos colegas, conhecidos, amigos, etc., com um peso que determine a importância da amizade ou, nesse caso, o valor da autoridade, mesmo porque esse costume geraria conflitos (ao concedermos maior importância a um amigo que a outro, ou ao concedermos um baixo valor ao nosso chefe), e também porque, de acordo com Wilson (1983), nem sempre as pessoas conseguem expressar a autoridade concedida a uma entidade em uma escala mensurável. Normalmente, essa autoridade é implícita e difícil de ser traduzida em um valor. Por essas razões, como já mencionado na subseção 4.2.2.1, ao possibilitar a avaliação da autoridade de uma entidade é necessário que o sistema proteja a privacidade dessa informação — o peso não pode ser visualizado de forma aberta, pode ser exposta a média de todos os pesos, mas sem demonstrar os pesos atribuídos individualmente.

No que se relaciona à atribuição do peso e a dificuldade em expressá-lo numa escala numérica, um esquema de avaliação utilizado por muitos sistemas (*Orkut*, *Folha On-line*, *Windows Media Player*) é a marcação por “estrelas”. Os avaliadores atribuem de 1 a 5 estrelas e a avaliação individual é exibida para os mesmos de forma privada, mas considerada para a exibição da avaliação geral. Entretanto, quando procuramos por uma referência em um assunto, mais importante que saber o valor da autoridade que uma entidade possui é saber **quem** a concedeu, e até mesmo **quantas** pessoas\entidades atribuíram essa autoridade. Deste modo, para fornecer meios mais relevantes de demonstrar a autoridade de uma entidade o sistema deve informar: a **popularidade** (quantas vezes a entidade recebeu autoridade), o **peso médio** (quantidade média de “estrelas” recebidas por categorização) e **quem são os concessores da autoridade** (entidades que concederam a autoridade exibida).

A concessão de autoridade ao ser utilizada para a recuperação de informação implica em possíveis conflitos entre os vocabulários dos usuários. A entidade que concede a autoridade possui seu vocabulário de termos utilizados para a categorização de informação, assim como a entidade que recebe a autoridade também possui. No entanto, para se obter resultados categorizados por autoridades na recuperação de informação é necessário que a entidade que recebe a autoridade utilize, para categorizar informações, a mesma *tag* que foi aplicada para lhe conceder autoridade — caso contrário, haverá uma autoridade que não possui nenhuma informação categorizada em sua área de domínio e, assim, nada a ser recuperado. Por mais que a autoridade concedida a uma entidade seja reflexo da opinião já existente sobre a mesma, a compatibilidade necessária para a recuperação de informação ocasiona uma maior carga de trabalho para o concessores e restringe o vocabulário de termos que ele pode utilizar para representar essa autoridade. Por outro lado, isso pode fazer com que ocorra uma maior compatibilidade de termos e de significados entre os vocabulários das entidades. Já mencionamos anteriormente sobre o efeito *long tail* das *tags* e, de acordo com

Russell (2005), não existem razões aparentes para não se esperar o mesmo das *tags* aplicadas para a concessão de autoridade. Entretanto, a proposta de Russell não menciona a recuperação de informação e, conseqüentemente, as considerações em torno da compatibilidade de vocabulários entre as entidades — o que acreditamos atenuar o *long tail* e reduzir problemas como polissemia e ambigüidades.

4.3.2 Considerações Sobre a Interação dos Usuários com a Ferramenta CAW

Nos sistemas baseados em folksonomia o foco no usuário é evidenciado. Deve-se pensar muito nos aspectos de interação com o sistema e na contribuição que o mesmo tem a oferecer. Se o **custo do processo de concessão de autoridade** for alto em termos de tempo e de carga de trabalho, então, os usuários não se sentirão motivados a utilizá-lo. Todo sistema deve ser projeto tendo o usuário em mente, “...*não se pode pensar em interfaces sem considerar o ser humano que vai usá-la...*” (ROCHA & BARANAUSKAS, 2003, p.13). A aceitação e o sucesso de um sistema depende diretamente da forma como o mesmo foi projetado, do que ele exige dos usuários e do que ele tem a oferecer em troca.

Ao refletirmos sobre assuntos relacionados à interação dos usuários com a ferramenta CAW, identificamos quatro pontos principais que a diferem de um sistema baseado em folksonomia convencional: i) autoridade negativa; ii) autoridade consentida; iii) manutenção da autoridade concedida; e iv) custo do processo de concessão de autoridade.

O sistema CAW não considerará a questão de **autoridade negativa** (ou anti-autoridade). Presume-se que se uma entidade recebe autoridade em determinado assunto ela a merece, pelo menos do ponto de vista de quem a concedeu e cuja opinião deve ser respeitada. O que caracteriza e representa a autoridade de uma entidade são as *tags* a ela atribuídas de forma que, as áreas em que essa autoridade não possui conhecimento, ou não é considerada referência, não estarão representadas por *tag* alguma. Esta política para preservação dos relacionamentos sociais evita conflitos desnecessários: se não consideramos um pesquisador

como uma autoridade em seu assunto de pesquisa não precisamos tornar isso explícito, basta não manifestarmos nossa opinião a respeito. Não estamos negando sua autoridade e nem estamos manifestando-a. Qualquer sistema de caráter social deve se preocupar com questões que possam gerar conflitos entre seus usuários — Farrell & Tessa (FARRELL *et al.*, 2006) mencionam experiências negativas com a utilização de *tags* para a descrição de colaboradores em um projeto desenvolvido pela IBM[®]. Essas experiências demonstram que recursos mal projetados podem ser definitivos para a rejeição dos usuários condenando todo um projeto ao fracasso.

Russell (2005) argumenta sobre a necessidade de uma entidade “concordar” com a autoridade recebida, chamando de: *blessed authority* ou **autoridade consentida**. O autor menciona sobre a ocorrência de casos em que uma entidade não concorda com a autoridade recebida, seja por questões de privacidade ou mesmo por não se considerar como tal. Voltamos a questão social envolvida no processo — a concessão de autoridade torna pública a opinião sobre a entidade que recebe a autoridade. Como qualquer outra entidade pode conceder-lhe autoridade em qualquer área, é compreensível que a entidade queira manter um gerenciamento do que está sendo dito sobre ela, seja para evitar falsas categorizações, seja para se preservar ou para preservar sua privacidade: uma pessoa que desenvolve estudos sobre a reprodução humana pode não concordar, ou não desejar, ser caracterizada como uma autoridade em “sexo”. O Orkut é um sistema social que aplica essa filosofia: quando um usuário adiciona outro usuário como amigo, essa ação somente estará publicamente exposta quando o usuário adicionado concordar com ela. Na concessão de autoridade a lógica é semelhante — a autoridade só deve estar publicamente disponível após a entidade que a recebeu ter concordado com a concessão. No entanto, enquanto no Orkut o relacionamento de “amizade” é bidirecional (“A” é amigo de “B”, logo, “B” é amigo de “A”) e deixa de existir caso “B” recuse a amizade de “A”, em um sistema para a concessão de autoridade isso não

acontece (ver Figura 17). A primeira diferença é que o relacionamento é unidirecional (“A” é autoridade de “B”, mas não implica em “B” ser autoridade de “A”) e a segunda é que, considerando o exemplo utilizado anteriormente, mesmo que aquela pessoa não concorde em ser uma autoridade em “sexo” isso não muda a opinião de quem lhe concedeu autoridade, e ambas estão corretas. “A” deseja manter sua privacidade e, portanto, recusa ser reconhecida explicitamente como tal; e “B” deseja manter sua opinião sobre “A” e utilizar essa concessão de autoridade na recuperação de informação. Deste modo, se “A” negar a autoridade concedida por “B”, essa autoridade será visualizada somente por “B” e nenhuma informação a respeito estará publicamente exposta; “A” estará na rede de autoridades de “B”, mas de forma privada. Caso “A” concorde com a autoridade que “B” lhe concede, essa autoridade será exibida publicamente, assim como seu concessor.

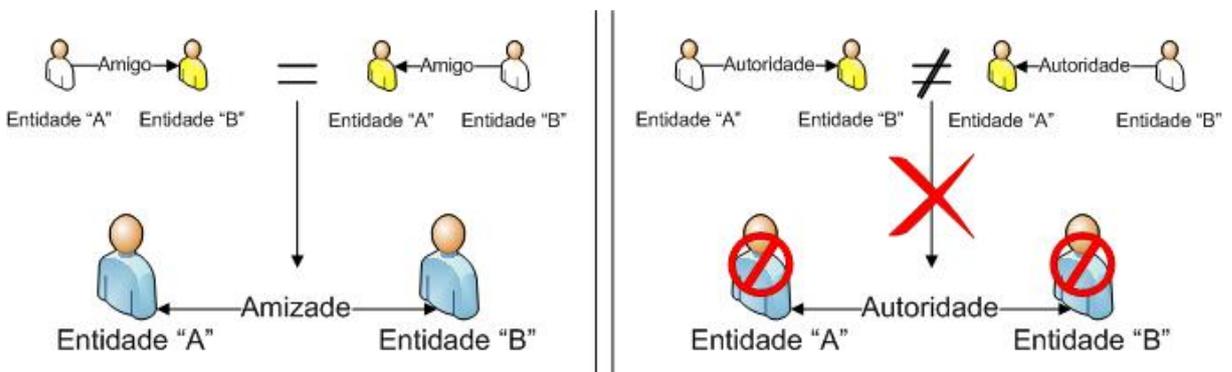


Figura 17 – Relacionamentos: amizade (bidirecional) Vs. autoridade (unidirecional).

A **manutenção da autoridade concedida** também é uma característica intrínseca da ferramenta proposta e de todo sistema social. Primo (2006) debate sobre os relacionamentos sociais na *web* e discute sobre a interação mediada por computador. De acordo com Wilson (1983), a autoridade sofre alterações com o passar do tempo influenciada por fatores subjetivos (experiência, objetivos, espaço, etc.). Nós estamos constantemente alterando nossa opinião sobre os outros e sobre a autoridade que eles representam — é assim quando uma criança enxerga em seus pais a autoridade absoluta em todas as coisas e, com o passar do tempo, e ao decorrer de suas experiências, começa a reconhecer autoridades novas e a

reavaliar as antigas; essa manutenção é necessária para refletir a opinião atual das entidades e para demonstrar a evolução das autoridades ao decorrer do tempo.

4.3.3 A Cadeia de Autoridades

A concessão de autoridade pelo sistema *CAW* torna explícita a autoridade já concedida pelas entidades umas às outras no mundo real — Alde *et al.* (2005) e Lopes (2005) argumentam que a autoridade dos produtores de informação na Internet e, conseqüentemente, da própria informação por eles produzida, reflete a autoridade já existente e já recebida por esses produtores no mundo real. Nesse sentido, podemos considerar que a metáfora da percepção citada por Wilson (1983) sobre o modo como as pessoas compreendem, entendem e interpretam as coisas também é válida para o que se conhece sobre os outros. Normalmente, temos uma visão limitada do que as pessoas sabem, de suas experiências e de suas áreas de interesse — nosso conhecimento está limitado às nossas percepções e experiências diretas com essas pessoas ou, então, ao o que ficamos sabendo por meio de outros (novamente, o conhecimento de segunda-mão). A cadeia de autoridades nos coloca em contato com a opinião e com a visão de muitas outras pessoas, uma visão explícita do que essas pessoas pensam ou consideram sobre os outros, às quais não teríamos acesso de outro modo. Considerando as afirmações de Russell (2005) sobre a concessão de autoridade por meio de folksonomia, podemos supor que, com uma quantidade suficiente de entidades concedendo e recebendo autoridade, é possível identificar uma autoridade para praticamente qualquer tópico almejado ou para algum tópico relacionado ao mesmo. Além disso, é possível também identificar as várias visões existentes sobre uma mesma entidade — as várias áreas nas quais ela é considerada autoridade por todas as entidades que a categorizaram.

Fundamentando-nos nas reflexões de Wilson (1983), Rieh & Belkin (1998; 2000; 2005) e Russell (2005) sobre a autoridade cognitiva no dia-a-dia das pessoas, podemos afirmar que a cadeia de autoridades apresenta explicitamente a autoridade concedida às

pessoas nos relacionamentos sociais e que a mesma é capaz de suportar boa parte da complexidade existente nesses relacionamentos. Normalmente, possuímos autoridades em diversos assuntos e, também, diversas autoridades em um mesmo assunto. Uma pessoa considerada referência em matemática raramente também será considerada como uma referência em direito, em medicina e em filosofia — e em caso positivo, é provável que o seja em níveis distintos. Dificilmente uma pessoa ou entidade é considerada autoridade absoluta em todos os assuntos — a autoridade é restrita a um domínio ou conjunto de áreas (esfera de interesse). Um veterinário é mais indicado para fornecer conselhos sobre como cuidar de um cão do que sobre como eliminar as formigas que estão estragando o jardim. A concessão de autoridade por meio de folksonomia oferece suporte para a existência de várias autoridades em um mesmo assunto, além de possibilitar que uma entidade receba autoridade em diferentes assuntos e com diferentes pesos.

Nem sempre a **maior** autoridade é a **melhor** autoridade. Principalmente nos casos em que buscamos por conselhos ou por opiniões, preferimos recorrer primeiramente a pessoas mais próximas do que àquelas com as quais não possuímos um contato direto — é também uma questão de confiança. Quando temos curiosidade em começar a aprender um novo assunto, por exemplo, é mais provável buscarmos uma orientação com o colega da sala ao lado que possui um conhecimento básico na área, do que entrarmos em contato com um respeitado pesquisador de outra universidade com o qual nunca conversamos. Assim, mesmo que esse pesquisador figure como a maior autoridade no assunto, em muitos casos é preferível uma autoridade em grau mais próximo, mesmo que esta possua menor peso. A cadeia de autoridades permite a identificação das autoridades mais populares e de maior peso, e dentro da cadeia, o grau mais próximo da entidade é justamente a rede de autoridades diretamente por ela construída.

Muitas vezes buscamos por referências em uma área, porém, não conhecemos diretamente ninguém que possa nos orientar sobre o assunto — ou não sabemos que alguém o pode fazer. Considere o desejo de nos informarmos sobre a influência dos *Beatles* na cultura inglesa: sabemos que um aluno possui um amigo que estudou durante vários anos na Inglaterra e que, talvez, possa nos fornecer informações úteis. Entretanto, desconhecemos que aquele nosso colega da sala ao lado é fã dos *Beatles* e que mantém um repositório de informações coletadas sobre os mesmos durante a sua adolescência. A cadeia de autoridades torna explícitas as diversas áreas nas quais uma entidade é considerada autoridade — um amigo próximo do nosso colega sabe que ele é fã dos *Beatles* e concedeu a ele autoridade nesse assunto; isso implica numa expansão da percepção limitada que as entidades possuem isoladamente. Por outro lado, mesmo que nosso colega não tenha ou não saiba as informações que procuramos, é possível percorrer a cadeia de autoridades em busca de alguém reconhecido como autoridade em tal assunto e que mantenha uma ligação, mesmo que remota, conosco. Desse modo, a cadeia de autoridades representa uma situação típica em nosso convívio social: podemos não conhecer diretamente alguém que saiba sobre o assunto, mas conhecemos alguém, que conhece alguém que sabe a resposta. A cadeia de autoridades permite encontrar a autoridade desejada mapeando um caminho entre nós e essa autoridade — os graus de separação do Oráculo de Bacon (VIRGINIA, 2007).

A opinião das pessoas ou entidades que consideramos autoridades exerce influência sobre nós também no julgamento do merecimento da autoridade de outras entidades. Se alguém que consideramos autoridade possui uma opinião positiva sobre a autoridade de uma entidade, então podemos nos sentir encorajados a concordar com ela. O contrário também é verdadeiro. A cadeia de autoridades não demonstra autoridade negativa, mas permite que seja traçada uma trajetória do peso da autoridade ao decorrer de seus graus, de modo que quanto

maior o grau de separação de uma entidade com uma autoridade, essa autoridade tenha maior ou menor importância. A isto chamamos de propagação e de retenção de autoridade.

A **propagação de autoridade** ocorre quando uma entidade, que é autoridade em determinado assunto, concede autoridade nesse mesmo assunto para outras entidades mantendo ou elevando o peso que possui — o peso que ela recebeu das entidades que a categorizaram. Se um professor que consideramos autoridade em pesquisa científica nos diz que determinado autor é uma grande autoridade nessa mesma área, tendemos a absorver tal consideração a menos que tenhamos motivos para pensar o contrário. Por outro lado, a **retenção de autoridade** representa o caso oposto, no qual uma autoridade não recomenda outra entidade como tal em sua área. Como a cadeia não considera autoridade negativa, o que ocorre é uma redução no peso da autoridade: uma autoridade em nível máximo localizada no primeiro grau da cadeia concede autoridade em nível médio para outra entidade (localizada então no 2º grau de separação). Houve desse modo, uma redução na autoridade encontrada no grau mais distante. Isso aconteceria se o professor mencionasse um ex-aluno como um bom estudioso na área — nós já estamos ligados diretamente a uma grande autoridade e, por meio dela, à uma autoridade de menor nível.

A Figura 18 ilustra a situação em que a autoridade é retida e a situação em que a autoridade é propagada. A entidade “X”, que possui um peso médio igual a 3 (três) em “usabilidade”, está retendo a autoridade ao conceder autoridade com peso 1 (um) em “usabilidade” para a entidade “A”. Por outro lado, a entidade “X” propaga a autoridade ao conceder peso 4 (quatro) de autoridade em “usabilidade” para a entidade “B”.

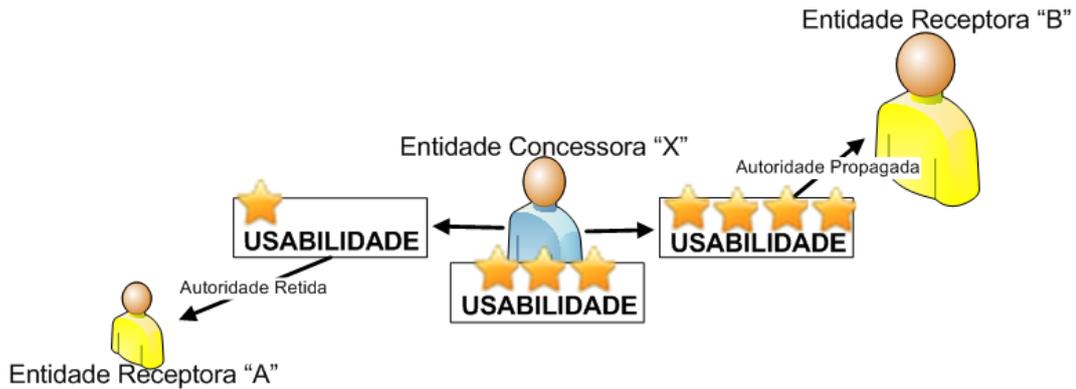


Figura 18 - Propagação e retenção de autoridade cognitiva.

Essa análise de retenção ou de propagação pode ser realizada por dois focos distintos:

i) quando a análise é geral e, portanto, considera-se o peso médio da autoridade da entidade concessora “X”; ou ii) quando a análise parte da entidade que concedeu autoridade para a entidade “X” e, deste modo, considera-se o peso da autoridade que ela concedeu para “X” em “usabilidade” em vez do peso médio de todas as concessões que “X” recebeu. A identificação da retenção e da propagação da autoridade permite ir além do cálculo da popularidade, pois torna possível demonstrar quem são as “maiores autoridades na cadeia” levando em conta o peso dos concessores da autoridade nesse processo.

4.4 Trabalhos Relacionados

Nesta seção, discute-se brevemente sobre as principais pesquisas relacionadas ao nosso trabalho, sobre os pontos em comum e sobre as diferenças mais importantes existentes entre eles.

