

FOLKSONOMIA E CLASSIFICAÇÃO DE ETIQUETAS: estudo de caso Flickr

Wagner Ferreira Caldas*
Manoel Palhares Moreira**

RESUMO

A evolução da web estabeleceu uma nova fase de interatividade com os usuários. Se antes todo o conteúdo armazenado era indexado por heurísticas de máquinas de busca, agora este processo conta com a participação daqueles que disponibilizam os documentos e daqueles que deles fazem uso. A Folksonomia, através da utilização de etiquetas, representa uma mudança nos processos de organização e tratamento da informação na web. O trabalho apresenta um estudo de classificação de conteúdos de etiquetas postadas pelos usuários, através de um estudo de caso com as etiquetas de fotografias do site Flickr e de acordo com a classificação de Cañada. Concluiu-se que a maioria das etiquetas corresponde a informações pertinentes ao conteúdo fotográfico, mostrando que para esse tipo de sistema, existe um interesse em contribuir para o compartilhamento de informações coerentes.

Palavras-chave: Flickr; Folksonomia; Web 2.0; Internet colaborativa; Cañada.

1 INTRODUÇÃO

A disseminação e a recuperação de informação encontraram no suporte digital e na web um campo fértil de expansão. Neste contexto, a publicação e o acesso à informação tornaram-se possíveis ao indivíduo comum, gerando hábitos e possibilidades pessoais ou comuns ao grupo onde encontra-se inserido. Como não se trata de algo pronto e acabado, a web tem evoluído seguindo os passos da própria evolução da tecnologia. Mas, notoriamente, desde seu surgimento, tem permitido que seus usuários participem ativamente na construção de seu conteúdo.

Com o passar dos anos e o aumento da interatividade entre sistemas e usuários, essa rede mundial passou a ser cada vez mais “um espaço interativo, de trocas, de criação e geração, além de armazenamento de informações, tornando-se uma importante ferramenta de colaboração entre os participantes do mundo digital...” (BLATTMANN, 2007, p.191).

Mais recentemente, o termo Web 2.0 passou a ser utilizado na comunidade de usuários e entre especialistas da área, dando indício que uma nova geração de serviços passaria a

* Graduado em Ciência da Informação pela Pontifícia Universidade Católica de Minas Gerais.
vhwagner@yahoo.com

** Doutor em Ciência da Informação Pontifícia Universidade Católica de Minas Gerais, Professor Colaborador do programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais – ECI.
palhares@pucminas.br

compor a web. O termo Web 2.0 possui um significado amplo, nem sempre de fácil compreensão e os limites de sua definição geralmente são inalcançáveis. Para O'Reilly (2005), responsável pela utilização primeira do termo, a Web 2.0 pode ser vista como uma plataforma onde são compartilhados recursos e serviços. Neste ambiente, os sites crescem em funcionalidades e perdem a característica estática, passando a ser atualizados automaticamente e possibilitando a execução de aplicativos no servidor. Na realidade a web é a mesma, mas há uma nova filosofia determinante de construção de páginas com uma interatividade maior com o usuário e um direcionamento na prestação de serviços.

Com o desenvolvimento dos sistemas e da liberdade de criação inicia-se a formação de uma inteligência coletiva, derivada da colaboração dos próprios usuários dos sistemas dos quais desfrutam. Trata-se de uma inteligência distribuída por toda a parte, valorizada a todo instante, resultando em uma mobilização efetiva das competências. Está distribuída por que a ninguém é conferido o conhecimento total das coisas, mas o saber está na humanidade e é construído por partes a partir do conhecimento de cada um. E como está distribuída por toda parte deve transformar-se em projeto coletivo para ser valorizada. A mobilização de competências requer anteriormente sua identificação, reconhecendo as diversidades e o fato de que o conhecimento do outro é uma forma de expressão de sua identidade social (LÉVY, 2002).

Essa liberdade de construção de conhecimento, seus prós e contras, é alvo de discussão em diversos casos, pois é difícil o consenso de quando ela é indicada e/ou quando se devem manter os padrões tradicionais onde os mediadores da informação devem possuir uma formação acadêmica ou serem conceituados no assunto.

Apesar da polêmica começam a despontar novos sites e sistemas de indexação, e a Web 2.0 caracteriza-se hoje, principalmente, pela presença de *Really Simple Syndication* (RSS), dos *wikis*, das redes sociais, dos *blogs* e da *folksonomia*. A tecnologia advinda do RSS permite que usuários inscrevam-se em sites que forneçam fontes de alimentação (*feeds*) em RSS. Usualmente são sites dinâmicos que alteram constantemente seu conteúdo e utilizam-se desses *feeds* para manter informados seus usuários destas atualizações. De forma geral, oferecem conteúdo em formato reduzido com *links* para as versões completas.

Os *wikis* são softwares colaborativos que permitem a edição coletiva dos documentos, através de um sistema de computação preparado para tal, sem uma revisão prévia para sua publicação. O conceito é bem simples: textos publicados na web que podem ser modificados por qualquer usuário, sem a necessidade de autorização prévia, apenas com o emprego de um sistema que registra todas as alterações, sua origem (o protocolo da rede) passando a publicá-

las imediatamente agilizando a construção do conhecimento no assunto. A *Wikipédia* é hoje o mais famoso *wiki* existente, mas já existem empresas utilizando o recurso para propagação de conhecimento entre seus empregados, como menciona Prescott (2008).

Os sites que adotam tecnologia de Redes Sociais possibilitam a formação de uma rede entre pessoas que em linhas gerais representa relacionamentos existentes entre elas, quer sejam eles afetivos, profissionais ou movidos por outro motivo mútuo qualquer. As redes são compostas por indivíduos que possuem alguma identidade comum e através delas dá-se o compartilhamento de idéias e opiniões. O crescimento de sites que são uma expressão das redes sociais é tanto que, dentre os dez sites mais visitados no mundo, de acordo com o Alexa (2008), quatro deles são sites construídos pelos próprios usuários: sites de relacionamento, blogs, e sites que disponibilizam vídeos e fotografias.

