

ORGANIZAÇÃO DA INFORMAÇÃO PARA RECUPERAÇÃO EM REDES DE PRODUÇÃO E COLABORAÇÃO NA WEB

Lidiane dos Santos Carvalho
Elaine R. de Oliveira Lucas
Lucas Henrique Gonçalves

Resumo: Este artigo apóia-se na interdisciplinaridade entre a Ciência da Informação e da Ciência da Computação (Recuperação da Informação) para fomentar a discussão sobre sistemas de representação da informação em ambientes de produção e colaboração na *web* e a organização de seus conteúdos. Apresenta a ascensão da *Folksonomia* como facilitador no modo de organizar e recuperar conhecimentos na *web*. A movimentação desta informação compartilhada demanda a necessidade de aprimoramento nos estudos referente ao tema por parte das disciplinas que compõe a Ciência da Informação.

Palavras-chave: Representação do Conhecimento; Recuperação da Informação; *Folksonomia*.

1 INTRODUÇÃO

A organização do conhecimento e a recuperação da informação passam pela história da classificação do conhecimento humano, desde o princípio da institucionalização da ciência e a necessidade de classificá-la visando sua posterior recuperação. Com a expansão da produção do conhecimento surge a necessidade da criação de métodos, técnicas e sistemas que possibilitassem à recuperação de informações.

Os sistemas de organização do conhecimento evoluíram consideravelmente desde o surgimento da internet. No contexto da *World Wide Web (web)* surgiram por exemplo, às redes *peer-to-peer*, cuja arquitetura caracteriza-se pela descentralização das

funções na rede e permite a classificação do conteúdo pelos pares que produzem e compartilham informações.

Neste ambiente de compartilhamento a intermediação de atores que organizam informações para disponibilizá-las de maneira relevante deu lugar a um novo cenário: a organização da informação é realizada pelos próprios usuários produtores e consumidores da informação. Surge então a popularização da *Folksonomia* em redes de produção e colaboração na *web*, uma prática cujo próprio usuário organiza a informação atribuindo termos indexadores.

Este artigo tem o objetivo fomentar a discussão sobre sistemas de representação da informação em ambientes de produção e colaboração e apresentar algumas considerações sobre a organização do conhecimento na *web*.

2 ORGANIZAÇÃO DA INFORMAÇÃO PARA A RECUPERAÇÃO

A interdisciplinaridade como essência da Ciência da Informação (CI) é tratada por Saracevic (1996) por concentrar relações entre Biblioteconomia, Ciência da Computação, Ciência Cognitiva (incluindo Inteligência Artificial) e Comunicação, sendo esta característica compartilhada por outros autores como Borko (1968), Pinheiro e Loureiro (1995), Targino (1995), Silva e Ribeiro (2002), Le Coadic (2004) muitas vezes desenhando outras disciplinas relacionadas conforme percebemos na revisão exaustiva feita por Pinheiro (2007, p.17).

Ferneda (2003) reconstitui a recuperação da informação como recuperação de documentos, pois, os sistemas não recuperam informações e sim textos e/ou documentos, que podem ser utilizados como informação em potencial. Saracevic (1996, p. 50) comenta que a recuperação da informação é a vertente tecnológica da Ciência da Informação, quando afirma que “a base da relação entre a CI e

Ciência da Computação reside na aplicação dos computadores e da computação na recuperação da informação, assim como nos produtos, serviços e redes associados”. Paes Cardoso (2000) define recuperação de informação como

uma subárea da ciência da computação que estuda o armazenamento e recuperação automática de documentos, que são objetos de dados, geralmente textos. Os componentes do sistema incluem documentos, necessidades do usuário, gera a consulta formulada, e finalmente o processo de recuperação que, a partir das estruturas de dados e da consulta formulada, recupera uma lista de documentos considerados relevantes.

A compreensão científica da estrutura do conhecimento para a organização da informação no campo da Ciência da Informação na descrição de Novellino (1996) constitui o desenvolvimento e avaliação de teorias, para análise de determinadas áreas de assunto, visando à elaboração de instrumentos e métodos para a representação das informações, sendo as pesquisas desenvolvidas voltadas para a entrada do sistema, projeta-se metodologias para a análise de assunto e para a geração de sistemas de classificação e linguagens documentárias.

