
Aplicação teórico-conceitual da folksonomia assistida para recuperação da informação

Aplicación teórico-conceptual de folksonomías asistidas para la recuperación de información

Application of theoretical-conceptual assisted folksonomies for information retrieval

Jose Eduardo SANTAREM SEGUNDO, Cinthia Scaldelai SIQUEIRA

Department of Education, Information and Communication, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brasil, santarem@usp.br, cinthiascaldelai@hotmail.com

Resumen

Se presenta una experiencia de validación teórico-conceptual de una folksonomía asistida, que consiste en el uso de un corpus de información digital para ayudar en el proceso de definición de las etiquetas con el objetivo de aumentar el grado de relevancia de estas en relación con el objeto depositado. Se construyó una matriz de adyacencia de todas las palabras-clave de cincuenta artículos publicados entre los años 2008 a 2010 en la revista *Ciencias de la Información* (Brasil); y, a partir de ella, se generó una red de etiquetas mediante Pajek y Vos Viewer, con el fin de resaltar las diferencias entre el uso o no del modelo. La aplicación del concepto de folksonomía asistida redujo el 12,79 % del total de las etiquetas asignadas inicialmente, aumentando la consistencia de la relación entre los artículos seleccionados y reforzando la relación entre los artículos publicados.

Palabras clave: Folksonomías asistidas. Web 2.0. Etiquetado colaborativo en red. Representación iterativa.

1. Introdução

A convergência de novas tecnologias comunicacionais se apropria da rede e nela emoldura novos espaços colaborativos que primam em fruir o conhecimento e solidificar a promoção da inteligência coletiva.

Redesenhando este cenário que agora vigora na rede está a Folksonomia: fenômeno de comunicação livre e marca da nova geração da web. Já identificada e utilizada em vários sites da Web, a Folksonomia que deriva do neologismo entre as palavras *folk* (povo) e *taxonomy* (taxonomia), é utilizada pelos usuários na categorização e indexação dos objetos e/ou informações digitais (Catarino, 2009).

De acordo com Silva (2009), se por um lado esta liberdade de atribuição de tags e de etiquetagem de recursos da Web representa o sucesso de sites colaborativos baseados no uso da Folksonomia, por outro lado, a proliferação do

Abstract

An experience of a theoretical-conceptual validation of an assisted folksonomy is presented, which involves the use of a corpus of digital information to assist in the process of defining labels so such labels increase their degree of relevance in relation to the articles they describe. An adjacency matrix of all keywords in fifty articles published between the years 2008-2010 in the journal *Information Sciences* (Brazil) was built; and, from it, a network of tags was generated using Pajek and Vos Viewer, in order to highlight the differences that appear when the model is used or not. Applying the concept of assisted folksonomy achieved a 12.79% fall in the number of the initially assigned labels, increasing the consistency of the relationship among the selected items and strengthening their relation to the published articles.

Keywords: Assisted folksonomies. Web 2.0. Network tags. Iterative representation.

uso de tags inconsistentes e não padronizadas podem afetar a recuperação da informação.

As variações linguísticas podem prejudicar de forma considerável o processo que compreende a busca do usuário pela informação desejada, afetando negativamente a concretização do ciclo informacional. Esta situação decorre da heterogeneidade de termos e tags atribuídos pelos usuários, que podem ser revestidos de ambiguidades devido ao uso de termos polisêmicos, de sinônimas ou do uso de plural e singular para indicar o mesmo tipo de objeto descrito, ocorrência de erros de digitação e casos de junção de palavras com separadores, como hífen, e também uso de iniciais maiúsculas, dentre outras inconsistências que podem impedir a eficiência na busca e a recuperação da informação.

Estas instabilidades vocabulares levam o usuário a não reutilizar termos já indexados em novas descrições no ambiente informacional, resultando no isolamento de seu documento digi-

tal ou na sua dispersão. Suscitando questões que apontam para novas estratégias de gerenciamento da informação em ambientes informacionais, frente a novas tecnologias de informação e de comunicação, é pertinente refletir sobre como aprimorar o uso da indexação, classificação e recuperação de informações no contexto da web.

