



XI Encontro Nacional de Pesquisa em Ciência da Informação
Inovação e inclusão social: questões contemporâneas da informação
Rio de Janeiro, 25 a 28 de outubro de 2010

GT 2: Organização e Representação do Conhecimento
Modalidade de apresentação: comunicação oral

**CLASSIFICAÇÃO COLABORATIVA DE CONTEÚDOS NÃO-TEXTUAIS NA INTERNET:
AS NOVAS FORMAS DE MEDIAÇÃO E ORGANIZAÇÃO DA INFORMAÇÃO DA
MÚSICA ATRAVÉS DA FOLKSONOMIA**

Rose Marie Santini
Rosali Fernandez de Souza
Instituto Brasileiro de Informação em Ciência e Tecnologia

Resumo: Este trabalho discute as novas formas de mediação e organização de conteúdos não-textuais na Internet. Os conceitos de folksonomia, social tags e classificação colaborativa para a recuperação da informação nos chamados Sistemas de Recomendação, cujo funcionamento é baseado na tecnologia de “Filtragem Colaborativa” (*Collaborative Filtering*). O caso analisado foi o Sistema de Recomendação Last.fm, plataforma amplamente utilizada em nível mundial para busca, recuperação e acesso ao material musical disponível na Internet. As análises foram realizadas em base das estatísticas de uso das tags neste sistema e os resultados encontrados demonstram que, em ambientes informacionais caracterizados pela abundância de informação, a folksonomia tende a apresentar utilidade e relevância social para as comunidades de prática. Os resultados indicam que o processo de classificação colaborativa da música resulta em meta-dados relevantes para a recuperação da informação. A folksonomia potencializa a inclusão de vocabulários e necessidades informacionais da maioria dos usuários que participam da classificação e enfraquece os julgamentos parciais ou particulares - que podem ser comerciais, políticos ou institucionais. Conclui-se que a classificação colaborativa faz surgir novos enunciados em proveito de outras instâncias de organização da informação que tendem a assegurar a multiplicidade de conexões entre pessoas e conteúdos na Internet.

Palavras-chave: Classificação colaborativa. Internet. Música. Sistemas de Recomendação. Folksonomia.



INTRODUÇÃO

Uma série de mudanças na tecnologia, na economia e nas práticas sociais caracteriza a emergência de novas formas de organização social na chamada sociedade da informação. Os padrões de organização social emergentes nas economias mais avançadas caracterizam-se pela produção, mediação, distribuição e uso da informação em um ambiente de redes colaborativas, principalmente através da Internet. No entorno digital, diferentes oportunidades surgem e modificam as formas de uso da informação, do conhecimento e da cultura.

Essas transformações ocorrem através dos esforços individuais e coletivos em uma ampla variedade de tipos de colaboração, desde as mais sutis às mais firmemente tecidas. A emergência dessas práticas sociais apresenta sucessos notáveis em diversas áreas, destacando-se entre outras o desenvolvimento de softwares livres, a criação das “folksonomias” e filtragens colaborativas de informação através dos chamados Sistemas de Recomendação.

Os Sistemas de Recomendação (SR) são sistemas informáticos de classificação, organização e recomendação de informação sobre bens culturais baseados nas práticas e gostos dos usuários. Estes sistemas incluem uma tecnologia chamada de “Filtragem Colaborativa”, também utilizada como sinônimo de SR, para referir-se a um tipo específico de software no qual a filtragem da informação é realizada com auxílio humano - ou seja, com a colaboração da rede de usuários. Atualmente, a tecnologia da Filtragem Colaborativa é utilizada pelos mais importantes websites de recomendação de conteúdos na Rede, tais como *Youtube, Last.Fm, Amazon, Pandora, NetFlix, Google Books, Yahoo Music, MyStrands, Hunch*, etc. Estes sistemas são grandes plataformas interativas onde se exibem e se recomendam bens culturais aos usuários de acordo com seus interesses e preferências - tais como música, filmes, vídeos, livros, textos, etc.

No contexto colaborativo da Internet, as ações individuais descentralizadas passam a ter um papel muito maior do que tinham e poderiam ter na “economia industrial da informação”. Isto ocorre especificamente através da cooperação e ação social coordenada. Mesmo quando inseridos dentro de entornos comerciais, a colaboração dos usuários-voluntários na Rede possui motivações não-mercadológicas que independem das estratégias proprietárias.



O que caracteriza essas novas formas de uso da informação, do conhecimento e da cultura são os esforços efetivos de colaboração em larga escala - processo radicalmente novo, porém difícil para os observadores de acreditar. O surgimento e popularização dos Sistemas de Recomendação (SR) na Internet demonstram a expansão desse modelo não somente no centro das plataformas dos softwares livres, mas entre todos os domínios da produção e mediação da informação e da cultura. O objetivo comum envolvido na ação colaborativa dos usuários-voluntários em torno dos sistemas de informação baseados na tecnologia de Filtragem Colaborativa é aumentar a relevância e a credibilidade nos processos de busca de informação para cada usuário.

Consideramos a classificação colaborativa de conteúdos na Internet como um processo inovador de mediação social, por primeira vez produzida pelos próprios usuários. Portanto, os Sistemas de Recomendação conformam um lugar privilegiado de observação dos padrões de mediação descentralizados e não-comerciais que emergem na Internet. A colaboração gera a possibilidade de representar, através de novas categorias, a multiplicidade de desejos e necessidades dos usuários; aproximar pessoas com interesses semelhantes e distribuir oportunidades de atuação voluntária de forma direcionada.

Folksonomia e *Social Tags*: a classificação colaborativa de conteúdos na Internet

Nos últimos anos, o surgimento de websites como *Flickr*¹, *Del.icio.us*² e dos Sistemas de Recomendação baseados em Filtragem Colaborativa inauguram discussões acadêmicas sobre os usos sociais das *tags* (também chamadas de “anotação coletiva”). As *tags* são palavras-chave, rótulos ou breves descrições criadas pelos usuários para classificar conteúdos disponíveis na Internet a partir de vocabulário próprio.

Os usuários deste tipo de websites ou SR podem etiquetar (ou “*taggear*”) a informação de acordo com seus usos e interesses, e também consultar e acessar as *tags* aplicadas por outros. Quanto mais um item é etiquetado com uma determinada *tag* mais

¹ Site de compartilhamento de fotos onde os usuários organizam e classificam as imagens postadas através de *tags*. Ver: <http://www.flickr.com/>.