Historicamente os sistemas de classificação, segundo Tristão, Fachin e Alarcon (2004), sinalizam uma variedade de métodos que organizam, gerenciam e recuperam a informação, em sua maioria, ligados a bibliotecas e outras organizações que se focam no gerenciamento da informação visando organizar, recuperar e disseminar a mesma. Os métodos compreendem os tesouros, ontologias bem como os tradicionais glossários e dicionários técnicos.

a) Tesouro: definido como um vocabulário de termos relacionados genérica e semanticamente sobre determinada área de conhecimento (MOTTA, 1987 *apud* TRISTÃO; FACHIN; ALARCON, 2004);

b) Ontologia: especificação formal e explícita de uma conceitualização compartilhada (GRUBER, 1993 *apud* TRISTÃO; FACHIN; ALARCON, 2004) em que: conceitualização se refere a um modelo de fenômeno abstrato no mundo por ter identificado os conceitos relevantes daquele fenômeno; explícito significa que o tipo dos conceitos usados e as restrições no seu uso são definidos explicitamente; formal se refere ao fato de que a ontologia deveria ser lida pela máquina; compartilhado reflete que ontologia deveria capturar conhecimento consensual aceito pelas comunidades.

Os três modelos clássicos dos sistemas de recuperação da informação, na visão de Baeza-Yates e Ribeiro Neto (1999) são:

a) Modelo Booleano: define a consulta como um conjunto de termos ligado por operadores booleanos como ‘*and*’, ‘*not*’ ou ‘*or*’. Os documentos recuperados são aqueles que contêm os termos que satisfazem à expressão lógica da consulta.

b) Modelo Vetorial: representa termos da consulta e do documento como vetores, tendo cada termo de indexação um peso associado, usando os termos e pesos na avaliação do grau de similaridade entre os documentos e a consulta do usuário. Existe também um grau de similaridade umbral a partir do qual o processo faz uma comparação com o grau de similaridade do documento, para determinar se este deve ser recuperado para o usuário. Sua principal vantagem é que os pesos associados aos termos permitem ter uma melhor performance de recuperação.

c) Modelo Probabilístico: define os documentos relevantes como um conjunto de resposta ideal baseado em um enfoque probabilístico. O sistema estima a probabilidade que um documento seja relevante para o usuário, e essa estimativa é comparada com uma probabilidade umbral para avaliar se o documento é ou não é recuperado. O usuário deve informar quais dos documentos recuperados são relevantes para sua consulta, para tanto o modelo precisa de um “*feedback*”.

Além dos modelos clássicos, existem outros modelos de recuperação de informação, descritos por diversos autores, dentro os quais se encontram o Modelo de Conjuntos Difusos, Modelo Booleano Estendido, Modelo Vetor Espaço, Redes Neurais, Redes Bayesianas, Modelo de Redes de Inferência, entre outros. (INGWERSEN, 1992); (BAEZA-YATES E RIBEIRO NETO, 1999); (PAES CARDOSO, 2000).

A *web* pode ser incluída como foco de interesse em questões de organização do conhecimento e representação da informação, com o objetivo primordial de usufruir ao máximo as vantagens de acesso ao conhecimento composta por distintos conhecimentos que utilizados na atualidade cria a necessidade de estruturação e recuperação das informações. (SOUZA, 2000).

3 ORGANIZAÇÃO DO CONHECIMENTO

Com a institucionalização do método científico, a consequente produção de conhecimento, e a sedimentação das academias de ciências de Galileu no ano de 1560, a academia francesa em 1635, a *Royal Society of London* no ano de 1645, e as academias na Alemanha e Rússia em 1724 forneceram, segundo Burke (2003), motivos suficientes para o homem pensar a necessidade de organizar o conhecimento, pois a informação, produto do conhecimento, precisava ser organizada de modo que fosse de fácil recuperação.

Na história do conhecimento surgem então os primeiros sistemas para a sua organização, alicerçado no tripé das estruturas de currículos, bibliotecas e enciclopédias (BURKE, 2003).

