Journal of Universal Computer Science

J.UCS

http://www.jucs.org

Proceedings of I-KNOW 2010

10th International Conference on Knowledge Management and Knowledge Technologies

Graz, Austria, September 1 – 3, 2010

Editors I-KNOW

Klaus Tochtermann • Graz, Austria and Kiel / Hamburg, Germany

Hermann Maurer • Graz, Austria

Editors-in-Chief of J.UCS

Hermann Maurer (Managing Editor) • Graz, Austria

Narayanan Kulathuramaiyer • Sarawak, Malaysia

Klaus Tochtermann • Graz, Austria and Kiel / Hamburg, Germany

Expert Recommender Systems: Establishing Communities of Practice Based on Social Bookmarking Systems

Tamara Heck
(Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany
tamara.heck@uni-duesseldorf.de)

Isabella Peters
(Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany
isabella.peters@uni-duesseldorf.de)

Abstract: Recommender systems have established mostly in e-commerce, whereas in companies or scientific institutions the recommendation of experts und possible colleagues has yet been discussed mostly theoretically. We propose a recommender system on the basis of Social Bookmarking Systems and Folksonomies, which may help to find communities of practice, where people share the same interests and support each other in their working or scientific field. The paper reports research in knowledge management and information retrieval, and therefore offers new insights and fields of studies in information science.
Keywords: recommender system, social bookmarking, communities of practice, folksonomies
Categories: M.5, M.7, H.3.3

1 Introduction

The Web 2.0 offers great innovative possibilities: By using 2.0 technologies, new behaviors and innovative collaboration methods can develop [Delic, 08]. Scientists and researchers can use these new ways to gain scientific knowledge, share it, offer their results and collaborate on Web 2.0 platforms. In scientific environments the term "Science 2.0" has established as a definition for new forms of communication. [Waltrop, 08] says that Science 2.0 isn't anymore a "competition" between scientists but gets even more "collaboration". On the other hand [McAfee, 06] coined the term "Enterprise 2.0" which describes the use of Web 2.0 in companies "to make visible the practices and outputs of their knowledge workers" [McAfee, 06, p. 23].

In the following we describe a current research project which aims at developing an expert recommender system, i.e. a software tool which recommends similar users based on same bookmarks and tags within Social Bookmarking Systems. This prototype should find possible colleagues for establishing Communities of Practice (CoPs). It's the basis for further research to analyze how Web 2.0 applications can be used more effectively for scientific and business communication. An important part will be the evaluation of our system. Therefore the development of the tool is accompanied by a qualitative survey in which researchers report their attitude towards CoPs and Social Bookmarking. First results of this survey and exemplary user recommendations will be presented in section 3.

1.1 Communities of Practice (CoPs)

It seems that beneath the recommendation of similar resources in a collaborative system the recommendation of similar or relevant users is far more important [Diederich, 06]. [Panke, 08] asked approximately 200 users about their tagging behaviour: Two-thirds of them use tags to socialize with other users. CoPs are groups of people, who have the same interests, change information and knowledge and cooperate with each other [Wenger, 00]. Three aspects are important: First, members of a CoP have a joint enterprise of what their community is about. Second, the community is built on mutual engagement, and third, the members produce a "shared repertoire of communal resources" such as language and tools. CoPs in companies and institutions can exist of division-intern or -extern members, members who work at the same location or on different places or even members, who don't belong to the same company. The important factor is that CoPs establish and organize themselves. CoP-members meet willingly and are not ordered by leadership because enforcement could lead to refusal of cooperation [Blair, 02]. That is the main difference between them and teams established by company managers [Wenger, Snyder, 00]. The modern web with its social networking systems or web-based tools like SBS may support the cooperation of CoP-colleagues. The "virtual meeting" cannot replace the face-to-face meeting of the members [Gust von Loh, 09]. But maybe the contact via internet is the first step to develop a CoP.

1.2 Social Bookmarking Systems (SBS) and recommender systems

SBS offer platforms, on which users could archive their references to have access to and manage them from any web-accessible device. Examples of SBS are BibSonomy, CiteULike and del.icio.us. Another important aspect of SBS is collaboration. The users' bookmark lists are made public and can be used by any other participant of the platform. Every bookmark can be tagged by keywords. Users can tag bookmarks of others and help each other to organize their database. So SBS is not only an individual resource management system, it's a collaborative system, where community users act through combined resources. SBS exploit the main features of social networking. That is why companies nowadays embed SBS in their intern systems [John, 06].