Contextual Authority Tagging (RUSSELL, 2005): Russell apresenta uma abordagem propondo a concessão de autoridade cognitiva por meio de folksonomia dentro de uma comunidade de usuários. O intuito dessa concessão é a identificação de autoridades e o reconhecimento de competências dos usuários. Russell desenvolveu uma investigação sobre autoridade cognitiva e sobre folksonomia apresentando reflexões relacionadas ao aspecto social de um sistema que permita a concessão de autoridades a seus usuários da mesma forma

que os sistemas baseados em folksonomia permitem a categorização de objetos. O trabalho de Russell motivou nossa pesquisa apresentando a proposta de utilizar folksonomia para a concessão de autoridades, o que permitiu nos aprofundarmos nas considerações sobre as particularidades sociais dessa proposta e direcioná-la para outro foco. Russell propõe a concessão de autoridade para a identificação de autoridades e para o reconhecimento de competências dentro de uma comunidade de usuários. Em nosso trabalho a proposta de Russell também é considerada, entretanto, o foco está na construção da cadeia de autoridades para melhorar o processo de recuperação de informação em sistemas baseados em folksonomia — a identificação de usuários e o reconhecimento de competências são conseqüências do processo de concessão de autoridade cognitiva.

Fringe Contacts: People-Tagging for the Enterprise (FARRELL & LAU, 2007):

Fringe Contacts é um projeto em desenvolvimento pela IBM® cujo foco é a categorização de pessoas para auxiliar o contato entre as mesmas, enquanto essas pessoas contribuem para a caracterização de seus amigos e colegas. O intuito é fornecer um meio de encontrar, manter contato, e aprender sobre outras pessoas em ambientes empresariais, visando aumentar a efetividade e reduzir o custo da criação e da manutenção de relacionamentos profissionais. Um projeto piloto foi desenvolvido para possibilitar que os colaboradores atribuíssem *tags* uns aos outros e *plug-ins* relacionados a software de mensagem instantânea e obtenção de informações dos bancos de dados internos da empresa começaram a ser projetados. O projeto *Fringe Contacts* propõe utilizar a folksonomia para a caracterização de pessoas (ou entidades), entretanto, para o projeto piloto não foram considerados aspectos das relações sociais como os discutidos por Russell (2005) ou mencionados em nosso trabalho. Ferrell e Lau (2007) chegam a mencionar a ocorrência de problemas relacionados à caracterização de pessoas que não concordaram com as *tags* utilizadas, demonstrando a complexidade desse processo e a necessidade de considerar tais aspectos. Outra diferença é que, pelo menos

inicialmente, o trabalho é direcionado para utilização interna em uma empresa ou organização, restrito aos seus colaboradores, com o principal objetivo de facilitar o contato e a manutenção dos relacionamentos profissionais entre os mesmos.

Expertise Classification: Collaborative Classification Vs. Automatic Extraction (BOGERS *et al.*, 2006): O trabalho desenvolvido pelo grupo ILK da Universidade de Tilburg, Holanda, é direcionado à classificação da especialidade de membros de uma comunidade. O grupo propõe um algoritmo para a extração automática de termos que representa as áreas nas quais os membros de um grupo possuem domínio — as áreas de especialidade —, e calcula o peso de cada área de acordo com dados obtidos automaticamente de publicações realizadas por esses membros. A avaliação dos resultados obtidos pela aplicação do algoritmo sobre os membros do próprio grupo de pesquisa é realizada comparando-os com dados coletados por meio de uma classificação colaborativa, em que os membros classificaram e avaliaram uns aos outros de acordo com uma lista de termos identificada previamente pelo algoritmo desenvolvido. A validação também solicitava que os membros informassem suas próprias áreas de especialidade para comparações com os termos extraídos automaticamente. Os experimentos demonstraram a viabilidade de realizar a comparação utilizando a classificação colaborativa, mas demonstraram também que ainda existe uma grande diferença de relevância entre os termos identificados automaticamente e aqueles sugeridos pelos próprios usuários. Apesar desse trabalho também ser direcionado para a identificação de competências e para uma distinção das mesmas por meio de avaliações, ele se foca em uma identificação automática utilizando a opinião das pessoas como uma etapa de avaliação e não como um processo central.

Autoridade de Usuários em Sistemas de Recomendação (MILANI & CAZELLA, 2006): As pesquisas coordenadas pelo Professor Silvio Cazella, Universidade Vale dos Sinos – RS, são direcionadas ao estudo e desenvolvimento de técnicas para a recomendação de

informação. O diferencial dos trabalhos coordenados por Cazella está na consideração da autoridade de quem expressa a opinião (uma opinião que possuirá maior ou menor importância de acordo com o peso da autoridade de quem a emitiu) — essa autoridade é obtida por meio de indicadores relacionados a publicações e a produções científicas do indivíduo disponíveis na plataforma *Lattes*. A proposta de trabalho atual é a consideração da opinião dos usuários sobre os resultados da recomendação de artigos científicos, diferenciando essa opinião de recomendação de acordo com a autoridade de seu emissor. Nesse foco, é utilizada a participação dos usuários na avaliação do que lhe foi recomendado, entretanto, a autoridade que determina o peso dessa avaliação é obtida de forma automática, o que, segundo as definições de Wilson (1983), não a caracteriza como autoridade cognitiva, mas sim como um esquema de representação da reputação do usuário em determinado assunto e com critérios previamente estabelecidos.

Inferring Binary Trust Relationships in Web-Based Social Networks (GOLBECK & HENDLER, 2006): Sobre a coordenação de Jennifer Goldbeck, Universidade de Maryland, estão sendo desenvolvidos projetos para inferência de confiança entre pessoas e outras propriedades dos relacionamentos sociais obtidas por meio da análise de redes sociais baseadas na Internet. A folksonomia não é citada como parte das pesquisas realizadas, entretanto, os estudos desenvolvidos por esses pesquisadores, principalmente no que se relaciona à redes sociais e suas particulares, apresentam contribuições úteis à qualquer tipo de software social. Comparando confiança com autoridade cognitiva, o trabalho de Goldbeck & Hendler é direcionado à inferência binária de confiança entre dois membros de uma rede social (confia ou não confia), entretanto, a noção de autoridade cognitiva que utilizamos nesse trabalho é mais ampla e mais complexa, não podendo ser determinada por meio de um mecanismo de inferência ou por um conjunto de regras padronizadas. Os experimentos realizados para a validação das propostas são executados sobre redes sociais simuladas,

Goldbeck & Hendler argumentam que “*Naturally occurring networks take a long time to gain a large number of users, and the topological properties are fixed.*”²¹ (GOLBECK & HENDLER, 2006, p.509), de modo que a simulação é a alternativa mais viável para validar os trabalhos desenvolvidos.

Neste capítulo, a concessão de autoridade cognitiva por meio de folksonomia foi discutida e considerada para melhorar o processo de recuperação de informação em um sistema baseado em folksonomia. Foram expostas reflexões e discussões sobre as particularidades e sobre as conseqüências de se utilizar a folksonomia para conceder autoridade cognitiva e, então, de se utilizar essa autoridade na recuperação de informação. O projeto do sistema CAW foi brevemente descrito, juntamente com a forma como o processo de concessão de autoridade cognitiva é suportado pelo mesmo. O próximo capítulo apresenta uma simulação dos processos de concessão de autoridade cognitiva e de recuperação de informação aqui discutidos. O objetivo da simulação é validar a hipótese de que no processo de recuperação de informação, priorizar as categorizações realizadas por autoridades colabora para um aumento da relevância e da qualidade dos resultados obtidos, e reduz o impacto causado pela sobrecarga de informação nos sistemas baseados em folksonomia.

²¹ Tradução do autor: “Redes construídas naturalmente levam muito tempo para obter um grande número de usuários, e as propriedades topológicas são fixas”

Capítulo V

Simulações: Concessão de Autoridade Cognitiva e Recuperação de Informação

Segundo Goldbeck & Hendler (2006), obter dados provenientes da utilização de um sistema baseado na *web* com usuários reais leva muito tempo e pode implicar num custo muito alto. Deste modo, eles defendem que simular esses dados e analisá-los seja uma alternativa adequada para identificar a viabilidade de uma proposta ou a validade de uma hipótese. Assim, para demonstrar o processo de *folkauthority* e a sua aplicação na recuperação de informação optamos por desenvolver um estudo simulado. As simulações executadas em nosso trabalho e apresentadas neste capítulo possuem o intuito de representar usuários reais na utilização de um sistema baseado em folksonomia ao efetuarem a categorização de objetos, a concessão de autoridade cognitiva e a recuperação de informação. Essas simulações visam a verificação da hipótese de que a aplicação do conceito de autoridade cognitiva na recuperação de informação apresenta resultados de maior precisão e maior qualidade, e ameniza o impacto da sobrecarga de informação em sistemas baseados em

folksonomia. Neste capítulo, expomos a metodologia utilizada para estas simulações, as informações geradas, as análises realizadas e os resultados obtidos.

5.1 Metodologia

O capítulo II apresentou conceitos e discussões sobre a técnica de Folksonomia. Nessas discussões, a categorização de objetos foi considerada como um processo no qual os usuários de um sistema baseado em folksonomia atribuem *tags* aos objetos, rotulando-os, descrevendo-os, expressando seus significados, conteúdos ou características. O ponto principal nesse processo é a dependência da participação e do envolvimento dos usuários do sistema, sem os quais não há dados sendo gerados e, conseqüentemente, não há informações para serem analisadas.

Para demonstrar a aplicação do conceito de autoridade cognitiva e a sua utilização na recuperação de informação em um sistema baseado em folksonomia, é necessário ter controle sobre os dados desse sistema, sobre seus usuários e sobre as informações geradas. Devido ao grande volume de dados requerido, a realização de uma simulação que ofereça dados para tal demonstração é uma alternativa viável, por adequar-se aos requisitos de tempo e custo e, também, por permitir completo controle sobre os dados gerados (GOLBECK & HENDLER, 2006). É necessário, entretanto, salientar que o processo de simulação precisa seguir uma metodologia rigorosa e respeitar um conjunto determinado de regras para que sejam gerados dados o mais próximo possível dos obtidos por meio de usuários reais. A Figura 19 apresenta os passos executados para a realização da simulação, os quais serão comentados a seguir (no Anexo I deste documento encontra-se maiores detalhes sobre o processo de simulação).

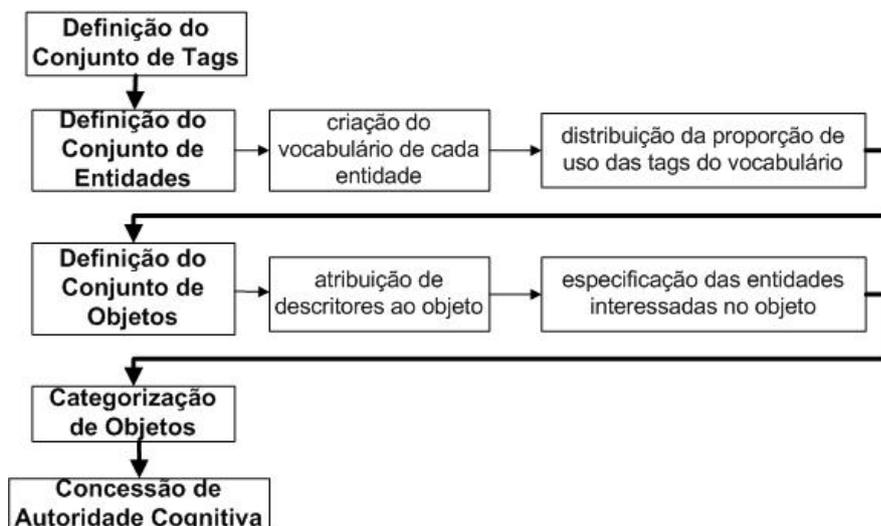


Figura 19 - Processo de simulação.

É interessante mencionar que, em um sistema baseado em folksonomia com usuários reais, as *tags* não são definidas previamente da forma como ocorre em nossa simulação. O vocabulário dos usuários e, conseqüentemente, do sistema de uma forma geral, é criado na medida em que as categorizações de objetos vão ocorrendo. Entretanto, como em uma simulação essa criação de vocabulários não ocorre por não haver usuários categorizando as informações gradativamente, é necessário definir previamente o conjunto de termos (*tags*) que serão utilizados para formar os vocabulários das entidades e para os demais processos (categorização de objetos, concessão de autoridade cognitiva e recuperação de informação).

5.1.1 A Definição dos Pivôs da Folksonomia

Nas três etapas iniciais do processo apresentado na Figura 19 são gerados os três pivôs principais de um sistema baseado em folksonomia: as *tags*, as entidades e os objetos. A primeira etapa consiste em definir o conjunto de *tags* que será utilizado durante todo o processo de simulação. Essas *tags* representam o vocabulário de termos que os usuários aplicam na categorização de objetos e que serão utilizados, também, para a concessão de autoridade cognitiva. Foram selecionadas 125 *tags* utilizadas na ferramenta *Delicious* (Delicious, 2007) e 125 *tags* representando áreas de conhecimento definidas pela CAPES e pelo CNPq (CAPES\CNPq, 2007).

A segunda etapa define o conjunto de entidades do sistema. Essas entidades representam os usuários de um sistema baseado em folksonomia, e estão vinculadas aos processos de categorização de objetos, concessão de autoridade e recuperação de informação — processos que os usuários reais executariam no sistema CAW. Para a simulação, uma entidade será representada por um descritor único que permita identificá-la, não sendo necessários maiores detalhes. O conjunto definido para a simulação possui uma população de 100 entidades, na qual cada uma possui um vocabulário composto por 20 termos aleatoriamente selecionados do conjunto de *tags* definido anteriormente — Oliveira *et al.* (2005) demonstram que uma amostra aleatória de 100 elementos apresenta um nível de confiança de 95% e uma taxa de erro máxima de 10% para uma população cujo tamanho não pode ser determinado.

Alguns experimentos e análises já realizados sobre a utilização de *tags* em sistemas baseados em folksonomia (SHAW, 2005) (KOME, 2005) (HALPIN, et al., 2007) demonstraram que existem diferenças nas proporções em que os termos são utilizados pelos usuários para a categorização de objetos, formando um *long tail* de termos utilizados esporadicamente. Essa proporção diferenciada é gerada no vocabulário das entidades por meio da aplicação de uma variação do Princípio de Pareto (NEWMAN, 2006) que simula esse *long tail*. O Princípio de Pareto, também conhecido como a regra dos 80-20 (NEWMAN, 2006), determina que 80% das causas são responsáveis por 20% dos efeitos, enquanto que os 20% das causas restantes são responsáveis por 80% dos efeitos observáveis. A variação utilizada em nosso trabalho corresponde à uma regra de 75-25, na qual 75% das *tags* são utilizadas em 25% das categorizações de objetos, enquanto que 25% das *tags* são empregadas nos outros 75% das categorizações. O Anexo II deste documento apresenta mais detalhes sobre o Princípio de Pareto e sobre a variação utilizada em nossa simulação.

A terceira etapa abrange a criação do conjunto de objetos. Esses objetos representam as informações que podem ser categorizadas pelas entidades para posterior recuperação. O conjunto é composto por 1000 objetos distintos, cada um possuindo 10 termos selecionados aleatoriamente do conjunto de *tags* para serem utilizados como descritores (esses descritores exercerão influência na etapa de categorização de objetos). Uma quantidade entre 0 e 20 entidades é relacionada ao objeto para representar entidades às quais objeto seria indicado na recuperação de informação. Essa quantidade foi definida aleatoriamente.

5.1.2 A Categorização de Objetos

A etapa de categorização de objetos é executada para cada uma das 100 entidades previamente definidas. O procedimento consiste em gerar um relacionamento modelado pela tupla {entidade, *tags*, objeto} representando o processo no qual os usuários de um sistema baseado em folksonomia categorizam um objeto atribuindo *tags* a ele. Na simulação, em média as entidades possuem 95 objetos²² categorizados com uma quantidade aleatória de *tags* — mínimo 2 e máximo 5 —, que de acordo com os experimentos de Gouvêa *et al.* (2007) corresponde à quantidade de *tags* que os usuários, normalmente, utilizam na categorização de notícias. A seleção dos objetos também ocorre de forma aleatória dentre os 1000 objetos gerados.

Os usuários de um sistema baseado em folksonomia, normalmente, atribuem *tags* que possuem alguma relação com o objeto desejado, descrevendo-o ou organizando-o. Os objetos utilizados na simulação possuem um conjunto de *tags* relacionadas aos mesmos que correspondem a termos descritores. Deste modo, ao construir a tupla {entidade, *tags*, objeto}, prioriza-se os termos presentes tanto no vocabulário da entidade envolvida no processo de categorização quanto nos descritores do objeto. No caso de não existir termos comuns em

²² A quantidade de objetos para a categorização por autoridade foi definida experimentalmente como 100 objetos. No momento da construção dos relacionamentos de categorização de objetos, a seleção do objeto ocorre aleatoriamente e descarta os relacionamentos repetidos entre entidade e objeto. Devido a isto, a média obtida foi de 95 objetos categorizados por entidade.

ambos os conjuntos em quantidade suficiente para a categorização, seleciona-se a quantidade restante aleatoriamente do vocabulário da entidade. A Figura 20 ilustra um exemplo no qual são selecionadas 3 *tags* para categorizar um objeto. A “*tag 1*” é priorizada nessa seleção por estar contida tanto no vocabulário da entidade quanto nos termos descritores do objeto. Como não existem mais *tags* comuns ao objeto e à entidade as outras duas *tags* restantes são selecionadas do vocabulário de *tags* que a entidade utiliza para efetuar categorizações.

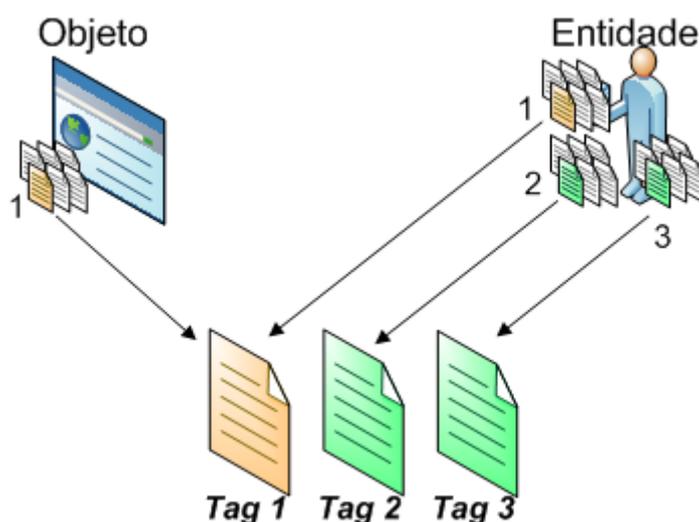


Figura 20 - Seleção de termos para a categorização de objetos.

Na categorização de objetos por usuários reais percebe-se outro comportamento que deve ser respeitado na simulação: a diferença na quantidade em que as *tags* são utilizadas. Como já mencionado anteriormente, uma pequena quantidade de *tags* é empregada em um grande número de categorizações, enquanto que uma grande quantidade de *tags* é aplicada poucas vezes para categorizar objetos. Durante a criação das entidades e de seus vocabulários a desigualdade na utilização das *tags* está prevista por uma distribuição que induz a essa diferenciação (a variação 75-25 detalhada no Anexo II). Deste modo, a seleção de termos do vocabulário da entidade para a categorização de objetos respeita uma proporção pré-definida, evitando que a seleção aleatória utilize todas as *tags* em quantidades iguais ou próximas gerando uma distribuição linear.

Tabela 2 - Frequência de utilização das *tags*.

Quantidade de <i>Tags</i>	Frequência de Utilização
1	33
1	32
1	25
1	24
2	23
1	20
1	18
1	16
1	15
3	14
2	13
1	12
2	11
1	10
1	9
3	8
5	7
4	6
7	5
9	4
15	3
31	2
114	1
Total	209
	725

A Tabela 2 demonstra a utilização de *tags* para a categorização de *bookmarks* por um usuário real na ferramenta *Delicious* (YAHOO, 2003) até a data de 20 de Março de 2008. A primeira coluna demonstra a quantidade “X” de *tags*, enquanto que a segunda coluna demonstra a frequência “N” de utilização dessas *tags* para a categorização de *bookmarks* pelo usuário. Os dados demonstram que o usuário possui 1 *tag* utilizada 33 vezes na categorização de *bookmarks* (ver os registros destacados na tabela), 5 *tags* utilizadas 7 vezes, e 114 *tags* aplicadas uma única vez. O gráfico da Figura 21 demonstra o *long tail* formado pela frequência diferenciada da utilização das *tags*.

