

# Uma Proposta para o Uso de Folksonomias como Conceitualizações Compartilhadas na Especificação de Modelos Conceituais

Josiane M. P. Ferreira<sup>1,2</sup>, Cesar Augusto Tacla<sup>1</sup>, Sérgio R. P. da Silva<sup>2</sup>

<sup>1</sup>CPGEI – Universidade Tecnológica Federal do Paraná (UTFPR)  
Av. Sete de setembro 3165, CEP 80230-901, Curitiba-PR

<sup>2</sup>Departamento de Informática – Universidade Estadual de Maringá  
Av. Colombo 5.790, CEP 87020-900, Maringá-PR

josiane@gmail.com, tacla@utfpr.edu.br, sergio.r.dasilva@gmail.com

***Abstract.** This work purposes to use data of collaborative tagging as shared conceptualization that can be useful to build conceptual models. The hypothesis assumes that the folksonomy induced from the collaborative tagging's data based on parameters of authorship and motivation of categorization can represent a shared conceptualization of a domain. Thus, it is expected that the utilization of this folksonomies generate the decrease of the divergences in the terms elicitation that will be part of the conceptual model when compared with algorithms of folksonomy induction that don't use this parameters.*

***Resumo.** Este trabalho propõe utilizar os dados de tagging colaborativo como conceitualizações compartilhadas que possam ser úteis na construção de modelos conceituais. A hipótese adotada é a de que a folksonomia induzida dos dados de tagging colaborativo com base nos parâmetros de autoria das tags e motivação das categorizações represente uma conceitualização compartilhada de domínio. Desta forma, espera-se que a utilização destas folksonomias provoque a diminuição de divergências na elicitação de termos que farão parte do modelo conceitual em comparação com algoritmos de indução de folksonomias que não utilizam estes parâmetros.*

## 1. Introdução

Guizzardi (2005, p. 2) adota o nome **conceitualização** para designar o conjunto de conceitos utilizados para articular abstrações do estado das coisas em um domínio. **Modelo** é uma abstração de uma porção da realidade articulada segundo uma conceitualização de um domínio. Ainda, para Guizzardi (2005), tanto conceitualizações como modelos existem somente nas mentes das pessoas. O que há de concreto são **especificações de modelos conceituais** feitas em uma **linguagem de modelagem** que permitem expressar (representar) conceitualizações. Desta forma, a especificação do modelo conceitual – denominada de modelo conceitual, é um artefato concreto que permite aos atores envolvidos no processo de construção do modelo compreender o domínio, atingir consenso sobre o significado das entidades representadas e se comunicar. Deste modo, uma **ontologia**, do ponto de vista de artefato, é um caso particular de modelo concreto.

Na passagem das conceitualizações e modelos abstratos para modelos concretos ocorre o problema descrito por Feigenbaum (1984) denominado de **gargalo de aquisição de conhecimentos** que diz respeito à dificuldade que os engenheiros de conhecimentos têm em capturar e representar conhecimentos a partir de interações com especialistas. As ontologias, modelos concretos destinados a comunidades de usuários, foco particular deste artigo, necessitam de uma aquisição de conhecimentos que envolvem também um grande número de atores, além de outras fontes de informação.

Realizar aquisição de conhecimentos em larga escala é demorado e custoso. Atingir consenso com um número elevado de atores torna-se difícil, pois aumentam as divergências, assim como o número de interações para resolvê-las. Há abordagens de aprendizado de ontologias que se utilizam de métodos e técnicas de processamento de linguagem natural, aprendizado de máquina e mineração de textos para extrair conceitos, relações e instâncias de fontes de informação processáveis (ex. *schemas* de bancos de dados, textos) [Maedche e Staab 2001]. Algumas destas abordagens têm utilizado dados dos sistemas baseados em *tagging* colaborativo como fonte de informação para estes algoritmos.