² Site onde as pessoas podem classificar através de *tags* os websites da internet e compartilhar suas classificações e *tags* com outros usuários. Ver: <http://delicious.com/>.



essa categoria de classificação é assumida como relevante pela rede de usuários e pelos próprios critérios de funcionamento dos sistemas. Uma coleção de *tags* em uma determinada comunidade de usuários é denominada “folksonomia”.

O termo folksonomia foi criado por Thomas Vander Wal em 2004 a partir de uma lista de discussão online entre membros do Instituto de Arquitetura de Informação (*IA Institute*)³ em que descreve a folksonomia como sendo “uma estrutura categórica criada pelos usuários, de baixo para cima, desenvolvida com um thesaurus emergente” (VANDER WAL, 2007).

No mesmo período, o fenômeno da classificação colaborativa na web foi denominado por outros autores como “folk-classificação”⁴, “indexação colaborativa”⁵, “classificação social”⁶, “etiquetas colaborativas”⁷, entre outros nomes. Porém, foi o conceito de folksonomia e “*social tag*” de Vander Wal que se consolidou no discurso acadêmico, impulsionado principalmente pela publicação do artigo de Gene Smith em 2004, intitulado “*Folksonomy: social classification*” (MATHES, 2004; WELLER, 2007; THOMPSON, 2008).

Segundo Lamere e Pampalk (2008), há uma diferença importante entre “*tags*” e “*social tags*”. As *tags* são as etiquetas individuais, ou seja, etiquetas pequenas e de texto livre aplicada a um conteúdo por um usuário específico. O vocabulário destas etiquetas não são estruturados e não possuem limites linguísticos. As *social tags*, também chamadas de folksonomia, se referem aos conjuntos de tags individuais compartilhados por uma comunidade de usuários.

Em um ambiente informacional onde a classificação dos conteúdos está baseada na folksonomia, o usuário pode consultar as tags dos outros para realizar uma busca ou classificar o mesmo conteúdo. Paralelamente, os sistemas gerenciadores de coleções organizadas por tags orientam os usuários a classificar os conteúdos de acordo com categorias comuns, utilizadas pela comunidade. Ou seja, a folksonomia (ou *social tags*)

³ Para mais detalhes sobre o nascimento do termo folksonomia, ver o site deste instituto de pesquisa no endereço: <http://iainstitute.org/>.

⁴ Tradução de “*folk-classification*”. Para mais informação ver a definição de Eric Scheid do *Instituto Ironclad Information Architecture*. Disponível em <http://www.ironclad.com.au/>.

⁵ Ver por exemplo Furner, Smith & Winget (2006).

⁶ Ver SMITH (2004).

⁷ Ver Golder & Huberman (2006).



corresponde ao conjunto das categorias criadas coletivamente e compartilhadas por uma determinada comunidade de prática em um dado domínio.

A classificação colaborativa (CC) - também conhecida como classificação social - é um *processo* no qual uma comunidade de usuários categoriza bens culturais de uma coleção para o uso dos próprios membros (MATHES, 2004). Quando há usuários e categorias suficientes, cada item disponível é classificado por múltiplas pessoas em um determinado período. O resultado potencial da classificação colaborativa é que cada recurso em uma dada coleção possa ser representado por um conjunto de etiquetas compartilhadas, porém geradas por diferentes pessoas.

Segundo Furner, Smith, & Winget (2006), a classificação colaborativa de recursos culturais é um método desenvolvido para aprimorar a interação entre os usuários e os itens disponíveis. As vantagens que justificam o uso desta abordagem em coleções online apóiam-se nas características distintivas da classificação colaborativa (CC) e definidas pelos autores, a saber:

a) CC é *distribuída*: nenhum usuário específico é responsável pela categorização de todos os recursos; e os recursos considerados isoladamente não precisam ser categorizados por todos os usuários.

b) CC é *barata*: a classificação dos itens de uma coleção é tipicamente realizada por usuários-voluntários que concentram seus esforços na categorização dos itens disponíveis, sem nenhum custo para os fornecedores/administradores dos conteúdos.

c) CC é *democrática*: os usuários que classificam os recursos não são selecionados por sua “expertise” pelos fornecedores/administradores dos conteúdos. Ao contrário, os usuários se auto-candidatam para classificar os recursos de acordo com seus interesses e objetivos.

d) CC favorece o “*empowering*”⁸: em um contexto de vocabulário controlado as pessoas estão acostumadas a fazer buscas em base de dados tentando prever os descritores e categorias usados pelos “especialistas”. A taxonomia demanda um

⁸ Não encontramos tradução adequada para a palavra “*empowering*”. Alguns autores traduzem para o português como “empoderamento”. Considerando que essa palavra não existe no dicionário da língua portuguesa, preferimos manter a palavra original em inglês. O termo “empoderamento”, em geral, significa a ação coletiva desenvolvida pelos indivíduos quando participam de espaços privilegiados de decisões e neste contexto se refere ao “ganho ou aumento de poder” por parte dos usuários em um domínio informacional.



conhecimento prévio e uma competência informacional por parte dos usuários. Com a CC, tanto o registro como a organização e a classificação dos recursos são desenvolvidos com base no conhecimento dos próprios usuários.

e) CC é *coletiva*: cada categoria de classificação criada nos sistemas colaborativos representa potencialmente o trabalho em conjunto de múltiplos usuários.

f) CC é *dinâmica*: a categorização de um dado recurso muda ao longo do tempo. Os usuários adaptam os sistemas de classificação na medida em que usam a informação. Ou seja, diferentes usuários classificam e fazem seus próprios julgamentos sobre a relevância de cada categoria enquanto utilizam os próprios sistemas de informação. A classificação colaborativa soma os conhecimentos de todos os usuários na criação de meta-dados que aumentam o desempenho e a eficácia dos sistemas nos processos de recuperação da informação.

A partir das características citadas por Furner, Smith, & Winget (2006), cabe acrescentar uma última vantagem da “classificação colaborativa” no âmbito das redes digitais. Na medida em que os próprios usuários criam os critérios para a organização da informação, os processos cognitivos passam a estabelecer uma correspondência direta com os esquemas classificatórios. Esse fenômeno aumenta a relevância dos resultados de busca e a credibilidade das recomendações de informação geradas pelos sistemas.

