Induced Tagging: Promoting Resource Discovery and Recommendation in Digital Libraries

J. Alfredo Sánchez, Adriana Arzamendi-Pétriz, Omar Valdiviezo
CIRIA/CENTIA/ICT
Universidad de las Américas Puebla
Santa Catarina Mártir, Cholula, Puebla, Mexico
+52-222-229-2666

{j.alfredo.sanchez, arzamendip, omar.valdiviezo}@gmail.com

ABSTRACT

We introduce the notion of "induced tagging" in the context of learning communities that are supported by digital libraries. We also describe an environment aimed to foster discovery and recommendation of digital library resources based on induced tagging.

Categories and Subject Descriptors

H.3.7 [**Digital Libraries**]: Dissemination, user issues.

General Terms

Management, Design, Experimentation, Human Factors.

Keywords

Tagging, recommendation, knowledge discovery, collaboration.

1. INTRODUCTION

Finding relevant resources in vast digital libraries to support specific knowledge tasks remains a challenge both for users and access facilitators. Approaches to address this challenge include novel information retrieval algorithms and heuristics, visualization techniques, and automatic resource classification. So-called *folksonomies* take advantage of collaborative tagging that spontaneously generates, after some period of activity, categories for highly varied contents.

The enormous success of tag-oriented websites such as *delicious.com* and *flickr.com,* as well as a number of studies show promising results regarding the potential of collaborative tagging to assist users in finding connections among information resources. As traditional categorization schemes are challenged, barriers are also perceived that could prevent collaborative tagging from being used effectively in the context of more formal academic settings. In the realm of librarianship, for example, controlled vocabularies are preferred and precision and collocation have been cited as important drawbacks in folksonomies [3]. Others have noted that the lack of user incentives or social awareness will have a negative impact on the effectiveness of tagging systems [2, 4].

We posit that collaborative tagging opens new opportunities for information services that support academic communities. Digital libraries require new skills but also allow for new ways of applying existing expertise. In particular, we consider that knowledge about resources and user needs that has been developed by personnel who regularly serve in reference desks and online information services can be applied to the construction of meaningful, consistent resource categorizations and provide the basis for helpful content recommendation and discovery in digital libraries. Thus, we have designed REC, an environment involving human and technological components which is aimed to promote collaborative tagging of digital library resources and to take advantage of the resulting tags for producing recommendations. A salient feature of REC is its tagging scheme, which we term "induced tagging." We explain this notion next.

2. INDUCED TAGGING

The context of our work is a university division named CIRIA (http://ciria.udlap.mx), which comprises three major areas: a department of innovation and information services (IIS), a unit in charge of a multi-disciplinary undergraduate program in knowledge technologies, and the university libraries. The staff of the IIS department includes information experts whose role has been to assist local and remote users in finding the most relevant resources, both physical and digital, to support their learning and research tasks.

After several years of service, our staff has become knowledgeable and capable of locating and recommending resources from vast and dynamic collections in a timely manner. In the process of helping users, staff and users often discover resources that might be useful for supporting current or future tasks. We consider it should be just natural for our team to bookmark these resources and share their findings with their colleagues. Unfortunately, existing tools for social bookmarking are not particularly suited for a digital library setting, and policies have not been in place that require our personnel to participate in collaborative tagging. REC, our proposed environment, addresses these issues.

Although tagging systems in general rely on spontaneous user participation, we decided that at least the service side of REC would include mandatory tagging from our information experts. We use the term *induced tagging* to refer to social bookmarking with two key characteristics: (1) a well-defined group of participants are knowledgeable on the available resources and the background of the user community; and (2) tagging is required as part of their regular responsibilities as a reference team.

Induced tagging is especially well suited for stable learning communities such as those of universities or research centers that use digital libraries intensively and afford a regular staff that supports their information inquiries, such as reference desks and specialized librarians. With appropriate tools, tagging meshes smoothly with the information experts' tasks as they explore resources at hand. Although all users are encouraged to tag, having a specialized group that uses the system continuously and applies labels consistently for extended time periods, addresses concerns on the advantages of controlled vocabularies as well as incentive issues. The latter aspect is also addressed if recommendations are generated for resources that are consistently tagged by multiple users and tags are weighed depending on whether they have been generated by an expert or the general public.

3. RELATED WORK

A survey of tagging systems and a discussion of their potential for knowledge organization and discovery is presented in [3]. Models and taxonomies of tagging systems are offered by [1] and [4]. [5] discusses work involving tag-based recommendations, though not in the context of digital libraries.