Souza (2000) relata que os bibliotecários foram os principais responsáveis em ordenar o conhecimento registrado para fins de guarda, arquivamento, recuperação e acesso, quando afirma que

[...] os Bibliotecários foram os principais responsáveis em ordenar o conhecimento registrado para fins de guarda, arquivamento, recuperação e acesso. Na verdade, estas necessidades básicas dos sistemas tradicionais de informação são as mesmas que se apresentam na Internet

Os bibliotecários possuem disciplinas que trabalham a descrição da informação para sua posterior recuperação. Como consideram Castro e Santos (2009, p. 76), os bibliotecários

[...] possuem as técnicas e habilidade necessárias ao tratamento documentário dos recursos informacionais. A catalogação, enquanto disciplina responsável pela representação descritiva bibliográfica da área, pautada em regras e esquemas de descrição, fornece subsídios para a construção de formas de representação e de descrição padronizadas que possibilita o armazenamento, a preservação, o uso e o reuso das informações de modo eficiente e eficaz.

A organização do conhecimento na interdisciplinaridade da Ciência da Informação envolve desenvolver e avaliar teorias para análise de determinadas áreas de conteúdo e aprimorar os métodos de representação das informações geradas.

Novellino (1996, p. 38) afirma que a organização do conhecimento na Ciência da Informação “desenvolve-se homologamente à compreensão científica de estrutura do conhecimento, projetando

metodologias para a análise de assunto e para a geração de sistemas de classificação e linguagens documentárias”.

A institucionalização do método científico desde então contribuiu para o estudo, aprimoramento e surgimento de variados métodos para organização do conhecimento. As Tecnologias da Informação e da Comunicação (TICs) contribuíram de forma significativa para a agilidade no processo de recuperação da informação.

4 ORGANIZAÇÃO DA INFORMAÇÃO POR USUÁRIOS NA WEB

Baeza-Yates e Ribeiro Neto (1999, p. 07) afirmam que existem três mudanças fundamentais acontecidas após a proliferação exorbitante da *web* e os avanços na tecnologia moderna de computadores:

- 1) Diminuiu-se visivelmente o custo de acesso a diversas fontes de informação;
- 2) Os avanços em comunicação digital permitem um melhor acesso às redes de trabalho locais e globais. Assim, as fontes de informação podem ser acessadas rapidamente e desde grandes distâncias; e
- 3) A liberdade para incluir a informação que alguém considerou relevante contribuiu à popularidade da *web*.

A partir destas mudanças, Baeza-Yates e Ribeiro Neto (1999), concluem que a *web* é altamente interativa, porque não é passiva perante tempo e espaço, e possibilita a criação de informações por diversas pessoas. Porém, embora haja interatividade as pessoas ainda possuem dificuldades para recuperar informações relevantes na *web*.

A interação com novas ferramentas tecnológicas, síncronas, no caso das informações instantâneas tais como conferências pela Internet (videoconferências), *skype*, *chat*, *webcam*, entre outras; e assíncronas, informações passíveis de edição como *wikis*, por

exemplo, permitem que várias pessoas possam trabalhar de maneira colaborativa, bem como compartilhar recursos de informação.

No que envolve as novas formas de classificação na *web*, Catarino e Batista (2007) consideram que o trabalho comum, de produção, consumo e compartilhamento de informações em um ambiente colaborativo, permitem que o usuário indexe os recursos informacionais.

Esse trabalho comum é apontado por Lévy (1999) como uma abordagem de caráter geral da vida em sociedade e de seu possível futuro, sendo que essa visão de futuro, esta organizada em torno de dois eixos complementares: o da renovação do laço social por intermédio do conhecimento e o da inteligência coletiva propriamente dita. O que passa a ser diferente são as novas dimensões das atividades de coleta, arquivamento, busca e acesso ao conhecimento (SOUZA, 2000).

O conceito de transferência de informações é aqui compreendido como a intervenção realizada por sistemas de organização do conhecimento e recuperação da informação em determinadas ações comunicativas, que são aquelas que se dão entre produtores e consumidores de conhecimento (NOVELLINO, 1996).