No contexto dos Sistemas de Recomendação (SR), a relevância é um critério “subjetivo” que julga a pertinência de um resultado gerado pela busca de informação com um fim particular, definido por um ou mais indivíduos. A credibilidade relaciona-se com a crença no valor social das categorias de representação dos conteúdos - ou seja, em que grau a classificação dos itens disponíveis é apropriada para os propósitos de uso e condizem com as formas de percepção coletiva que se conformam entre a rede de usuários (BENKLER, 2006). A distinção desses dois conceitos pode parecer artificial, porém a utilidade de uma informação nos SR dependerá da combinação da avaliação dos critérios de credibilidade e relevância por parte dos usuários.

Muitos investigadores se perguntam se é possível que os parâmetros de relevância e credibilidade sejam construídos através do modelo colaborativo, com a participação descentralizada de milhões de usuários de um determinado sistema de informação. A resposta é dada através da observação de sistemas comerciais como *Google*, *Amazon*, *Youtube*, *Last.fm* entre outros, que confiam na organização e produção colaborativa da



informação de seus usuários para orientar o funcionamento e a desempenho de seus próprios sistemas.

Conforme mencionado, a arquitetura e a tecnologia da Internet viabilizam condições sociais para o surgimento de práticas baseadas em uma “organização social em rede” onde a emergência das ações individuais descentralizadas passa a ter um papel muito maior do que tinham na “economia industrial da informação”.

A classificação colaborativa como fenômeno social

Alguns autores, apesar de reconhecerem a emergência dos fenômenos colaborativos na Rede, argumentam que a participação nos processos de colaboração são minoritários. Porém, as estatísticas mostram uma realidade oposta: as práticas colaborativas de classificação da informação, do conhecimento e da cultura demonstram tendências majoritárias entre os usuários da Internet.

Em dezembro de 2006 a consultoria *Pew Internet & American Life Project*⁹, em uma pesquisa realizada nos EUA sobre o comportamento das pessoas em relação ao fenômeno das “*social tags*” demonstrou que 28% dos usuários da Internet criam tags, categorizaram ou classificaram conteúdos online em algum momento, e esse número tende a ser maior ao longo dos anos.

De acordo com o relatório divulgado, diariamente 7% dos usuários da Internet dizem classificar conteúdos online através de tags. A Tabela 1 indica os resultados do estudo realizado pela *Pew Internet* (2006) sobre o perfil dos usuários-*taggers*.

⁹ Ver http://www.pewinternet.org/~media/Files/Reports/2007/PIP_Tagging.pdf.pdf

Tabela 1: Perfil demográfico dos usuários *taggers*

Perfil Demográfico dos Taggers	
Proporção do total de norte-americanos no grupo de usuários-<i>taggers</i>	
Homens	29%
Mulheres	27%
Raça/etnia	
Brancos, não hispânicos	26%
Negros, não hispânicos	36%
Hispânicos que falam inglês	33%
Idade	
18-29 anos	32%
30-49 anos	31%
50-64 anos	23%
+ 65 anos	18%
Nível de escolarização	
Diploma de primeiro grau	24%
Diploma de segundo grau	28%
Ensino superior ou +	31%
Renda familiar anual	
< 30 mil dólares	28%
30 mil dólares – 49,9 mil dólares	28%
50 mil dólares – 74,9 mil dólares	27%
75 mil dólares ou +	36%
Conexão à Internet em casa	
Discada	23%
Banda Larga	38%
<i>Nota da fonte: Pesquisa realizada com 1.623 usuários da Internet. A margem de erro é de aproximadamente 3%.</i>	

Fonte: PEW INTERNET, 2006.

Apesar de ter sido realizada nos EUA no ano de 2006, a pesquisa conduzida pela *Pew Internet* oferece uma perspectiva inovadora sobre o fenômeno da colaboração entre os usuários da Internet. Este relatório indica que os *taggers* possuem um perfil semelhante aos adeptos precoces de tecnologia.

Os dados da pesquisa revelam que a maioria dos usuários *taggers* possui menos de 40 anos e alto nível de educação e renda. Tanto homens como mulheres apresentam



a mesma propensão para as atividades de classificação colaborativa através das *tags*. A maior parte dos usuários possui conexão à Internet banda larga em casa, ao invés de conexão discada (*dial-up*). A pesquisa indica que nos EUA as “minorias” sociais estão mais inclinadas a classificar os conteúdos (“negros”, “latino-americanos”, “descendentes de hispânicos”) do que os “brancos” (PEW INTERNET, 2006).

Para Lamere & Pampalk (2008), a motivação das pessoas para classificar os conteúdos na Internet se deve à facilidade oferecida pelas *tags* para a busca e recuperação de informação. A classificação colaborativa ajuda a memorizar e contextualizar o uso dos conteúdos.

Outro aspecto é a utilidade das *tags* para a organização de informação relevante através da sinalização social sobre os gostos e interesses dos usuários. Segundo a *Pew Internet* (2006), as *tags* tendem a ser incorporadas aos hábitos da maioria da população mundial na medida em que mais indivíduos acessam e conectam-se à Internet.

Neste aspecto, há um intenso debate na Ciência da Informação sobre a utilidade da folksonomia para a classificação, catalogação, geração de meta-dados ou mesmo para as atividades tradicionais de busca e recuperação da informação (DYE, 2006; GRUBER, 2007; MATHES, 2004; PETERSON, 2006). A folksonomia, de acordo com seus críticos, é notoriamente imprecisa, composta por *tags* que são “geralmente ambíguas, excessivamente personalizadas e inexatas”, resultando em “um conjunto de termos caóticos e incontroláveis que dificultam os processos de busca” (GUY & TONKIN, 2006).

Elaine Peterson (2006) avança ainda mais nas críticas e afirma que “o universo da folksonomia permite a coexistência de discursos falsos e verdadeiros”. E denuncia que “visões idiossincráticas podem coexistir e se consolidar através das *tags*, apesar de suas inconsistências”. A autora acusa a folksonomia de ameaçar sua própria utilidade.

Algumas pesquisas identificam que muitas das estruturas “folksonômicas” possuem suas próprias relações hierárquicas e estruturais que se aproximam das estruturas da taxonomia. Estudos recentes como o de Kome¹⁰ (2005) argumentam que, pelo fato da classificação ser “uma atividade cognitiva humana fundamental”, as relações semânticas de hierarquia seguem existindo e se expressam nos meta-dados encontrados na folksonomia.

¹⁰ Kome analisou o website *Del.icio.us* e suas *tags*.