4. INDUCED TAGGING IN PRACTICE

REC is the Ajax-based platform we have developed to explore induced tagging in the context of CIRIA. The REC software provides a toolbar that can be added to a web browser so users may label resources in a non-disruptive manner while they navigate around a digital library. Additionally, users may manage tags and request recommendations using the main REC interface, illustrated in Figure 1 (an English interface is being produced at the time of this writing). The current version of REC is available for public evaluation at http://ict.udlap.mx:9090/reduc.

In the figure, user "waldo" has requested recommendations on available resources about "OAI". He also indicates that four-star resources are preferred, since users may rate resources easily using up to five stars at the time of tagging. A list of OAI papers is presented which previously have been rated by CIRIA staff or other users. The number of users who have used the same label for a given resource as well as related labels are displayed. Resources labeled by a particular user may also be selected. Other sections of the interface include helpful information as follows (details not shown for space reasons): "Common topics" displays other labels shared by the resources being recommended; "my documents" lists resources tagged by the user on the subject (OAI in this case); "my subjects" shows a list of all the tags that have been assigned by the user; and "Popular" shows the top resources in terms of global tagging frequency and rating. Both "common topics" and "my topics" are presented as *clouds* in which font size is proportional to the weight of each terms. Term weight is determined as a combination of global term frequency, resource rating, and user category (regular or expert). All sections of the interface are dynamically updated upon selection of any of its elements (e.g., specific users or tags).

We have trained the staff of CIRIA to start labeling resources in our digital collections using REC and have started pilot testing with selected users. Six information experts and 20 regular users

are formally participating in this stage, though the system is already open to the public for evaluation and feedback.
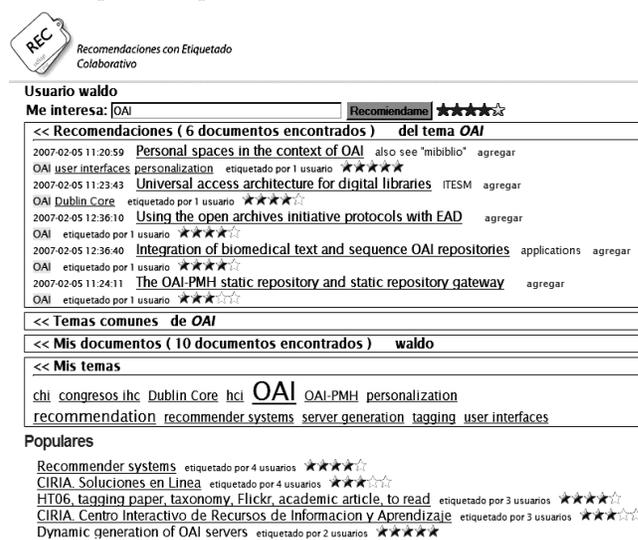

Figure 1. Managing tags and recommendations in REC

Initial observations confirm our premises regarding the potential of induced tagging, as our information professionals have been very enthusiastic about their role as recommenders and authoritative source for resource tagging, whereas end users have started to rely on REC for bookmarking and recommendations. We are in the process of collecting quantitative and qualitative data to further support our approach. Our long-term target user base is a learning community comprising around 8,000 students and some 450 faculty.

5. CLOSING REMARKS

We believe induced tagging has an important role to play in bringing together existing expertise, proven methodologies and social bookmarking in order to assist users in obtaining recommendations and discovering digital library resources. The work on REC shows promising results in that direction.

REFERENCES

[1] Golder, S., Huberman, B. A. The structure of collaborative tagging systems. HP Labs Tech. Report. 2005. (hplabs.com).

[2] Lee, K. What goes around comes around: An analysis of delicious as a social space. *Proc. CSCW'06.* 2006. 191-194.

[3] Macgregor, G., McCulloch, E. Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review, 55, 5* (Feb.), 2006.

[4] Marlow, C., Naaman, M., boyd, d., Davis, M. 2006. HT06, tagging paper, taxonomy, flickr, academic article, to read. *Proc. Hypertext 2006*, ACM Press, NY, 2006, 31-39.

[5] Stoilova, L., Holloway, T., Markines, B. GiveALink: Mining a semantic network of bookmarks for web search and recommendation. *Proc. 3rd Intnl. Workshop on Link Discovery* (LinkKDD '05, Chicago), 2005, 66-73.