Um blog, ou um *webblog*, é uma página da web onde as atualizações de conteúdos encontram-se organizadas cronologicamente de forma inversa. O conteúdo pode pertencer ao próprio proprietário ou a algum de seus visitantes. Geralmente os blogs são páginas pessoais, mas existem organizações que utilizam os recursos para diversos fins. Os sites que hospedam blogs contam com sistemas de computação para criar facilidades aos usuários na construção de seu próprio blog. Eles possuem ferramentas de consultas aos conteúdos armazenados assim como possibilitam aos proprietários o controle do número de acessos, das páginas visitadas, a localização do usuário no ciberespaço, do tempo que os usuários da rede despenderam no blog, a origem e o destino da navegação do usuário, entre muitos.

A folksonomia surgiu da necessidade do usuário disseminar e recuperar com maior facilidade seus próprios documentos. Neste contexto, um documento pode ser uma foto, um filme, um slide ou mesmo um documento de texto. Neste trabalho este conjunto será sempre referenciado como documento.

Os sites que armazenam os documentos digitais possuem seus próprios instrumentos de busca que facilitam o acesso dos usuários aos conteúdos disponibilizados. Entretanto, era necessária uma ferramenta que facilitasse esse processo de recuperação aproximando a tecnologia aos próprios usuários.

Realmente tudo evoluiu porque o usuário tinha desejo de que seus conteúdos fossem conhecidos além de também desejar conhecer conteúdos semelhantes aos seus. Então, surge uma nova forma de indexação na web, a folksonomia: uma indexação social, como é conhecida por muitos, gerada a partir do próprio usuário detentor dos documentos. A solução veio com algo simples conhecido como *tag*, ou etiqueta. Os usuários informam o conteúdo de suas etiquetas (*tags*) de acordo com sua própria orientação. Elas norteiam a classificação de

seus documentos e é através delas que os mesmos são recuperados. Por isso, a folksonomia é uma indexação a partir de linguagem natural, embasada na realidade do proprietário do documento ou da comunidade que ele imagina que irá recuperá-lo. É a interatividade advinda da Web 2.0 possibilitando que o usuário dite os rumos da web. Entretanto, por tratar-se de uma ferramenta que utiliza linguagem natural, alimentada diretamente pelos usuários, a folksonomia não se preocupa com aspectos computacionais ou da ciência da informação. Nesta filosofia sobressai um caráter extremamente social. Ou seja, o usuário está mais preocupado em estar bem inserido em um grupo de usuários e ali ser aceito do que necessariamente fazer com que a recuperação de seus documentos, ou de documentos, alheios seja facilitada ou coerente. Assim criou-se um “sistema nervoso social”, onde o processo de socialização se torna a chave e o determinante dos caminhos para onde a internet segue; um sistema que comanda as ações do mundo físico, como propõe Ross (2009).

A internet e sua gama de documentos, imagens, textos e vídeos era há alguns anos uma fonte de conhecimento e entretenimento para a sociedade, mas isto não determinava de forma direta os rumos desta. A cada dia que passa a sociedade está mais envolvida com a rede e agora ela não se altera simplesmente no que tange cada indivíduo, mas sim no poder de interação entre eles.

Na realidade, para a Ciência da Informação a Folksonomia é um novo desafio. As teorias que sustentam todo o processo de classificação e organização da informação têm agora uma nova alternativa advinda do próprio usuário. Mas aonde tudo isto chegará?

Este trabalho elegeu a Folksonomia como seu objeto de estudo. A grande motivação foi verificar como as relações sociais se fazem presente na forma como os usuários nomeiam e organizam suas etiquetas ao armazenar documentos. Baseando-se na classificação de etiquetas de Cañada (2008), analisou-se como se encaminham as etiquetas de documentos em um site de fotografias.

1.1 O problema investigado e o objetivo do trabalho

Seria possível a organização de conteúdos na web? Como organizar o que nasceu sob a ordem de não haver ordem alguma? Entre os significados do verbo ‘organizar’ no dicionário Aurélio (FERREIRA, 1993) encontra-se o de “das às partes de (um corpo) a disposição necessária para as funções a que ele se destina”. Talvez esteja nesta frase a justificativa da necessidade desta organização.

A forma utilizada para a nomeação de etiquetas em ambientes folksonômicos tem sido objeto de estudo de diversas pesquisas. A Ciência da Informação, neste contexto, volta-se

para a compreensão de como reagem os usuários frente a esse novo processo de indexação. Se antigamente o profissional da informação era quem se colocava entre o produtor da informação e aqueles que dela se utilizariam, agora os papéis desses atores se fundem no usuário: é o proprietário da informação quem passa a dirigir e sugerir a indexação de seus documentos. Mais ainda, o usuário dessas informações também pode participar desse processo.

Assim, o problema eleito para este trabalho passa pela compreensão desse processo e pode ser expresso através das perguntas: as etiquetas informadas por um usuário proprietário podem ser classificadas quanto a sua natureza no que tange à sua usabilidade e à sua função frente a outros usuários? Chegará a um ponto onde todo o conteúdo desse processo terá um caráter de integração social, uma vez que a tendência da *world wide web* seja esta?

Existem problemas clássicos quando se trata de um tipo de classificação de etiquetas baseada em uma linguagem natural. A polissemia, a sinonímia, o grau, o gênero e a ambigüidade de termos, por exemplo, são pontos difíceis de serem tratados, mas com soluções já conhecidas nas linguagens de indexação. Mesmo assim, apesar da desordem, é possível conviver com estes problemas ao se utilizar os mecanismos de busca disponíveis na web e alguns já possuem soluções não completas a estes problemas. O que não significa que esta convivência ou mesmo estas soluções consigam sufocar o anseio de que a evolução e o uso destas ferramentas caminhem para uma recuperação de documentos com uma maior precisão e uma diminuição no grau de revocação.