Para a Ciência da Informação, é importante que a folksonomia seja considerada um fenômeno social e não um esforço para substituir os vocabulários controlados, estruturados e hierarquizados. Esses dois métodos de classificação dos recursos culturais podem ser complementares e úteis para as análises de discurso de um domínio e estudos de usuários, dependendo do contexto informacional.

Weller (2007) argumenta que a principal propriedade da folksonomia é poder capturar o uso ativo de uma linguagem dentro de um determinado domínio ou comunidade. Por outro lado, a liberdade dos usuários na escolha do vocabulário utilizado nas tags constitui o chamado “problema de vocabulário” (FURNAS, LANDAUER, & GOMEZ, 1987), considerado como a principal dificuldade encontrada nos sistemas baseados em tags (GOLDER & HUBERMAN, 2006; FURNAS, et al., 2006; MATHES, 2004): ou seja, diferentes pessoas usam diferentes palavras para descrever os mesmos itens ou documentos.

Na folksonomia, tanto os sinônimos linguísticos e translinguísticos, variações no plural ou abreviações não se aglutinam em uma mesma definição ou categoria. Homônimos, por exemplo, são difíceis de distinguir. Erros de ortografia limitam a decodificação e constituem graves problemas para a recuperação da informação a partir da folksonomia (GUY & TONKIN, 2006). Todos esses entraves geram insuficiente precisão e revocação quando executadas as buscas de informação.

Por outro lado, a flexibilidade da escolha das tags é uma grande vantagem quando considerada a possibilidade de múltiplas perspectivas para um mesmo recurso cultural. O vocabulário controlado está sempre relacionado apenas com um ponto de vista específico e estabilizado. Os usuários da folksonomia criam rapidamente tags em resposta a novos usos, valores e mudanças contextuais para as terminologias dos bens simbólicos (KROSKI, 2006). Dessa forma, as tags oferecem novas opções para representar o conhecimento tácito dos usuários através de sistemas de classificação colaborativa (WELLER, 2007).

Mesmo quando tratamos de “*social tags*”, as intenções que motivam a elaboração das etiquetas nem sempre são coletivas. Os usuários que classificam conteúdos na Internet não necessariamente o fazem com o objetivo de facilitar o encontro e a busca de informação relevante para outros. Muitos usuários simplesmente usam as tags por interesses pessoais.



Entretanto, é importante ressaltar que a força da folksonomia repousa na *serendipity* (MATHES, 2004) - ou seja, na possibilidade de descoberta de informação por diferentes caminhos (WELLER, 2007). Através de mecanismos de busca fáceis de operar, a folksonomia pode promover encontros surpreendentes ou imprevisíveis entre usuários e conteúdos. Esse aspecto é essencial quando tratamos de bens culturais e das possibilidades de promoção da diversidade cultural na Internet.

Por contraste, a taxonomia depende da “competência cultural” – ou seja, do conhecimento prévio dos indivíduos - e impacta diretamente nos resultados de busca e de descoberta da informação: quando o usuário não sabe representar o que busca, normalmente tem dificuldade de encontrar o que deseja ou necessita. Neste sentido, a folksonomia ajuda a superar o “gap cognitivo” presente nos processos de busca de informação através do compartilhamento dos conhecimentos dos usuários – fenômeno denominado por Pierre Levy (2000) como “inteligência coletiva”. As principais diferenças entre a folksonomia e a taxonomia são resumidas a seguir:

Tabela 2: Características da Taxonomia *versus* Folksonomia

Taxonomia	Folksonomia
Rótulos fortes	Rótulos fracos
Vocabulário estruturado	Vocabulário não-estruturado
Vocabulário fixo	Vocabulário livre

Fonte: TURBULL, BARRINGTON, & LANCKRIET, 2008

A folksonomia oferece diferentes entradas para cada item em uma dada coleção de documentos. Dessa forma, baseia-se em três tipos básicos de entidades documentárias - os documentos, as tags e os usuários - que oferecem diferentes informações quando combinadas. O usuário pode manusear as relações entre esses três tipos de entidades documentárias caso a plataforma em questão os permita.

No caso dos Sistemas de Recomendação, os três tipos de entidades documentárias combinam-se entre si: os documentos (arquivos) estão relacionados através das tags e através dos usuários que os classificam. Os usuários estão relacionados quando utilizam as mesmas tags ou quando classificam o mesmo documento (PETERS & STOCK, 2007). Dessa forma, o funcionamento da folksonomia nos sistemas de informação ajuda a identificar comunidades de prática (DIEDERICH &



IOFCIU, 2006). No caso da classificação colaborativa da música, a relação entre estas entidades também apóia a formação de redes de interesse e a identificação de comunidades de gosto. A seguir, apresentamos o estudo de caso da classificação colaborativa em um Sistema de Recomendação de música na Internet – o site Last.fm.

Estudo de caso: a classificação colaborativa da música na Last.fm

Com o objetivo construir uma rede social de música na Internet, o Sistema de Recomendação Last.fm foi criado em 2003 como rádio online, porém em agosto de 2005 incorporou tanto o software Audioscrobbler (que registra os hábitos de escuta dos ouvintes) como a interface de acesso as músicas através das tags criadas por seus usuários, inspirada nas inovações dos sites *Flickr* e *Del.icio.us*.

A Last.fm estimula seus usuários a criarem tags para: 1) construir *playlists* de canções baseadas na classificação colaborativa da música pela comunidade; 2) categorizar o perfil e gosto de cada ouvinte; 3) melhorar as recomendações do sistema baseado nas tags 4) que a comunidade de usuários colabore na classificação do “maior catálogo global de música” (LAST.FM, 2009).

Portanto, a Last.fm permite que os ouvintes do sistema classifiquem artistas, álbuns e canções que escutam e apreciam dentro de sua plataforma de recomendação. Este SR usa a classificação colaborativa baseada em tags para agrupar artistas e canções. A associação do material musical através das tags determina o desempenho das recomendações do sistema, que é baseado também nos hábitos de escuta de cada usuário cadastrado em sua base de dados.

A Last.fm estimula ativamente em seu website que a classificação criada pelos usuários ultrapasse as tradicionais fronteiras de gêneros musicais disseminados pela Indústria. Ou seja, que os ouvintes usem as tags como um método de organização de seus gostos e preferências, e que categorizem o conteúdo disponível baseando-se na percepção musical individual e nos usos pessoais e/ou coletivos.



Os conteúdos classificados na Last.fm são principalmente canções, artistas e álbuns. A porcentagem da classificação colaborativa na Last.fm se distribui conforme mostram as Tabelas 3 e 4.