No sentido de aprimorar o processo de atribuição de conceitos aos documentos no momento de sua indexação no repositório digital para sua posterior recuperação, Santarém Segundo (2010) propõe como base o provimento do processo de Folksonomia Assistida.

A Folksonomia Assistida integra-se ao modelo de Representação Iterativa, sugerido por Santarém Segundo (2010) na busca pela viabilização e incorporação de qualidade na descrição, armazenamento e recuperação de recursos informacionais na web em repositórios digitais. O processo que se inicia no usuário, através da extração de informações de um determinado documento, auxilia a descrição do objeto digital para concretizar seu depósito no repositório digital.

Assim, o propósito e objetivo deste trabalho é homologar de forma teórico-conceitual a Folksonomia Assistida, a partir da construção de uma estrutura de rede de tags, aplicada especificamente a repositórios digitais, para aprimoramento da recuperação da informação, na busca pela estruturação de relações entre os termos associados.

Destaca-se neste trabalho a Folksonomia Assistida, a qual permitirá que os usuários tenham uma orientação guiada e eficiente para a indexação de seu objeto informacional, aprimorando e elevando o nível de qualidade de descrição do mesmo em repositórios digitais.

A partir do modelo conceitual de Folksonomia Assistida, será possível refletir sobre a pertinência de se realizar estudos mais profundos sobre a organização de uma estrutura formal de orientação guiada ao usuário na catalogação de objetos digitais. Além disso, a construção da rede de tags, à luz dos novos recursos e possibilidades surgidas com as novas tecnologias de informação e de comunicação e o acesso aberto, implicará no favorecimento do estudo das redes de colaboração científica entre autores.

2. Folksonomia

Elemento vital da Web 2.0, e popularizada por sua ilustre atuação em ambientes que corroboram para a criação de interconexões colaborativas através da utilização de tags, a Folksonomia

configura-se como uma inovação no âmbito da descrição e gerenciamento de conteúdos na web.

O conjunto de dados criado a partir da associação de palavras-chave aos recursos dos participantes é chamado de Folksonomia. O serviço de marcação permite que o usuário associe livremente palavras-chave (tags) com um recurso. O recurso pode ser uma enorme variedade de coisas, tais como fotografias, URLs, podcasts, jogos de computador, música e vídeos (Sinclair e Cardew-Hall, 2008).

O conceito de usar livremente determinadas palavras-chave atribuídas por não-especialistas não é novo, muitas revistas científicas possibilitam aos autores criarem suas palavras-chave para completar a indexação de seus artigos. Utilizar um vocabulário descontrolado neste caso é menos oneroso do que a contratação de um especialista para realizar uma análise de domínio para determinar um esquema de classificação apropriado (Sinclair e Cardew-Hall, 2008).

De acordo com Guedes e Dias (2010), a Folksonomia vem sendo empregada pelos usuários para representar suas ações de classificação, categorização e indexação de objetos digitais na web.

Ao enfatizar a Folksonomia como novo processo social que atribui novas significâncias sobre a organização e a recuperação da informação no ambiente digital, é pertinente refletir também sobre como aprimorar o processo de atribuição de conceitos aos documentos no momento da indexação dos mesmos no ambiente digital, para posterior recuperação dos conteúdos. A escolha de um ponto de acesso, ou seja, a atribuição de uma determinada palavra-chave para a descrição de um documento é guiada por razões pessoais de cada indivíduo, regidas por valores, visões e opiniões que se agregam no momento de indexação, e caracterizam a Folksonomia como taxonomia popular.

De acordo com Silva (2009), se, por um lado, esta liberdade de atribuição de tags e de categorização de recursos da Web representa o sucesso de sites colaborativos baseados no uso da Folksonomia, por outro lado, a proliferação do uso de tags inconsistentes e não padronizadas, pode afetar a descrição do recurso e da recuperação da informação.

O uso de um vocabulário controlado minimiza os ruídos comunicacionais entre os usuários e o sistema informacional durante o processo de busca, pois permite o controle dos termos indexados, porém constitui restrição à liberdade e

também exige um esforço cognitivo maior no momento da categorização do documento, dificultando o processo de depósito pelo usuário.