Recommender systems make use of collaborative filtering, that means to restrain a quantity of information with the help of a user community: "Collaborative filtering simply means that people collaborate to help one another perform filtering by recording their reactions to documents they read" [Goldberg, 92]. This is based on the idea of a "referral chain", i.e. a user requires on his relations and contacts to get relevant information or find an expert to solve a problem. [Kautz, 97] alludes to a great advantage of collaborative filtering systems: "A user is only aware of a portion of the social network to which he or she belongs. By instantiating the larger community, the user can discover connections to people and information that would otherwise lay hidden over the horizon." In information retrieval we use the tripartite structure of terms, resources and users to filter information. The principle herewith is: "Co-occurrence means similarity." Collaborative filtering systems are often used in e-commerce, for example in the online catalogue Amazon (user similarity based on similar resources).

A folksonomy [Peters, 09] is a set of user-generated keywords, called tags, in a collaborative information system. Folksonomy-based filter- or recommender-systems are able to suggest two kinds of information: "Recommendation systems may, in general, suggest either people or items" [John, 06]. The recommended resources are generated with similarity measures and clustering methods [Shardanand, 95].

2 Establishing CoPs based on SBS

The question is: How can the process of establishing CoPs be supported and initiated? How can companies and institutions advise their colleagues to each other without enforcing their cooperation? [Wenger, 2000, p. 144] claims: "The task is to identify such groups and help them come together as communities of practice." Collaborative information systems with their inherent networking structure [Peters, 09], such as SBS and folksonomies, are able to visualize CoPs and make the sharing of knowledge more effective. The claim to get relevant information is now: "More like me!" – find users, who are similar to me so that I may get relevant information from them. We suppose, that users are similar to each other when they use equal or similar tags for indexing a resource ("thematic linkage"), or when they index, edit and save the same resources ("bibliographic coupling") [Kessler, 63].

So far there are a few empirical studies, which examine the adoption of recommender systems based on folksonomies and their advantage in information retrieval and knowledge management. It should be mentioned the study by [Jäschke, 07], which analyzes the effect of recommended search-tags on retrieval efforts (based on BibSonomy and the music recommendation system Last.fm).

2.1 Method

We first chose the SBS CiteULike to develop an expert recommender system which analyzes the tripartite relation between users, tags and bookmarks. As we cooperate with the Forschungszentrum Jülich (FZJ), whose scientists helped us setting a relevant database, we chose a set of 45 relevant journals of solid-state physics (bookmarked between 2004 till 2008). Physicists don't use SBS more often than other scientists, but they are used to bibliography reference systems like JabRef. This was one relevant aspect for us, as the physicists will evaluate our recommender tool.

The first step to recommend users is to measure the similarity using one of the common coefficients Dice, Cosine or Jaccard-Sneath, which calculate the similarity between one user and the other users of a SBS. Using the Dice coefficient [Dice, 45] for example, D_i and D_i are users, a the number of bookmarks (or of tags) of user D_i, b the number of bookmarks (or of tags) of user D_j and g the number of bookmarks (or of tags) which both users applied:

$$S_{D_i,D_j} = \frac{2g}{a+b}$$

As we suppose an elaborated method and not the k-nearest-neighbour algorithm, the second step is to calculate the coincidence of the users and transfer the results into a cluster structure applying the single-link-, complete-link- or group-average-link-

method [Knautz, 10]. In this way we gain similar users who can be recommended to each other in order to establish a user-community, a possibly first step to build a CoP.

After the implementation of the recommender system an evaluation, which focuses on the quality of the system, will follow. Thus, we send a survey to FZJ-employees, which got us a first impression of the handling and understanding of SBS and CoPs. The same participants will also evaluate our recommender tool.

