Dynamic Tagging for Enterprise Knowledge Sharing and Representation

Emergent Research Forum Papers

Dazhi Chong
Department of Information Technology & Decision Sciences,
Old Dominion University
e-mail address: dchon001@odu.edu

Hui Shi
Department of management & information sciences,
University of southern Indiana
e-mail address:hshi@usi.edu

Harris Wu
Department of Information Technology & Decision Sciences,
Old Dominion University
e-mail address: hwu@odu.edu

Wu He
Department of Information Technology & Decision Sciences,
Old Dominion University
e-mail address: whe@odu.edu

Abstract

The development of Web 2.0 technology provides an easy way for people to transfer and share knowledge. As a collaborative tagging tool, folksonomy is an efficient indexing method in Web 2.0 environments. After analyzing the pros and cons of current knowledge management systems, this research proposes a dynamic Tagging system that combines folksonomy technologies with other approaches including automatic schema enrichment and training. The proposed system improves access to a large, growing collection by supporting users collaboratively contribute to the building of tags. In addition, the proposed system provides an efficient way for firms to represent knowledge and share knowledge with customers and other firms.

Keywords

Tagging, folksonomy, knowledge, collaboration.

Introduction

The key problem of enterprise knowledge sharing and representation is how to find and retrieve knowledge effectively. The rise of Web 2.0 technology provides an easy way for people to transfer and share knowledge. Popular Web 2.0 tools like wikis, blogs, and social tagging systems all contain a large amount of user-contributed information. Enterprise users have adopted social software to store and share information on a large scale. Among these tools, tagging systems demonstrate special value. A tagging system enables users to assign tags, to online resources such as web pages, videos etc. Tagging systems don't have to rely on a specific vocabulary or a previously defined structure. The tags reflect users' understanding of online resources and facilitate users to characterize the resource based on their own needs. These user-dependent tags also reflect users' social or cultural backgrounds and perception of the world (Sharif, 2009). Users can easily create, rename, merge and delete these tags. With the contributions of multiple users, a system of classification is created from the collaboratively creating tags (Peters, 2009). This practice is also called “folksonomy”. As a flexible knowledge organization method, folksonomy can be easily applied to different contexts. Since there are no specific skills needed for tagging and categorization in folksonomy practice, tremendous amount of information can be generated by lots of users in a short time. Thus, it provides a short cut for knowledge sharing and representation.

In business world, changes happen unexpectedly and continuously. In order to obtain more market shares and profits, firms have to respond as fast as possible to the market changes. Under this circumstance, organizational agile reflects firms' capability to quickly detect opportunities and take actions (Trinh et al., 2012). In order to improve this capability, more organizations are focusing on improving organizational processes, increasing the corporation among employees and making use of internal and external information. A research conducted by Economist Intelligence Unit (2009) found that organizations come into frequent contact with people from outside the organizational boundaries are much more agile than organizations that are not. This conclusion may suggest that interaction with the customer is the solution to become more agile. At this point, folksonomy provides an efficient way for firms to interact with customers and collect market information. By using a folksonomy, businesses can move from traditional time-consuming interaction approach to a quicker, deeper and more open approach. For example, many companies use Twitter as a powerful tool to interact with customers and intra-enterprise information sharing. Twitter hashtags allow users to tag online contents. Using the folksonomy of hashtags, these customers' feelings and opinions in twitter messages can be easier detected by businesses, which enables firms to quickly react to changes in customer preferences.

However, folksonomy also shows some drawbacks. Lack of specific vocabulary results in ambiguity. Flat organization of tags can't reflect hierarchal structure of reality (Pan, Taylor and Thomas, 2008; Passant, 2007). All these drawbacks lead to lack of precision when sharing and representing knowledge.

This paper attempts to remedy drawbacks of folksonomy by introducing a dynamic tagging system. The remainder of the paper is structured as follows: first, we will investigate main solutions of knowledge sharing and representation on internet. Secondly, by analyzing the pros and cons of current approaches, we introduce the framework of our system. A case study is used to evaluate the usefulness and efficiency of our system. Finally, we draw some conclusions of this study.

Knowledge sharing and representation

The development of technologies dramatically changed the approaches in knowledge sharing and representation. Basically, there are two main approaches (Weller, 2007). The first one is ontology. As a static, system-dependant model of knowledge management system, ontology focuses on identifying main categories in the domain and specifying the constraints on the ways in which the relationships can be used. (Park et al. 2013). More specifically, ontology uses formal semantics to specify conceptualization and make it understandable to machines (Sharif, 2009). A lot of attempts have been made to categorize and represent information and relationships using ontologies. Park et al. (2013) proposed a framework for mining and understanding sentiment in online content. Garcia-Crespo et al.(2010) adopted ontology and TF-IDF to analyze customer emotion relationship. However, many studies tended to develop their own ontologies based on specific conditions. The obstacles to a generalization of ontologies are mainly on their cost of design and maintenance (Limpens et,al 2008).