Sistemas de *tagging* colaborativo são aplicações ditas sociais que permitem aos seus usuários atribuírem etiquetas (*tags*) a recursos da Web. Um recurso pode ser etiquetado por vários usuários com as *tags* que acharem convenientes. O fato interessante é que, apesar de não existir um vocabulário controlado, depois de certo tempo as *tags* utilizadas pelos usuários para etiquetar um recurso parecem se estabilizar [Robu, Halpin e Shepherd 2009]. Este trabalho pretende utilizar os dados do *tagging* colaborativo para induzir estruturas, denominadas de folksonomias, que possam ser consideradas como representativas de conceitualizações compartilhadas de um domínio. A hipótese adotada é a de que estruturas que emergem da dimensão social do *tagging* atenuam o gargalo de aquisição que ocorre na especificação de modelos conceituais de domínios, por representarem uma conceitualização compartilhada em uma comunidade de usuários.

Sendo assim, o objetivo deste trabalho é determinar se as folksonomias que emergem dos dados do *tagging* colaborativo são úteis na construção de modelos conceituais. Especificamente, pretende-se construir um algoritmo que leve em conta informações de autoria das *tags* e de motivação de etiquetagem, e avaliar se as folksonomias produzidas com base nestes parâmetros realmente auxiliam a atenuar o gargalo da aquisição de conhecimento na construção de modelos conceituais. Espera-se que a utilização destas folksonomias provoque a diminuição de divergências na elicitação de termos que farão parte de um modelo conceitual em comparação com algoritmos de indução de folksonomias que não utilizam estes parâmetros.

A seção 2 justifica o uso dos dados de *tagging* colaborativo e comenta sobre outras abordagens neste sentido. A seção 3 discute como este trabalho pretende induzir folksonomias como conceitualizações compartilhadas dos dados de *tagging* colaborativo. A seção 4 descreve a metodologia e a seção 6 comenta as contribuições desta proposta.

## **2. Abordagens que utilizam dados de *tagging* colaborativo**

Em sistemas baseados em *tagging* os usuários podem associar quantas e quais *tags* quiserem para um recurso. Ao associarem as mesmas *tags* aos mesmos recursos, os

usuários constroem um “vocabulário consensual” para um determinado conjunto de recursos que pode ser representativo da conceitualização de um domínio. Este fato é mencionado por vários autores, tais como, Robu, Halpin e Shepherd (2009), Angeletou *et al.* (2007), Jäschke *et al.* (2008), Mika (2007), entre outros.

Alguns autores chamam os dados do *tagging* colaborativo de folksonomia. Neste artigo, o termo folksonomia designa a **estrutura** coletiva (lista de termos, taxonomia, categorização) que emerge do *tagging* colaborativo por meio de um algoritmo de indução de folksonomias [Strohmaier *et al.* 2012].

Os termos resultantes do *tagging* colaborativo carregam uma dimensão social de uso. Por isso, várias abordagens de aprendizado de ontologias, ou que simplesmente objetivam derivar alguma taxonomia ou conjunto de conceitos, utilizam estes dados para construir algum tipo de estrutura “consensual” a partir destes dados. As *tags* derivadas do *tagging* colaborativo apresentam uma estrutura plana, ou seja, a única relação explícita entre duas *tags* é a relação de coocorrência – duas *tags* coocorrem se elas fazem parte de uma mesma etiquetagem. Como esta é a única relação entre duas *tags* várias abordagens que identificam alguma estrutura coletiva das *tags* a utilizam como ponto de partida. Entre estas abordagens encontram-se: Begelman, Keller e Smadja (2006), X. Wu, Zhang e Yu (2006), Jäschke *et al.* (2008), Schmitz (2006), Mika (2007), Cattuto *et al.* (2008), Specia e Motta (2007), Angeletou *et al.* (2007) e Hamasaki *et al.* (2007).