Tabela 3: Distribuição das *Tags* na Last.fm

Total de <i>Tags</i>	> 50 milhões
<i>Tags</i> em Canções	> 50% (> 25 milhões)
<i>Tags</i> em Artistas	> 40% (> 20 milhões)
<i>Tags</i> em Álbuns	< 5% (> 2,5 milhões)
<i>Tags</i> em Gravadoras	< 1% (< 500 mil)
"Outros"	< 4% (< 2 milhões)

Fonte: LAMERE & PAMPALK, 2008.

Tabela 4: Estatísticas de uso das *Tags* na Last.fm

<i>Tags</i> únicas	> 1,2 milhões
<i>Tags</i> únicas aplicadas a mais de 10 itens	> 130 mil
Itens <i>taggeados</i>	> 3.8 milhões
<i>Tags</i> criadas por mês	> 2.5 milhões
Usuários <i>Taggers</i> por mês	> 300 mil

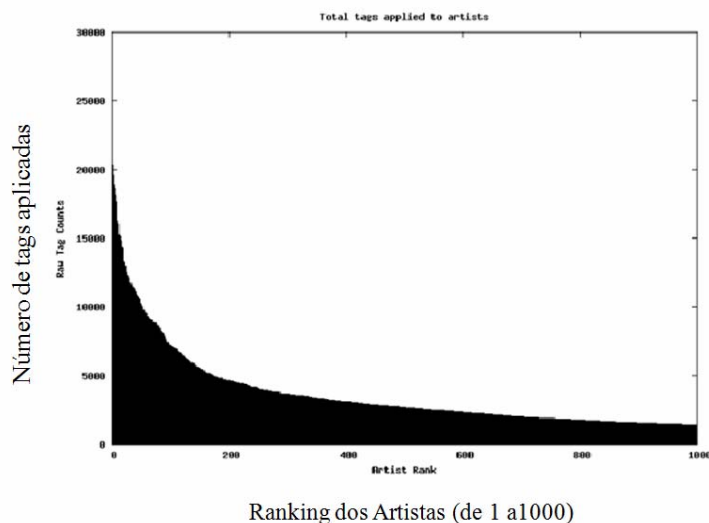
Fonte: LAMERE & PAMPALK, 2008.

A partir da observação destas tabelas, percebe-se que uma pequena parte do catálogo da Last.fm está classificada. Ou seja, apenas 3,8% dos itens disponíveis na coleção de música receberam tags, em um universo de 100 milhões¹¹ de canções registradas na Last.fm em 2008. Ao mesmo tempo, alguns itens recebem um número excessivo de etiquetas, o que demonstra que os artistas mais populares concentram o maior número de tags, que pode chegar à ordem de milhares de categorias para um único

¹¹ A proporção calculada de 3.8% das músicas que receberam tags em um universo de 100 milhões de canções foi calculada por LAMERE & PAMPALK em 2008. Em 2009 foram registradas 150 milhões de músicas na Last.fm e o número de itens classificados não foi atualizado – ou seja, não sabemos se, com o aumento paralelo do número de músicas identificadas pelo sistema em sua base de dados, a proporção se mantém ou se foi alterada - para mais ou para menos.

artista. O Gráfico 1 a seguir mostra a concentração das tags de acordo com ranking de popularidade dos artistas:

Gráfico 1: Número de Tags *versus* ranking dos artistas



Fonte: LAMERE & PAMPALK, 2008.

A Tabela 4 anterior indica que o número de tags únicas - criadas por apenas um usuário - é bastante significativo e está relacionada com a motivação dos participantes em classificar os conteúdos a partir de critérios de uso pessoal. A última linha da mesma Tabela 4 indica o número usuários que criam tags por mês. São 300 mil *taggers* mensais em um universo de 30 milhões de usuários da Last.fm - o que, a princípio, indica uma participação mensal de apenas 1% dos ouvintes¹².

Em uma primeira leitura parece que o número de usuários envolvidos na classificação colaborativa representa uma porcentagem minoritária, porém a classificação das músicas através das tags não pode ser contabilizada por sua frequência mensal. A maioria dos usuários classifica ou classificaram conteúdo musical na Last.fm, conforme mostra a Tabela 5 subsequente.

Segundo estudos preliminares realizados pelo grupo *MIR (Music Information Research)*¹³ o número total de pessoas que criaram algum tipo de tag para classificar as

¹² O mesmo comentário da nota nº11 vale para este parágrafo. O cálculo do número de usuários *taggers* por mês foi realizado em 2008 com base no número total de usuários, que na época, correspondia a 30 milhões (ver nota anterior). Os últimos dados divulgados indicam que a Last.fm possui um total de 37,8 milhões de usuários (LAST.FM, 2009).

¹³ Ver resultado do estudo do grupo *MIR* em: http://mir-research.blogspot.com/2008/06/lastfms-api-python-and-tagging_29.html.



músicas no sistema da Last.fm corresponde a aproximadamente 60,6% de sua rede de usuários. A seguir a Tabela 5 mostra a porcentagem de usuários *taggers* e o tamanho médio do vocabulário dos usuários por idade, de acordo com os resultados encontrados no estudo realizado pelo grupo de pesquisa MIR em 2008 sobre o SR da Last.fm:

Tabela 5: Idade e Comportamento dos usuários na Last.fm

Idade dos Usuários	% de Usuários-Taggers	Média do Tamanho do Vocabulário dos Taggers
14 – 19 anos	51.9 – 58.7 %	6 tags
19 – 22	56.4 – 62.5	7
22 – 25	52.7 – 59.4	9
25 – 30	58.6 – 65.2	8
30 – 60	64.0 – 71.6	13
Média Total	60,6%	8.6 tags

Fonte: MIR, 2008.

A Tabela 5 mostra que, ao contrário do que se poderia imaginar sobre a participação das audiências jovens, são os usuários mais velhos que mais “*taggeam*” e também os que possuem vocabulário mais amplo para criação das tags. Este fenômeno pode estar relacionado com o fato de que quanto mais velhos os usuários, maiores são as chances de possuírem conhecimento e competência musical para classificar.

O estudo realizado por Lamere & Pampalk (2008) ilustra essa situação e indica que os usuários com menos de 19 anos são os que mais utilizam categorias pessoais, “absurdas” ou ininteligíveis para classificar os itens musicais. Por contraste, os usuários com mais de 35 anos são os que aplicam categorias mais sofisticadas e precisas para classificar a música, conforme mostram os exemplos das Tabelas 6 e 7 a seguir.