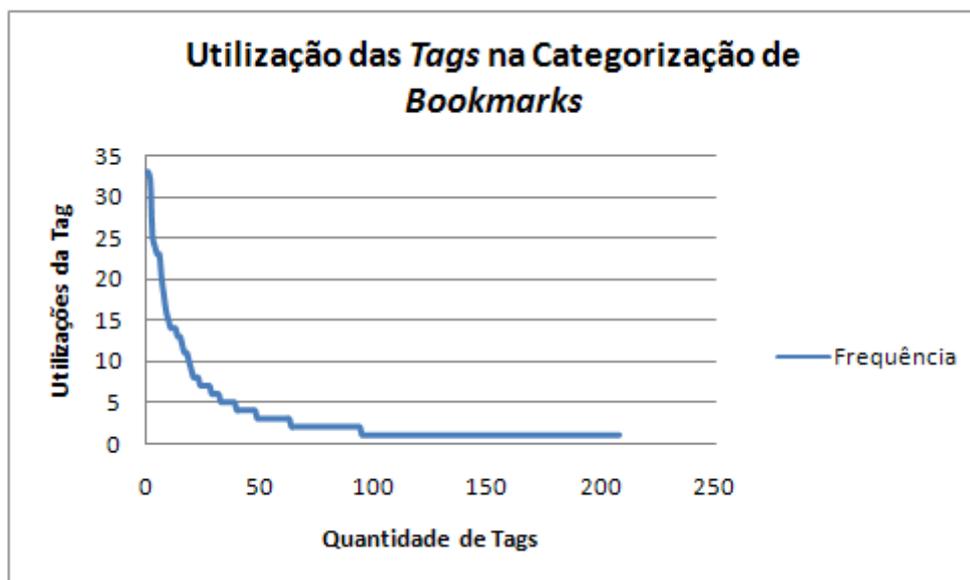


Figura 21 - *Long tail* formado pela frequência da utilização das *tags*.

5.1.3 A Concessão de Autoridade Cognitiva

A concessão de autoridade cognitiva entre as entidades é um procedimento diferenciado da categorização de objetos. Conceder autoridade para uma entidade equivale a reconhecer as áreas, assuntos ou mesmo termos nos quais a entidade é considerada referência. No caso de um sistema baseado em folksonomia no qual a concessão de autoridade será aplicada na recuperação de informação pode-se considerar que:

- i) A entidade que concede a autoridade tem um conhecimento prévio, mesmo que superficial, das competências, habilidades ou dos interesses da entidade categorizada. Toda entidade possui seu vocabulário de termos utilizados para categorização de objetos e, nesse vocabulário, as *tags* mais utilizadas representam as áreas de maior interesse ou de maior domínio da entidade — ou pelo menos, os assuntos mais utilizados ou mais vistos pela mesma. Deste modo, as *tags* utilizadas para a concessão de autoridade são as *tags* que fazem parte do vocabulário da entidade que “recebe” a autoridade cognitiva, uma vez que não existe sentido em conceder autoridade em um assunto “desconhecido” da própria entidade.

- ii) Como o intuito é utilizar a concessão de autoridade cognitiva no processo de recuperação de informação, não convém conceder autoridade a uma entidade numa área em que a mesma não possui, ou possui poucos, objetos categorizados. Os termos contidos no *long tail* do vocabulário das entidades não são considerados na concessão de autoridade às mesmas, porque esses termos, isoladamente, não representam e não informam muito sobre a autoridade, suas competências ou suas áreas de interesse.
- iii) Os níveis de autoridade são utilizados para distinguir o peso de uma autoridade e para servir como parâmetro na ordenação dos resultados obtidos em um processo de recuperação de informação. Esses pesos são considerados na simulação do processo de recuperação de informação para: a) distinguir a importância das diferentes áreas nas quais uma entidade é considerada autoridade; e, b) distinguir a importância de duas ou mais entidades, com relação à autoridade possuída pelas mesmas em determinado assunto, área ou tópico.
- iv) Na simulação não são considerados aspectos sociais como: autoridade consentida, privacidade, falsas autoridades, etc. Esses aspectos são fundamentais em um sistema com usuários reais, mas não exercem influência nos resultados com entidades simuladas nos quais não se pretende demonstrar os riscos\problemas que podem ser originados no processo de concessão de autoridade cognitiva.

A etapa de concessão de autoridade, assim como a categorização de objetos, também é realizada para cada uma das 100 entidades existentes. Essa etapa consiste em um procedimento no qual cada entidade concede autoridade à uma quantidade aleatória de entidades (entre 1 e 10). A autoridade é atribuída por meio de 5 *tags* selecionadas aleatoriamente do vocabulário da entidade que recebe a autoridade (excluindo-se os termos situados no *long tail*) — cada *tag* é associada a um valor aleatório entre 1 e 5 que corresponde ao seu peso, representando a distinção de níveis de autoridade. O relacionamento pode ser

representando pela tupla {"entidade 1", tags, "entidade 2"}, indicando que a "entidade 1" concedeu autoridade à "entidade 2" com o conjunto de termos "Y" nos diferentes pesos de "1" a "5". A Figura 22 ilustra um exemplo no qual uma entidade "X" concede autoridade a 3 entidades quaisquer.

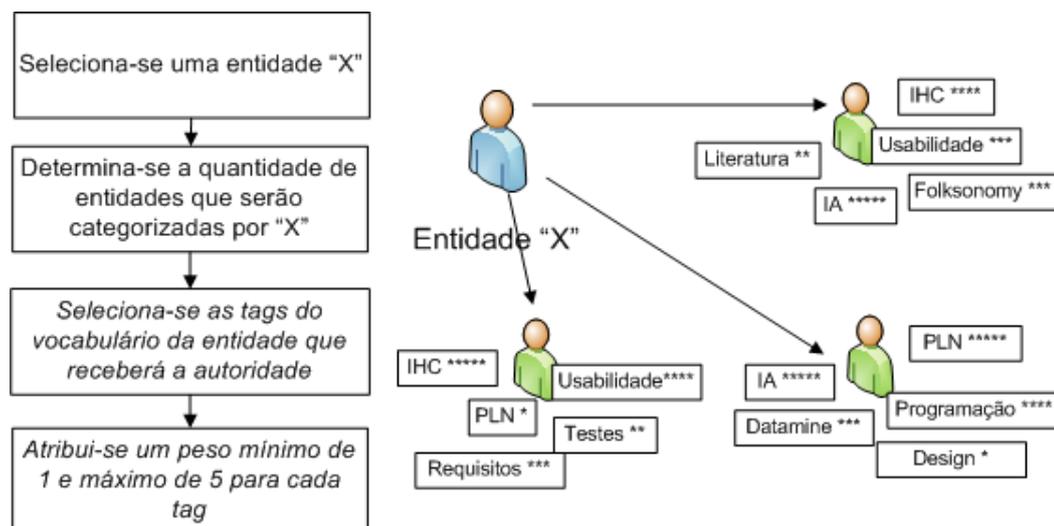


Figura 22 - Simulação do processo de concessão de autoridade cognitiva

Esta etapa finaliza a geração de dados para análise. As categorizações de objetos e as concessões de autoridade estão estabelecidas permitindo a simulação do processo de recuperação de informação.

5.2 A Recuperação de Informação

Os sistemas baseados em folksonomia apresentam os mesmos problemas de sobrecarga de informação e a mesma demanda por solução que os buscadores tradicionais. Essa sobrecarga representa a dificuldade que os usuários possuem em encontrar os resultados (objetos) desejados — objetos relevantes e cuja qualidade possa ser justificada, ou fundamentada, de algum modo. A hipótese defendida neste trabalho é a de que a priorização das informações categorizadas por autoridades eleva a precisão e a qualidade das informações recuperadas e reduz as dificuldades ocasionadas pela sobrecarga de informação.

5.2.1 Objetivos da Análise

A hipótese defendida em nosso trabalho será confirmada caso a priorização das categorizações realizadas pelas autoridades melhore o posicionamento dos objetos relevantes ao se realizar uma recuperação de informação, melhorando assim a precisão da recuperação.

Considerando as categorizações de objetos realizadas pelas entidades na simulação, ao se realizar um processo de recuperação de informação sobre essas categorizações serão retornados todos os objetos categorizados com o termo especificado na busca. A ordem de exibição desses objetos respeitará a ordem cronológica inversa de sua categorização, apresentando os objetos categorizados mais recentemente primeiro e os mais antigos por último, conforme adotado no sistema *Delicious*. A proposta do nosso trabalho consiste em priorizar os objetos que foram categorizados por autoridades no assunto de forma que, quanto maior a quantidade de autoridades que os categorizaram e quanto maior o peso dessas autoridades, melhor posicionado será o objeto.

Para a análise, realizamos os dois tipos de recuperação (priorizando e não priorizando as categorizações efetuadas pelas autoridades), verificamos o posicionamento dos objetos e, então, comparamos a variação das suas posições. Como podem existir vários termos que recuperem um mesmo objeto, existirão termos mais (ou menos) adequados para recuperá-los. A análise considera a melhor posição em que os objetos aparecem ao serem recuperados, independentemente do termo utilizado.

É importante salientar que, na recuperação convencional em que se considera a ordem cronológica inversa das categorizações, quanto mais recente for a categorização do objeto com a *tag* utilizada na recuperação, melhor será o posicionamento do objeto. A simulação realizada corresponde à esse comportamento respeitando a ordem em que foram geradas as categorizações dos objetos pelas entidades — os objetos categorizados por último aparecem na frente dos objetos categorizados no início do processo.

5.2.2 Análise sobre o posicionamento dos objetos

Para realizar a análise do posicionamento dos objetos na recuperação de informação é necessário que uma seqüência de procedimentos seja executada. Os procedimentos estão ilustrados pela Figura 23 e brevemente descritos a seguir.

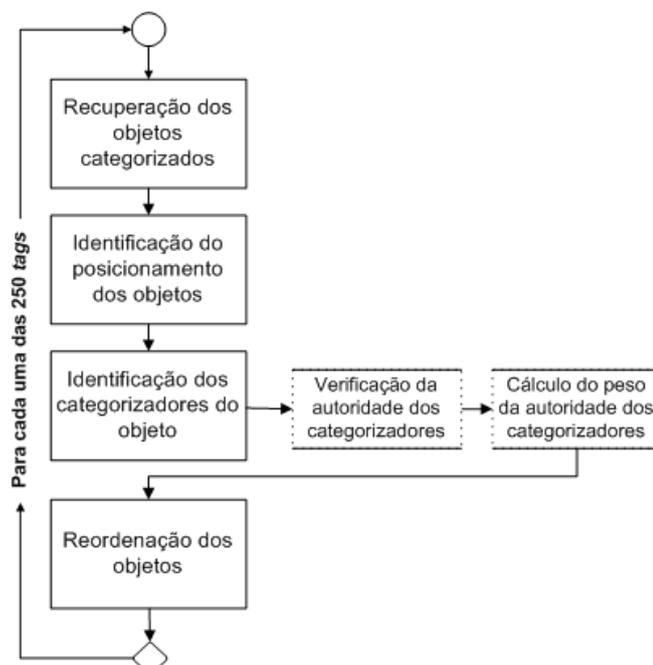


Figura 23 - Processo de recuperação de objetos

Para cada uma das 250 *tags* utilizadas na simulação:

- i) Recupera-se os objetos categorizados com a *tag* (processo de recuperação convencional em que os objetos são ordenados cronologicamente);
- ii) Em cada recuperação, verifica-se a posição em que os objetos recuperados aparecem;
- iii) Identifica-se quem foram os categorizadores de cada objeto:
 - a. Verifica-se quais desses categorizadores são autoridades no assunto;
 - b. Calcula-se o peso da autoridade de cada categorizador;
- iv) Reordena-se os objetos, decrescentemente, de acordo com a quantidade de autoridades que os categorizaram e com o respectivo peso dessas autoridades. Deste modo, quanto maior o número de autoridades que categorizaram o objeto e quanto maior o peso dessas autoridades, melhor será a posição que o objeto ocupará após a reordenação dos resultados.

A recuperação dos objetos utilizando todas as *tags* demonstra as prováveis posições em que os objetos podem aparecer ao se realizar uma busca por cada uma das *tags*. Como toda entidade possui objetos de interesse, é possível identificar quais *tags* melhor recuperam esses objetos. Levando em conta que todos os 1000 objetos foram categorizados e, portanto, podem ser recuperados, haverá pelo menos 1 *tag* que recupera cada objeto de interesse de cada entidade. Em nossa simulação, a recuperação de todos os objetos categorizados pelas 100 entidades com as 250 *tags* possíveis correspondem à um total de 256.580 registros que compõem a coleção de dados para análise de posicionamento dos objetos.

5.2.3 Análises e Resultados

Os resultados obtidos na simulação de recuperação de todos os objetos por todas as *tags* servem como base para identificar, para cada um dos 1000 objetos, quais as *tags* que melhor recuperam um objeto no processo convencional e no processo em que se priorizam as categorizações realizadas pelas autoridades.

A Figura 24 ilustra o posicionamento do “Objeto 993”. A melhor posição na qual esse objeto aparece na recuperação normal é na 15ª posição quando recuperado pela *tag* “learning” — e na 13ª posição utilizando a mesma *tag* e priorizando as categorizações realizadas pelas autoridades. Esse mesmo objeto aparece na 128ª pela recuperação convencional quando recuperado pela *tag* “books” — porém, na 1ª posição pela recuperação em que se priorizam as categorizações efetuadas pelas autoridades ainda considerando o termo “books”. Nesse experimento, o foco não está na *tag* utilizada, mas nas posições em que os objetos aparecem quando se realiza uma recuperação utilizando a *tag* como parâmetro.

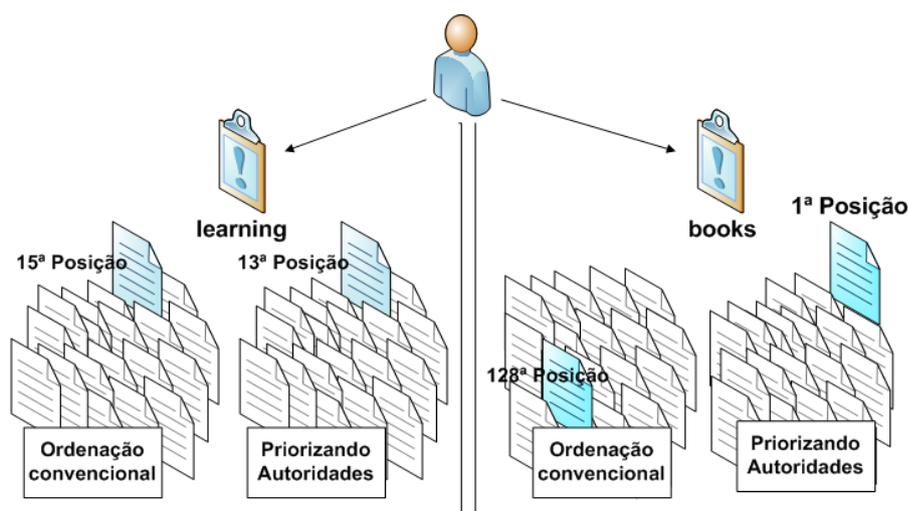


Figura 24 - Posicionamento dos objetos na recuperação de informação.

Recuperação Convencional

Para a recuperação de informação convencional em que não se priorizam as categorizações efetuadas pelas autoridades, identificou-se as melhores posições em que cada um dos 1000 objetos pôde ser encontrado. A Tabela 3 apresenta a síntese dos resultados obtidos exibindo a quantidade de objetos encontrados em 6 faixas de posições distintas. Ela apresenta também o posicionamento desses objetos ao realizar a recuperação pelo mesmo termo, porém, priorizando as categorizações que as autoridades efetuaram.

Tabela 3 - Posicionamento dos Objetos na Recuperação Convencional.

Posição	Convencional	(%)	Autoridade	(%)
Até a 10 ^a	251	25,1%	244	24,4%
Da 11 ^a à 20 ^a	247	24,7%	235	23,5%
Da 21 ^a à 30 ^a	182	18,2%	170	17,0%
Da 31 ^a à 40 ^a	151	15,1%	144	14,4%
Da 41 ^a à 50 ^a	88	8,8%	86	8,6%
Da 51 ^a em diante	81	8,1%	119	11,9%

É possível observar na Tabela 3 que aproximadamente 25% dos objetos possuem pelo menos um termo que os posicione entre os 10 primeiros resultados na recuperação de informação e aproximadamente 50% entre os 20 primeiros resultados. Por outro lado, ao priorizar as categorizações efetuadas por autoridades o percentual de objetos encontrados

entre os 10 e 20 primeiros resultados reduz-se para aproximadamente 24 e 48% respectivamente. A princípio, esses resultados aparentam mostrar que a priorização das categorizações realizadas pelas autoridades piora o processo de recuperação de informação. No entanto, essa redução não é significativa, ocorrendo porque muitos dos termos que melhor posicionam os objetos na recuperação normal são pouco utilizados (ou não o são) por autoridades no assunto, e o experimento considera a melhor posição do objeto na recuperação e não a melhor forma para recuperá-lo.

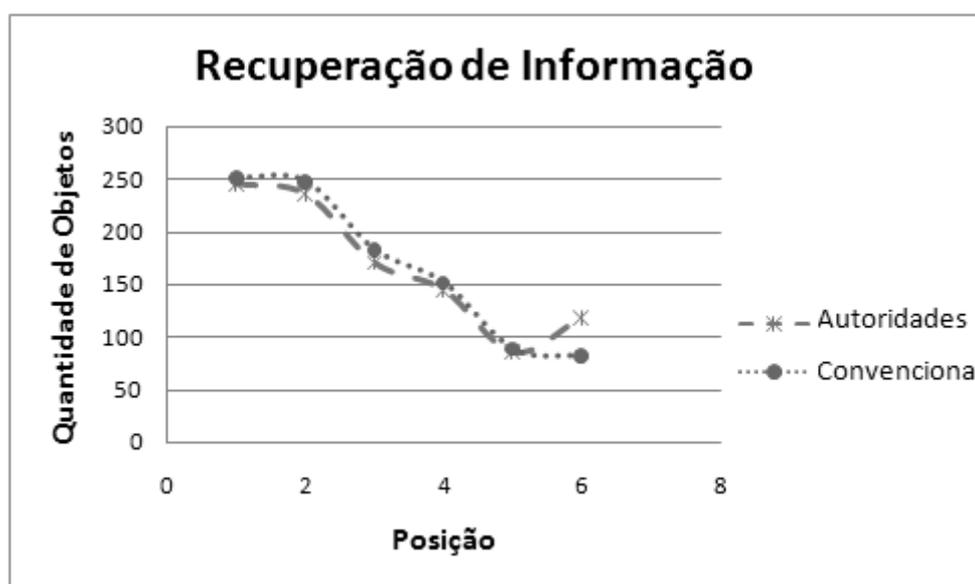


Figura 25 - Posicionamento dos objetos na recuperação convencional.

Por outro lado, o gráfico da Figura 25 procura ilustrar que não existe uma diferença significativa nos resultados obtidos com os termos que melhor posicionam os objetos na recuperação convencional quando esses mesmos termos são aplicados na recuperação que prioriza as categorizações realizadas por autoridades. Podemos observar a existência de uma maior concentração de objetos somente após os 50 primeiros resultados priorizando-se as autoridades. No entanto, mesmo com essa concentração diferenciada, é possível afirmar que não existe variação significativa para a maioria dos objetos e, conseqüentemente, que a priorização das autoridades não prejudica o processo de recuperação de informação ao utilizar

tags que não figurem entre as mais representativas ou mais utilizadas nos processos de categorização.