Para Capurro e Hjørland (2007), o termo informação não respeita os limites das áreas de conhecimento e não encontra consenso sobre qual deveria ser sua definição, que varia de uma área do conhecimento para outra e em relação a diferentes contextos. A organização da informação produzida em ambiente colaborativo na *web* tem essa característica, onde o usuário ao indexar determinado conteúdo informacional, utiliza uma linguagem natural sem limitações das classificatórias, para expressar o entendimento a respeito do conteúdo do documento.

Catarino e Baptista (2007) consideraram um novo paradigma para a organização dos conteúdos dos recursos digitais na *web*, a *Folksonomia*. A possibilidade de os próprios usuários participarem na organização desses conteúdos é, em especial, uma questão que

vale ser pesquisada e implementada, esta nova abordagem relaciona-se com a indexação dos recursos digitais da *web*. (CATARINO; BAPTISTA, 2007)

5 FOLKSONOMIA

Criada por Thomas Vander Wal a *Folksonomia* inclui o prefixo “*folks*”, palavra da língua inglesa que significa pessoas. Seu criador, “[...] define a prática como o resultado de processos livres de ‘etiquetamento’ de páginas e objetos, realizados em um ambiente social, por pessoas que consomem as informações, cujo objetivo é a recuperação. (AMARAL; AQUINO, 2008).

Ou seja, conforme os autores (*op. cit.*) as “*tags*” ou etiquetas utilizadas pelos usuários em ambientes colaborativos têm a função de palavras-chave para recuperar as informações e compartilhá-las com outros usuários. A principal característica das “*tags*”, segundo Amaral e Aquino (2008, p. 02) é que estas podem ser

[...] criadas por qualquer usuário na *web*, ao contrário do que ocorria no início da *web* onde apenas programadores e conhecedores de linguagens de programação editavam *hipertextos*, e porque são criados com base no significado das informações etiquetadas”

Moura (2009, p. 63) afirma que o termo *Folksonomia* não é amplamente utilizado e que diversos autores adotam termos como “*ethnoclassification, collaborative tagging, social bookmarking, social tagging, narrow folksonomy, taxonomia popular, classificação social, entre outros*”. Contudo, após analisar o conceito da recente *Folksonomia*, cunhada por diversos autores, em documentos científicos, Catarino e Baptista (2007) apresentam um conceito a partir da análise de diversos conceitos: *Folksonomia*

‘resultado da etiquetagem dos recursos...’, ‘um vocabulário’, ‘lista de termos’, ‘conjunto de termos’, ‘resultado de um sistema de classificação socialmente construída’, ‘coleção de conceitos’, ‘conjunto informal e orgânico de terminologia relacionada’, e ‘conjunto de etiquetas’.

As principais preocupações que giram em torno da organização de informações em redes colaborativas, onde quem produz e classifica as informações são os usuários (linguagem natural), dizem respeito aos problemas que influem na precisão da informação recuperada.

Golder e Huberman (2006) apontam alguns problemas como por exemplo, a polissemia, sinonímia, e variação de nível básico como plurais, abreviações e erros ortográficos, pois, a liberdade na atribuição tem como resultado um conjunto de palavras-chave que muitas vezes podem ser ambíguas, inexatas e conseqüentemente afetam a precisão na recuperação da informação.

No entanto a vantagem desses sistemas, conforme apontam Golder e Huberman (2006) é o caráter colaborativo, onde a colaboração mediada por computador proporciona aos seus utilizadores novas formas de criação, compartilhamento e organização do conteúdo informacional. Esse caráter colaborativo na *web* possibilita a criação de qualquer tipo de informações, sejam estas verídicas ou não, sendo assim, possibilitam, também, a utilização de quaisquer palavras-chave para a recuperação da informação em alguns sistemas colaborativos.

Os usuários e idealizadores desses sistemas procuram alternativas para resolver a falta de organização da informação em redes de produção e colaboração mediadas por computador, uma delas é a utilização de “*Folksonomia* Controlada”, em que ao atribuir “*tag*” a um determinado documento, a ferramenta sugere as palavras-chave, utilizadas por outros usuários para indexar o mesmo recurso informacional.