3 Results

3.1 Use of SBS and CoPs in Science

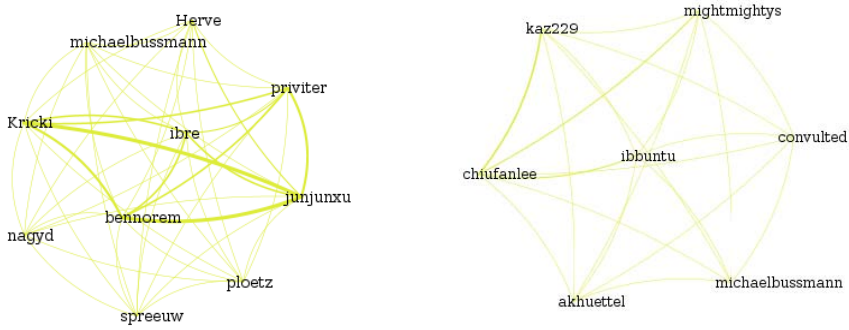
The survey sent to 363 employees yields to following results: 43 employees (25 till 62 years old) attended the survey, which is 11,85 %. The questions to SBS showed appalling, but also revealing results: Only one participant really uses one SBS (del.icio.us). We also asked why they use/don't use SBS. One answer was: "For me, search engines, normal bookmarks and literature databases like JabRef work fine." It seems that several systems for information management and retrieval have established and the new services have difficulty to convince users of their advantage. 7 respondents thought that SBS are "less important", 14 that they are "not important".

2 out of 27 said that they work in a CoP. One CoP was founded through "contact through a conferences or a colleague", the other one "through connections by people formerly working at the same institute." This shows us, that CoPs, although not common between researchers, have found their way into scientific work. 8 % out of 25 said, that CoPs and working groups could facilitate their work well, 16 % said that they could. We also asked if SBS could help to socialize with other colleagues and researchers and to establish working groups, which 4 out of 12 answered with "yes". One comment was that the "communication is made easier." 10 out of 28 thought a recommendation system that proposes scientists with same interests for possible cooperation could be helpful. Comments on this questions show, that most participants prefer the personal contact as the most important factor for cooperation. But a "serious structured" recommendation system could support the work, "especially for younger scientists."

3.2 Clustering Experts

In the SBS CiteULike there were 1,006 users and 2,861 bookmarks for our database. We forgo an aggregation of both similarity values (user-bookmarks and user-tags) in order to better recognize the different users who would be recommended by the system based on the different calculations. Users who have put only one bookmark into the SBS were left out of calculation because they would show improper results, i.e. the similarity between them and other users would be calculated very high. To visualize our results and gain recommendation candidates we built up clusters. Figure 1a shows the resulting cluster (with a threshold of 0.1) for the user "michaelbussmann" based on similar bookmarks. It can be seen that this user belongs to a group, where the users "Kricki", "bennorem" and "junjunxu" show the strongest similarity (shown by the thickness of the edges between the users). In figure 1b the complete-link cluster for user "michaelbussmann" based on tags (threshold set on 0.1)

is displayed. It is obvious that both generated clusters show great differences in the recommended users. Users who were recommended because they put the same bookmarks into the SBS do not appear in the cluster based on tag similarity, and conversely users with similar tag-behaviour often have not indexed the same bookmarks. It is striking that some users seem to be "tired of tagging", i.e. they often use one tag, which makes it difficult to calculate similarities based on tags.



Figures 1a/b: Complete-link cluster based on similar bookmarks (left) and based on tags (right) for the user "michaelbussmann", threshold = 0.1, similarity measure: Dice, source: CiteULike

4 Conclusion and future work

Using users-tags-resources relations as in SBS for analyzing similarities provide a lot of opportunities for establishing CoPs and improve cooperation and socialisation of members. Yet, our survey confirms results of other studies [Bernius, 09] that usage ratio of SBS in Science 2.0 is rather rare. Enterprises 2.0 show a different picture [Lai, 08] and can document slightly higher usage statistics. For proper expert recommendations as well as establishment and support of CoPs more intensive user activity in SBS is necessary and will bring better results.

To forego the lack of user activity we will use the publications and references of our interviewees as basis for similarity calculations. The interviewees become "simulated users" as we assume that they have entered their own publications and references into the SBS. Resulting recommendations will suggest our interviewees "real users" of the SBS. Evaluation of the recommender system will follow via half-standardized interviews with the FZJ researchers.

Still open to research are analysis of threshold values for cluster-size regulations and the implementation of weighted similarity relations. The roles of concepts like "centrality", "betweeness" and "degree" [Wasserman, 94] and their representation in CoPs have not yet been discussed as well. Furthermore there is to be discussed the conditions to build a real CoP and how the members should cooperate to make their group work effectively.

Acknowledgements

We would like to the thank the researchers at the FZJ, the members of the team project "SoBoCops" and Wolfgang G. Stock for their support in preparing this paper. Parts of this work are supported by the DFG (STO 764/4-1) and the Strategischer Forschungsfonds of the Heinrich-Heine-Universität Düsseldorf.