Diferentemente dos vocabulários controlados, a linguagem natural utilizada pelos usuários no momento da criação da tag, é impregnada de inconsistências que tendem a dificultar o processo de identificação e escolha de uma palavra-chave adequada e pertinente ao que o usuário procura, sendo este um dos pontos de conflito relacionado à Folksonomia.

Guy e Tonkin (2006, p. 1) afirmam que,

Começamos por olhar para a questão das "tags malfeitas", um problema para o qual os críticos da Folksonomia fazem questão de aludir, e perguntar à comunidade que pesquisa sobre Folksonomia se há maneiras de compensar esses problemas [...]

[...]Provavelmente, a grande falha dos sistemas de folksonomia atuais, é que os termos de marcação utilizados nesses sistemas são imprecisos. Os usuários dos sistemas que utilizam Folksonomia inserem livremente as tags, o que significa que as tags são muitas vezes ambíguas, excessivamente personalizadas e inexatas.

O uso de separadores de palavras compostas, erros ortográficos representados pela troca de letras e palavras com vírgula no final, além da mescla de palavras no singular e plural, podem acarretar na impossibilidade de recuperação dos documentos informacionais (Silva, 2009).

Ao mesmo tempo em que a Folksonomia oferece ao usuário a possibilidade de atribuir de forma livre tags/termos que sirvam como referência a seu objeto digital depositado, este procedimento acaba acarretando uma grande quantidade de termos semelhantes e que dificultam a associação de conteúdo semanticamente relacionado assim como o posterior processo de recuperação da informação.

2.1. Folksonomia Assistida

Santarém Segundo (2010) propõe como base a implementação do processo de Folksonomia Assistida, que visa a oferecer um eficiente gerenciamento das tags, ao proporcionar uma arquitetura baseada em uma nova interface de busca e inserção de palavras-chave.

O processo que se inicia no usuário através da extração de informações para a catalogação de um determinado documento auxilia-o na escolha de termos mais adequados para a realização da descrição do objeto digital e a concretização de seu depósito em um ambiente específico da web: um repositório digital (Santarém Segundo e Vidotti, 2011).

Para que se efetive o processo de Folksonomia Assistida, a interface de comunicação entre o usuário e o repositório digital deve ser modificada, visto que o processo apoia-se em informações anteriormente inseridas no sistema por outros usuários, proporcionando assim, como uma retroalimentação, um processo gerenciado de atribuição de tags e uma indexação orientada.

Na primeira etapa, o modelo de Folksonomia Assistida proposto por Santarém Segundo (2011) irá coletar e se apropriar das tags já indexadas no repositório digital, para auxiliar na escolha adequada e guiada de termos que serão representados pelos termos já inseridos, além de uma estrutura de representação da área do conhecimento associada ao repositório digital. Independentemente de o usuário aceitar, ou não, as sugestões de termos que poderão descrever seu documento digital, no segundo passo o sistema acolherá as informações inseridas pelos usuários e irá proceder a uma busca por relacionamentos entre a informação digitada e o acervo de tags que se encontra indexado no repositório.

Posteriormente a este processo, o sistema novamente irá apresentar ao usuário novos termos que poderão ser avaliados e escolhidos totalmente ou parcialmente pelo usuário, ou optar por descartar definitivamente as sugestões. Pode-se evidenciar que as decisões estão totalmente concentradas no usuário, e é este quem decidirá qual o termo mais apropriado para seu documento.

Estas ações desenvolvidas pelo sistema demonstram claramente que o objetivo da Folksonomia Assistida não é tornar rígido o processo cognitivo do usuário em efetivar a descrição do objeto informacional, e nem modificar as concepções inerentes à Folksonomia de construção coletiva do conhecimento, mas se concentra em propiciar uma forma guiada, livre e eficiente de atribuição de tags para a indexação e recuperação da informação científica, aprimorando e elevando o nível de qualidade de descrição de documentos em repositórios digitais.