The second approach, folksonomy, has been treated as an easy way to make knowledge accessible and retrievable (Sharif, 2009). Folksonomy allows all users to tag documents with freely chosen words. Thus, it's easier to collect users' opinions and get consensual agreement on online contents. As a cheap method of indexing, Folksonomy don't need users to develop their own domain and scope of the content. The search function of folksonomy simplifies the procedures to find and share knowledge in online materials. However user-created tags in folksonomy may have an initial quality problem compared to ontologies. The quality of freely chosen words rely on users' understanding to online content. Lack of vocabulary control will definitely result in redundant and low quality tags.

The semantic relations between tags are not will utilized by most folksonomy systems. Peters (2006) pointed out that it is infeasible to use folksonomy as the only way to process data in professional environments. Mixing folksonomy with other indexing methods may result in better output. One of these solutions is to combine folksonomy with other knowledge organization systems (Weller, 2007). Peters & Stock (2007) argued that treating tags as elements of natural language and adopting automatic methods of natural language processing will result in better retrieval results. Another approach focuses on the training and education of users. (Peters and Weller, 2007) The training will bring a deeper understanding of indexing contents, as a result, it will improve the quality of tags.

Based on Peters and Weller's(2007) advices, we propose a system that improves access to a large, growing collection by supporting users collaboratively contribute to the building of tags . Different from a single hierarchy, this system allows assignment of multiple classifications to an object, supporting multiple user perspectives in search and exploration.

Dynamic hierarchical tagging system

The features of folksonomy decide that the quality of tags is relied on users' experience and understanding to the content. However, we believe that the continuous training and design improvement of content may lead to better quality. Compared to other tagging systems, Categories in Dynamic Hierarchical Tagging System (DHTS) can be considered as a hierarchy of tags along certain dimensions, users can associate documents with a category structure, therefore more thoughts are put into tagging and classification; Compared to a simple hierarchical schema, DHTS gives the users the ability to find items more effectively along different dimensions. Also, DHTS allows the resulting classification schema to be the guideline of future tagging. the guided tagging feature prompts users with tag suggestions, which will improve correctness, consistency, unambiguity of tags. For example, when a user wants to tag a picture, the system will ask user to provide the category tag and subcategory tag of this picture, such as "Location -> Washington", these two tags will be treated as guided tags which can help the future tagging.

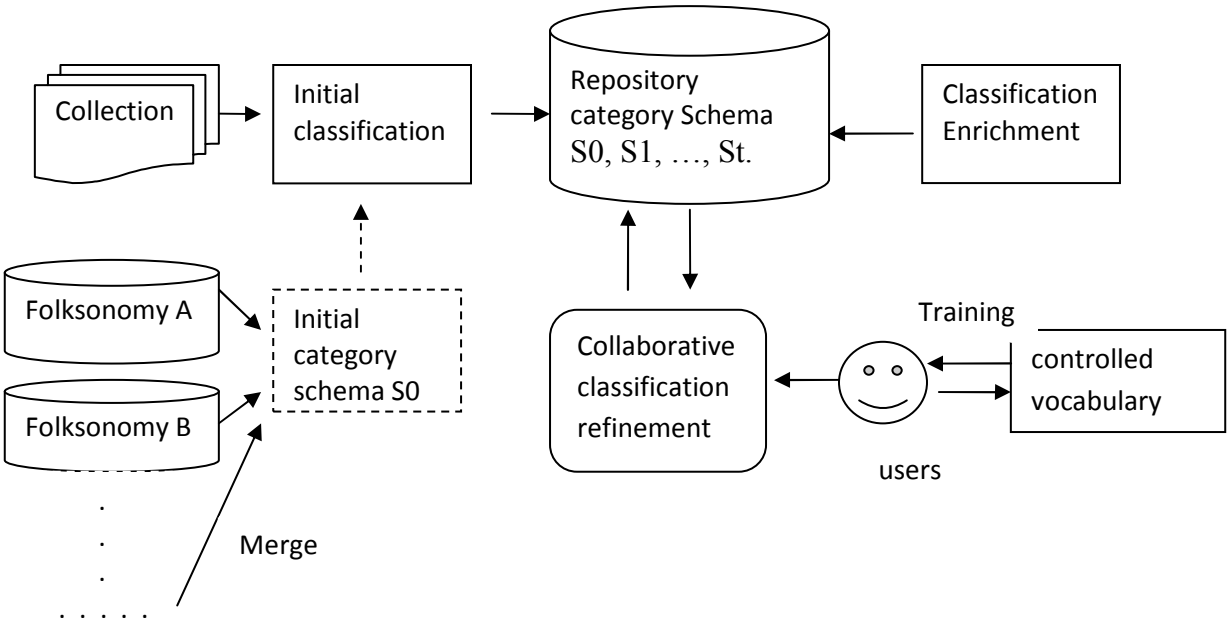

Figure 1 The framework of dynamic hierarchical tagging system

The following elaborates the tagging process of DHTS.