Recuperação Priorizando as Autoridades

A análise da recuperação de informação priorizando as categorizações realizadas pelas autoridades segue a mesma lógica da análise realizada para a recuperação convencional. Deste modo, identificou-se a melhor posição em que cada um dos 1000 objetos pode aparecer (a posição ocupada pelos objetos na recuperação convencional também foi identificada). Os resultados obtidos são apresentados na tabela abaixo:

Tabela 4 - Posicionamento dos objetos na recuperação priorizando autoridades.

Posição	Convencional	(%)	Autoridade	(%)
Até a 10 ^a	125	12,5%	890	89,0%
Da 11 ^a à 20 ^a	91	9,1%	94	9,4%
Da 21 ^a à 30 ^a	94	9,4%	14	1,4%
Da 31 ^a à 40 ^a	88	8,8%	1	0,1%
Da 41 ^a à 50 ^a	100	10,0%	1	0,1%
Da 51 ^a em diante	502	50,2%	0	0,0%

A Tabela 4 permite verificar que 89% dos objetos podem ser apresentados entre os 10 primeiros resultados, elevando-se para 98% ao considerar os 20 primeiros resultados. Isto equivale a dizer que apenas 2% de todos os objetos não possuem nenhuma *tag* que os posicione entre os 20 primeiros resultados. Neste caso, ao considerar as posições em que estes objetos são apresentados na recuperação convencional, percebe-se uma significativa diferença de posicionamento, sendo somente 21,6% dos objetos recuperados entre as 20 primeiras posições com as *tags* utilizadas pelas autoridades.

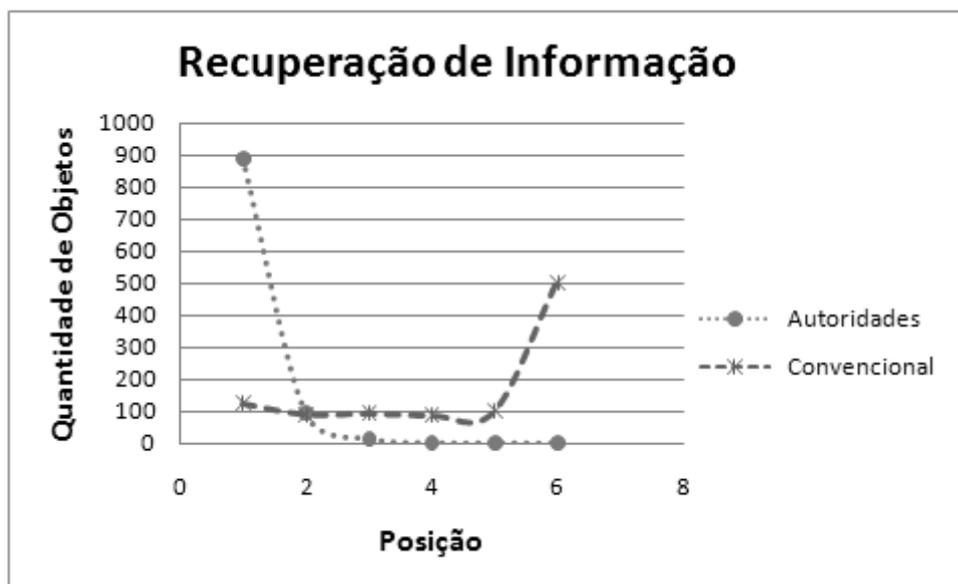


Figura 26 - Posicionamento dos objetos priorizando as autoridades (gráfico com linhas).

O gráfico da Figura 26, semelhante ao gráfico da Figura 25, permite observar mais facilmente a forma como os objetos são apresentados na recuperação que prioriza as categorizações realizadas por autoridades e, também, na recuperação convencional. O gráfico demonstra dois *long tails*: ao priorizar as autoridades a maioria dos objetos (98%) é apresentada nos 20 primeiros resultados (formando um *long tail* a partir do 21º resultado); na recuperação convencional pelos mesmos termos é formado um *long tail* invertido no qual a maior concentração de objetos (50%) aparece após os 50 primeiros resultados.

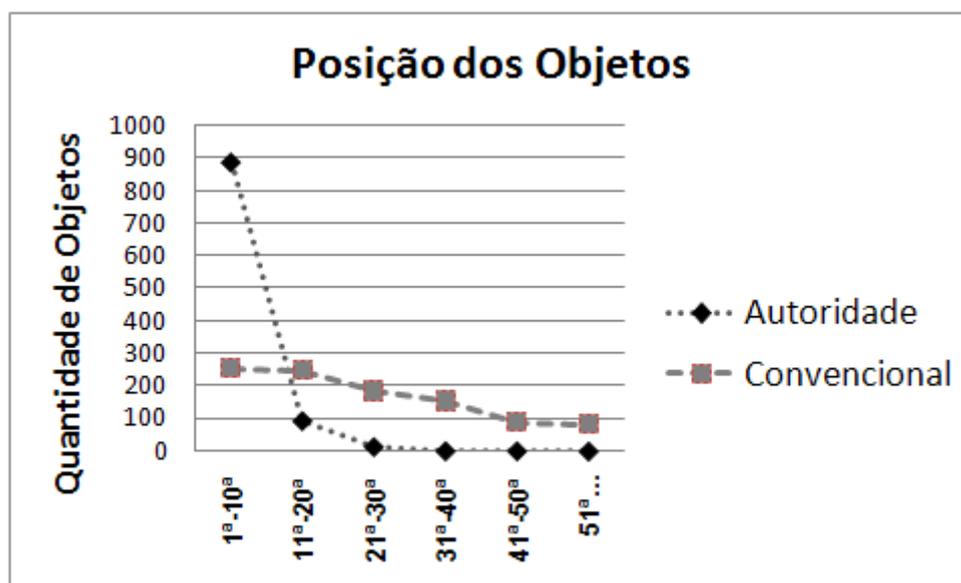


Figura 27 - Recuperação normal Vs. recuperação considerando autoridades.

O gráfico da Figura 27 compara os melhores resultados do processo de recuperação convencional com os melhores resultados do processo que prioriza as categorizações das autoridades. Podemos observar um expressivo número de objetos (89%) que possui pelo menos 1 termo que os inclua entre os 10 primeiros resultados de uma busca quando a recuperação prioriza as categorizações efetuadas pelas autoridades, enquanto que um número bem menor (25%) ocorre na recuperação convencional. Podemos atribuir essa diferença de posicionamento, principalmente, ao fato de que o que possibilita uma alteração da posição de um objeto são as *tags* utilizadas para categorizá-lo, por exemplo: um objeto categorizado com a *tag* “*hci*” aparecerá entre os últimos resultados caso a categorização seja antiga e a *tag* seja muito utilizada. No entanto, se as entidades que categorizaram esse objeto forem autoridades na *tag* “*hci*”, o objeto será priorizado no processo de recuperação e apresentado antes de objetos categorizados mais recentemente (mas não por autoridades). Essa priorização permite que objetos categorizados com um conjunto de *tags* muito utilizadas e, portanto, com uma grande quantidade de resultados, sejam exibidos entre os primeiros resultados dependendo de quem foram os responsáveis por sua categorização. Existe, desse modo, uma maior possibilidade de alteração e de melhora no posicionamento dos resultados na recuperação de informação do que existe no processo de recuperação convencional.

A comparação dos resultados obtidos nas duas análises demonstra que a priorização das informações categorizadas pelas autoridades pode trazer consideráveis benefícios no processo de recuperação de informação. A iProspect (2007) afirma em sua pesquisa que a maioria dos usuários considera apenas os primeiros resultados em um processo de busca e, deste modo, não encontra muitos dos resultados que lhes seriam interessantes pelo fato de os mesmos não figurarem entre os resultados principais. De acordo com os dados obtidos nas simulações, informações que seriam apresentadas nas últimas posições podem ser priorizadas e apresentadas entre os primeiros resultados ao se utilizar um termo adequado para a

recuperação. Como a maioria dos objetos possui pelo menos 1 termo suficientemente adequado para posicioná-lo entre os 10 primeiros resultados, esses objetos possuem maiores chances de serem encontrados pelos usuários.

Além disso, a análise dos gráficos apresentados demonstra também que essa priorização das categorizações efetuadas pelas autoridades não possui a tendência de prejudicar os resultados quando termos menos significativos são utilizados como parâmetro para a recuperação de informação. Desta forma, podemos ver que os benefícios da aplicação do conceito de autoridade cognitiva podem ser obtidos em um sistema baseado em folksonomia, com usuários reais, sem comprometer seus resultados na recuperação de informação convencional.

Os experimentos realizados e apresentados nessa seção visaram demonstrar a viabilidade da aplicação do conceito de autoridade cognitiva em um sistema baseado em folksonomia, cuja aplicação é motivada pela necessidade de solução de dois problemas de grande importância atuais, a dizer: a sobrecarga de informação e a ausência de um controle de qualidade das informações categorizadas.

A priorização das categorizações efetuadas por autoridades auxilia na redução dos problemas ocasionados pela sobrecarga ao aumentar a possibilidade dos objetos figurarem entre os primeiros resultados em um processo de recuperação de informação. Por sua vez, existe uma forte relação entre a autoridade cognitiva e a qualidade da informação recuperada (já discutidas no capítulo III). Assim, seguindo as argumentações de que uma informação categorizada por autoridades tende a ser informação de qualidade e a possuir maior relevância para aqueles que concedem ou que concordam com a autoridade dos categorizadores (WILSON, 1983) (RIEH, 2005), podemos afirmar que priorizar essas categorizações possibilita recuperar informações relevantes e de qualidade apresentando-as em posições mais favoráveis. Desta forma, isso indica a consistência da hipótese de que aplicar o conceito de

autoridade cognitiva apresenta informações mais precisas e de maior qualidade, e reduz as dificuldades da sobrecarga de informação em sistemas baseados em folksonomia.

5.3 Cadeia de Autoridades

A simulação do processo de concessão de autoridade possibilita, também, a realização de uma série de outras análises sobre a rede social formada pela cadeia de autoridades (uma cadeia gerada pela atribuição de autoridade às entidades) e sobre outras informações que podem ser obtidas desta cadeia, tais como: busca por autoridades, reconhecimento de competências, formação de grupos, entre outros.

A simulação do processo de concessão de autoridade cognitiva constrói as redes de autoridades das entidades envolvidas. As ligações entre essas redes de autoridades formam uma rede gigante, aqui denominada de **cadeia de autoridades**, em que não existe uma figura central ou dominante, e na qual as entidades encontram-se direta ou indiretamente conectadas umas às outras por meio do relacionamento de concessão de autoridade cognitiva.

A Figura 28 exibe a cadeia de autoridades obtida da ligação das redes de autoridades das 100 entidades utilizadas na simulação.

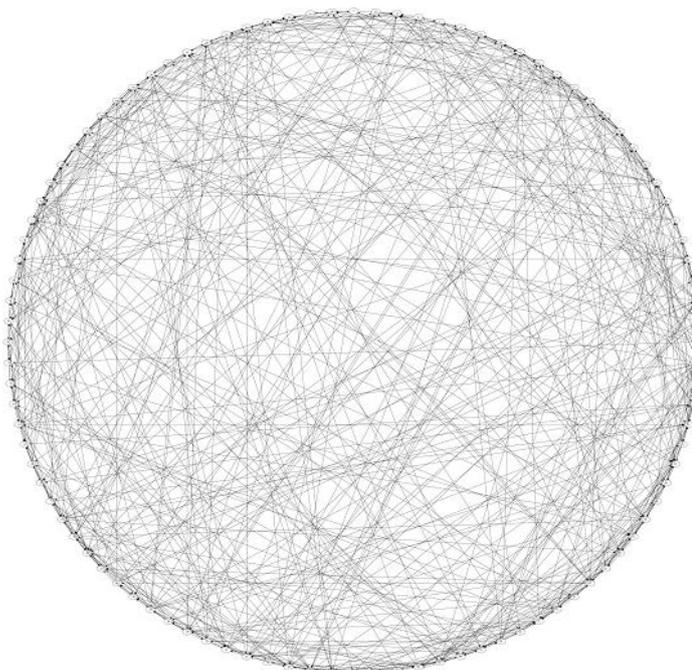


Figura 28 - Cadeia de autoridades formada pelas redes de autoridades das entidades.

Todas as entidades encontram-se interligadas por meio dos relacionamentos gerados no processo de concessão de autoridade cognitiva. Seguindo as definições da teoria dos grafos (CORMEN *et al.*, 2002) (FEOFILOFF *et al.*, 2007), podemos dizer que a cadeia de autoridades é um grafo no qual as entidades são vértices e os relacionamentos originados pelas concessões de autoridade são as arestas que os conectam, formando assim, um multigrafo orientado. Um grafo é denominado de multigrafo quando o mesmo possui arestas paralelas; e é considerado orientado se, e somente se, o relacionamento {vértice A, vértice B} for diferente do relacionamento {vértice B, vértice A}. Como a concessão de autoridade cognitiva da entidade “A” para a entidade “B” não implica em uma relação inversa na qual “B” também concede autoridade para “A” — mesmo sendo isto perfeitamente possível —, têm-se um grafo orientado cuja incidência do vértice “A” para o vértice “B” representa essa concessão de autoridade. A existência de um relacionamento mútuo entre “A” e “B” implica na existência de arestas paralelas entre os vértices “A” e “B” e, deste modo, caracteriza a cadeia como um multigrafo. Além disso, como a autoridade sempre é concedida à outra entidade, não existe a ocorrência de auto-concessões de autoridade pelas entidades e, conseqüentemente, a existência de laços. A Figura 29 demonstra a cadeia de autoridades na forma de um grafo percebido a partir da entidade 3.

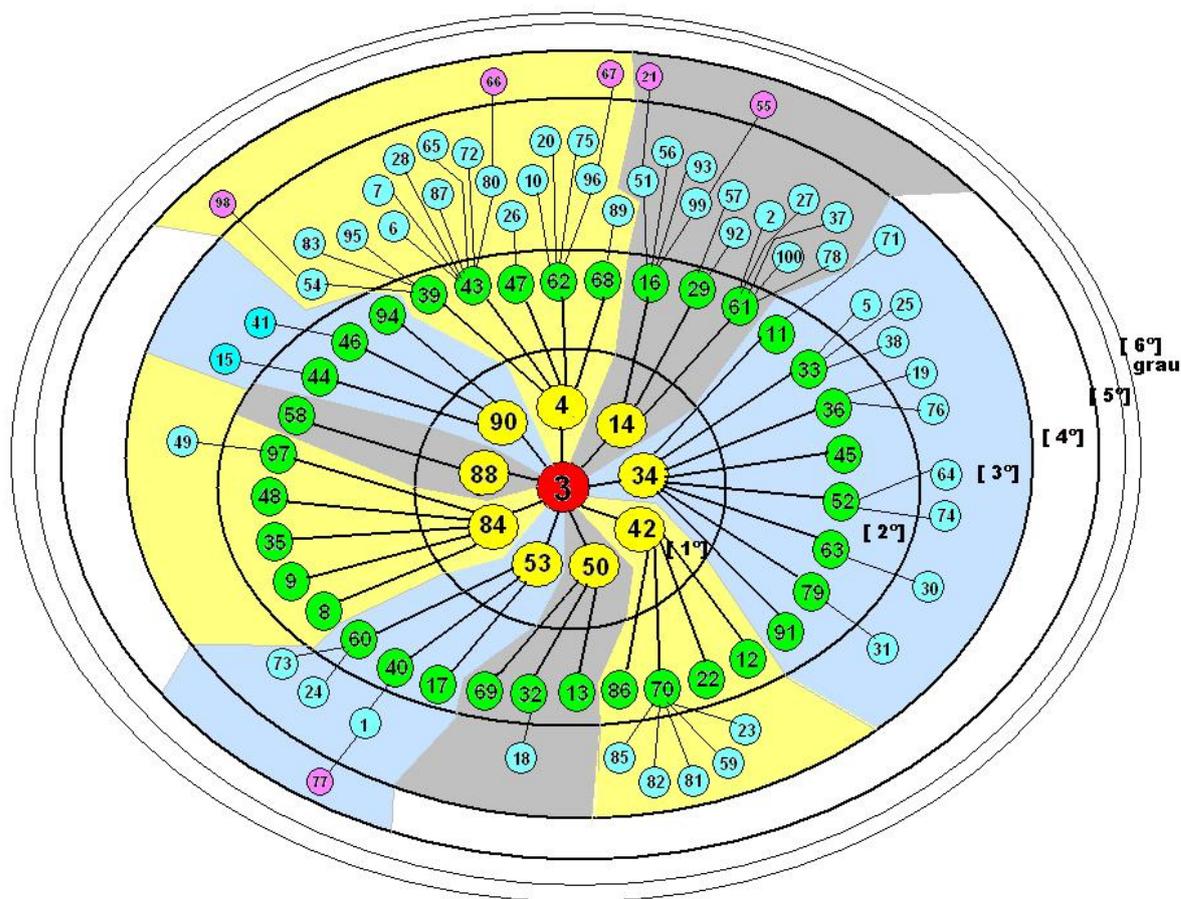


Figura 29 - Cadeia de autoridades – “entidade 3”.

A “entidade 3” aparece ao centro, rodeada pelas entidades às quais ela concedeu autoridade cognitiva — a sua rede de autoridades, ou o 1º grau de separação da cadeia (elipse central marcada com o rótulo [1]). O 2º grau de separação é representado pela elipse marcada com o rótulo [2] e corresponde às redes de autoridades das entidades situadas no 1º grau de separação. O 3º grau é marcado pelo rótulo [3] e, assim sucessivamente até o 6º grau de separação (observe que nos graus 5 e 6 não se encontra nenhuma entidade). O grafo exibe apenas o caminho mais próximo entre a “entidade 3” e as demais entidades com as quais ela encontra-se conectada. Isso quer dizer que, como ilustrado pela Figura 30, se houver mais de um caminho entre a “entidade 3” e a “entidade X”, a cadeia de autoridades demonstrará apenas o caminho mais curto entre as duas entidades.

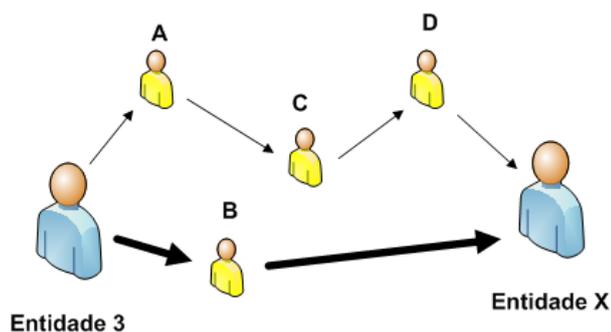


Figura 30 - Identificação do menor caminho entre duas entidades.

Além do menor caminho, o fato do relacionamento construído por meio da concessão de autoridade ser assimétrico cria uma distância diferente entre as entidades. Por exemplo: a “entidade 3” concede autoridade para a “entidade 88” (entidade 3 ► entidade 88); porém, a “entidade 88” concede autoridade para a “entidade 58” que, por sua vez, concede autoridade à “entidade 3” (88 ► 58 ► 3). Assim, o caminho para se chegar à “entidade 8” partindo da “entidade 3” é menor que o caminho oposto que parte da “entidade 88” para chegar à “entidade 3”. A Figura 31 demonstra essa diferença que existe na distância entre duas entidades dependendo da entidade escolhida como ponto de partida.

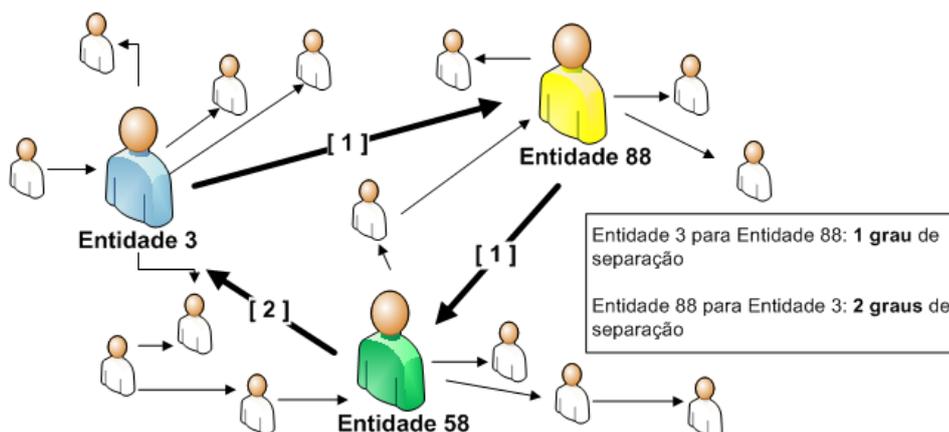


Figura 31 - Distância entre entidades.