Diante da possibilidade ilimitada de criar tags, algumas categorias se popularizam e se estabilizam entre os usuários no domínio da Last.fm. Na Tabela 8 a seguir apresentamos as doze tags mais utilizadas pelos usuários para classificar e buscar as músicas neste Sistema de Recomendação.

Tabela 6: Comportamento dos usuários menores de 19 anos

Tipos de Tags	Média de idade dos usuários
<i>lame lame lame lame</i> ¹⁴	18.2
<i>tokio hotel</i>	18.5
<i>xd</i> ¹⁵	18.7
<3	18.7
<i>tr00 satanic black metal</i>	18.8

Fonte: LAMERE & PAMPALK, 2008.

Tabela 7: Comportamento dos usuários maiores de 35 anos

Tipos de Tags	Média de idade dos usuários
<i>zydeco</i> ¹⁶	37.3
<i>soukous</i> ¹⁷	36.6
<i>cajun</i> ¹⁸	35.9
<i>zimbabwe</i>	35.3
<i>jazz saxophone</i>	35.1

Fonte: LAMERE & PAMPALK, 2008.

Ao observar as doze tags mais utilizadas pelos usuários da Last.fm na Tabela 8, percebe-se que a maioria corresponde a um gênero musical. Entretanto, a tag “*seen live*”¹⁹ é um exemplo de uso social e de valor que a música pode adquirir dentro de uma comunidade de prática. Esta, que corresponde à segunda tag mais utilizada pelos usuários, representa uma categoria difícil de ser incorporada em uma taxonomia.

¹⁴ A palavra “lame” significa em inglês coloquial algo “desinteressante” ou “chato”.

¹⁵ A sigla XD (abreviação de XD emoticon) é usada para expressar estado de ânimo ou expressão facial pelos usuários da Internet, principalmente em chats, mensagens instantâneas e jogos online.

¹⁶ Zydeco (zydeco music) é um gênero da música popular surgido no início do século XX a partir da mistura cultural entre os negros crioulos e os descendentes de colonizadores franceses localizados no Sudoeste de Louisiana - região dos EUA localizada entre Texas e New Orleans. Mais detalhes ver: <http://worldmusic.about.com/od/northamerican/p/Zydeco.htm>

¹⁷ O Gênero Musical *soukous* (também conhecido como *Lingala* ou *Congo*), anteriormente chamado genericamente de “*Rumba Africana*” surgiu na fronteira entre dois países na década de 1930: o Congo Belga e o Congo Francês, e ganhou popularidade na África nesta época.

¹⁸ A Música *cajun* (*cajun music*) também é um estilo emblemático da região de Louisiana - EUA, porém se difere do *zydeco* por ser cantada em francês e ser menos dançante. A presença da língua francesa se deve a história do lugar, que foi uma antiga colônia francesa nos EUA. O gênero musical é um desdobramento do *zydeco*. O *cajun* influenciou a música popular americana por décadas, especialmente na criação do atual e popular gênero musical chamado “*country music*”. Mais detalhes ver: <http://worldmusic.about.com/od/northamerican/p/Zydeco.htm>

¹⁹ A tradução para esta tag “seen live” em português seria “visto ao vivo”.

Tabela 8: Ranking das Tags mais utilizadas na Last.fm

Ranking	Tag	Nº de usuário que criaram a tag	Nº de vezes que a tag foi usada na Last.fm
1	<i>rock</i>	220.754	2.310.160
2	<i>seen live</i>	77.839	1.274.563
3	<i>alternative</i>	144.219	1.171.277
4	<i>indie</i>	139.013	1.080.697
5	<i>electronic</i>	121.320	975.884
6	<i>pop</i>	104.846	875.596
7	<i>metal</i>	90.804	676.244
8	<i>female vocalists</i>	76.169	616.712
9	<i>classic rock</i>	72.047	548.693
10	<i>alternative rock</i>	88.148	546.356
11	<i>jazz</i>	75.862	539.684
12	<i>punk</i>	7.207	513.106

Fonte: LAST.FM, 2009²⁰.

Sobre os hábitos de classificação dos artistas na Last.fm, os usuários tendem a classificá-los por gêneros musicais e locais de origem. Em contrapartida, quando se trata de processos de busca, os usuários priorizam os estilos musicais, sub-gêneros e épocas, conforme mostram as Tabela 9 e 10.

Por outro lado, quando examinada a frequência da classificação por tags nas canções, a importância do gênero musical é menor. Mais de 38% das categorias atribuídas às músicas correspondem a “estado de ânimo” ou a “opinião” dos usuários em relação às canções, conforme apresenta a Tabela 11.

²⁰ Dados de 08 de fevereiro de 2009. Disponível em: <http://www.lastfm.com.br/charts/toptags>

Tabela 9: Frequência da classificação de Artistas na Last.fm

Tipo de Tag	Frequência	Exemplos
Gênero Musical	68%	Heavy metal, punk
Local	12%	<i>French, Seattle</i>
Mood (estado de ânimo)	5%	<i>Chill, party</i>
Opinião	4%	<i>Love, favorite</i>
Instrumentalização	4%	Piano, <i>female vocal</i>
Estilo/sub-gênero	3%	<i>Political</i> , humor, psicodélico
Variado	2%	Coldplay, <i>composers</i>
Pessoal	1%	<i>Seen live, I own it</i>
Organizacional	1%	<i>Check out, to buy</i>
TOTAL	100%	

Fonte: LAMERE, 2008.

Tabela 10: Classificação versus frequência de busca dos artistas na Last.fm

Tipo de Tag	Frequência de classificação	Frequência de busca por tags
Gênero Musical	68%	51%
Local	12%	7%
Estado de ânimo	5%	4%
Opinião	4%	2%
Instrumentalização	4%	5%
Estilo /sub-gênero	3%	26%
Pessoal	1%	0%
Organizacional	1%	0%
Época	1%	3%
Outros	1%	2%

Fonte: BOSTEELS, KERRE, & PAMPALK, 2008.

Tabela 11: Frequência da classificação de canções na Last.fm

Tipo de Tag	Frequência	Exemplos
Gênero Musical	23,8%	Heavy metal, punk
Local	3,9%	<i>French, Seattle</i>
Estado de ânimo/Opinião	38,8%	<i>Party, Love, favorite, Relax</i>
Estilo	10,7%	Piano, <i>female vocal</i>
Outros	22,8%	

Fonte: THOMPSON, 2008.