1. A collection C has n documents with no existing category classification.

2. An initial category schema, S0, is developed from the refinement of multiple folksonomy websites. A portion of the collection, T, is automatically classified into S0. In this stage, the refinement is based on the simple hierarchical schema of multiple folksonomies. These schemas represent different understanding of the same records in various folksonomies. By combining tags with same or similar meanings and refining tags with different meanings, an initial category schema will be created.

3. The system-made classifications are called initial classification. The initial classification will be available to user search and exploration, it provides guidance and suggestions to users' classification efforts. The users can accept, reject or modify the initial classifications.

4. As a prerequisite of efficient classification, users get training from experts. The training includes the introduction to key controlled vocabularies and key terms in the collection. As more folksonomy data added to collection, users will accept further training which focuses on new vocabularies and terms.

5. Over time, users continue to classify documents into the category schema. Users collaboratively edit the initial category schema, by add, modify categories, or delete the existing ones. The schema evolves over time: S0, S1, ..., St.

6. As the category schema changes, the system will tentatively classify existing documents to new categories that belong to F't, a "stable" subset of the revised category schema Ft. At this stage, a statistical co-occurrence model (Sanderson and Croft, 1999) will be used to identify parent-child relationships: X subsumes Y if: $P(x|y >= 0.8)$ and $P(y|x < 1)$. For example, suppose X = "glass", and Y = "stained glass". If most documents tagged with "stained glass" are also tagged with "glass", then "glass" subsumes "stained glass. Future users will accept or reject these classifications as they tag the collection.7. As new documents are added to the collection, the system will tentatively classify the document if any metadata is available.

8. As new users come to the collection, and as new documents are added to the collection, and as users' interests change, users will bring in new perspectives and evolve the category schema. With the contribution of numerous users, the quality and accuracy of facet schema will improve continuously

Case study

For the experiment, we imported collections from two popular folksonomy websites (BibSonomy and CiteULike). The reasons why we chose these two websites were: first, the individual collections have limited metadata, it's very hard to get common metadata standard across different collections. For CiteUlike and BibSonomy, both of them are the most popular social bookmarking website. They allow researchers and scholars to download their datasets. Also, users of these two websites can store , organize, tag their bookmarks and publication references. Large amount of active users on these two websites make it possible to retrieve high quality metadata from their collections. Second, in order to merge collections from multiple folksonomy websites, it's important to find common records among them. Both CiteUlike and BibSonomy focus on academic collections which will facilitate the building of initial category schema. Third, our system is based on continuous training and design improvement, the improved tagging quality will have an immediate, tangible social value by facilitating know sharing and representation among academic studies.

For the first step, we retrieved two collections from CiteUlike and BibSonomy. The time span of CiteUlike is from November 4, 2004 to June 18,2014. The collection of CiteUlike contains 4,755,674 articles and 873,873 tags. The time span BibSonomy is from June 30, 2006 to January 1, 2014. The collection of BibSonomy contains 522,134 articles and 177,281 tags. In order to simplify data merging, we used DOI to retrieve common articles among these two collections. We found 10,955 common articles published by both BibSonomy and CiteULike. By analyzing the tags attached to these articles, we found that some articles don't contain any tags. Lack of tags make it impossible to build initial category schema, thus, none-tag articles were eliminated. Finally, we got 7766 articles, all of them were tagged by BibSonomy users and CiteULike users.

In this study, 213 students at a large public university were recruited to further tag and classify the articles using our system. These students were assigned to 43 groups, 20 different articles were assigned to each group.

Training is a key factor which can affect tagging quality. Most articles retrieved from CiteUlike and BibSonomy are ACM-related papers. To conduct the training, first, we obtained a controlled vocabulary from the list of ACM-approved categories (http://dl.acm.org/ft_gateway.cfm?id=2371137&ftid=1290923&dwn=1). Then experts with computer science master degree introduced the meaning and differences of keywords in taxonomy. During the tagging, students in each group were asked to read the abstract of each article, and collaborate with each

other to revise initial category schema of these articles. Finally, the initial category schema was revised to a new version.