As inovações em algumas redes de compartilhamento de informações sugerem uma grande categoria de classificação pré-definida onde o sistema não sugira mais de uma palavra-chave para determinado assunto, ou seja, o criador já tem termos definidos, só necessita adequá-los a seu documento, o que facilita na recuperação da informação.

6 CONSIDERAÇÕES FINAIS

As formas colaborativas de produção do conhecimento movimentam atualmente a ordem do trabalho e das relações humanas. O compartilhamento de informações nestes ambientes envolve aspectos sociais e comportamentais das pessoas envolvidas.

A classificação de conteúdos envolvendo a participação direta dos usuários exprime o pensamento humano e a significação dos conteúdos nos descritores atribuídos, em que se cria uma linguagem comum.

Os sistemas empregando marcação de forma livre são extremamente sensíveis as necessidades dos usuários e estão incentivando os usuários a organizar as informações e envolve-los ativamente no sistema organizacional. Transformar um recurso isolado em uma atividade comunicativa e compartilhada é um importante fenômeno que deve ser explorado e considerado para os sistemas futuros. (MATHES, 2004).

A movimentação desta informação compartilhada demanda a necessidade de aprimoramento nos estudos referente ao tema por parte das disciplinas que compõe a Ciência da Informação, principalmente a Ciência da Computação.

A Ciência da Informação, ao discutir a questão da organização do conhecimento, contribui para o aprimoramento dos sistemas computacionais no que se refere a tecnologias de recuperação da informação fortalecendo a autonomia da organização do

conhecimento pelos usuários abre caminhos para questionamentos quanto a métodos de catalogação e indexação para posterior recuperação da informação; quanto à utilização de linguagem natural em documentos na *web*; e quanto à importância da colaboração de conhecimento na sociedade atual.

Neste contexto vê-se a importância da colaboração na *web* para melhor criação e disseminação de conhecimentos, além da formação de usuário ativos, que participam das etapas de criação, armazenamento, descrição (metadados e etiquetamento), disponibilização e recuperação da informação.

7 REFERÊNCIAS

AMARAL, Adriana; AQUINO, Maria Clara. Práticas de Folksonomia e Social Tagging no Last.fm. In: Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais, 8. 2008. *Anais...* Porto Alegre, 2008.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. Modeling. In: _____. *Modern information retrieval*. New York: Addison Wesley, 1999. Cap. 2. Disponível em: <<http://www2.dcc.ufmg.br/livros/irbook/chapters/chap2.html>> Acesso em: 12 ago. 2009.

BURKE, Peter. *Uma história social do conhecimento: de Gutenberg a Diderot*. Rio de Janeiro: Zahar, 2003.

CAPURRO, Rafael; HJORLAND, Birger. O Conceito de Informação. *Perspectivas em Ciência da Informação*, v. 12, n 1, p. 148-207, jan./abr. 2007. Disponível em: <http://www.eci.ufmg.br/pcionline/index.php/pci/article/view/54> Acesso em: 12 ago. 2009.

CASTRO, Fabiano Ferreira; SANTOS, Plácida Leopoldina Ventura Amorin da Costa. Uso das tecnologias na representação descritiva: o padrão de descrição bibliográfica semântica MarcOnt *Initiative* nos ambientes informacionais digitais. *Ci. Inf.*, Brasília, v. 38, n. 1, p. 74-85, jan./abr. 2009. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/viewArticle/1082>
Acesso em: 12 ago. 2009.

CATARINO, Maria Elisabete; BAPTISTA, Ana Alice. *Folksonomia: um novo conceito para a organização dos recursos digitais na Web*. *DataGramZero*, v. 8, n. 3, jun. 2007. Disponível em: http://dgz.org.br/jun07/F_I_art.htm Acesso em: 12 ago. 2009.

FERNEDA, E. *Recuperação de informação: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação*. 2003. Tese (Doutorado em Ciências da Comunicação). Escola de Comunicação e Artes – USP, São Paulo, 2003. Disponível em: <http://www.teses.usp.br/teses/disponiveis/27/27143/tde-15032004-130230> Acesso em: 12 ago. 2009.