O fato é que várias destas abordagens **pressupõem** que as folksonomias ajudam no desenvolvimento de modelos consensuais pelo fato de resultarem de um processo humano e coletivo sem, no entanto, verificar com profundidade a natureza do conhecimento existente no *tagging* (quem o fez, ou por qual motivo, por exemplo). A maioria das abordagens citadas procura avaliar o algoritmo utilizado que induz a estrutura coletiva dos dados de *tagging*, sem, no entanto, avaliar a utilidade da estrutura derivada, ou a origem dos dados de entrada.

### 3. Folksonomias como conceitualizações compartilhadas

Praticamente nenhuma das abordagens de indução de folksonomias citadas avalia a origem dos dados do *tagging*, como, por exemplo, qual o conhecimento/especialidade do usuário que fez a etiquetagem e o motivo que o levou a etiquetar.

A **motivação** do usuário ao realizar uma etiquetagem pode ser reveladora do significado pretendido para a *tag*, o que é importante no momento de se construir um modelo conceitual. Neste trabalho, defende-se a ideia de que a motivação para criar uma *tag* tem influência no seu uso (ou não) durante a criação de um modelo conceitual. Körner *et al.* (2010) abordam a motivação dos usuários durante a etiquetagem e tentam identificá-la automaticamente separando-as em dois grandes grupos: *tags* de categorização e *tags* de descrição de recursos. Quando as *tags* são utilizadas para categorizar, há pouco uso de sinônimos (o que deve facilitar o consenso entre os atores envolvidos na especificação do modelo conceitual) e a estrutura induzida dos dados de *tagging* se aproxima de uma taxonomia. Quando as *tags* são utilizadas para descrever recursos, então há uso mais proeminente de sinônimos e o vocabulário é, portanto, frequentemente maior, dificultando o consenso na especificação do modelo conceitual.

Outro ponto importante a ser considerado é saber quem realizou a etiquetagem. Segundo Wilson (1983), entidades consideradas autoridades em determinado assunto

tendem a organizar melhor suas informações, possuem conteúdos de qualidade e manterem contato com pessoas que entendam ou tenham interesse no mesmo assunto. O autor define o conceito de **autoridade cognitiva** – uma autoridade fundamentada na competência e nas capacidades intelectuais de quem a recebe e cuja concessão é compreendida como o reconhecimento e o mérito por estas capacidades – uma autoridade que define “quem sabe o quê sobre o quê”. Desta forma, acredita-se que os dados de *tagging* elaborados por usuários que são considerados autoridade cognitiva sobre o domínio de interesse tendem a ser mais informativos sobre o domínio do que os dados de *tagging* elaborados por um usuário leigo.

As abordagens de X. Wu, Zhang e Yu (2006), Jäschke *et al.* (2008), Schmitz (2006), Mika (2007) e Hamasaki *et al.* (2007) para derivar estruturas dos dados de *tagging* colaborativo utilizam informações sobre a autoria das *tags* (em termos de qual usuário utilizou qual *tag* para etiquetar qual recurso) para extrair a relação de coocorrência entre as *tags*, mas sem avaliar o conhecimento do usuário sobre o recurso que está sendo categorizado.

Portanto, propõe-se levar em consideração a autoria das *tags* (em termos de autoridade cognitiva) e a motivação na criação das mesmas para melhor utilizar as folksonomias como fonte de informação na construção de modelos conceituais.

#### 4. Metodologia proposta

Um algoritmo para induzir folksonomias que considere a autoria e a motivação das etiquetas a serem utilizadas na modelagem conceitual está sendo construído. O algoritmo deve selecionar *tags* sobre o domínio para o qual se pretende construir o modelo conceitual. Estas *tags* devem ter sido criadas por usuários categorizadores e considerados autoridades no domínio em questão. Para fins de comparação, um algoritmo de indução de folksonomias que não utiliza as informações de autoria e de motivação, servirá de referência na avaliação (em princípio, será implementado o algoritmo de [Hamasaki *et al.* 2007]), bem como, um segundo algoritmo de controle fundamentado na técnica *TF-IDF*.