Entender qual a expectativa/perspectiva de um usuário, ou a falta dela, pode se tornar chave para alguma organização comercial que lida com produtos derivados do processo de socialização da Internet. O mundo dos negócios na web pode canalizar seus esforços para a busca de uma possível clientela se souber de maneira aproximada o que lhe agrada, e como atingir o interesse dessa clientela sem que muitas vezes ela perceba.

Assim como alguns sistemas de compra *online* combinam preferências de compra, companhias voltadas para o relacionamento (tais como compartilhamento de vídeos, documentos, fotografias, *websites* ou *sites* de relacionamento) podem analisar a natureza das etiquetagens em seus respectivos sistemas e, assim, atingir de maneira mais precisa seu usuário. A maioria dos sites de comércio eletrônico organiza de alguma forma seu conteúdo favorecendo o usuário no encontro daquilo que busca.

A Ciência da Informação encontra na Web 2.0 alguns desafios que surgem pelo fato do usuário, neste contexto, não somente ser aquele que deseja e busca a informação, mas também por ser um agente de produção e disseminação dela. Um pouco mais até, pois através

dos recursos que nela estão disponíveis, ele é capaz de influenciar a relevância do conteúdo produzido nos processos de recuperação.

Entre muitos, a Folksonomia representa um desses desafios. O ato de nomear etiquetas que servirão de base em pesquisas de conteúdos remete a mecanismos já conhecidos pelos cientistas da informação, pela semelhança que existe entre este processo e os de classificação de informação. A nomeação de palavras-chave em textos científicos é exemplo disto. E também na Folksonomia é através da linguagem natural que os mecanismos de busca se sustentarão. Então, retorna-se ao confronto das vantagens e desvantagens da linguagem natural e da linguagem de indexação em processos de indexação e recuperação de informação.

O objetivo principal deste trabalho tornou-se então compreender e analisar as formas de classificação de conteúdos de etiquetas nomeadas pelo usuário dispostas na teoria, mais especificamente as propostas por Cañada (2008).

Como objetivos específicos podem ser citados: compreender o processo de nomeação de etiquetas em sistemas folksonômicos; compreender e avaliar os métodos de classificação de etiquetas propostos na teoria; verificar em estudo de caso se as classificações propostas estão adequadas para análise de objetos do tipo fotografia.

A Web 2.0 tem multiplicado as possibilidades de interação do usuário de diversas maneiras. *Websites* passaram a ter uma usabilidade cada vez mais voltada a quem os acessa, de maneira que a inserção e a classificação de informação tornaram-se algo mais amigável e próximo dos usuários. Entender como os consumidores manipulam essas ferramentas e de que maneira eles estão interessados em contribuir (quer seja para eles próprios, para o sistema/*website* ou para a comunidade de usuários) é de fundamental importância para o sucesso desses sistemas em termos de uso, acesso, confiança e contribuição. A organização e a coerência informacional nesses casos ficam em segundo plano, uma vez que o sucesso de sistemas folksonômicos se dá muito mais pela liberdade de criação, usabilidade e inserção de documentos do que necessariamente o alto grau de precisão das buscas realizadas. Dentro dessa perspectiva, o usuário passa a criar uma expectativa sobre esses sistemas, e passam a depositar sua confiança, tempo e dinheiro nos mesmos.

Em termos informacionais, os sistemas que possuem um alto índice de uso podem ter uma chance maior de desenvolvimento de possibilidades em melhorias na recuperação e na disseminação da informação, contribuindo assim para o próprio uso. No que tange ao contexto social, o entendimento das características de etiquetagem dos sistemas pode ser um norte de como gerenciá-lo. Saber qual a preferência dos usuários na etiquetagem, como manifesta seu interesse de integração social, ou mesmo o interesse em facilitar a sua própria

recuperação de informação, pode auxiliar na criação de métodos para a orientação de novos e experientes usuários desse sistema. E nisto justifica-se o estudo proposto.

2 FOLKSONOMIA

Folksonomia é a classificação de documentos ou objetos a partir de uma linguagem natural, indexada pelos próprios usuários de certo sistema. É um neologismo criado por Tomas Wander Val com a junção dos termos *folks* (pessoas) e *taxonomy* (taxonomia). Esta própria união sugere seu significado como sendo “taxonomia pelas pessoas”.

Referenciada algumas vezes como “vocabulário descontrolado”, a folksonomia é um sinal claro do momento atual onde a grande rede deixa de ser estática para ser cada vez mais dinâmica. A Web 2.0 trás um conjunto de novas características à *world wide web* que busca uma maior interação do usuário com o sistema no qual ele navega, através de personalização de interfaces e conteúdos colaborativos. Como afirma Bosch (2007), a inovação vem pelo uso, e a diferença entre um internauta dos primeiros tempos da web para o de agora não está tanto nas ferramentas técnicas, mas em seu modo de uso, que não mais se limita à aplicação das mesmas formas previstas por quem as concebeu. É a construção da informação com cada vez mais mediadores.

Como trata-se de um instrumento que serve a uma comunidade de usuários, observa-se que cada grupo de usuários possui suas próprias características de indexação e suas prioridades nas facetas escolhidas para indexação dos objetos. Assim, em comunidades diferentes um mesmo documento pode ser tratado de maneira distinta, dentro das características da inteligência coletiva desenvolvida naquele grupo.

A folksonomia engloba uma série de facilidades aos usuários, entre elas, a possibilidade de se recuperar os documentos por sua autoria, pela data de postagem, pelo título do documento ou por parte dele, a possibilidade de recuperação por termos informados nas tags e a Nuvem de Tags (*tagcloud*). Uma nuvem de tags é um conjunto de etiquetas exibidas em tamanhos diferentes, sendo que o destaque no tamanho de exibição é obtido por critérios específicos do sistema. Alguns atribuem este privilégio pela frequência com que determinada etiqueta é requisitada na busca de informação, outros, e de forma mais comum, à frequência com que uma etiqueta é utilizada pelos usuários do sistema. Ou então, um destaque dentro da frequência de utilização das etiquetas por determinado usuário quando a consulta é localizada no conjunto de documentos de um usuário específico. As nuvens de tags são dinâmicas e representam um meio de navegação entre um conjunto de documentos, uma vez que é possível acessar diretamente documentos por etiquetas informadas nessa nuvem.