As a dynamic tagging system, continuous evolving and design improvement are necessary for high quality of category schema. In next stages of this study, we will recruit more students to contribute to the evolving of tagging. Also, we will evaluate the effectiveness of the proposed system and the deployment methodology.

Conclusion

This research attempts to address problems of current knowledge sharing and representation on online contents. Folksonomy was proved to be an easy and cheap way to retrieve information from internet. At the same time, it also has some drawbacks such as lack of vocabulary control and hierarchal structure. By combining folksonomy with other approaches, we propose a system that improves access to a large, growing collection by supporting users collaboratively contribute to the building of tags. One implication of this study is that the dynamic hierarchical tagging system can be expanded into larger collections which have millions of items and users. Sambamurthy et al. [2003] argue that information technology facilitates firm's agility through the digitization of knowledge and the business process. Our framework provides an efficient way for firms to quickly compile and analyze information, detect intra and inter organizational relationships. In the next stage, we will continue the refinement of our category schema. The evaluation of our system will be conducted, for example, it's necessary to check the effectiveness of our system in discovering quality problems with categories and document classifications and efficiency.

REFERENCES

Alexiev, V., Breu, M., de Bruijn, J., Fensel, D., Lara, R., & Lausen, H. (Eds.). 2005. "Information integration with ontologies: Experiences from an industrial showcase". Chichester, England: Wiley & Sons.

Bhatt, G., A. Emdad, N. Roberts, and V. Grover .2010. “Building and Leveraging Information in Dynamic Environments: The Role of IT Infrastructure Flexibility as Enabler of Organizational Responsiveness and Competitive Advantage”, Information & Management (47:7–8), pp. 341–349.

Economist Intelligence Unit. 2009. "Organizational agility: How business can survive and thrive in turbulent times". The economist, pp. 1-25.

Limpens, F., Gandon, F., & Buffa, M. 2008, September. "Bridging Ontologies and Folksonomies to leverage knowledge sharing on the social web: a brief survey". Automated Software Engineering-Workshops, 2008. ASE Workshops 2008. 23rd IEEE/ACM International Conference . IEEE. pp. 13-18

Pan, J. Z., Taylor, S., & Thomas, E. 2009. "Reducing ambiguity in tagging systems with folksonomy search expansion". The Semantic Web: Research and Applications . Springer Berlin Heidelberg. pp. 669-683

Passant, A. 2007, March. "Using ontologies to strengthen folksonomies and enrich information retrieval in weblogs". International Conference on Weblogs and Social Media.

Park, E., Storey, V., and Givens, S. 2013. "An Ontology Artifact for Information Systems Sentiment Analysis". ICIS 2013.

Peters, I. 2006. "Against folksonomies: Indexing blogs and podcasts for corporate knowledge
management". Proceedings of Preparing for Information 2.0, Online Information 2006 (pp. 93-97). London: Learned Information Europe.

Peters, I. 2009. "Folksonomies: Indexing and retrieval in Web 2.0". Vol. 1. Walter de Gruyter.

Sambamurthy, V., A. Bharadwaj, and V. Grover. 2003. “Shaping Agility Through Digital Options: Reconceptualizing the Role of Information Technology in Contemporary Firms”, MIS Quarterly (27:2), pp. 237–263.

Sanderson, M. and B. Croft. 1999. "Deriving concept hierarchies from text". Proceedings of the 22nd ACM Conference of the Special Interest Group in Information Retrieval , pp.206-213.

Sharif, A. 2009, August. "Combining ontology and folksonomy: an integrated approach to knowledge representation". World Library and Information Congress (IFLA General Conference and Assembly), Milan, Italy.

Stock, W. G. 2007. "Information retrieval". Informationen suchen und finden. München, Germany: Oldenbourg.

Weller, K., Peters, I., & Stock, W. G. 2009. "The Collaborative Knowledge Organization System". Handbook of Research on Social Interaction Technologies and Collaboration Software: Concepts and Trends: Concepts and Trends, pp. 132-144.

Trinh, P., Molla, A., & Peszynski, K. 2012. "Enterprise systems and organizational agility: a review of the literature and conceptual framework". Communications of the Association for Information Systems, (31:1), pp.167-193.

Weller, K. 2007. "Folksonomies and ontologies: Two new players in indexing and knowledge representation". Online Information Conference Proceedings .pp. 108-115. London: Learned Information Europe.

Vander Wal, T. 2005, February 21. "Explaining and showing broad and narrow folksonomies".
Retrieved February 21, 2015, from http://www. vanderwal.net/random/entrysel.php?blog=1635