References

[Bernius, 09] Bernius, S., Hanauske, M., Dugall, B.: Von traditioneller wissenschaftlicher Kommunikation zu "Science 2.0", ABI-Technik 29, 2009, 214-226.

[Blair, 02] Blair, D.: Knowledge Management: Hype, Hope, or Help, In Journal of the American Society for Information Science and Technology, 53/2, 2002, 1019-1028.

[Delic, 08] Delic, K. A., Walker, M.A.: Emergence of The Academic Computing Clouds In ACM Ubiquity, 9/31, http://www.acm.org/ubiquity/volume_9/v9i31_delic.html.

[Dice, 45] Dice, L.R.: Measures of the Amount of Ecologic Association between Species, In Ecology, 26/3, 1945, 297-302.

[Diederich, 06] Diederich, J.; Iofciu, T.: Finding Communities of Practice from User Profiles Based on Folksonomies, In (E. Tomadaki; P. Scott, eds.) Innovative Approaches for Learning and Knowledge Sharing, EC-TEL 2006 Workshops Proceedings, Crete, Greece 2006, 288-297.

[Goldberg, 92] Goldberg, D.; Nichols, D.; Oki, B. M.; Terry, D.: Using Collaborative Filtering to Weave and Information Tapestry, In Communications of the ACM, 1992, 35(12), 61-70.

[Gust von Loh, 09] Gust von Loh, S.: Evidenzbasiertes Wissensmanagement, Wiesbaden: Gabler, 2009.

[Jäschke, 07] Jäschke, R.; Marinho, L.; Hotho, A.; Schmidt-Thieme, L.; Stumme, G.: Tag Recommendations in Folksonomies, In Lecture Notes in Artificial Intelligence, 2007, 506-514.

[John, 06] John, A.; Seligmann, D. (2006): Collaborative Tagging and Expertise in the Enterprise, In Proc. 15th Int. Conf. on World Wide Web, Edinburgh, Scotland 2006. ACM Press, New York, NY, 2006.

[Kautz, 97] Kautz, H.A., Selman, B., Shah M.A.: Referral Web: Combining Social Networks and Collaborative Filtering, In ACM 40/3, 1997, 63-65.

[Kessler, 63] Kessler, M. M.: Bibliographic Coupling between Scientific Papers, In American Documentation, 1693, 14, 10–25.

[Knautz, 10] Knautz, K., Soubusta, S., Stock, W.G.: Tag clusters as information retrieval interfaces, In Proceedings of the 43th Annual Hawaii International Conference on System Sciences (HICSS-43), IEEE Computer Society Press, 2010.

[McAfee, 06] McAfee, A.P.: Enterprise 2.0: The Dawn of Emergent Collaboration, In MIT Sloan Management Review, 47/3, 20–28.

[Panke, 08] Panke, S.; Gaiser, B.: Nutzerperspektiven auf Social Tagging – Eine Online Befragung, 2008, www.e-teaching.org/didaktik/recherche/goodtagsbadtags2.pdf.

[Peters, 09] Peters, I.: Folksonomies: Indexing and Retrieval in Web 2.0., De Gruyter, Saur, München, 2009.

[Peters, 08] Peters, I.; Stock, W.G.: Folksonomies in Wissensrepräsentation und Information Retrieval, In Information - Wissenschaft & Praxis, 59/2, 2008, 77-90.

[Shardanand, 95] Shardanand, U.; Maes, P.: Social Information Filtering: Algorithms for Automating "Word of Mouth", In Proc. Human Factors in Computing Systems, Denver, Colorado, USA 1995, 210-217.

[Waltrop, 08] Waldrop, M.M.: Science 2.0 - Is open access science the future? In Scientific American. 2008, http://www.sciam.com/article.cfm?id=science-2-point-0.

[Wasserman, 94] Wasserman, S.: Faust, K.: Social Network Analysis: Methods and Applications, Cambridge University Press, Cambridge, 1994.

[Wenger, 00] Wenger, E.: Communities of Practice and Social Learning Systems, In Organisation, 7/2, 2000, 225-246.

[Wenger,Snyder, 00] Wenger, E., Snyder, W.: Communities of Practice? The Organizational Frontier, In Harvard Business Review, 78/1, 2000, 139-145.