Também através desse recurso o usuário obtém facilidades para visualizar quais as etiquetas são mais usadas em uma comunidade. Isso funciona como uma orientação para futuras postagens de documentos. O problema desse tipo de orientação é que o usuário passa a marcar suas fotos tendo como primeiro princípio o uso das etiquetas que sobressaem, e não necessariamente pelo grau de relação entre o termo indicado e o documento postado.

O primeiro site a utilizar a folksonomia foi o *del.icio.us* (www.delicious.com), lançado em 2003, que permite ao usuário indicar e encontrar sites relativos a qualquer assunto que estejam em suas listas de favoritos ou na de outros usuários. Trata-se de um compartilhamento de endereços favoritos ou mesmo de uma organização pessoal para endereços favoritos, pois ao usuário é permitido liberar ou não o acesso à sua lista.

Com o passar dos anos e o sucesso desse sistema, outros sites começaram a usar a folksonomia e se destacaram. Dentre os grandes hoje existentes além do *del.icio.us* destacam-se ainda o Digg, o YouTube, o Technorati, o Last.fm e o Flickr:

O Digg (www.digg.com) é um site colaborativo onde seus usuários postam incessantemente websites com as mais diversas notícias. O site possui divisão por assuntos e por tempo de postagem, mas não permite ao usuário a inserção de etiquetas relacionadas. Os *websites* com notícias interessantes são conseqüentemente acessados com mais freqüência, e isso lhes dá certa posição de destaque na página inicial. É possível visualizar um breve resumo da notícia, a quantidade de comentários a respeito daquele site e o usuário que o postou.

O YouTube (www.youtube.com) disponibiliza vídeos online a qualquer usuário que o acesse. O site possui uma ferramenta de *upload* fácil e permite que o usuário adicione etiquetas para facilitar a recuperação, manifeste sua opinião sobre os vídeos e crie grupos de usuários orientados por interesses comuns.

O Technorati (www.technorati.com) é uma máquina de busca em tempo real para dados gerados por usuários, o que incluem os blogs por etiquetas ou palavras-chave. Disponibiliza também um *ranking* de popularidade entre eles.

Na Last.fm (www.last.fm) os usuários criam rádios pessoais e podem indexar seus artistas ou estações preferidas e compartilhar com usuários que buscam uma mesma afinidade musical.

Esleveu-se o Flickr (www.flickr.com) como objeto de pesquisa deste trabalho e ele é um site utilizado como uma fonte de documentos fotográficos por usuários de diversos países. Possui um alto grau de visitas, estando na 38ª posição do *ranking* de sites mais acessados de acordo com o site Alexa (2008). Foi desenvolvido no Canadá pela empresa *Ludicorp* e

lançado em fevereiro de 2004, mas em 2006 teve seus direitos comprados pela *Yahoo! Inc.* Dentre os sites que usam da folksonomia como modo de indexação documental, o Flickr tem seu lugar de destaque, pois é um site bastante amplo, com usuários espalhados por todo mundo, e com um imenso volume de fotografias.

A busca básica no Flickr possui duas opções: uma pelo “texto completo”, onde o termo informado é buscado no título da foto, na descrição, nas etiquetas e nas comunidades às quais a foto pertence. A outra busca é baseada simplesmente nas etiquetas que os usuários determinaram às suas respectivas fotos, que foi a busca realizada por esta pesquisa. Também é possível encontrar certo documento passando não tão somente pelo texto e etiquetas, mas também pelo Exchangeable Image Files (EXIF). O EXIF é um padrão de metadados de especificação técnica, tais como a câmera usada para a foto, a data da foto, o formato e o tamanho do arquivo, o tipo de exposição, a abertura das fotos. Estes dados são gravados no arquivo automaticamente pela maior parte das câmeras digitais modernas, e sua exibição no site é submissa à permissão do autor. Os dados EXIF não são considerados como etiquetas, mas influem na busca gerada. O site disponibiliza ainda buscas pré-determinadas por câmeras baseadas somente pelo EXIF.

O Flickr é aberto a qualquer usuário na internet, mas apenas os cadastrados podem postar fotos, inserir etiquetas e se integrar a comunidades. Sua interação quer ir além do acesso do computador tradicional facilitando a adição e visualização de fotos “de todas as maneiras possíveis: da Web, de dispositivos móveis, computadores domésticos ...”, de acordo com o próprio site.

No Flickr a indexação de documentos é feita através de dados do documento como título, autor, ano, ou como já dito, pelos dados do EXIF ou pelas etiquetas. Para um documento postado, o usuário e outros membros podem adicionar etiquetas que, para eles, tenham alguma relação com o tema do documento. Embora nesse sistema cada documento possa receber até setenta e cinco etiquetas, nem sempre esse número é alcançado e, muitas vezes, existe uma grande incidência de sinônimos.

Compartilham o Flickr fotógrafos profissionais e amadores (que postam incansavelmente seus documentos), e usuários que gostam de apreciar ou ter uma ótima base fotográfica. Assim, possui diversificados tipos de usuários, cada um considerando uma forma própria de classificação e organização de documentos, com base em sua própria orientação e conhecimento, ou pelo aprendizado e observação de como fazem outros usuários.