3. Metodologia, aplicação e resultados

A metodologia utilizada no desenvolvimento deste trabalho foi desenvolvida a partir de uma análise teórica da Folksonomia. A seguir foi desenvolvido um processo que tem como princípio aplicar o modelo Folksonomia Assistida de Santarém Segundo (2010).

Para a concretização deste trabalho, foram selecionados 50 artigos publicados entre os perío-

dos de 2008 a 2010 na revista *Ciência da Informação* (1), do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Os artigos coletados referem-se aos periódicos de volume 37, número 2 de 2008, ao volume 39, número 3 de 2010.

De forma a homologar o processo da Folksonomia Assistida, desenvolveu-se uma tabela, Figura 1, que revela a quantidade de relacionamentos existentes entre as tags, que são as palavras-chave utilizada nos textos publicados. A partir desta, foi arquitetada a matriz de adjacências para demonstração de como as tags ficariam armazenadas no banco de dados de um repositório digital.

CÓDIGO	DESCRIÇÃO	QUANTIDADE	ITEM (ARTIGO)	IDIOMA
9	Aprendizagem organizacional	2	26;36	Português
10	Aprendizes de inglês	1	5	Português
11	Archivos de televisión	1	19	Espanhol
12	Arquivo & administração	1	13	Português
13	Arquivologia	1	13	Português
14	Arquivos abertos	1	22	Português
15	Atuação profissional	1	47	Português
16	Avaliação da qualidade	1	18	Português
17	Belém	1	29	Português
18	Bibliometria	2	32; 45	Português
19	Bibliometria	3	1; 23; 24;	Espanhol

Figura 1. Parte da tabela de relacionamento entre as tags

A primeira tabela se estrutura a partir da descrição das palavras-chave, configuradas como tags e identificadas a partir da atribuição de um código numérico, associando-as aos artigos das quais se originam, denominados itens. Além disso, há também descrita nesta tabela a quantidade de vezes em que há a presença de cada palavra-chave em diferentes artigos e o idioma referente a cada uma.

Como forma de suprimir as inconsistências sintáticas e a coexistências de palavras com o mesmo significado semântico, aplicou-se de forma teórico-conceitual o processo de Folksonomia Assistida. Portanto, como resultado, obteve-se a agregação de alguns termos, aprimorando a qualidade e consistência dos documentos recuperados.

A matriz de adjacências, apresentada em parte como exemplo na Figura 2, constituiu-se fundamentada nos dados inseridos na tabela em que constam os códigos numéricos atribuídos às palavras-chave de cada artigo e a quantidade de relacionamentos entre elas. A base da construção de uma rede dar-se-á através do relacionamento horizontal entre palavras-chave do mesmo artigo e de forma vertical através do relacionamento entre termos semelhantes de artigos diferentes.

	127	126	125	124	123	122	121	120	119	118
	Integração de sistemas	Instituições bancárias	Inovação	Informetria	Informetria	Informação social	Informação estratégica	Informação e Tecnologia	Informação corporativa	Informação científica
1	Arquivos abertos									1
2	Atuação profissional									
3	Avaliação da qualidade	1						1		
4	Belém						1			
5	Bibliometria				2					
6	Bibliometria			2						
7	Biblioteca escolar									
8	Biblioteca universitária									
9	Bibliotecas digitais							1		

Figura 2. Parte da matriz de adjacências constituída

Para apresentação dos resultados ao olho humano foram geradas quatro imagens, em dois softwares diferentes, de forma a apresentar de forma gráfica os resultados obtidos com a matriz de adjacências. Construído com base nos dados da matriz de adjacências, a Figura 3 representa a imagem gráfica gerada pelo software VOS Viewer (2). O software efetivou mapeamentos gerando concentrações de tags, denominadas de clusters.