As relações geradas entre as entidades possibilitam, também, que seja observada a forma como as autoridades e os objetos de interesse de uma entidade estão distribuídos ao longo da cadeia de autoridades. Na próxima seção expomos algumas verificações realizadas sobre a distribuição de objetos e de autoridades na cadeia.

5.3.1 Distribuição das autoridades na cadeia: graus de separação

Russell (2005) afirma que com uma quantidade suficiente de usuários e de *tags* quase todos os assuntos imagináveis possuirão uma entidade que figure como autoridade neles. Entretanto, a posição em que essa entidade aparece na cadeia pode afetar positiva ou negativamente a importância de sua autoridade. A teoria dos Seis Graus de Separação (BALANCIERI *et al.*, 2005) (YU *et al.*, 2003), amplamente estudada em redes sociais e utilizada por recursos como o Orkut (GOOGLE, 2005) e o Oráculo de Bacon (VIRGINIA, 2007), propõe que, normalmente, um indivíduo está conectado a qualquer outro em até um máximo de 6 graus de separação. Yu *et al.* (2003) citam 6 graus como a separação média entre duas entidades quaisquer.

Para demonstrar a distribuição apresentada pelas concessões de autoridade realizadas na simulação é necessário percorrer a cadeia de autoridades gerada a partir de cada uma das 100 entidades. Na verificação da cada cadeia, contabiliza-se a quantidade de entidades encontradas nos seis graus considerados — o gráfico da Figura 32 demonstra uma visão de como as autoridades estão distribuídas na presente simulação.

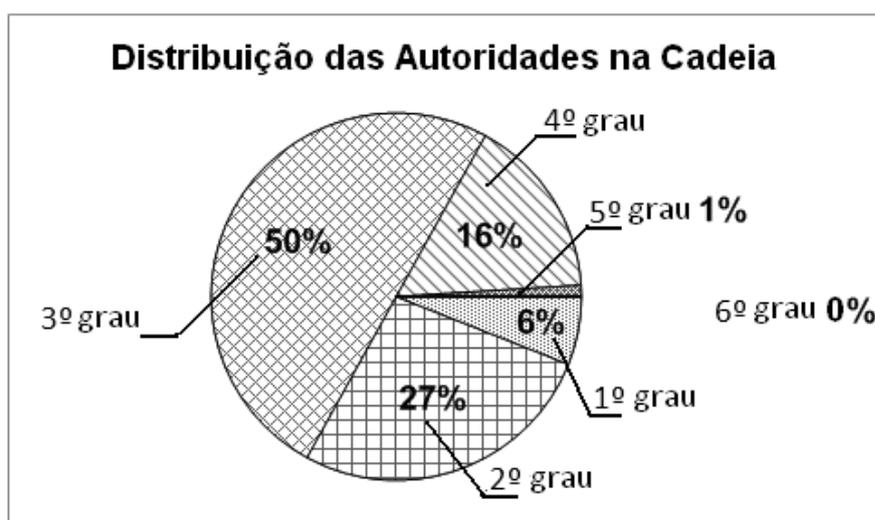


Figura 32 - Distribuição das autoridades por graus na cadeia.

É possível perceber, de acordo com o percentual do 1º grau de separação, que as entidades estão diretamente conectadas, em média, a 6 autoridades (rede de autoridades). O

gráfico da Figura 32 demonstra também que uma entidade qualquer está ligada à aproximadamente 83% das demais entidades até o 3º grau de separação, e que apenas 1% das entidades se encontra além do 4ª grau. Não há ocorrências de entidades cujo grau de separação seja superior ao 5º grau.

É possível afirmar, observando que 83% das entidades estão conectadas até o terceiro grau, que a maioria das autoridades possui uma ligação próxima com a entidade foco. Uma ligação próxima implica que: i) Existe uma grande quantidade de autoridades com as quais o usuário não manteria contato (mesmo estando próximo) e não teria acesso considerando apenas sua rede, mas que podem ser localizadas por meio da cadeia. ii) O usuário consegue mapear o caminho que o conecta à autoridade e, assim, verificar quem concedeu autoridade a quem. Este mapeamento oferece maiores condições para o julgamento da plausibilidade das autoridades encontradas — como citado no capítulo II, Wilson (1983) expõe o teste da plausibilidade como um dos principais critérios de julgamento da autoridade de uma entidade. E, iii) o vocabulário do usuário tende a ser compatível com o vocabulário da autoridade. Como a “distância” entre eles é considerada pequena, um termo utilizado por ambos possui maior probabilidade de expressar o mesmo significado. Exemplo: se a “entidade 1” concedeu autoridade no assunto “Y” para a “entidade 2”, que concedeu autoridade nesse mesmo assunto para a “entidade 3” que, por sua vez, concedeu para a “entidade 4”, é mais factível considerar que a “entidade 1” e a “entidade 4” estejam falando sobre o mesmo assunto, do que se estivéssemos considerando duas entidades quaisquer que não possuem nenhum tipo de ligação.

5.3.2 Distribuição dos objetos de interesse das entidades

Da mesma forma que é possível identificar a distribuição das autoridades entre os graus de separação na cadeia de autoridades, também é possível demonstrar como se apresenta a distribuição dos objetos de interesse das entidades. Os dados sobre a distribuição

dos objetos de interesse de uma entidade são obtidos ao construir sua cadeia de autoridades identificando em quais graus da cadeia encontram-se as entidades que categorizaram esses objetos de interesse. As 100 entidades e todos os seus objetos de interesse foram considerados na análise.

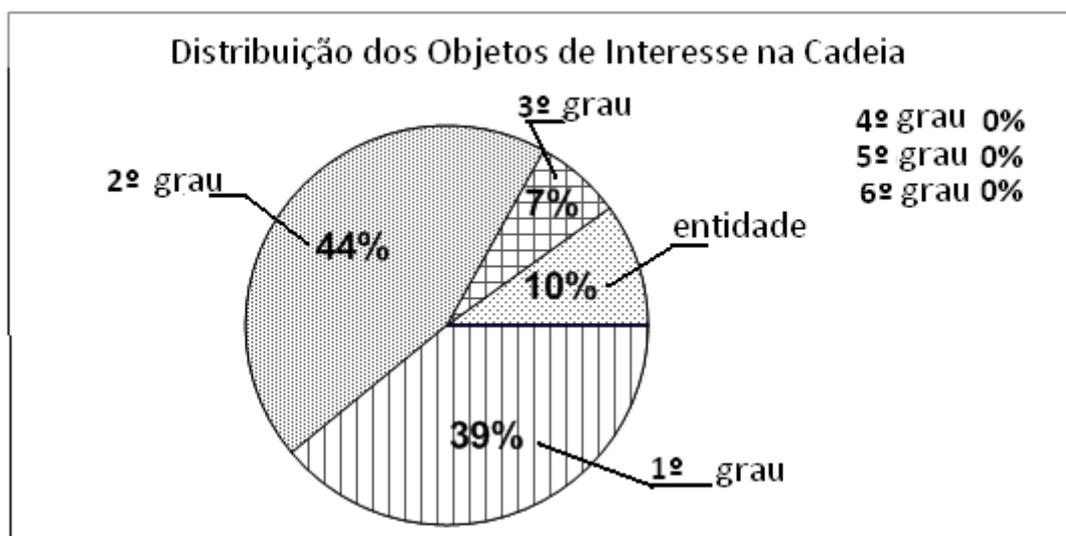


Figura 33 - Distribuição dos objetos de interesse das entidades na cadeia de autoridades.

O gráfico da Figura 33 demonstra que dentre os objetos categorizados pelas próprias entidades e por suas autoridades diretas estão concentrados 49% de seus objetos de interesse. Incluindo as categorizações realizadas por autoridades situadas até o 2º grau de separação esse percentual evolui para 93% do total de objetos de interesse das entidades. Apenas 7% dos objetos são encontrados categorizados por autoridades situadas no 3º grau. Esses dados demonstram que a maioria dos objetos de interesse das entidades encontra-se relativamente próxima às mesmas, estando categorizados ou por suas autoridades, ou pelas autoridades de suas autoridades. Eles também indicam que uma pequena parcela dos objetos (7%) é encontrada categorizada apenas por autoridades situadas além do 2º grau de separação.

Ao compararmos o gráfico da Figura 32 (distribuição das autoridades) com o gráfico da Figura 33 (distribuição dos objetos de interesse das entidades) percebe-se que, em média, 33% das autoridades e 83% dos objetos de interesse (desconta-se aqui os 10% categorizados

pela própria entidade) são encontrados até o 2º grau de separação. Isto é, 33% das autoridades categorizaram 83% dos objetos de interesse de uma entidade qualquer. Também é possível perceber que 50% das autoridades encontradas no 3º grau categorizaram apenas 7% dos objetos de interesse que não foram categorizados por nenhuma outra autoridade situada em um grau de separação menor. Essas comparações não indicam que as autoridades em graus mais afastados realizam “piores” categorizações, mas sim que as autoridades em graus mais próximos abrangem a maioria dos objetos de interesse da entidade e, deste modo, apresentam melhores contribuições na recuperação de informação.

5.3.3 Consultas: extração de informação na cadeia

Os experimentos realizados para a concessão de autoridade cognitiva possibilitam que sejam realizadas novas simulações com o intuito de obter informações complementares sobre a cadeia de autoridades, sobre as entidades e seus comportamentos. Nesta seção, são expostos e exemplificados alguns recursos desenvolvidos para a extração de informação sobre os dados gerados e que foram projetados para a aplicação em um sistema baseado em folksonomia com usuários reais.

1. **Procura global por autoridades:** um dos principais benefícios que o processo de concessão de autoridade cognitiva por meio de folksonomia pode trazer é a construção de uma rede social na qual os usuários estão interligados por relações que expressam e reconhecem suas competências. Russel (2005) expõe vários benefícios proporcionados pela cadeia de autoridades e pode-se dizer que um dos principais é a possibilidade de localizar autoridades em determinado tópico\assunto. Compreende-se que apenas a rede de autoridades de uma entidade não é grande o suficiente para ostentar autoridades numa grande gama de áreas. No entanto, ao considerar todas as concessões de autoridade realizadas, torna-se possível encontrar novas autoridades (e em novas áreas) com as quais a entidade não possui contato direto ou não tem

conhecimento sobre as competências. A vantagem da execução de uma busca independente da cadeia é que não ocorre o problema de “ilhas” ou de pequenas redes isoladas às quais não seja possível chegar por meio de uma ligação encadeada, uma vez que esse recurso apresenta todas as entidades que foram categorizadas com determinado termo.

Na simulação do processo de concessão de autoridade não houve ocorrência de formação de ilhas — esse evento é particularmente comum em sistemas com usuários esporádicos ou entre grupos pequenos e fechados.

2. **Procura por autoridades usando a cadeia de autoridades:** além da busca mencionada no item anterior, a busca por autoridades pode ocorrer de outras duas formas: **encadeada local** (na rede da entidade) ou **encadeada completa** (na cadeia de autoridades). Em comparação com a busca descrita no item anterior, a busca **encadeada** limita-se ao 6º grau de separação entre a entidade foco e as demais entidades interconectadas, porém, ela apresenta a vantagem de retornar entidades com uma ligação, mesmo que remota, com a entidade foco. Além disso, experimentos demonstram que seis graus são suficientes para abranger resultados relevantes (YU *et al.*, 2003). Essa ligação identificada permite uma melhor análise da autoridade, da sua reputação, dos seus méritos e até mesmo de questões relacionadas à confiança merecida pelas entidades recuperadas. O recurso desenvolvido para procurar por autoridades dentro da cadeia de autoridades permite que se realize uma busca considerando no mínimo o 1º (rede) e no máximo o 6º grau de separação, além de considerar o nível da autoridade recebida por essas entidades. Deste modo, pode-se procurar por autoridades no assunto X que estejam até o Nº grau de separação na cadeia e que possuam no mínimo o nível Y nesse assunto. Por exemplo, se a “entidade 27” procurar por autoridades em “internet” que estejam até o 5º grau de separação e

que possuam, no mínimo, peso 2 nesse assunto, ela receberá os resultados apresentados na Tabela 5.

Tabela 5 - Autoridades encontradas pela cadeia de autoridades.

Autoridade	Peso	Grau	Caminho	Quantidade
Entidade 94	3	3°	27 -> 95 -> 58 -> 94	4
Entidade 68	2	4°	27 -> 95 -> 35 -> 11 -> 68	1

A Tabela 5 exibe as duas entidades que satisfazem os critérios utilizados na busca: são localizadas até o 5° grau e receberam pesos 3 e 2 quando foram categorizadas com a tag “internet” — a coluna “quantidade” demonstra quantas vezes a entidade recebeu autoridade nessa tag.

3. **Reconhecimento das competências da entidade “X”:** as concessões de autoridade a uma entidade permitem que se observe todas as áreas nas quais a entidade é considerada autoridade (*tags* que a entidade recebeu). Identificar os assuntos nos quais determinada entidade é reconhecida como uma autoridade serve de base para que outras entidades que não possuam contato direto com a mesma a reconheçam como tal, ou pelo menos, percebam a opinião dos outros sobre ela e decidam se concordam ou não. O recurso implementado para o reconhecimento de competências permite saber quais termos foram atribuídos à determinada entidade, a quantidade de vezes que o termo foi utilizado e quem foram os concessores da autoridade. *Exemplo:* verificar em quais áreas a “entidade 27” é considerada uma autoridade (ver Tabela 6).

Tabela 6 - Autoridade recebida pela “entidade 27”.

Termo	Quantidade	Entidades Concessoras
Funny	6	28, 32, 50, 63, 65, 88
Geografia	6	28, 41, 50, 63, 65, 88
Fisiologia	6	28, 41, 50, 63, 65, 88
Cooking	7	28, 32, 41, 50, 63, 65, 88

Educação	7	28, 32, 41, 50, 63, 65, 88
Humanas	2	32,41
Pediatria	1	32

O peso correspondente à cada *tag* utilizada para a concessão de autoridade cognitiva não é exibido publicamente, apenas os concessores da autoridade podem visualizá-lo.

4. **Descoberta do menor caminho entre “A” e “B”:** como a cadeia de autoridades é resultante de várias redes interconectadas, uma entidade “A” pode estar conectada à entidade “B” por vários caminhos diferentes, alguns mais longos, outros mais curtos. Este recurso identifica a menor distância entre duas entidades na cadeia considerando um mínimo de 1 grau (quando a entidade “B” estiver na rede de autoridades da entidade “A”) e um máximo de 6 graus (quando houver outras cinco entidades entre “A” e “B”). Ex: “A” possui em sua rede “C” e “D”; “C” possui “B”, enquanto que “D” possui “F” e “F” também possui “B”. Assim, os caminhos até “B” podem ser: A ► C ► B (dois graus, menor caminho) ou A ► D ► F ► B (três graus, maior caminho).

Exemplo: o menor caminho entre as entidades 27 e 83 corresponde a uma separação de três graus (27 ► 51 ► 56 ► 83) — a entidade 27 concedeu autoridade para 51, que concedeu para 56 que por sua vez, concedeu autoridade para a entidade 83. Também é possível encontrar a entidade 83 no 6º grau de separação (27 ► 51 ► 19 ► 20 ► 87 ► 53 ► 83), porém, as ligações mais próximas prevalecem sobre as mais distantes.

5. **Identificação da rede de autoridades da entidade “X”:** ao simular o processo de concessão de autoridade, uma entidade “X” concede autoridade para outras entidades formando sua rede de autoridades. Esta rede corresponde à todas as entidades categorizadas pela entidade “X”, estando, portanto, diretamente ligadas à mesma. A identificação da rede de autoridades de uma entidade consiste em demonstrar todas as entidades às quais ela concedeu autoridade — independentemente dos termos e dos níveis utilizados (ver Figura 34).

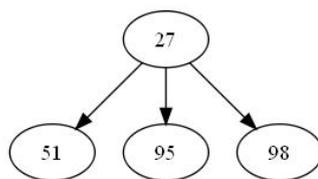


Figura 34 - Rede de autoridades da “entidade 27”.

Exemplo: A execução do recurso sobre a base de dados gerada para a simulação identifica que a “entidade 27” possui três entidades em sua rede de autoridades (51, 95, 98) e demonstra as *tags* utilizadas para a concessão de autoridade. O peso atribuído para cada *tag* é exibido apenas para a entidade que o concede.

6. **Identificação dos fãs da entidade “X”:** processo em sentido inverso ao da identificação da rede de autoridades. Neste recurso, identifica-se todas as entidades que concederam autoridade para “X”, independentemente dos termos e dos níveis utilizados. É importante lembrar que o fato da entidade “X” ser considerada autoridade coloca-a na rede de autoridades das entidades que a qualificaram como tal, mas que isso não implica no inverso, ou seja, essas entidades também estarem na rede de autoridades de “X”.

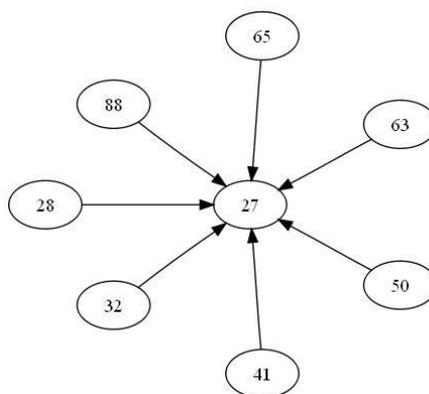


Figura 35 - Fãs da “entidade 27”.

Exemplo: A Figura 35 demonstra que a “entidade 27” é considerada autoridade em determinada(s) área(s) (e com determinado peso) por sete entidades (28, 32, 41, 50,

63, 65, 88). Essas sete entidades possuem a “entidade 27” em suas redes de autoridades.

7. **Criação da cadeia de autoridades da entidade “X”:** a cadeia de autoridades origina-se das ligações entre as redes de autoridades existentes. A entidade “X” atribui autoridade para outras entidades formando sua rede de autoridades; cada uma dessas entidades também possui suas próprias redes e, assim, sucessivamente. Essas redes quando interligadas constroem naturalmente uma cadeia de autoridades cuja aresta entre dois vértices representa o menor caminho que os conecta. Entretanto, o fato da cadeia ser um grafo orientado ocasiona a variação da cadeia gerada de acordo com o ponto de partida adotado — gerar a cadeia iniciando-se pela entidade “X” apresenta um grafo diferente do observado ao iniciá-lo pela entidade “Y”. A Figura 36 ilustra a cadeia de autoridades a partir da “entidade 27”.