Além da facilidade de recuperação de imagens a partir de uma única etiqueta, os usuários do Flickr criam grupos de etiquetas específicas para facilitar a interação entre os

mesmos, ou mesmo compartilhando fotos de seus computadores pessoais através de softwares de mensagens instantâneas (STURTZ, 2004). Outras etiquetas são criadas especificamente para grupos de usuários de certas marcas ou modelos de máquinas fotográficas. Grupos relacionados a essas máquinas ou acessórios para as mesmas são criados a fim de agrupar usuários com as mesmas afinidades, discutindo problemas e soluções. Nesse caso, as etiquetas facilitam a recuperação de fotografias com a mesma câmera, a mesma marca, filme, etc. Para facilitar o encontro de usuários, os próprios autores etiquetam suas fotografias com seus próprios nomes ou apelidos. Essas etiquetas citadas correspondem a uma derivação do conceito original das *tags*: etiquetar o documento de acordo com o assunto do qual ele trata. Assim, utilizam etiquetas específicas desse site, dessa comunidade, que facilitam a recuperação de documentos a usuários mais experientes.

3 TIPOLOGIA DAS ETIQUETAS

A maneira com que cada grupo de usuários realiza a etiquetagem (*tagging*) de seus documentos pode determinar o desempenho de um determinado sistema. O que irá determinar essa maneira é a unidade de interesse daquele grupo e suas características. A soma dos comportamentos individuais em prol do grupo caracterizará uma maior democratização e beneficiamento daquele sistema.

De acordo com Cañada (2008), existem quatro tipologias específicas de etiquetagem. Cada uma delas possui um determinado benefício social e uma motivação para que seja usada. Para o autor, de acordo com seu conteúdo elas podem ser descritas como etiquetas egoístas, etiquetas amigáveis, etiquetas altruístas e etiquetas populares.

Uma etiqueta egoísta é aquela voltada para o uso pessoal de seu criador. O conteúdo dessas etiquetas possui bastante significado para quem o criou, sem necessariamente ter um sentido fora de seu contexto individual. Costumam auxiliar o criador para a recuperação de arquivos pessoais, ou sistemas de classificação que ele próprio criou para sua organização de documentos, quando seu acervo se torna grande. Por exemplo: “feriasjan3”, ou “para ler”.

O benefício social de tal etiquetagem é de médio a baixo, pois o uso de diversas classificações individuais em um sistema trará a este uma grande quantidade de ruído, e depende da maturidade de quem etiqueta. Já a motivação para realizá-la é alta, uma vez que quem a faz, a faz para benefício próprio.

A etiquetagem amigável é aquela reservada a um grupo fechado. Usam-se termos consensuais e geralmente desconhecidos para quem está fora desse grupo. Por exemplo: “Bienal da tecnologia 2009”. Seu benefício social é alto, pois, embora para grupos muito

grandes este tipo de etiquetagem possa perder sua razão, é extremamente útil para grupos menores. Possui também uma alta motivação para que seja usada, pois ela reforça e estreita os laços entre os membros daquele grupo.

Altruísta é uma etiquetagem com o maior propósito de compartilhamento com o resto do mundo. Busca a descrição real do que está presente no documento, para que quem procure aquele determinado tema possa ter acesso a um conteúdo coerente com a pesquisa realizada. Por exemplo: “Cristo Redentor”, “esporte”. O benefício social dessa é bastante alto, uma vez que são escolhidos conteúdos conhecidos pela comunidade usuária para as etiquetas e que estas são as mais descritivas possíveis para aquele documento. Além disso, é o tipo mais útil para a Ciência da Informação, uma vez que facilita a recuperação da informação contribuindo diretamente para o aumento da precisão e da diminuição da revocação de uma busca (CAÑADA, 2008). Embora de todas as tipologias seja a mais colaborativa, a motivação para usá-la é baixa, já que é trabalhoso realizar uma indexação coerente principalmente quando o benefício próprio não é direto.

A etiquetagem popular é aquela que busca apenas crescer o número de acessos àquele documento. Nesse aspecto, qualquer termo que possa trazer popularidade ao documento poderá ser usado, mesmo que não exista coerência entre a busca e o documento. Por exemplo, numa foto de um carro Mercedes-Benz colocar etiquetas com conteúdo “Fiat”, “Ferrari” e “Alfa Romeo”. Assim, o benefício social é praticamente nulo, uma vez que dificulta a precisão e revocação. Já a motivação passa a ser alta, pois quem a usa busca um benefício próprio.

4 METODOLOGIA

A metodologia escolhida foi a de estudo de caso. A escolha do Flickr como site para estudo neste trabalho deu-se pela preferência de seus autores por fotografias e por este site disponibilizar um conjunto de serviços que facilitaria esta pesquisa. Além disso, embora outros sites usem a folksonomia como maneira de indexação de documentos, o fato do Flickr lidar com fotografias facilita este estudo, uma vez que para se saber o conteúdo de vídeos, músicas, *websites* ou artigos, é necessário um tempo maior para lê-los, assisti-los ou ouvi-los. A fotografia é de simples observação, e o tempo gasto em sua visualização é mínimo.

Após a recuperação e a leitura de bibliografias, foram seguidos os seguintes passos: a escolha do site de estudo de caso e a escolha da classificação a ser averiguada no trabalho: optou-se pelo Flickr por razões já expostas e pela classificação de Cañada(2008); a escolha

das etiquetas para análise dos dados, melhor detalhada a seguir; levantamento, análise dos dados e conclusão.

As consultas foram feitas em datas diferentes, utilizando sempre o mesmo tipo de busca. A análise dos dados deu-se depois de efetivada todas as consultas.

Anteriormente à busca dos documentos, estabeleceu-se que a busca de documentos seria por etiquetas, serviço esse disponibilizado no Flickr, e que não consideraria nesta opção a presença da palavra no título, na descrição e no EXIF do documento.

A escolha das etiquetas a serem buscadas seguiu os seguintes passos:

- Passo 1: Foram entrevistados aleatoriamente dez alunos do curso de Ciência da Informação da PUC Minas, e a eles foram solicitados sete temas, que neste trabalho foram encarados como etiquetas para uma busca fotográfica. Ao final desse primeiro passo, setenta etiquetas foram citadas, numeradas de um a setenta, seguindo a ordem de entrevista do aluno e da resposta de cada um deles;
- Passo 2: Optou-se por realizar a busca em dez etiquetas deste conjunto, que foram escolhidas iniciando o intervalo através do sorteio de um número entre um e sete equivalendo à primeira etiqueta escolhida por equivalência direta na lista. A partir deste número somou-se o número sete por nove vezes. A cada soma, tomava-se uma etiqueta correspondente àquele resultado e assim sucessivamente.
- Passo 3: uma vez escolhidas as etiquetas, realizou-se uma busca do tipo “por etiquetas” no Flickr, sendo considerados apenas o primeiro documento retornado de cada busca;
- Passo 4: as etiquetas presentes em cada documento encontrado foram classificadas de acordo com o sistema proposto por Cañada (2008).