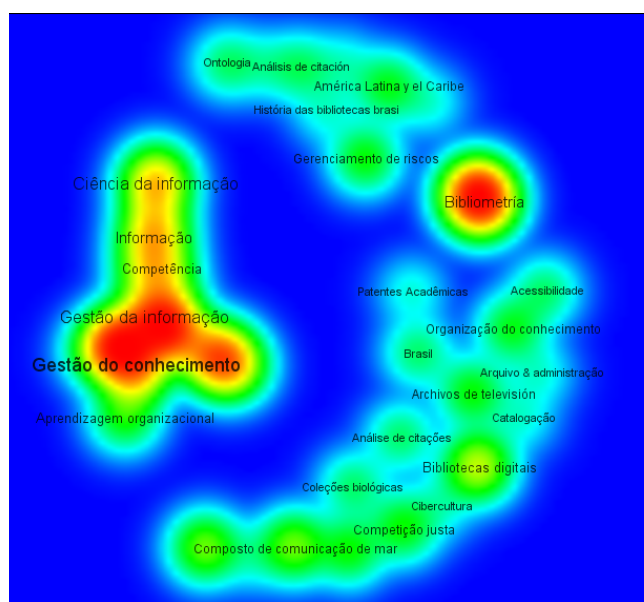


Figura 3. Densidade apresentada pelo VOS Viewer sem aplicação da Folksonomia Assistida

Verifica-se na Figura 3 o primeiro mapeamento, no qual o software VOS Viewer identificou trinta e cinco clusters, apresentados em quatro grupos de imagens, que representam os relacionamentos entre as tags e a frequência de citações referentes às mesmas. Quanto às relações entre os termos, temos no lado direito da imagem e também na parte superior da figura, tags

dispersas, citadas com menor frequência, que revelam mínimas relações existentes entre os termos, destacados na cor verde. Isso se caracteriza principalmente por erros de polissemia e sinonímia na seleção dos termos chave.

No lado esquerdo da Figura 3, estão em evidência tags com maior número de citações e maior densidade de relações, representadas pelas cores que gradativamente se intensificam e se agrupam, porém, este agrupamento de tags se mostra inconsistente e de forma isolada.

Diante das observações, infere-se que quanto mais alta a quantidade de clusters gerados, maior será a frequência de tags dispersas e, conseqüentemente, menor quantidade de relações são estabelecidas entre elas.

Ao expressar o resultado gerado a partir da aplicação da Folksonomia Assistida, a Figura 4, gerada pelo VOS Viewer produziu agrupamentos entre as tags mais citadas, evidenciando relações mais sólidas e consistentes entre as mesmas, concentradas no lado esquerdo da imagem. Outro grupo de tags foi gerado no lado direito da imagem, porém, representados com menor frequência de citações e relacionamentos.

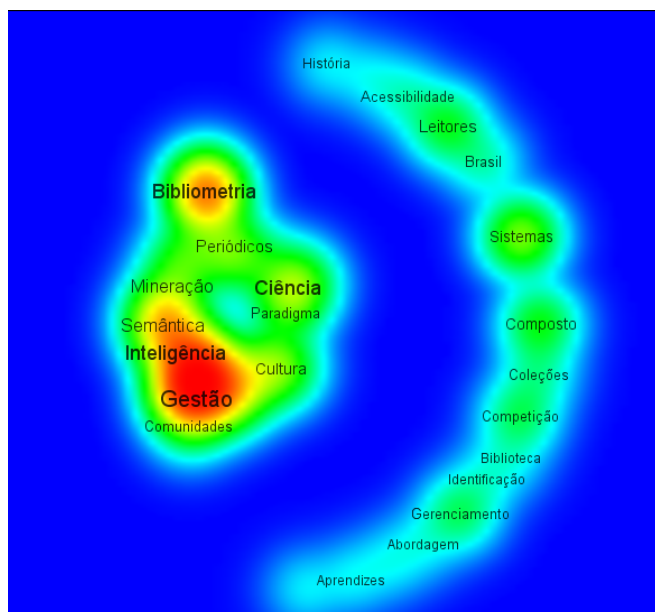


Figura 4. Densidade apresentada pelo VOS Viewer com aplicação da Folksonomia Assistida

O VOS Viewer identificou 28 clusters na geração da Figura 4, sete a menos que na geração da Figura 3. Este fato distingue-se pela aplicação do modelo de Folksonomia Assistida, que primou pela agregação dos termos que anteriormente estavam dispersos, gerados pela eliminação dos termos repetidos e que apresenta-

vam irregularidades vocabulares gramaticais como as já citadas anteriormente.