Pretende-se realizar experimentos com três grupos: o grupo de teste, que utiliza a folksonomia produzida pelo algoritmo proposto neste trabalho; o grupo de controle I, que utiliza a folksonomia produzida pelo algoritmo de Hamasaki; e o grupo de controle II, que utiliza o conjunto de termos obtidos por *TF-IDF* a partir de um corpus. Cada grupo deve ser formado por pelo menos 10 pessoas. Os grupos de teste e de controle I e II utilizarão um conjunto de termos/*tags* como ponto de partida para modelagens conceituais de domínios variados. No caso do grupo de teste e do grupo de controle I, estes conjuntos de *tags* representam uma folksonomia em uma estrutura plana. Os algoritmos de teste e controle (I e II) utilizarão o mesmo conjunto de anotações como entrada. Os algoritmos de teste e controle I geram folksonomias por meio de suas heurísticas, enquanto que o algoritmo de controle II gera um conjunto de termos utilizando como corpus as *URLs* encontradas nas mesmas anotações. Os dados para induzir as folksonomias/gerar o conjunto de termos serão extraídos de aplicações sociais, tais como, *Delicious*<sup>®</sup> e *Bibsonomy*<sup>©</sup>.

Espera-se que o grupo de teste se depare com um número menor de divergências durante os experimentos de modelagem em relação aos grupos de controle. Portanto, pretende-se utilizar como métrica para efeito de comparação o número de divergências

geradas por cada grupo durante os experimentos. Para controlar as divergências geradas será utilizado o método *CoFolkconcept* [Hauagge *et al.* 2011]. O processo de modelagem no *CoFolkconcept* é colaborativo e se desenvolve da seguinte maneira: *i*) cada usuário constrói um modelo conceitual individualmente utilizando-se de um conjunto de *tags*/termos, produzindo, desta forma, um modelo conceitual particular; *ii*) os diferentes modelos conceituais de cada usuário são comparados a fim de se detectar divergências nas *tags*/termos escolhidos por cada usuário quanto ao tipo (conceito, instância ou relação) e à posição taxonômica (quando forem conceito ou instância); *iii*) resolvem-se as divergências por meio de discussões estruturadas de acordo com a metodologia *DILIGENT* [Tempich *et al.* 2005]; *iv*) gera-se uma nova versão do modelo conceitual que é consensual e repete-se o processo modificando-se individualmente o modelo consensual.

Serão realizados experimentos com diferentes parâmetros de geração das folksonomias a fim de determinar em quais condições o algoritmo de teste produz folksonomias que podem ser consideradas como conceitualizações compartilhadas em função do tipo de modelo conceitual almejado (se mais especializado ou menos especializado).

## 5. Contribuições da proposta

As contribuições desta proposta interessam aos pesquisadores que lidam com modelagem conceitual, em particular, com a atenuação do gargalo de aquisição de conhecimentos na modelagem conceitual, bem como no entendimento da utilização e dos limites de uso das folksonomias como fonte de informação na modelagem conceitual. Particularmente, propõe-se melhorar os algoritmos de indução de folksonomias pelo uso de autoria (autoridade cognitiva) e motivação das etiquetagens.

## 6. Agradecimentos

Agradecemos à Fundação Araucária pela bolsa de doutorado concedida a Josiane M. P. Ferreira durante o seu doutorado, no qual esta proposta será desenvolvida, e ao financiamento firmado no convênio 10/2011-FUP18520.