Os resultados demonstrados nas Tabelas 9 e 10 nos permitem observar que os usuários tendem a classificar os artistas principalmente por grandes gêneros e em segundo lugar pelo local de origem. Entretanto as estratégias de busca privilegiam significativamente os estilos e/ou sub-gêneros como categorias cognitivas relevantes para a recuperação da informação na web.

Porém, quando analisamos a frequência de classificação das obras em si - ou seja, das canções e não dos artistas - através das *social tags* mais utilizadas pelos usuários, observamos que as categorias que correspondem à faceta “opinião” e “outros/nenhuma categoria” possuem uma representação considerável, sendo a faceta “opinião” a mais utilizada para classificar canções.

Ao interpretar os dados indicados na Tabela 11, percebe-se que as tags que correspondem a categorias “subjetivas” (como as que se referem a “opinião” e “estado de ânimo”) são frequentemente criadas pelos usuários, porém algumas delas são excessivamente pessoais e não podem ser compartilhadas pela comunidade da Last.fm. Por outro, os dados demonstram que os hábitos de escuta de determinadas canções tendem a ser mais determinados por contextos de uso do que pela classificação dos artistas em determinados gêneros musicais.

Conclusões

Com a popularização dos Sistemas de Recomendação a maior parte da informação na Internet passa a ser anotada manualmente de forma não-estruturada em um processo colaborativo, dinâmico e “desintermediado” – ou seja, através da relação direta entre os usuários e os conteúdos ou obras.

Cabe ressaltar que a anotação e descrição da maior parte dos arquivos multimídia disponíveis na web não seriam possíveis de alcançar sem a participação de um “exército” de usuários-voluntários conectados em rede. A anotação manual por especialistas é custosa e consome um tempo incalculável diante do número de obras disponíveis na Rede, cuja ordem de grandeza alcança a cifra de bilhões de arquivos musicais. Apesar dos esforços contínuos de cientistas e empresas no desenvolvimento de tecnologias para a anotação automática de arquivos de áudio, a utilidade deste método é ainda restrita a



alguns domínios da música. Até o momento, os métodos de anotação automática disponíveis não estão maduros suficientemente para classificar e rotular sons e arquivos de áudio digitais.

Além disso, as máquinas possuem limitações (que não substituem os processos cognitivos humanos) para a classificação de bens culturais, sendo o método de anotação automática complementar ao manual, porém insubstituível. Os sistemas informáticos são incapazes de interpretar os valores sociais das obras, a constituição dos gêneros musicais e as dinâmicas sociais em questão. Portanto, as tecnologias que viabilizam a classificação colaborativa de arquivos não-textuais na Internet, como é o caso da música, responde a demandas sociais latentes em relação aos sistemas de recuperação de informação.

Diante desse cenário, milhões de usuários conectam-se diariamente a websites e sistemas de recomendação como o da Last.fm e através de descritores livres (tags) classificam as músicas que escutam, que gostam ou não. As tags de cada ouvinte são disponibilizadas a todos os membros da comunidade.

Os sistemas de recomendação baseados em tags constroem assim um repositório colaborativo de conhecimento musical e esta classificação colaborativa influencia diretamente a forma como os usuários de uma comunidade de prática entendem, reconhecem, descrevem e escutam música em decorrência do uso da folksonomia para a recuperação de material musical.

De acordo com os dados apresentados nas Tabelas 3 e 4, observa-se que quando uma grande comunidade de usuários determina os meta-dados utilizados para a recuperação da informação em música (ao invés de uma pequena comunidade de especialistas), estas categorias tendem a resultar em meta-dados mais diversificados e, ao mesmo tempo, mais específicos.

Ao investigar a linguagem utilizada nas tags da Last.fm (Tabelas 6, 7, 10 e 11) percebe-se também que a folksonomia inclui uma quantidade elevada de “meta-dados semânticos”. Ou seja, que os usuários captam principalmente os atributos sensíveis da música e a partir de uma “percepção sensorial” criam tags. Isso explica porque a maior parte das tags são “carregadas” de uma “dimensão semântica” (ver Tabela 11) – ou seja, os usuários tendem a utilizar categorias de classificação que representam sentimentos, estados de ânimo, contextos de uso e opiniões.



Ao analisar a folksonomia encontrada na Last.fm (Tabela 9), conclui-se que os dados gerados pelos usuários através das tags são mais confiáveis na representação dos gêneros musicais do que se especulava previamente, na medida em que 68% das tags criadas pelos usuários correspondem a gêneros musicais estabilizados e reconhecidos, tanto pela indústria da música como por instituições especializadas. Portanto, a folksonomia pode oferecer parâmetros relevantes na criação de meta-dados adequados para a representação de material musical.

Por outro lado os resultados indicam que a folksonomia potencializa a inclusão dos vocabulários e das necessidades de todos os usuários que participam da classificação e enfraquece os julgamentos parciais ou particulares - que podem ser comerciais, políticos ou institucionais. Neste contexto, os interesses de grupos minoritários são incluídos por seus próprios membros, que podem representar e classificar a informação de forma autônoma e afastar-se dos estigmas dominantes. A questão mais importante é que a classificação colaborativa oferece a possibilidade de representar a informação a partir de visões alternativas. Dessa forma, a rede de usuários se beneficia da folksonomia.

Mesmo considerando que todavia existem contradições e polêmicas em relação ao fenômeno da classificação colaborativa, a folksonomia não deve ser considerada como um substituto de padrões tradicionais de classificação baseados em hierarquia de classe. Os processos de classificação colaborativa precisam ser considerados no contexto da Internet, onde há uma demanda social latente por novas possibilidades de representação, busca e descoberta de informação.

Com o surgimento da folksonomia, começam a operar condições no nível social que criam possibilidades de escapar da organização da cultura realizada tradicionalmente por especialistas ou pelas Indústrias Culturais. A classificação colaborativa “desorganiza” a cultura para fazer surgir novos enunciados em proveito de outras instâncias de organização que assegurem a multiplicidade de conexões entre pessoas e conteúdos.