O software Pajek (3) não realiza neste trabalho nenhum tipo de inferência, ele foi utilizado, pois apresenta, de forma clara, imagens de representação gráfica de grafos gerados pelas matrizes de adjacências. Para visualização foram construídas redes que apresentam o grau de popularidade das tags a partir das variações de frequência e relacionamento nos artigos. Os graus de relacionamentos entre as tags são identificados a partir das arestas que as conectam, de acordo com as características referentes ao diâmetro das mesmas. Quanto maior o diâmetro das arestas, mais ligações são existentes entre as tags.

A análise das redes de tags permite que sejam visualizadas conexões de baixa, média e alta intensidade entre os artigos, demonstradas através da alteração do tamanho dos nós que as conectam. A Figura 5, sem aplicação da Folksonomia Assistida, revela três termos mais citados, destacados na cor verde, com vértices de maior diâmetro. Estes termos na cor verde são citados mais do que três vezes nos artigos selecionados para este trabalho.

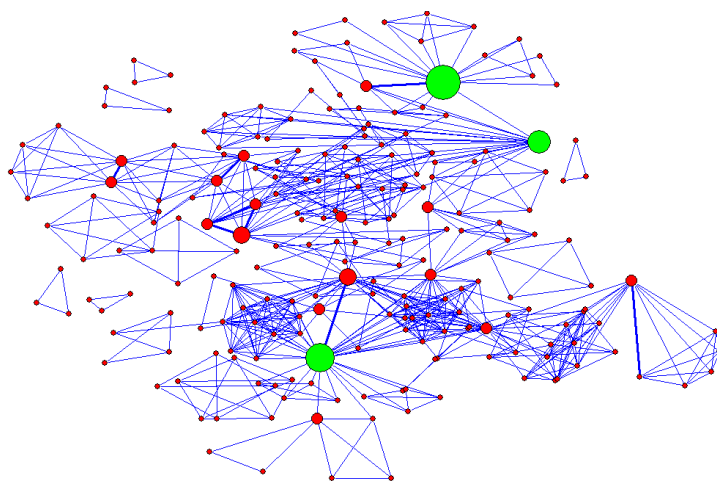


Figura 5. Rede de tags gerada pelo software Pajek, sem aplicação da Folksonomia Assistida

Verifica-se na figura 6, um aumento de vértices com diâmetros maiores em relação Figura 5, ao contar com oito termos mais citados. Esta situação ocorre pela agregação dos termos através processo de Folksonomia Assistida, em decorrência da eliminação e/ou substituição de termos sintaticamente inconsistentes. Observa-se também o aumento da quantidade de termos próximos e com maior número de ligações e a ocorrência de maior concentração de relacionamentos ao redor dos vértices.

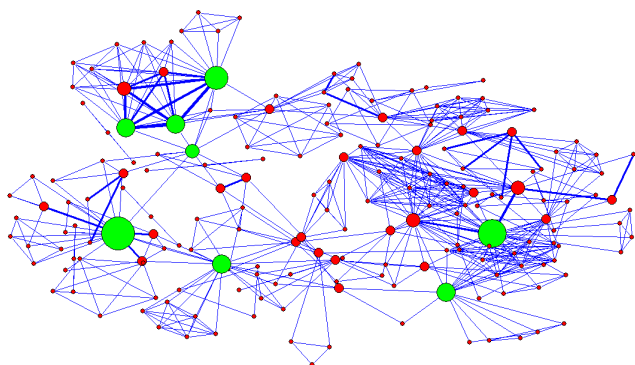


Figura 6. Rede de tags gerada pelo software Pajek, com aplicação da Folksonomia Assistida

Sistematizadas pelo processo de Folksonomia Assistida, as conexões das tags que fazem parte da segunda rede, evidenciam relações e fluxos informacionais consistentes, demonstrados pela diminuição de 12,79% das tags existentes inicialmente.