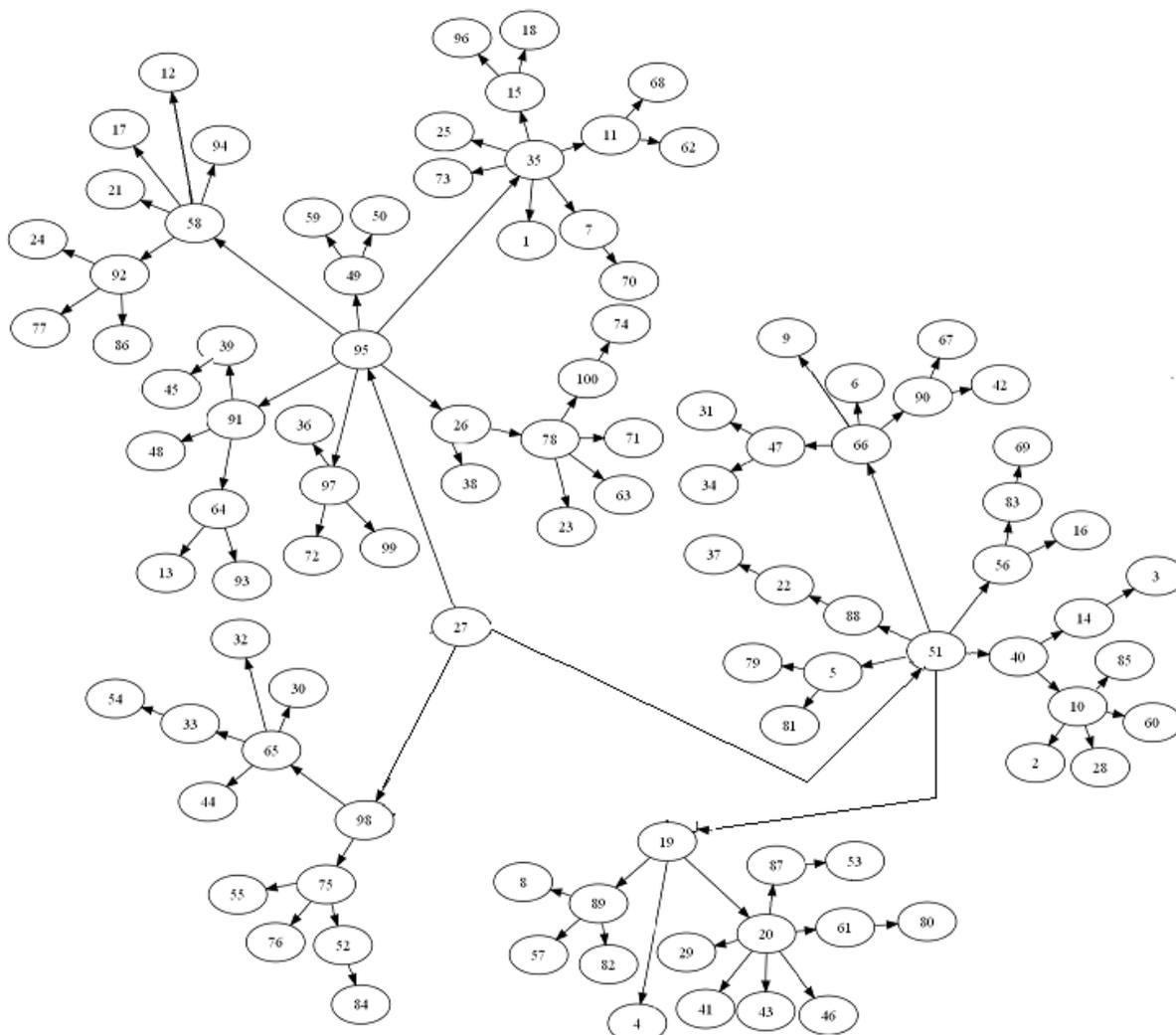


Figura 36 - Cadeia de autoridades da “entidade 27”.

Neste capítulo foram expostos os experimentos e as análises efetuados para demonstrar a validade da hipótese defendida no trabalho. Também foram apresentados alguns recursos desenvolvidos para a extração de informação sobre a cadeia de autoridades e alguns exemplos de sua utilização. No próximo capítulo, serão expostas as considerações finais sobre o nosso trabalho, também serão apresentadas as conclusões obtidas, as limitações das simulações desenvolvidas e do sistema proposto, assim como direções para pesquisas futuras.

Capítulo VI

Conclusão

Como dissemos inicialmente, a grande quantidade de recursos, conteúdos e informações disponibilizados na *web* gera uma sobrecarga de informação, tornando impossível a realização de uma avaliação de sua qualidade por profissionais especializados no assunto. Esses fatores contribuem para que a recuperação de informação nesse ambiente seja um processo cada vez mais complexo. Os sistemas baseados em folksonomia surgiram procurando facilitar a organização do conteúdo presente na *web*, porém eles também sofrem com problemas de qualidade e de sobrecarga de informação quando o número de usuários se torna muito grande.

Este trabalho vem ao encontro das discussões expostas na literatura por Halpin *et al.* (2007), Wal (2005), Russell (2005) e Sturtz (2004), sobre a necessidade de desenvolvimento de recursos que colaborem para a redução do impacto de tais problemas, sem que a liberdade de trabalho dos usuários seja comprometida e sem a adoção de esquemas controlados de classificação. Nossa proposta implica em uma meta-categorização — uma folksonomia em

nível duplo, na qual o primeiro nível abrange a categorização de recursos disponíveis na internet, e o segundo nível abrange a categorização das próprias entidades que desempenham o processo de categorização (*folkauthority*). A hipótese que levantamos é a de que, na recuperação de informação em sistemas baseados em folksonomia, a priorização das categorizações realizadas por fontes reconhecidas como autoridades apresenta resultados mais relevantes e de qualidade, amenizando, assim, o impacto causado pela sobrecarga de informação. Para validar nossa hipótese e para demonstrar sua coerência, apresentamos no capítulo III a relação entre autoridade cognitiva e qualidade; discutimos no capítulo IV o processo de concessão de autoridade cognitiva por meio de folksonomia e os benefícios dele esperados; e descrevemos no capítulo V experimentos relacionados aos processos de concessão de autoridade cognitiva e de recuperação de informação em um sistema baseado em folksonomia.

Existem alguns trabalhos na literatura que desenvolvem abordagens com objetivos semelhantes aos nossos, principalmente os trabalhos de Russell (2005) e Farrell *et al.* (2006). Embora nenhum desses trabalhos seja direcionado para a utilização da autoridade cognitiva na recuperação de informação, o trabalho de Farrel & Lau (2006) possui um enfoque relacionado e o trabalho de Russell (2006) serviu como base para que chegássemos à nossa proposta. Existem outros trabalhos relacionados à confiança em redes sociais (GOLBECK *et al.*, 2006), à reputação em comunidades de prática (CRUZ *et al.*, 2007) e à consideração da autoridade dos usuários na recomendação de informação (MILANI & CAZELLA, 2006) que também abordam questões relacionadas com nossa proposta, porém, esses trabalhos não aplicam a técnica de folksonomia e nem utilizam os conceitos da teoria da autoridade cognitiva de Wilson (1983).

Apesar de assuntos relacionados à autoridade cognitiva serem menos conhecidos e menos discutidos que assuntos relacionados à reputação, eles são mais abrangentes e mais

relacionados com aspectos subjetivos das pessoas envolvidas. Normalmente, como em (CRUZ *et al.*, 2007) (GOLBECK *et al.*, 2006) e (MILANI & CAZELLA, 2006), a reputação é inferida por meio de um conjunto de critérios pré-estabelecidos. Entretanto, de acordo com Muller (2003) e com Russell (2005), nesses processos de inferência não se mantém nenhuma informação relacionada ao contexto da reputação identificada. A nossa proposta de conceder autoridade cognitiva por meio de folksonomia busca, justamente, demonstrar a opinião dos concessores de autoridade preservando a informação gerada no processo de concessão de autoridade.

Os experimentos realizados para a validação da nossa hipótese demonstraram que no processo de recuperação de informação convencional, em um sistema baseado em folksonomia, 25,1% dos objetos categorizados possuem pelo menos um termo que os posicione entre os dez primeiros resultados. Todavia, se a ordem de apresentação dos resultados for determinada de acordo com as autoridades que realizaram as categorizações, 89% dos objetos possuirão pelo menos um termo que os coloque entre os dez primeiros resultados. Isso corresponde a um aumento expressivo no total de objetos que poderão ser encontrados entre os primeiros resultados em um processo de recuperação de informação.

Como citado anteriormente, 81,7% dos usuários de ferramentas de busca não consideram os resultados apresentados além da terceira página ao executarem um processo de recuperação de informação (IPROSPECT, 2007). Isso faz com que muitos resultados úteis não sejam encontrados por esses usuários, uma vez que jamais são apresentados nas primeiras posições. Considerando os resultados obtidos em nossos experimentos, na recuperação convencional, 74,9% dos objetos nunca serão apresentados entre os primeiros resultados e, portanto, nunca serão encontrados pela maioria dos usuários. Por outro lado, aplicando a teoria da autoridade cognitiva esse percentual cai para 11%. Portanto, podemos concluir que, na recuperação de informação em sistemas baseados em folksonomia, a priorização das

categorizações realizadas pelas autoridades resulta em um melhor posicionamento dos objetos, levando os melhores resultados a serem apresentados nas melhores posições, melhorando assim a relevância e a qualidade dos resultados e amenizando os problemas da sobrecarga de informação. Deste modo, os resultados obtidos em nossos experimentos demonstram a validade da hipótese do nosso trabalho.

É importante salientar que ao compararmos o posicionamento dos objetos na recuperação convencional com o posicionamento que esses objetos ocupam ao priorizar as categorizações realizadas por autoridades utilizando um mesmo termo, em sistemas baseados em folksonomia, o percentual de objetos apresentados entre os 10 primeiros resultados é de 25,1% e 24,4%, respectivamente. Verificamos, portanto, que a diferença não é significativa, resultando do fato de que, normalmente, os termos que melhor posicionam um objeto nesse tipo de recuperação são os termos situados no *long tail*. Como esses termos geralmente não possuem autoridades relacionadas eles dificilmente proporcionarão um bom posicionamento para um objeto quando as categorizações de autoridades forem priorizadas. Deste modo, mesmo havendo uma redução de 0,7%, podemos afirmar que a priorização das categorizações realizadas por autoridades não implica em prejuízos na recuperação de informação quando termos menos significativos e sem autoridades relacionadas foram utilizados como parâmetro.

No entanto, ao compararmos o posicionamento dos objetos na recuperação que prioriza as categorizações realizadas por autoridades com o posicionamento que esses objetos apresentam ao utilizar o mesmo termo na recuperação convencional, identificamos uma diferença expressiva: 89% dos objetos aparecem entre os 10 primeiros resultados na recuperação que considera as autoridades; e apenas 12,5% na recuperação convencional. Essa diferença deve-se, principalmente, ao fato de que na recuperação convencional os objetos com categorizações mais antigas que não possuem nenhuma *tag* pouco utilizada não aparecerão entre os primeiros resultados. O oposto acontece ao se considerar as autoridades, pois quanto

maior for o consenso em torno da categorização de um objeto, e quanto maior for a quantidade de autoridades que o categorizaram, melhor será a posição na qual esse objeto será apresentado.

Como resultado secundário, porém não menos importante, de nosso trabalho, verificamos que o processo de concessão de autoridade cognitiva gera uma cadeia de autoridades. Este fato cria uma gama muito rica de possibilidades de análise sobre a rede social formada por esta cadeia.

Uma das análises que realizamos permitiu identificar que nenhuma autoridade está distante com uma separação superior a cinco graus da entidade foco e, também, que 33% das autoridades são encontradas até o segundo grau e 83% até o terceiro grau de separação. Esses resultados demonstram que é possível identificar uma ligação não muito distante entre a maioria das entidades, reforçando assim a confiança e as bases para julgar a plausibilidade da autoridade que cada uma recebeu. Além disso, esses resultados encontram-se de acordo com a teoria dos seis graus de separação e com os experimentos realizados por Yu (2003), que determinam que a maioria das ligações entre duas entidades quaisquer pode ser construída até um máximo de seis graus de separação entre elas. Em outra análise observamos que a identificação da distribuição das autoridades torna possível verificar como os objetos por elas categorizados encontram-se distribuídos na cadeia de autoridades. Percebemos que, em média, 83% de todos os objetos de interesse de cada entidade são encontrados até o segundo grau de separação na cadeia de autoridades. Como 33% das autoridades estão situadas até o segundo grau de separação na cadeia, temos que 33% das entidades categorizaram 83% dos objetos de interesse de cada entidade. Esses resultados corroboram com os estudos de Rieh e Belkin (1998; 2000) sobre a relação que existe entre autoridade cognitiva, relevância e qualidade. Além disso, é possível afirmar que para uma entidade, as suas autoridades

apresentam melhores contribuições na recuperação de informação do que os demais categorizadores.

Existem algumas considerações relacionadas à utilização de folksonomia para a concessão de autoridade e aos experimentos desenvolvidos em nosso trabalho que devem ser destacadas. A primeira delas é que os mesmos problemas encontrados na aplicação da técnica de folksonomia para a categorização de objetos (exemplo: sinônimos e polissemia) serão encontrados ao utilizar a folksonomia para a concessão de autoridade. Não existem indícios que nos levem a concluir o contrário, porém, acreditamos que o fato da categorização representar uma opinião mais consistente sobre o objeto (visto que é uma opinião sobre as competências ou capacidades de uma entidade) faça com que os problemas encontrados na categorização de objetos ocorram em menor escala. Além disso, para que seja possível a obtenção de benefícios na recuperação de informação, é necessário que a concessão de autoridade cognitiva seja realizada com *tags* que a entidade que recebe a autoridade utiliza para categorizar objetos. Isso implica em uma carga de trabalho cognitivo maior para os concessores de autoridade, pois eles deverão considerar o vocabulário de outra entidade. Entretanto, acreditamos também existir uma tendência de maior compatibilidade entre os vocabulários de uma entidade com os de suas autoridades, mesmo que essa compatibilidade seja relacionada apenas ao foco de interesse.

A utilização de simulações para a geração de dados apresenta a desvantagem de não demonstrar dados relacionados à interação dos usuários com o sistema, e à interação social entre os próprios usuários. Entretanto, como argumentado por Golbeck & Hendler (2006), uma simulação que segue uma metodologia bem definida e coerente se torna uma alternativa para estudos com softwares\redes sociais, visto que o desenvolvimento e o povoamento de um sistema com usuários reais implicam em um alto custo em termos de tempo, esforço e investimento, podendo comprometer a viabilidade do trabalho.

Outra questão em torno da concessão de autoridade por meio de folksonomia é a de que a natureza excessivamente manual do processo pode resultar em um custo alto para o usuário. Alguns fatores determinantes na aceitação de um sistema proposto para essa finalidade estão na forma como esse sistema é projetado e nos benefícios que ele tem a oferecer. O que pode, e deve, ser feito é a disponibilização de recursos de orientação e de apoio que ofereçam alternativas e informações relevantes aos usuários como, por exemplo: i) sugerir *tags* durante a concessão de autoridade cognitiva baseando-se na relação entre o vocabulário da entidade e de sua autoridade, em *tags* que outras entidades utilizaram para conceder autoridade à mesma entidade, ou então, na identificação de relacionamentos entre as *tags*; ii) sugerir termos alternativos para o processo de recuperação de informação — termos que possuam um relacionamento semântico ou que sejam utilizados em conjunto com frequência; iii) disponibilizar meios práticos para o gerenciamento do vocabulário e das categorizações realizadas pelos usuários. A tomada de decisões deve sempre ser realizada pelo usuário, mas ao sistema cabe o papel de otimizá-las e de apoiá-las.

Ao decorrer de nossa pesquisa compreendemos a folksonomia como uma técnica que tem muito a contribuir para a organização de informação na Internet e, também, para estudos sociais e comportamentais sobre os usuários. Existe ainda uma carência de análises e de resultados sobre os sistemas baseados em folksonomia e sobre os softwares sociais de uma forma geral. Ainda não foram definidas métricas que nos permitam afirmar se um software social (especialmente os baseados em folksonomia) obtém êxito e alcança seu objetivo. Somos induzidos a tal conclusão de acordo com o número de usuários e com o volume de informação produzida em tais sistemas, mas não se pode dizer que isso é suficiente. Quanto à liberdade e ao poder de trabalho concedido aos usuários, acreditamos que o melhor da folksonomia esteja em um equilíbrio: nem tão flexível, nem tão rígida. O desenvolvimento de recursos que apoiem os processos de categorização e de recuperação de informação, e de

estudos que nos permitam entender mais sobre a dinâmica da participação coletiva nesses sistemas podem demonstrar se esse realmente é o caminho mais adequado.

Com a constatação de que a aplicação do conceito de autoridade cognitiva apresente-se viável por meio de folksonomia e de que proporciona contribuições para o processo de recuperação de informação, temos as indicações necessárias para o desenvolvimento de um sistema que apoie essa aplicação e que possibilite a análise sobre dados obtidos com usuários reais. Acreditamos que a cadeia de autoridades seja o grande benefício a ser obtido com a concessão de autoridade cognitiva, pois, diferentemente da maioria das redes sociais existentes, a cadeia forma uma rede social baseada na confiança e na autoridade entre seus membros. Deste modo, percebemos um campo no qual novas aplicações podem ser desenvolvidas, e no qual podem ser obtidas contribuições para diversas áreas de pesquisa, tais como: teoria de redes sociais, interação humano-computador, inteligência artificial, além de estudos relacionados à psicologia, lingüística e sociologia. O desenvolvimento de um sistema com as características do sistema *CAW*, proposto em nosso trabalho, é o primeiro passo para a criação de um ambiente que possibilite essas novas pesquisas.

*Capítulo VII***Referências**

AIRES, R. V. X. **Linguarudo - Uma Arquitetura Lingüisticamente motivada para Recuperação de Informação de Textos em Português**. Tese de Doutorado em Ciência da Computação e Matemática Computacional. USP - São Carlos, 2003.

ALDE, A; CHAGAS, V. **Blog de Política e Identidade Jornalística: Transformações na Autoridade Cognitiva e na Relação entre Jornal e Leitor**. UERJ. 2005.

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

BALANCIERI, Renato; BOVO, Alessandro B.; PACHECO, Roberto C. S.; BARCIA, Ricardo M.; KERN, Vinicius M. **A Análise de Redes de Colaboração Científica sob as Novas Tecnologias de Informação**. Ciência da Informação. 34.1, Disponível em: <<http://www.ibict.br/cionline/viewarticle.php?id=680>>. Acesso em: 26 fev. 2008. Brasília. 2005.

BOGERS, Toine; THOONEN, Willem; BOSCH Antal Van Den. **Expertise Classification: Collaborative Classification Vs. Automatic Extraction**. 17th SIG/CR Classification Research Workshop. Tilburg. 2006.

BRYNJOLFSSON, Erik; HU, Yu; SIMESTE, Duncan. **Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales**. SSRN. 2007. Disponível em <<http://ssrn.com/abstract=953587>>. Acesso em 04 mar. 2008.

CAPES\CNPq. **Tabelas das Áreas do Conhecimento - Capes/Cnpq** [Online]. Disponível em <<http://www.rc.unesp.br/igce/pos/areas.html>>. Acesso em 16 out. 2007.

CHUNG, L.; NIXON, B. A.; YU, E.; MYLOPOULOS, J. **Non-Functional Requirements in Software Engineering**. Boston : Kluwer Academic Publishers. 1999.

CORMEN, Thomas H.; LEISERSON, Charles E.; RIVEST, Ronald L.; STEIN, Clifford. **Algoritmos: Teoria e Prática**. Rio de Janeiro. Elsevier. 2ª Ed. 2002.

CRUZ, C. C. P. ; Motta, C. L. R. ; SANTORO, F. M. . **Applying Reputation Mechanisms in Communities of Practice**. In: XIII Simpósio Brasileiro de Sistemas Multimídia e Web WebMedia 2007. Porto Alegre : SBC - Sociedade Brasileira de Computação, 2007. v. 1. p. 17-24

DAVENPORT, T. H. **Conhecimento Empresarial: Como as Empresas Gerenciam seu Capital Intelectual**. Rio de Janeiro. 1ª Ed. Campus. 1998.

DELICIOUS. **Delicious Social Bookmarks**. Disponível em <<http://del.icio.us/tags>> Acesso em: 27 out. 2007.

DELICIOUS. **Delicious.** - **Yahoo**, 2006. Disponível em <<http://blog.delicious.com/blog/page/3>> Acesso em 30 jan. 2008.

DIGG Digg. **Digg.** – **Digg**. 2004. Disponível em <<http://www.digg.com>>. Acesso em 06 jan. 2008.

ECHARTE, Francisco; ASTRAIN, José J.; CÓRDOBA, Alberto; VILLADANGOS, Jesús. **Ontology of Folksonomy: A New Modeling Method**. Proceedings of the Semantic Authoring, Annotation and Knowledge Markup Workshop (SAAKM-2007). Canada. 2007.