GOLDER, Scott A.; HUBERMAN, Bernardo A. The Structure of Collaborative tagging systems. *Journal of Information Science*, 2006. Disponível em: <http://arxiv.org/ftp/cs/papers/0508/0508082.pdf>
Acesso em: 12 ago. 2009.

LEVY, P. *Cibercultura*. São Paulo: Editora 34, 1999. 260 p.

MATHES, Adam. *Folksonomies: Cooperative Classification and Communication Through Shared Metadata*. *Computer Mediated Communication*. University of Illinois Urbana-Champaign, dez. 2004. Disponível em :

<http://adammathes.com/academic/computer-mediated-communication/folksonomies.pdf> Acesso em: 05 set.. 2009.

MOURA, Maria Aparecida. Informação, ferramentas ontológicas e redes sociais ad hoc: a interoperabilidade na construção de tesouros e ontologias. *Informação & Sociedade: Estudos*, v.19, n.1, p. 59-73, jan./abr. 2009. Disponível em:
<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/2396/2688>
Acesso em: 12 ago. 2009

NOVELLINO, Maria Salet Ferreira. Instrumentos e metodologias de representação da informação. *Inf.Inf.*, Londrina, v.1, n.2, p.37-45, jul./dez. 1996. Disponível em:
<http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/1603/1358> Acesso em: 12 ago. 2009

PAES CARDOSO, O. N. Recuperação da Informação. *Infocomp: Journal of Computer Science*, v. 2, n. 1. 2000. Disponível em:
<http://www.dcc.ufba.br/infocomp/artigos/v2.1/art07.pdf> Acesso em: 12 ago. 2009

Pinheiro, Liliane Vieira. *As redes cognitivas e a produção do conhecimento em Ciência da Informação no Brasil: um estudo nos periódicos da área*. 140 f. 2007. Dissertação (Mestrado em Ciência da Informação) Universidade Federal de Santa Catarina, Florianópolis, 2007. Disponível em:
<http://www.cin.ufsc.br/pgcin/Pinheiro,%20Liliane.pdf> Acesso em: 05 set. 2009

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

<http://www.eci.ufmg.br/pcionline/index.php/pci/article/viewFile/235/22> Acesso em: 12 ago. 2009

SOUZA, R. F. A Classificação como Interface da Internet. *DataGramaZero*, v. 2, n. 2, abr. 2000. Disponível em: http://dgz.org.br/abr00/Art_01.htm Acesso em: 12 ago. 2009

TRISTÃO, Ana Maria Delazari; FACHIN, Gleisy Regina; ALARCON, Orestes Estevam. Sistemas de classificação facetados e tesouros: instrumentos para organização do conhecimento. *Ci. Inf.*, Brasília, v. 33, n. 2, dez. 2004. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/viewArticle/88> Acesso em: 12 ago. 2009

THE ORGANIZATION OF INFORMATION TO RECOVER IN SOCIAL NETWORKS OF PRODUCTION AND COLLABORATION IN THE WEB.

Abstract: *This study presents the historical-conceptual contextualization of the recuperation of information in the scope of Science Information and in the scope of knowledge management systems aiming to present collaboration and sharing information networks. It also approaches the Folksonomy, natural language of classification used by the web users to index information in collaborative environments mediated by computer and their implications to recover information.*

Keywords: *Knowledge Classification; Information Recuperation; Folksonomy.*

Lidiane dos SantosCarvalho
Mestre em Ciência da Informação (UFSC)
Contato: carvalho.ldn@gmail.com

Elaine R. de Oliveira Lucas

Mestre em Eng. de Produção (UFSC); Bacharel em Biblioteconomia (UFSC); Docente do departamento de Biblioteconomia e Gestão da Informação da UDESC
Contato: lani@udesc.br

Lucas Henrique Gonçalves

Acadêmico do Curso de Graduação em Biblioteconomia (UDESC);
Contato: luc4s_hg@hotmail.com

Artigo: Recebido em: 14/08/2009 Aceito em: 11/09/2009