Ao resgatar e agregar as tags inseridas pelo processo de Folksonomia Assistida, o modelo em rede legitima um ciclo de conexão entre os assuntos dos artigos, ao aproximar termos que se relacionam frequentemente. A rede de tags enseja a existência de novas abordagens que refletem inclusive a visualização da transdisciplinaridade, a partir dos relacionamentos entre os assuntos abordados nos artigos no contexto da informação científica.

4. Considerações finais

A arquitetura do modelo denominado Folksonomia Assistida por Santarém Segundo (2010) ascende e soma novas contribuições e perspectivas significativas para o armazenamento, organização, e recuperação da informação em repositórios digitais, ao ir além do contexto da Folksonomia relacionado a simples categorização de objetos através do uso de tags.

Em virtude das irregularidades sintáticas que ocorrem na atribuição de termos pelos indivíduos em ambientes informacionais e no anseio por adaptações de interfaces de busca em sistemas colaborativos que emergem no âmbito da Web 2.0, a Folksonomia Assistida se torna conceitualmente, uma ferramenta que promove através da orientação guiada, a eficiente categorização do documento científico em repositórios digitais.

Os resultados homologam a vantagem de uso da Folksonomia Assistida, ao possibilitar a visualização da congregação dos relacionamentos

entre as tags anteriormente dispersas pelas inconsistências. De forma cooperativa, desempenham um papel valioso no gerenciamento do vocabulário utilizado na representação informacional, em aspectos que contribuem para a não redundância de informação, e na confiabilidade da pesquisa e recuperação dos artigos pelos usuários do repositório digital.

Destaca-se, ainda, o dinamismo que o modelo apresentado, já identificado na Folksonomia, oferece a um ambiente informacional, permitindo com que haja um processo de iteratividade e retroalimentação rica para o vocabulário de um sistema de informação.

Notas

- (1) <http://revista.ibict.br/ciinf/index.php/ciinf>.
- (2) <http://www.vosviewer.com/>
- (3) <http://pajek.imfm.si/>

Referencias

- Catarino, M. E. (2009). Integração da folksonomia nos metadados: identificação de novos elementos como contributo para a descrição de recursos em repositórios. Tesis doctoral. Braga: Universidade do Minho, Departamento de Sistemas de Informação. http://repositorium.sdum.uminho.pt/bitstream/1822/9564/1/Tese_CatarinoMElisabete.pdf (2013-03-25).
- Guedes, R. M.; Dias, E. J. W. (2010). Indexação social: abordagem conceitual. // Revista ACB. 15:1 (jan./jun. 2010) 22-38. http://revista.acbsc.org.br/index.php/racb/article/view/686/pdf_17 (2013-03-25)
- Guy, M.; Tonkin, E. (2006) Folksonomies: tidying up tags? // D-Lib Magazine. 12:1 (Jan. 2006). <http://www.dlib.org/dlib/january06/guy/01guy.html> (2013-02-19)
- Santarém Segundo, J. E. (2010) Representação iterativa: um modelo para repositórios digitais. Tesis doctoral. Faculdade de Filosofia e Ciências, Universidade Estadual Paulista. Marília, Brasil. http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/santaremsegundo_je_do_mar.pdf (2013-03-10).
- Santarém Segundo, J. E. ; Vidotti, S. A. B. G. (2011) Rede de tags para recuperação da informação no contexto da representação iterativa. // InCID - Revista de Ciência da Informação e Documentação. 2:1 (jan./jun. 2011) 86-109. <http://revistas.ffclrp.usp.br/incid/article/view/50/pdf> (2013-03-10)
- Silva, J. V. (2009) Gerenciamento do vocabulário de tags do usuário em sistemas baseados em folksonomia Dissertação. Maringá: Universidade Estadual de Maringá. <http://www.din.uem.br/pos-graduacao/mestrado-em-ciencia-da-computacao/arquivos/dissertacoes-1/Jose%20Valderlei%20da%20Silva.pdf> (2013-02-01).
- Sinclair, J., & Cardew-Hall, M. (2008). The folksonomy tag cloud: when is it useful?. Journal of Information Science, 34:1, 15-29.

Enviado: 2013-05-22. Segunda versão: 2013-07-15.
 Aceptado: 2013-09-09.