Com relação às etiquetas encontradas nos documentos selecionados, os critérios que orientaram sua seleção foram:

- etiquetas que possuíam relação de sinonímia com outras etiquetas, etiquetas com diferenciação por número, grau e gênero, ou etiquetas de mesma origem com diferentes conjugações foram consideradas como sendo etiquetas distintas (exemplo: “explore” e “explored”);
- etiquetas sinônimas, mas escritas em idiomas diferentes foram consideradas como sendo etiquetas distintas;

- etiquetas que são abreviaturas ou junções de termos compostos foram consideradas individualmente, embora com o mesmo conteúdo ou significado. Ex: “Rio”, “de”, “Janeiro”, “RJ” e “riodejaneiro”;
- etiquetas informadas em duplicidade por usuários distintos foram mantidas na lista inicial.

5 RESULTADO E ANÁLISE DE DADOS

Foi encontrado um total de 306 etiquetas nos dez documentos selecionados. A Tabela 1 apresenta os termos selecionados no sorteio o número de etiquetas presentes em cada documento. Observa-se que os termos foram empregados conforme informados pelos usuários entrevistados, ou seja sem tratamento de gênero e número.

As etiquetas de cada termo foram então analisadas segundo a classificação de Canadá(2008) e a Tabela 2 apresenta o total de etiquetas encontradas em cada uma dessas categorias. Cada etiqueta foi enquadrada apenas em uma categoria de Canadá. A Figura 1 ilustra estes dados.

Tabela 1: Etiquetas sorteadas e número de etiquetas por documento

Termo	Número de etiquetas do primeiro documento recuperado
Antiguidade	7
Paisagens	3
Crianças	48
Artistas	48
Balões	27
Deserto	17
Água	19
Exposição	48
Natureza	60
Carros	29
TOTAL	306

5.1 Etiquetagem altruísta

Numa primeira análise, nota-se uma grande diferença nos valores relacionados à etiquetagem do tipo altruísta frente às demais etiquetagens, uma vez que o número de etiquetas com esse caráter é maior que o dobro do segundo colocado, a etiquetagem do tipo amigável. Um número alto para este tipo de etiquetas era esperado. Sistemas que utilizam Folksonomia exploram de seus usuários a vontade de que seus objetos sejam conhecidos. Quanto mais organizada a comunidade que se forma em torno do objetivo do sistema, maior o grau de especialização que a própria comunidade exige, por sua utilização, nos conteúdos informados.

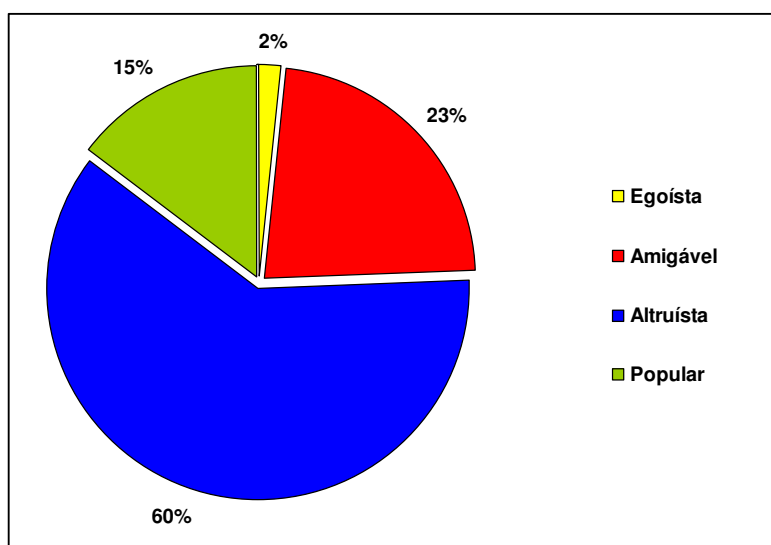


Figura 1: Gráfico com porcentagem de tipologia das etiquetas

Os números demonstram então um grande interesse na coerência da etiquetagem, ou seja, as etiquetas necessitam ter um conteúdo que seja de fato representativo da fotografia. Isto é de extrema utilidade para o crescimento do sistema e para o seu bom uso, uma vez que as buscas então têm uma maior probabilidade da satisfação do usuário. A recuperação torna-se mais fácil, aumenta-se a precisão e revocação de imagens relacionadas à busca realizada por conteúdo de etiquetas, uma vez que o Flickr oferece outras opções de busca.

5.2 Etiquetagem amigável

Na segunda colocação em números de etiquetas, vem a etiquetagem do tipo amigável, representando 23% do número total de etiquetas recuperadas. Este é um valor considerável, que representa o interesse que a comunidade de usuários do Flickr tem por uma interação social. A maioria das etiquetas assim classificadas refere-se a nomes de comunidades internas

do Flickr, o que mostra que um grande número de usuários elege esta forma como um modo de recuperação de suas imagens. Isso é explicável por dois motivos: o primeiro é que à medida que um usuário participa de uma comunidade, este vai estreitando os laços com os membros daquela comunidade, e assim pode se desenvolver alguma afinidade de gostos fotográficos ou possíveis amizades. Essa interação pessoal é uma das bases do Flickr, uma vez que é um sistema baseado principalmente na arte, na expressão pessoal de seus usuários.