## Referências

- Angeletou, S., Sabou, M., Specia, L., Motta, E. (2007). “Bridging the Gap between Folksonomies and the Semantic Web: An Experience Report”. In: *Proceedings of Bridging the Gap between Semantic Web and Web 2.0 Workshop, European Semantic Web Conference*.
- Begelman, G., Keller, P., Smadja, F. (2006). “Automated Tag Clustering: Improving search and exploration in the tag space”. In: *Proceedings of Collaborative Web Tagging Workshop at WWW’06*. Edinburgh, Scotland.
- Cattuto, C., Benz, D., Hotho, A., Stumme, G. (2008). “Semantic Analysis of Tag Similarity Measures in Collaborative Systems”. In: *Proceedings of 3rd Workshop on Ontology Learning and Population OLP3*, (pp. 39-43). Patras, Greece.
- Feigenbaum, E. A. (1984). “Knowledge Engineering”. *Annals of the New York Academy of Sciences*, 426: 91–107. doi: 10.1111/j.1749-6632.1984.tb16513.x

- Guizzardi, G. (2005). "Ontological Foundations for Structural Conceptual Models". Telematica Instituut Fundamental Research Series no. 15, Universal Press, The Netherlands, 2005, ISBN 90-75176-81-3.
- Hamasaki, M., Matsuo, Y., Nishimura, T., Takeda, H. (2007). "Ontology Extraction using Social Network", In *Proceeding of International Workshop on Semantic Web for Collaborative Knowledge Acquisition*, vol. 18700163, no. 18700163.
- Hauagge, J. M., Tacla, C. A., Freddo, A. R., Molinari, A. H., Paraiso, E.C. (2011). "The Use of Well-founded Argumentation on the Conceptual Modeling of Collaborative Ontology Development". In: *International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2011, Lausanne. Proceedings of the 2011 15th CSCWD. Piscataway: IEEE, 2011. v. 1. p. 113-119.
- Jäschke, R., Hotho, A., Schmitz, C., Ganter, B., Stumme, G. (2008). "Discovering shared conceptualizations in folksonomias". In: *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 6, Issue 1, February 2008, p. 38-53.
- Körner, C., Kern, R., Grahl, H., Strohmaier, M. (2010). "Of Categorizers and Describers: An Evaluation of Quantitative Measures for Tagging Motivation", In: *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, pp. 157-166.
- Maedche, A., Staab, S. (2001). "Ontology Learning for the Semantic Web". *IEEE Intelligent Systems*, 16(2), 1-18. doi: 10.1109/5254.920602.
- Mika, P. (2007). "Ontologies are us: A unified model of social networks and semantics". In: *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(1), 1-15. Springer. doi:10.1016/j.websem.2006.11.002
- Robu, V., Halpin, H., Shepherd, H. (2009). "Emergence of consensus and shared vocabularies in collaborative tagging systems". *ACM Transactions on the Web*, 3(4), 1-34. doi:10.1145/1594173.1594176
- Schmitz, P. (2006). "Inducing ontology from Flickr tags". In: *Proceedings of Collaborative Web Tagging Workshop, 15th WWW Conference*, Edinburgh.
- Specia, L., Motta, E. (2007). "Integrating Folksonomies with the Semantic Web". In: *4th European Semantic Web Conference (Vol. 4519, pp. 624-639)*. Berlin Heidelberg, Germany: Springer-Verlag.
- Strohmaier, M., Helic, D., Benz, D., Orner, C. K., and Kern, R. (2012). "Evaluation of Folksonomy Induction Algorithms". To appear. *Transactions on Intelligent Systems and Technology*.
- Tempich, C., Pinto, H. S., Sure, Y., Staab, S. (2005). "An argumentation Ontology for Distributed, Loosely-controlled and evolving Engineering processes of ontologies (DILIGENT)". *The Semantic Web: Research and Applications – Lecture Notes in Computer Science* (pp. 241-256). Springer.
- Wilson P. (1983) *Second-hand knowledge: An Inquiry into Cognitive Authority*. WestPort: Greenwood Press.
- Xu, X., Zhang, L., & Yu, Y. (2006). "Exploring social annotations for the semantic web". In: *Proceedings of the 15th international conference on World Wide Web*, 417. New York, New York, USA: ACM Press. doi:10.1145/1135777.1135839