ABSTRACT: This work discusses the new mediation and organization forms of no-textual contents in the Internet. The folksonomy concepts, social tags and collaborative classification methods are considered for information retrieval in Recommendation Systems, which works on basis of Collaborative Filtering technology. The case study analyzed was the Recommendation Systems of Last.fm, a platform widely used in the



world for information seeking, retrieval and access to the available musical material in the Internet. The analyses were based on the tag statistics of this system and the results demonstrate that, in this kind environment, characterized by the abundance of information, the folksonomy tends to present great usefulness and social relevance for the communities of practice. The results indicate that the process of collaborative classification of music achieves relevant results in information retrieval. In that way, the folksonomy potentiates the inclusion of vocabularies and information needs of most users that participate in the classification and it weakens partial or private judgments - that can be commercial, political or institutional. The collaborative classification gives rise to new statements for the benefit of other instances of information organization that tend to assure multiplicity of connections between people and Internet contents.

Keywords: Colaborative classification. Internet. Music. Recommendation Systems. Folksonomy

Referências:

- BENKLER, Y. *The Wealth of Networks: how social production transforms markets and freedom*. New York: Yale University Press, New Haven and London, 2006.
- BOSTEELS, K.; KERRE, E.; PAMPALK, E. Music Retrieval Based on Social Tags: a case study. In: *Proceedings in ISMIR 2008*. Vienna, Austria: [s.n.], 2008.
- DIEDERICH, J.; IOFCIU, T. Finding Communities of Practice from User Profiles Based on Folksonomies. In: *Proceedings of the 1st International Workshop on Building Technology Enhanced Learning Solutions for Communities of Practice (TEL-CoPs 06)*. [s.l. : s.n.], 2006.
- DYE, J. Folksonomy: a game of high-tech (and high-stakes) tag. *E-Content Magazine*, v. 29, n. 3, p. 8-43, 2006.
- FURNAS, G. et al. Why Do Tagging Systems Work? In: *CHI 06 Extended Abstracts on Human Factors in Computing Systems*, New York: ACM, 2006. p. 36–39.
- FURNAS, G. et al. The Vocabulary Problem in Human-System Communication: an analysis and a solution. *Communications of the ACM*, New York, v. 30, n. 11, p. 964-971, 1987.
- FURNER, G.; SMITH, M.; WINGET, M. Collaborative Indexing of Cultural Resources: some outstanding issues. *Digital Humanities*, p. 69-71, 2006.
- GOLDER, S. A.; HUBERMAN, B. The Structure of Collaborative Tagging Systems. *Journal of Information Science*, v. 32, n. 2, p.198-208, 2006.
- GRUBER, T. Ontology of folksonomy: a mash-up of apples and oranges. *International Journal on Semantic Web & Information Systems*, v. 3, n. 2, 2007. Disponível em: <<http://tomgruber.org/writing/ontology-of-folksonomy.htm>>.



XI Encontro Nacional de Pesquisa em Ciência da Informação
Inovação e inclusão social: questões contemporâneas da informação
Rio de Janeiro, 25 a 28 de outubro de 2010

GUY, M.; TONKIN, E. Folksonomies: tidying up tags?. *D-Lib Magazine*, v. 12, n. 1, 2006. Disponível em: <<http://www.dlib.org/dlib/january06/guy/01guy.html>>.

KOME, S. *Hierarchical subject relationships in folksonomies*. Chapel Hill: University of North Carolina, 2005.

KROSKI, E. *The Hive Mind: folksonomies and user- based tagging*. [s.l. : s.n.], 2006. Página web. Disponível em: <<http://infotangle.blogspot.com/2005/12/07/the-hivemind-folksonomies-and-user-based-tagging>>.

LAMERE, P. Social Tagging and Music Information Retrieval. *Journal of New Music Research*, v. 37, n. 2, p. 101-114, jun. 2008.

LAMERE, P.; PAMPALK, E. Social Tags and Music Information Retrieval. In: *Proceeding of International Symposium of Music Information Retrieval, ISMIR*. [s.l. : s.n.], 2008.

LAST.FM. LTD. Seção Perguntas frequentes: O que são tags? [s.l. : s.n.], 2009. Disponível em: <<http://www.lastfm.com.br/help/faq?fagsearch=tag&submit=Pesquisar+FAQ>>.

LÉVY, P. *A inteligência coletiva: por uma antropologia do ciberespaço*. São Paulo: Loyola, 2000.

MATHES, A. *Folksonomies: cooperative classification and communication through shared metadata*. Urbana-Champaign: University of Illinois, 2004. Technical Report.

MUSIC INFORMATION RETRIEVAL. MIR Research. 2008. Disponível em: <http://mir-research.blogspot.com/2008/06/lastfms-api-python-and-tagging_29.html>

PETERS, I.; STOCK, W. Folksonomy and Information Retrieval. In: *Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology*, vol. 44. Milwaukee, WI: ASIS&T. 2007. (CD-ROM).

PETERSON, E. Beneath the metadata: some philosophical problems with folksonomy. *D-Lib Magazine*, v. 12, n. 11, 2006. Disponível em: <<http://www.dlib.org/dlib/november06/peterson/11peterson.html>>.

SMITH, G. *Folksonomy: social classification*. *Information Architecture*. August, 2004. Página web. Disponível em: <http://atomiq.org/archives/2004/08/folksonomy_social_classification.html>.

THOMPSON, A. E. *Playing Tags: an analysis of vocabulary patterns and relationship within a popular music folksonomy*. Chapel Hill, North Caroline: University of North Carolina: 2008.

TURBULL, D.; BARRINGTON, L.; LANCKRIET, G. Five Approaches to Collecting Tags for Music. In: *Proceedings in ISMIR 2008*. Vienna, Austria: [s.n.], 2008. p. 225-230.

VANDER WAL, T. *Folksonomy*. [s.l. : s.n.], 2 fev. 2007. Página web. Disponível em: <<http://www.vanderwal.net/folksonomy.html>>. Acesso em: 15.01.2009.



XI Encontro Nacional de Pesquisa em Ciência da Informação
Inovação e inclusão social: questões contemporâneas da informação
Rio de Janeiro, 25 a 28 de outubro de 2010

WEINBERGER, D. 28% of Online Americans Have Used the Internet to Tag Content. Forget Dewey and His Decimals, Internet Users are Revolutionizing the Way We Classify Information – and Make Sense of It. By Lee Rainie, Director. *Pew Internet & American Life Project*, 2007. Disponível em: <http://www.pewinternet.org/~media/Files/Reports/2007/PIP_Tagging.pdf.pdf>.

WELLER, K. Folksonomies and ontologies: two new players in indexing and knowledge representation. In: *Proceedings of Online Information*. London: Great Britain, 2007. p. 108-115.