O segundo motivo dá-se pelo fato de que existem diversas comunidades controladas pelos seus administradores foto a foto, as chamadas comunidades “fechadas”. Nessas comunidades, para que alguém se candidate a membro, os administradores e moderados devem investigar o trabalho fotográfico daquele membro e julgar se este é digno ou não de participar daquela comunidade. Em outras comunidades observa-se que, mesmo sendo livre a entrada de usuários, as fotos candidatas à galeria daquela comunidade devem passar pelo crivo de seus administradores. Algumas galerias de comunidades possuem um nível de qualidade fotográfico tão alto, que apenas o fato de uma foto pertencer àquela galeria, já se considera um prêmio para o seu autor.

Existem ainda as comunidades com apenas fotografias convidadas, ou seja, os administradores navegam pelo sistema e assim que encontram uma boa foto, pedem autorização ao autor da foto e, se autorizados, a adicionam às suas respectivas comunidades. Nesse contexto, é importante o usuário ter etiquetas que mostrem a quais comunidades sua foto pertence, pois isso pode representar um certo *status* dependendo de quais comunidades aquela foto pertence.

5.3 Etiquetagem popular

Essa etiquetagem é talvez a que menos possa variar em termos de folksonomia, uma vez que é inerente a qualquer sistema aberto o interesse na promoção pessoal por parte de alguns membros. Esta etiquetagem visa a um aumento no número de visitas a um documento para a popularização do mesmo. Para que isto ocorra, o autor (ou quem apenas posta um documento sem necessariamente ter a autoria do mesmo) recorre a diversas etiquetas que podem ou não especificar o conteúdo da foto.

A porcentagem de etiquetas que se enquadraram nessa classificação foi de 15%, uma quantidade considerável, levando-se em conta o sistema usado para a busca (melhor especificado no item 2.2). Acredita-se que o ideal seria que este número sempre se aproximasse de zero, uma vez que ele é prejudicial à boa recuperação de informação, mas 15% é um total tolerável, posto que o Flickr é um sistema aberto.

Para a Ciência da Informação etiquetas classificadas neste grupo tornam-se um desafio, pois não representam o conteúdo dos documentos armazenados. Vence-lo exige esforços e medidas que passem pela conscientização de usuários para promoção de um alto grau de precisão em buscas realizadas. Enquanto o usuário não se conscientizar disto, o desejável será praticamente impossível.

5.4 Etiquetagem egoísta

Este tipo de etiquetagem foi o que retornou o menor número de etiquetas: cinco em 306 (aproximadamente 2%), observando que apenas cinco usuários a usaram e em cinco documentos distintos (uma etiqueta egoísta por documento). Para os mesmos documentos, estes usuários ainda informaram também etiquetas de outras categorias. Pode não ser de interesse de diversos usuários a popularização de seus documentos, e assim eles criam etiquetas que somente eles próprios usam. Nesse aspecto, existem aqueles que postam documentos e que colocam uma quantidade pequena de etiquetas (de três a seis) e os que não etiquetam seus documentos. No primeiro caso, se existem documentos onde a quantidade de etiquetas é menor, e ainda do tipo egoísta, diminuem-se as chances desses documentos serem recuperados e também exibidos no topo da busca. Já no segundo caso, esses documentos simplesmente não serão recuperados. Embora 2% seja um valor pequeno numa classificação de apenas quatro categorias, acredita-se que este valor na realidade pode ser maior, e devido à dificuldade de recuperação dessas etiquetas, pode ter sido mascarado neste trabalho.

Considerando que um usuário crie um sistema de classificação levando em consideração a inicial do nome do autor, o ano, o mês e o dia de criação do documento, então um exemplo seria a etiqueta P8412, sendo P a inicial do nome do autor e o número 8412 referente à data 12/04/2008. Embora essa etiqueta seja classificada como egoísta, não há como simplesmente adivinhar uma busca por P8412 para que esse documento seja recuperado e essa etiqueta classificada. Portanto, existem etiquetas egoístas que acredita-se que jamais serão recuperadas e contabilizadas numa pesquisa.

No caso específico desta amostra, como estas etiquetas foram informadas junto a outras de categorias distintas, acredita-se que essa utilização seja apenas como forma facilitadora a seus proprietários para a recuperação.

5.5 Considerações acerca dos resultados

O número de etiquetas altruístas foi massivo em relação aos outros tipos de etiquetas. Porém não se pode, a partir disso, declarar que a folksonomia tem um caráter de total

altruísmo, onde os membros de um sistema que a utilizam têm total propósito e comprometimento com a recuperação coerente de informação. O total de etiquetas classificadas como altruístas representa 60% do total de etiquetas recuperadas e embora seja um valor significativo, pode não ser suficiente para realizar tal declaração.

Existe também um fator agravante para a boa classificação de etiquetas: a ambigüidade, no caso mais específico, a ambigüidade morfológica, “quando não é possível classificar determinada forma quanto à categoria gramatical” (BRÄSCHER, 2002). A palavra “compra”, por exemplo, pode representar um substantivo ou um verbo. Como a etiqueta é na maioria das vezes formada por palavras isoladas, não há como saber a intenção real do autor. Além da ambigüidade morfológica, podem também ser consideradas aqui as resultantes de homonímias, como ‘manga’, ‘fio’, etc.

Outro fator a ser considerado nesses dados é a fonte do qual eles foram retirados. O Flickr é um *site* no qual sua maior parte de uso é feita por fotógrafos e amantes da fotografia artística. Por certo existem usuários que não têm esse perfil e usam o sistema para postagens pessoais, buscando promoção pessoal, ou mesmo o simples armazenamento de imagens. Esse público certamente não representa a maioria dos usuários. Assim sendo, é bem provável que em outros sistemas folksonômicos os resultados de uma classificação, como esta, podem ser bastante distintos dos aqui encontrados.