FARRELL, Stephen; LAU, Tessa. **Fringe Contacts: People-Tagging for the Enterprise**. WWW2006. Edinburgh. Scotland. San Jose. 2006.

FEOFILOFF, Paulo; KOHAYAKAWA, Yoshiharu; WAKABAYASHI Yoshiko. **Uma Introdução Sucinta à Teoria dos Grafos**. Departamento de Ciência da Computação. USP - Universidade de São Paulo. 2007. Disponível em <<http://www.ime.usp.br/~pf/teoriadosgrafos/>>. Acesso em 17 jan. 2008.

FILHO, Antonio Mendes da Silva. **Serendipismo: Descoberta por Acidente e Sagacidade**. Revista: Espaço Acadêmico. Vol. 13. Maringá. 2002. Disponível em <<http://www.espacoacademico.com.br/013/13mendes.htm>>

FOGG, B. J. **Prominence-Interpretation theory: Explaining how people assess credibility online**. Proceedings of ACM CHI 2003 Conference on Human Factors in Computing Systems (pp. 722-723). New York: ACM Press. 2003.

GOLBECK, Jennifer; HENDLER, James. **Inferring Binary Trust Relationships in Web-Based Social Networks**. ACM Transactions on Internet Technology. Vol. VI. New York. 2006.

GOOGLE Orkut. **Orkut.** – **Google**. 2005. Disponível em <<http://www.orkut.com>>. Acesso em 25 fev. 2008.

GOOGLE YouTube. **YouTube.** – **Google**. 2005. Disponível em <<http://www.youtube.com>>. Acesso em 06 jan. 2008.

GOUVÊA Cleber; LOH, Stanley; GARCIA, Luis Fernando Fortes. **Folksonomias: Identificação de Padrões na Seleção de Tags para Descrever Conteúdos**. XIII Webmedia. Gramado - RS. Brasil. ACM Press. 2007.

GRUBER, Thomas. **Ontology of Folksonomy: A Mash-up of Apples and Oranges**. Int'l Journal on Semantic Web & Information Systems. ACM. 2007.

HALPIN, Harry; ROBU, Valentin; SHEPERD, Hana. **The Complex Dynamics of Collaborative Tagging**. Proceedings of the 16th international conference on World Wide Web. pg 211-220. New York. 2007.

HIMMA, Kenneth Einar. **The Concept of Information Overload: A Preliminary Step in Understanding the Nature of a Harmful Information-Related Condition**. Ethics and Information Technology. Seattle. Springer. 2007.

IPROSPECT. **iProspect Search Engine User Attitudes**. iProspect Search Engine Marketing Firm. 2007. Disponível em <<http://www.iprospect.com/premiumPDFs/iProspectSurveyComplete.pdf>>. Acesso em 15 oct. 2007.

JOHNSON, Steven. **Emergência: A Dinâmica de Rede em Formigas, Cérebros, Cidades e Softwares**. Rio de Janeiro. Jorge Zahar. 1ª Ed. 2003.

KDNUGETS News. **Poll Results: The Most Popular Social Bookmarking Sites**. KDnuggets. 2007. Disponível em <<http://www.kdnuggets.com/news/2007/n08/1i.html>>. Acesso em 15 may 2007.

KOME, Sam, H. **Hierarchical Subject Relationships in Folksonomies**. Master's Degree Paper. Chapel Hill : University of North Carolina. 2005.

LANGVILLE, Amy N.; MEYER, Carl D. **Deeper Inside PageRank**. Internet Mathematics. Raleigh. A. K. Peters. Vol. I. 2004.

LEVY, D. M. **More, Faster, Better: Governance in an Age of Overload, Busyness, and Speed**. The Emergence of Governance in Global Cyberspace. 2006. Disponível em <http://www.firstmonday.org/issues/special11_9/>. Acesso em 19 jan. 2008.

LEVY, D. M. **To Grow in Wisdom: Vannevar Bush, Information Overload, and the Life of Leisure**. Dever. ACM. JCDL'05. 2005.

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

LOPES, M. I. **A Internet e a Busca de Informação em Comunidades Científicas: Um Estudo Focado nos Pesquisadores da UFSC**. Dissertação de Mestrado. Florianópolis : UFSC, 2005.

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

LYMAN, P. **How Much Information?**. University of California. USA. 2003. Disponível em <<http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>>. 31 mar. 2008.

MATHES, A. **Folksonomies - Cooperative Classification and Communication Through Shared Metadata**. University of Illinois Urbana-Champaign. 2004. Disponível em <http://blog.namics.com/archives/2005/Folksonomies_Cooperative_Classification.pdf>. Acesso em 04 may. 2007.

MATSUO, Y; TOMOB, H; HASIDA, K; ISHIZUKA, M. **Finding Social Network for Trust Calculation**. Japanese National Institute of Advanced Industrial Science and Technology (AIST). Web Semantics: Science, Services and Agents on the World Wide Web archive Vol 5. Pg 262-278. 2004.

MEYER, Paul L.; **Probabilidade: Aplicações à Estatística**. LTC. 2ª Ed. 2000.

MCKENZIE P. J. **Justifying Cognitive Authority Decisions: Discursive Strategies of Information Seekers**. Vol. 73. Library Quarterly, 2003.

MILANI, Fábio; CAZELLA, Silvio; **Um modelo para determinar a autoridade de usuários em Sistemas de Recomendação**. In: III Forum de Inteligência Artificial. Canoas. 2006.

MULLER, Paul. **The Role of Authority in the Governance of Knowledge Communities**. DRUID Winter. 2003 Conference. Aalborg. 2003.

NEWMAN, M. E. J. **Power Laws, Pareto Distributions and Zipf's Law**. Statistical Mechanics. - Contemporary Physics 46, 323-351. 2006. Disponível em <<http://arxiv.org/abs/cond-mat/0412004>>. Acesso em 11 dec. 2007.

OHMUKAI, I. A. **Proposal of Community-Based Folksonomy with RDF Metadata**. - Galway : 4th International Semantic Web Conference (ISWC2005), 2005.

OLIVEIRA, Ely F. T. de; GRÁCIO, Maria C. C. **Análise a Respeito do Tamanho de Amostras Aleatórias Simples: Uma Aplicação na Área de Ciência da Informação**. GramaZero - Revista de Ciência da Informação - v.6 n.3 jun/05. Disponível em <http://www.dgz.org.br/ago05/Art_01.htm>. Acesso em 17 oct. 2007.

O'REILLY, Tim. **What Is Web 2.0**. O'REILLY. 30 09 2005. Disponível em: <<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>>. Acesso em: 13 apr. 2008.

PEREIRA, Roberto; SILVA, Sérgio Roberto P. **The Application of the Cognitive Authority Concept in a Folksonomy-Based System**. XIII Brazilian Symposium on Multimedia and the Web: WTDWeb 2007 - VII Workshop on Ongoing Thesis and Dissertations. Gramado. Brasil. 2007.

PREECE J, Rogers Y.; SHARP, H. **Design de Interação: Além da Interação Homem-Computador**. Porto Alegre. Artmed. 1ª Ed. 2005.

PRIMO, Alex. **O Aspecto Relacional das Interações na Web 2.0**. XXIX Congresso Brasileiro de Ciências da Comunicação. v. 9, p. 1-21. Brasília. 2006.

RECUERO, Raquel C. **Redes Sociais na Internet: Considerações Iniciais**. E Compós. - Porto Alegre. Vol. v. 2. 2005.

RIDDLE, P. **Tags: What are They Good For?**. School of Information. University of Texas. 2005. Disponível em <http://www.ischool.utexas.edu/~i385q/archive/riddle_p/riddle-2005-tags.pdf>. Acesso em 19 may. 2007.

RIEH, S. Y. **Cognitive Authority**. Medford, NJ: Information Today - Theories of Information Behavior: A Researchers' Guide. pp. 83-87. 2005.

RIEH, S. Y.; BELKIN, N. J. **Interaction on the: Scholars' Judgement of Information Quality and Cognitive Authority**. 63st ASIS Annual Meeting. p. 83-87. 2000.

RIEH, S. Y.; BELKIN, N. J. **Understanding Judgement of Information Quality and Cognitive Authority in the WWW**. 61st ASIS Annual Meeting (Vol 35). p. 279–289. 1998.

ROCHA, Heloísa V.; BARANAUSKAS, Maria Cecilia C. **Design e Avaliação de Interfaces Humano-Computador**. Campinas. Nied_Unicamp. 2003.

ROSSI, Adriana. **Taxonomia ou Sistemática**. Conte Revista On-Line. 2000. Disponível em <http://br.geocities.com/anna_rgs/taxonomia.html>. Acesso em 29 jan. 2008.

RUSSELL, T. **Cloudalicious: Folksonomy Over Time**. School of Information and Library Science. University of North Carolina at Chapel Hill. 2006.

RUSSELL, T. **Contextual Contextual Authority Tagging: Cognitive Authority Through Folksonomy**. School of Information and Library Science. University North Carolina. 2005. Disponível em <<http://www.terrellrussell.com/projects/contextualauthoritytagging/conauthtag200505.pdf>>. Acesso em 19 may. 2007.

SCHAFER, J. BEN. **The Adaptive Web: Collaborative Filtering Recommender Systems**. Science Lecture Notes in Computer. Vol. 4321/2007. Berlin. Springer, 2007.

SMYTH, Barry; BALFE, Evelyn; FREYNE, Jill; BRIGGS, Peter; COYLE, Maurice; BOYDELL, Oisín. **Exploiting Query Repetition and Regularity in an Adaptive Community-Based Web Search Engine**. User Modeling and User-Adapted Interaction. - [s.l.] : Springer, Vols. 14: 383–423. 2004.

SHAW, Blake. **Building a Better Folksonomy: Web-based Aggregation of Metadata**. Technical Report. 2005.

SHEN, K; WU, L. **Folksonomy as a Complex Network**. Departament of Computer Science. Shanghai. Fudan University. 2005.

SOMMERVILLE, Ian. **Engenharia de Software**. São Paulo. Addison Wesley. 6ª Ed. 2003.

STURTZ, D. N. **Communal Categorization: The Folksonomy**. INFO622: Content Representation. 2004.

TAIT, Tania F. C; TRINDADE, J. T. P. **Aspectos Sociais da Informática**. Maringá. EDUEM. 2003.

TECHNORATI. **Technorati.com [Online]**. 2002. Disponível em <<http://technorati.com/>>. Acesso em 11 jan. 2008.

VIRGINIA. **The Oracle of Bacon at Virginia**. University of Virginia. Disponível em <<http://oracleofbacon.org/how.html>>. Acesso em 17 oct. 2007.

VASCONCELOS, F. C. **Racionalidade, Autoridade e Burocracia: as Bases da Definição de um Tipo Organizacional Pós-Burocrático**. Encontro de Estudos Organizacionais. PRODAP/UFPE. 2002.

WAL, T. V. **Folksonomy: Explaining and Showing Broad and Narrow Folksonomies.** vanderwal.net. 2005a. Disponível de <<http://www.vanderwal.net/random/entrysel.php?blog=1635>>. Acesso em 19 jan. 2008.

WAL, T. V. **Folksonomy. Online Information.** vanderwal.net. 2005b. Disponível <<http://www.vanderwal.net/random/entrysel.php?blog=1622>>. Acesso em 18 jan. 2008.

WEBER, Max. **Ensaio de Sociologia.** Rio de Janeiro. Zahar. 1963.

WILSON, P. **Second-hand Knowledge: An Inquiry into Cognitive Authority.** Westport : Greenwood Press. 1983.

WU, Harris. **Harvesting Social Knowledge from Folksonomies.** 17th conference on Hypertext and Hypermedia. Pg. 111–114. 2006.

WU X., ZHANG L.; YU Y. **Exploring Social Annotations for the Semantic Web.** Proceedings of the 15th International Conference onWorldWideWeb. New York. Pg. 417-426. ACM. 2006.

YAHOO. **Delicious – Yahoo!.** 2003. Disponível em <<http://del.icio.us>>. Acesso em 25 fev. 2008.

YAHOO. **Flickr. – Yahoo!.** 2004. Disponível em <<http://www.flickr.com>>. Acesso em 25 fev. 2008.

YU, Bin; MUNINDAR, P. S. **Searching Social Networks.** Melbourne. Australia : AAMAS 2003. Pg. 65-72. ACM. 2003.

Anexo I – Processo de Simulação

Objetivo deste documento: especificar a metodologia utilizada para a realização das simulações do processo de recuperação de informação e de concessão de autoridade cognitiva.

Componentes Necessários para a Simulação:

Entidade: uma entidade representa um indivíduo, um grupo de pessoas, uma empresa ou qualquer outro tipo de organização que deseje recuperar informações e/ou criar uma rede de autoridades. Uma entidade realizará a categorização de informações (objetos), concederá e estará apta a receber autoridade cognitiva de outras entidades.

Entidade	
Identificador único	Código único – seqüencial
Nome\descrição	Identificação da entidade
Termos	<i>Tags</i> que serão utilizadas para categorizar objetos pela entidade (vocabulário de <i>tags</i> da entidade)

Tag: as *tags* utilizadas na simulação são termos simples, são palavras-chaves aplicadas para a categorização de objetos (descrevendo-os) e para a concessão de autoridade cognitiva (representando áreas de conhecimento, habilidades, competências ou assuntos nos quais as entidades são consideradas referências)

Tag
Termos – texto livre

Objeto: Na simulação realizada, o objeto representa uma informação de interesse da entidade e que será categorizada pela mesma. Para a realização de simulações foram necessários poucos dados para identificar o objeto. No processo atual, não importa o que esse objeto é, ou representa, esses dados não exercem interferência nos resultados que pretendem ser observados na recuperação de objetos pelos usuários (o que imposta é o objeto como uma unidade de informação).

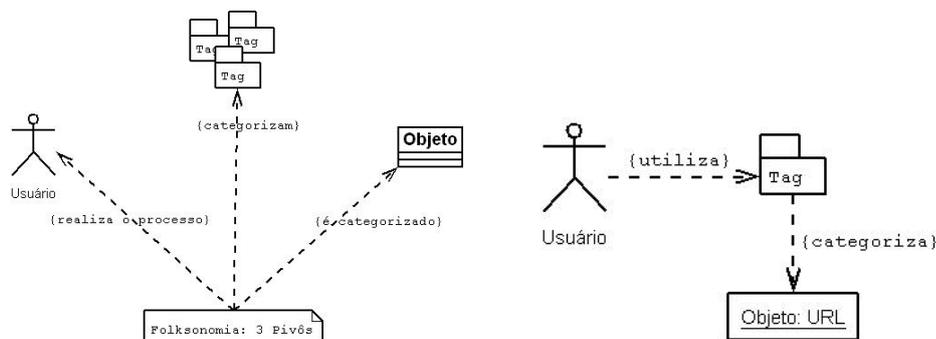
Objeto	
Identificador único	Código único – sequencial
Nome\descrição	Descrição do objeto – qualquer informação sobre o objeto
Interessados	Entidades interessadas no objeto.
Termos	<i>Tags</i> que poderão ser utilizadas para categorizar o objeto

Processos de Simulação:

O processo de simulação está dividido em **dois níveis**: no primeiro, realiza-se a simulação do processo de categorização de informação pelas entidades; enquanto que no segundo, simula-se o processo de concessão de autoridade cognitiva pelas entidades e sobre as entidades.

1 – Categorização de Informação

A categorização de informação realizada pelas entidades representa um processo de atribuição de *tags* aos objetos. Esse processo implica num relacionamento envolvendo os 3 pivôs (Figura abaixo) principais da folksonomia: entidade, *tag* e objeto, cujo objetivo é simular a atribuição de significado, a descrição, ou a categorização propriamente dita desse objeto.



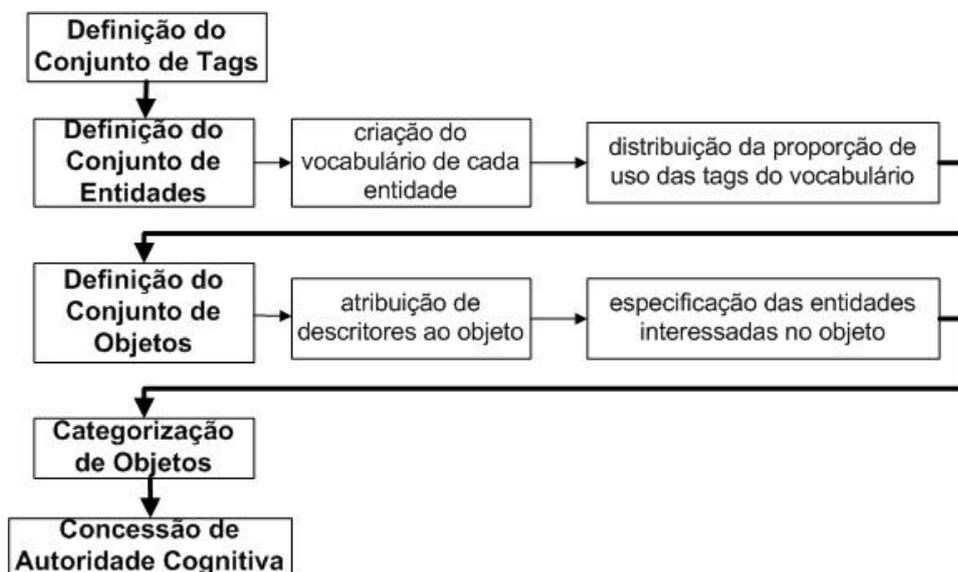
Processo de categorização

Deste modo, o relacionamento é construído por: {Entidade; *Tags*; Objeto}. O que corresponde à dizer que: a entidade “A” utilizou as *tags* “1”, “2” e “3” para categorizar o objeto “Y”.

Categorização*
Entidade
<i>Tags</i> {...}
Objeto

* uma entidade poderá categorizar um mesmo objeto uma única vez

Para a simulação da categorização de informação pelas entidades executou-se o conjunto de passos descritos na seqüência:



Fluxograma do Processo de Simulação

1 – **Definição do conjunto de tags:** o conjunto de *tags* representa os termos utilizados na simulação tanto no processo de categorização de informação quanto no processo de concessão de autoridade cognitiva. O conjunto utilizado é composto por 250 termos, sendo destes, 125 correspondentes às *tags* mais utilizadas pelos usuários da ferramenta *Delicious* (Delicious, 2007) e 125 representando áreas de conhecimento definidas pelas CAPES\CNPq (CAPES/CNPq, 2007).

2 – **Definição do conjunto de entidades:** para a realização das simulações foram geradas 100 entidades (n_0) que, segundo cálculo realizado utilizando a fórmula proposta por Oliveira (OLIVEIRA *et al.*, 2005), possui um erro amostral (E) tolerável de 10% e um nível de confiança de 95% para **uma amostra aleatória simples em que não se conhece o tamanho total da população** ($n_0 = 1/ E_0^2$). Cada entidade recebeu, aleatoriamente e com igual probabilidade de distribuição, 20 termos selecionados do conjunto de *tags* definido no item 1 para representar o vocabulário de termos utilizados por ela na categorização de objetos.

2.1 – **Sobre a distribuição da utilização dos termos:** estudos realizados sobre a utilização de palavras na comunicação verbal (NEWMAN, 2006) e sobre a utilização de *tags* em sistemas baseados em folksonomia (WU, 2006), demonstram a existência de grandes diferenças quanto à quantidade de vezes que cada palavra é utilizada tanto na comunicação quanto na categorização de objetos. Deste modo, a quantidade de

vezes que cada termo do vocabulário de uma entidade foi utilizado para a categorização de objetos foi determinada por uma variação do Princípio de Pareto (NEWMAN, 2006). A variação utilizada determina que 25% dos termos do vocabulário da entidade são utilizados em 70% das categorizações; outros 25% dos termos são utilizados em 15%, enquanto que 50% dos termos são utilizados em apenas 10% do total de categorizações efetuadas pela entidade (a distribuição original refere-se à uma distribuição em que 20% das “causas” são responsáveis por 80% dos “efeitos”). Essa diferença de distribuição permite uma simulação próxima ao que acontece com usuários reais, em que poucos termos são utilizados muitas vezes e muitos termos são utilizados poucas vezes nos processos de categorização — o que forma um *long tail* de termos esporádicos ou pouco utilizados.