6 CONCLUSÃO E TRABALHOS FUTUROS

Analisando algumas características da folksonomia foi possível constatar um deslocamento da orientação de indexação de documentos, manifestada no ambiente web. Desde o início das bibliotecas, entre aquele que produzia algo e os usuários do que foi produzido, existe a pessoa do indexador, responsável por ler e analisar esta produção, traduzindo seu conteúdo para uma linguagem específica que favorecesse sua disseminação. Isto não mudou, mas a folksonomia de forma ímpar muda agora o papel do produtor e do usuário da informação. No instante que algo passa a ser disponibilizado na web, o proprietário, ou o produtor, do documento, passam a informar palavras-chave, conhecidas como etiquetas nesta filosofia, pelas quais os usuários passam a recuperar a informação. Mais ainda, ao usuário é facultado o direito de inserir outras palavras-chaves como que validando a garantia do usuário sempre respaldada nos processos de indexação.

Assim, um estudo como o realizado por Cañada, cuja filosofia apoiou este trabalho, é de profunda importância e aponta a necessidade de que profissionais de informação conheçam

tendências com que os usuários - produtores e usuários de produtos – nomeiam suas etiquetas neste novo ambiente de disseminação que nos trouxe a Folksonomia.

Se estes usuários, em sua grande maioria, desconhecem sistemas de classificação, indexação ou taxonomia, e são os responsáveis pela disponibilização de documentos na web, que conseqüências passaremos a ter na recuperação da informação neste ambiente?

Com os diversos perfis de uso e a impossibilidade de orientação completa a todos os disponibilizadores de documentos, torna-se impossível uma total organização do conteúdo da web, no que tange documentos regidos pela folksonomia. Mas, por mais que os sistemas possam ser eficientes e orientados ao usuário, possibilitando um uso que possa ir de acordo com princípios organizacionais, haverá uma parcela de usuários que não seguirão tais princípios, quer seja por desinteresse, por promoção social ou por simples escolha. A integração social é inerente a sistemas assim regidos (haja vista que a maioria das etiquetas pesquisadas se enquadra na classificação como sendo altruísta). Se o inverso ocorrer, ou seja, se prevalecerem etiquetagens egoístas, populares, ou somente para grupos fechados, a web 2.0 perderá uma de suas principais características: a interação e com ela, a de inclusão social. E ainda, com o aumento de documentos disponíveis em cada sistema, tornar-se-á cada vez mais inviável uma consulta que possa trazer uma alta precisão e alta revocação.

Cada sistema folksonômico possui uma característica distinta de comportamento de seus usuários, o que pode levar à pesquisas futuras sobre como seria o comportamento de um sistema folksonômico para blogs, sites de armazenamento de vídeo ou *bookmarks*.

Os sistemas de indexação evoluíram ao longo dos séculos tendo em vista uma maior capacidade de recuperação e disseminação de informação. A web evoluiu por seu alto grau de utilização, mas mantêm-se como desafio à recuperação de informação. A Folksonomia veio com a força do usuário, e assim, pode alterar e reverter tudo isto.

FOLKSONOMY AND CLASSIFICATION OF LABELS: case study Flickr

ABSTRACT

The web's evolution has established a new level of users interactivity. If some time ago all the stored content was indexed by search engines heuristics, now this process counts with the participation of the ones who make the documents available and the ones who use them. Folksonomy, through tagging usage, represents a shift of the organization process and information retrieve on the web. This approach presents a content classification study of tags posted by users. Using the case study, photography tags of the Flickr website were classified according with Cañada's classification. To conclude, most of the tags correspond to the correct information about the photography content, which shows that to this kind of system, there is an interest in sharing and contribution of coherent information.

Keywords: Flickr. Folksonomy. Web 2.0. Collaborative Internet. Cañada

REFERÊNCIAS

ALEXA: **The Web Information Company**. Disponível em <www.alexa.com>. Acesso em 20 de maio de 2008.

BLATTMANN, Úrsula; SILVA, Fabiano Couto Corrêa da. **Colaboração e interação na Web 2.0 e biblioteca 2.0**. Revista ACB: Biblioteconomia em Santa Catarina, Florianópolis, v.12, n.2, p. 191-215, jul./dez., 2007.

BRÄSCHER, Marisa. **A ambigüidade na recuperação de informação**. DatagramaZero – Revista de Ciência da Informação – v.3 n.1 fev/02. Disponível em: <http://repositorio.bce.unb.br/bitstream/10482/924/1/ARTIGO_AmbiguidadeRecuperacaoInformacao.pdf>. Acesso em 05 de Abril de 2009.

BOSCH, Mela. **Entre la heurística y la hermenéutica: las competencias y actitud de los profesionales de la información para la web emergente**. Fundación Ciencias de la Documentación. Mar., 2007.

CAÑADA, Javier. **Tipologías y estilos em el etiquetado social**. Disponível em <<http://www.terremoto.net/tipologias-y-estilos-en-el-etiquetado-social>>. Acesso em: 10 de Novembro de 2008.

FERREIRA, Aurélio Buarque de Holanda. **Minidicionário da língua portuguesa**. 3.ed. Rio de Janeiro: Nova Fronteira, 1993.

LÉVY, Pierre. **A inteligência coletiva**. São Paulo: ed. Loyola, 1998.

O'REILLY, Tim. **What Is Web 2.0?: design patterns and business models for the next generation of software**. 2005. Disponível em: <<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>>, Acesso em: 20 de Março de 2009.

PRESCOTT, Roberta. **A nova fronteira da gestão do conhecimento**. Disponível em IT Web <http://www.itweb.com.br/noticias/index.asp?cod=46550&utm_source=itweb&utm_medium=portal&utm_content=tickersite>, Acesso em 14 de Novembro de 2008.

ROSS, Joshua-Michéle. **The rise of the social nervous system**. Disponível em: <<http://www.forbes.com/2009/03/09/internet-innovations-hive-technology-breakthroughs-innovations.html>>. Acesso: 25 de Março de 2009.

STURTZ, David N. **Communal categorization: the folksonomy**. INFO1622: Content Representation, Dez., 2004. Disponível em <http://davidsturtz.com/drexel/622/sturtz-folksonomy.pdf>. Acesso: 28 de Outubro de 2008.