3 – Definição do conjunto de objetos: os objetos representam as informações que podem ser categorizadas pelas entidades. O conjunto de objetos é composto por 1000 objetos distintos e, para cada objeto, atribui-se uma lista de descritores contendo 10 termos selecionados aleatoriamente do conjunto de *tags* definido no item 1 — termos que descrevem o objeto. Um conjunto de entidades também é relacionado à cada objeto representando, assim, as entidades às quais o objeto em questão é indicado\recomendado ou que seria de interesse das mesmas.

3.1 – Sobre o relacionamento de entidades aos objetos: para selecionar as entidades relacionadas ao objeto, um termo do conjunto de descritores do objeto é capturado e, então, identifica-se uma quantidade entre 0 e 20 entidades que possuem esse termo em seus vocabulários (gerados no item 2) — essas são as entidades relacionadas ao objeto. O intuito desse relacionamento é identificar as posições em que os objetos aparecem no processo de recuperação de informação e, conseqüentemente, verificar a ordem na qual os objetos de interesse de uma entidade serão apresentados à ela.

3.2 – Sobre a geração do conjunto de objetos: para a simulação da recuperação de informações utilizando a cadeia de autoridades, maiores características dos objetos utilizados não se fazem necessárias; eles são utilizados para demonstrar a forma como as informações podem ser apresentadas às entidades no processo de recuperação de informação em sistemas baseados em folksonomia que apliquem (e não apliquem) o conceito de autoridade cognitiva. Esses objetos representam as informações que são

categorizadas e recuperadas pelas entidades, sem se preocupar com o conteúdo ou com outras características desses objetos.

4 – **Categorização de objetos:** o processo de categorização de objetos é executado para cada uma das 100 entidades criadas e envolve basicamente 3 passos:

4.1 – para cada entidade seleciona-se aleatoriamente 100 objetos. Caso o par {entidade, objeto} já exista a contagem é considerada, porém, o par repetido é ignorado e um novo objeto é selecionado aleatoriamente. A coincidência na seleção aleatória está em aproximadamente 5% dos pares selecionados – isso resulta em uma média de 95 categorizações por entidade.

4.2 – uma quantidade aleatória de termos (mínimo 2, máximo 5) é considerada. Para a seleção dos termos, são priorizados aqueles que estiverem presentes no vocabulário da entidade e nos descritores do objeto simultaneamente. Caso não existam termos suficientes, os demais termos são selecionados do vocabulário da entidade utilizada como responsável pela categorização. Experimentos realizados por Gouvêa *et al.* (2007) sobre a utilização de *tags* em notícias, demonstra uma utilização média de 5 *tags* por notícia.

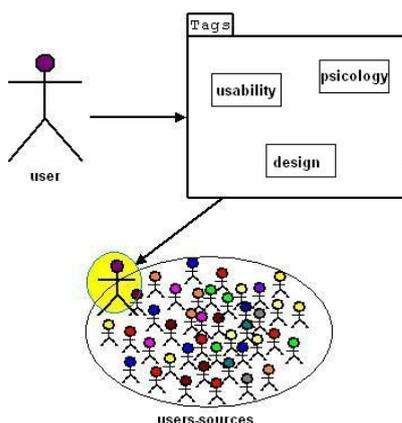
4.2.1 – **Sobre a proporção diferenciada dos termos:** durante o processo de categorização respeita-se a proporção diferenciada de utilização das *tags* (de acordo com a definição do item 2.1). Esta proporção diferenciada faz com que determinadas *tags* sejam mais utilizadas que outras, representando o que acontece em sistemas baseados em folksonomia com usuários reais.

4.3 – cada relacionamento {entidade, *tags*, objeto}, único entre entidade e objeto, indicando que a entidade “X” utilizou o conjunto de termos “Y” para categorizar o objeto “Z” é então construído.

Concessão de Autoridade Cognitiva

O processo de concessão de autoridade cognitiva se difere do processo de categorização de informações basicamente em dois pontos: i) os objetos são as próprias

entidades, o que corresponde à entidades categorizando entidades; ii) os objetos não são categorizados com relação ao seu conteúdo mas, sim, com relação ao seu conhecimento\competência. A técnica de folksonomia aparece em dois níveis nessa abordagem: enquanto no primeiro nível a categorização ocorre sobre a informação, no segundo nível a categorização ocorre sobre as entidades que realizam a categorização da informação.



Processo de concessão de autoridade cognitiva

Deste modo, um relacionamento que represente a concessão de autoridade cognitiva pode ser compreendido como: a entidade “A” considera que a entidade “B” é uma autoridade nos assuntos/áreas representados pelas *tags* “1”, “2” e “3”.

Categorização
Entidade
<i>Tags</i> {...}
Objeto (outra entidade)

* uma entidade poderá conceder autoridade à uma outra entidade uma única vez, porém, em diversas áreas (limite de 5 termos no processo de simulação).

5 – Concessão de autoridade cognitiva: alguns passos necessários para a simulação do processo de concessão de autoridade cognitiva já foram executados na simulação da categorização de objetos, tais como: “1 – Definição do conjunto de *tags*” e “2 – Definição do conjunto de entidades”; portanto, esses dois conjuntos são os mesmos em ambos os processos de simulação. Os passos realizados para a concessão de autoridade cognitiva entre as entidades são executados para cada uma das 100 entidades criadas e estão descritos na seqüência:

5.1 – cada entidade recebe um número aleatório (entre 1 e 10) que representa a quantidade de outras entidades que serão categorizadas pela mesma.

5.2 – todas as entidades são selecionadas de forma aleatória para a concessão de autoridade, desde que o par {entidade1 – concessora, entidade2 – receptora}, nesta ordem, seja inédito.

5.3 – os termos utilizados foram divididos em níveis (1 a 5), assim, um nível mais alto corresponde aos termos de mais impacto, ou de maior peso para uma entidade. Deste modo, toda entidade pode receber termos em 5 níveis distintos. Em cada concessão de autoridade, para cada um dos 5 níveis, um termo é selecionado aleatoriamente dentre os termos mais utilizados para a categorização de objetos pela entidade que esta recebendo autoridade. Desconsideram-se nessa seleção os termos que na proporção de utilização encontram-se no *long tail*.

5.3.1 – **Sobre a seleção dos termos para concessão de autoridade:** os termos do vocabulário de uma entidade que estão situados no *long tail* não são considerados para a concessão de autoridade por representarem interesses esporádicos de uma entidade. Esses termos utilizados um pequeno número de vezes podem representar termos ruidosos no vocabulário de uma entidade real, muito específicos ou que não são mais utilizados pela mesma. Conceder autoridade utilizando esses termos ocasiona uma caracterização de uma entidade como autoridade em áreas nas quais ela possui poucas informações (pouco interesse?!) e, conseqüentemente, em áreas nas quais não haverá muito que se aproveitar da autoridade concedida.

5.4 – cada relacionamento {entidade1, *tags*, entidade2}, único entre entidade1 e entidade2, indicando que a entidade “A” concedeu autoridade à entidade “B” com o conjunto termos “Y” nos diferentes níveis “X1” a “X5” é então construído.

As informações essenciais para análise do processo de recuperação de informação utilizando o conceito de autoridade cognitiva em sistemas baseados em folksonomia estão geradas e armazenadas ao término da execução desse processo.

Anexo II - Princípio de Pareto, *Long Tail* das Tags

De acordo com Newman (2006), quando a probabilidade de medir um certo valor de alguma quantidade variar inversamente a potência daquele valor, diz-se que a quantidade segue uma **lei de potência** (*power law*), também conhecida como a lei de Zipf ou distribuição de Pareto. Newman afirma que as leis de potência aparecem amplamente na física, biologia, ciências da terra e planetárias, economias e finanças, informática, demografia e ciências sociais. Como exemplo temos as distribuições dos tamanhos de cidades, terremotos, chamadas solares, crateras de lua, guerras e as fortunas pessoais de pessoas que, de acordo com Newman, parecem seguir leis de potência.

Brynjolfsson *et al.* (2007) expõe o Princípio de Pareto e a sua ligação com o *Long Tail* proposto e discutido por Anderson (2006). Nesse mesmo foco, Halpin *et al.* (2007) detectou e demonstrou a tendência de uma lei de potência na utilização das tags em sistemas baseados em folksonomia. Essa lei de potência faz com que seja gerado um *long tail* (cauda longa) de tags pouco utilizadas, enquanto que uma minoria de tags é utilizada na maioria das categorizações. Brynjolfsson *et al.* (2007) discute que o *long tail* apresentado por Anderson (2006) e percebido por Halpin *et al.* (2007) na distribuição da utilização de tags em sistemas baseados por folksonomia reforçam o que o economista Vilfredo Pareto afirmava em sua teoria conhecida como o Princípio de Pareto. Segundo Brynjolfsson *et al.*, o Princípio de Pareto determina que 80% dos efeitos são causados por 20% das causas, enquanto que 80% das causas são responsáveis por 20% dos efeitos (distribuição 80-20).

Para que na simulação do processo de categorização de objetos fosse respeitada a diferença na proporção de utilização das tags que, de acordo com Halpin *et al.* acontece com usuários reais, aplicamos uma variação do Princípio de Pareto. Nossa variação consiste em uma distribuição de 75-25: 75% dos efeitos são causados por 25% das causas. A variação para 75-25 foi utilizada devido ao tamanho dos vocabulários gerados para as entidades ser limitado a 20 termos. Utilizar essa variação evita que a distribuição aleatória da quantidade de utilização das tags gere uma distribuição linear de iguais proporções na qual todas as tags são utilizadas igualmente. Os passos para a geração da distribuição estão descritos na seqüência:

- 1- Selecionou-se uma entidade
- 2- Selecionou-se as 20 tags que do vocabulário de entidade
- 3- Determinou-se o número total de aplicação dessas tags para categorizar 100 objetos: Para cada uma das 100 categorizações, gerou-se uma quantidade aleatória

(entre 2 e 5) *tags* ainda sem determinar quais serão as *tags* aplicadas. A quantidade total é distribuída entre as *tags* de acordo com a variação 75-25.

4- Na seqüência, dividiu-se o conjunto de *tags* utilizadas em três partes: A=25%|B=25%|C=50% das *tags* (A=5 *tags*, B=5 *tags*, C=10 *tags*).

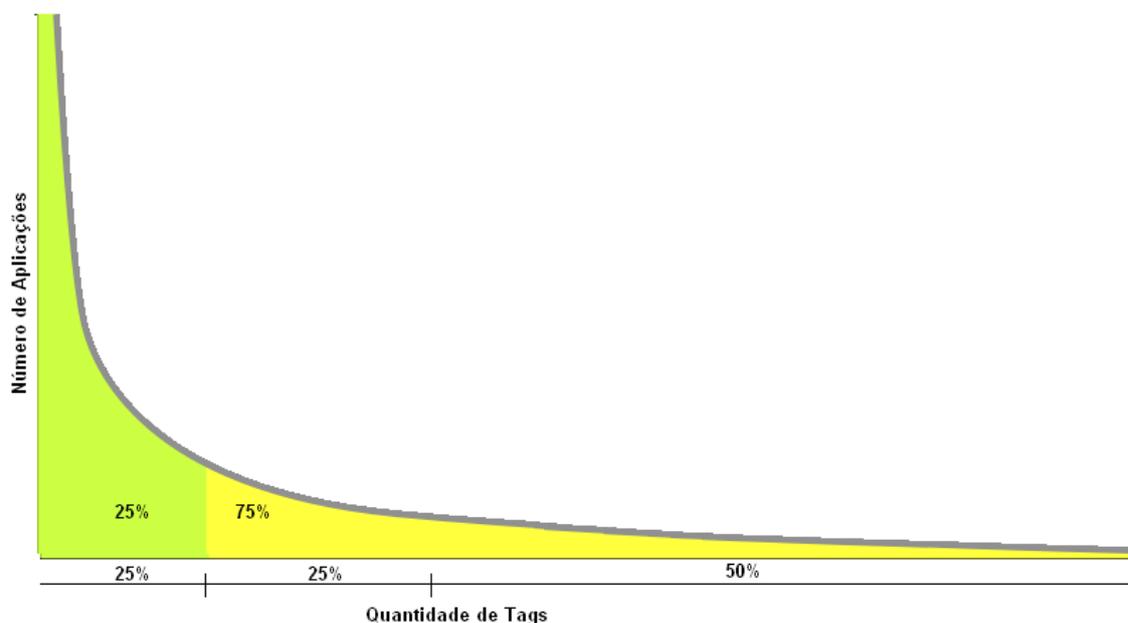
5- E então, distribuiu-se 75% do total de aplicações para o subconjunto A, 15% para o subconjunto B e 10% para o subconjunto C.

Duas regras principais foram consideradas:

- 1- Nenhum elemento deve possuir quantidade inferior a 1
- 2- Nenhum elemento do subconjunto B deve possuir valor superior ao valor do menor elemento do subconjunto A.

Quantidade mínima de aplicação do conjunto de *tags*

Considere um conjunto composto por 20 *tags*. Precisamos definir um número **mínimo** de aplicações dessas 20 *tags* de modo que seja possível efetuar a distribuição 75-25.



Distribuição gerada para simular o *long tail* das *tags*

A quantidade mínima do número total de aplicações das *tags* deve considerar que todos os elementos devem ser utilizados pelo menos uma vez. Deste modo, como ilustra a figura acima, temos que:

- 25% das *tags* serão utilizadas em 75% das categorizações;
- 75% das *tags* serão utilizadas em 25% das categorizações – em que 25% das *tags* são utilizadas em 15% das categorizações, e 50% das *tags* nos outros 10% restantes. Esta subdivisão ocorre para proporcionar a geração do *long tail* na distribuição.

A quantidade mínima deve considerar que 50% das *tags* serão utilizadas uma única vez, e que isso corresponde a 10% do total de utilizações. Deste modo, a quantidade mínima pode ser obtida por:

$$(I) \quad x = \text{quantidade de tags} \quad \begin{array}{l} x - 100 \\ y - 50 \end{array}$$

Assim: $y = 50x/100$, em que y é a quantidade de elementos encontrados no subconjunto C (50%).

Considerando o conjunto de 20 *tags* teríamos $x=20$:

$$\begin{array}{l} 20 - 100 \\ y - 50 \end{array}$$

Então, $y=20*50/100$, $y=10$ (10 termos serão utilizados apenas uma vez).

Agora é necessário definir a quantidade de elementos existentes no subconjunto B em função da quantidade mínima de C, de modo que 25% das *tags* correspondem a 15% das utilizações.

$$(II) \quad \begin{array}{l} y - 10 \\ z - 15 \end{array}$$

Assim: $z = 15y/10 = 15*10/10$, $z=15$; em que z é o número de vezes que os 5 elementos do subconjunto B serão utilizados.

Para o subconjunto A, temos que 25% dos elementos serão utilizados 75% das vezes. Deste modo:

$$(III) \quad \begin{array}{l} y - 10 \\ w - 75 \end{array}$$

Então: $w=10*75/10=75$.

Considerando y o número de utilizações para 10% dos elementos, z o número de utilizações para 15% e w para 75% dos elementos, o total mínimo pode ser obtido por $y+z+w = 10+15+75 = 100$.

Em que “total” representa o número mínimo de aplicações das *tags* com relação à quantidade de *tags* utilizadas para que a distribuição forme um *long tail*.

Como estamos utilizando 100 objetos por entidade para serem categorizados, a quantidade de aplicações mínimas deve ser igual (como no caso demonstrado) ou superior ao total de objetos que está sendo categorizado. Se o resultado da quantidade mínima for menor que o total de objetos, deve-se multiplicá-la por um índice i , obtido por meio do cálculo:

$$i(\text{quantidade mínima obtida}) - \text{total de objetos} = 0$$

Supondo que a quantidade obtida foi de 100 *tags* para 150 objetos:

$$i(100) - 150 = 0$$

$$i = 150/100 = 1.5$$

De modo que 100*i* equivalerá ao total de objetos existentes.

Distribuição da quantidade de utilização de *Tags*

Como exemplo para a simulação da distribuição de *tags* utilizadas, considere um conjunto composto por 20 *tags* aplicadas num total de 150 vezes.

	Quantidade de <i>Tags</i> (20)		Total de Utilização (150)	
25%	5	75%	112	5 <i>tags</i> serão utilizadas num total de 112 vezes
25%	5	15%	23	5 <i>tags</i> serão utilizadas num total de 23 vezes
50%	10	10%	15	10 <i>tags</i> serão utilizadas num total de 15 vezes

Utilizando estes percentuais é possível gerar uma distribuição com os seguintes dados (abaixo da tabela a distribuição gerada é exibida em formato gráfico):

Distribuição das *Tags* e suas respectivas quantidades

Tag	Quantidade
Tag1	19
Tag2	17
Tag3	15
Tag4	13
Tag5	12
Tag6	4
Tag7	3
Tag8	3
Tag9	3
Tag10	2
Tag11	1
...	...
Tag17	1
Tag18	1
Tag19	1
Tag20	1

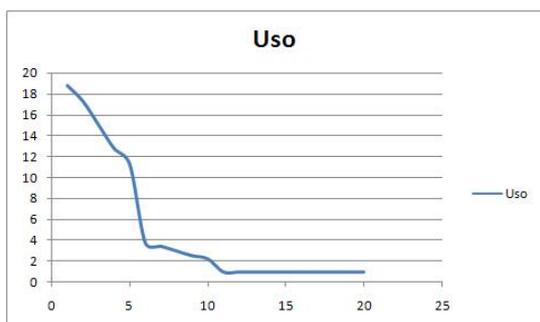


Gráfico 1 – Distribuição gerada (20 tags)

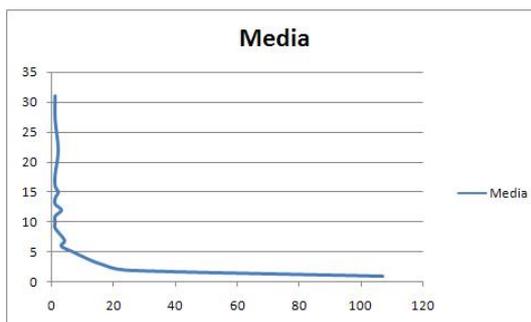


Gráfico 2 – Distribuição de um usuário real (188 tags)

Observe que a distribuição gerada e apresentada no Gráfico 1 mantém o formato do *long tail*. O gráfico 2 apresenta uma distribuição mais fiel *ao long tail* (corresponde à dados obtidos de um usuário real do sistema *Delicious*), entretanto, à medida em que aumentarmos a quantidade de *tags* utilizadas para a geração da distribuição, os dois gráficos tendem a apresentar uma forma mais semelhante. O gráfico 3 apresenta um aumento proporcional de 20 para 100 *tags*, e de 150 pra 750 utilizações considerando a mesma lógica de distribuição aplicada no gráfico 1.

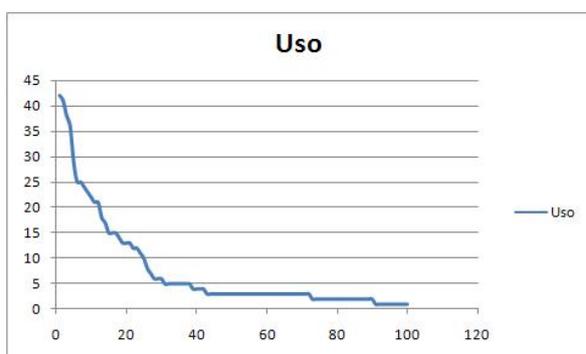


Gráfico 3 – Distribuição gerada (100 tags)

O Gráfico 3 possibilita observar melhor o *long tail* formado pela distribuição da variação 75-25 utilizada em nosso trabalho. Isso demonstra que essa distribuição é adequada para a simulação em nossos processos de categorização de informação.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)