

REC: Improving the Utilization of Digital Collections by Using Induced Tagging

J. Alfredo Sánchez, Omar Valdiviezo, Emmanuel Aquino, and Rosa Paredes

Laboratory of Interactive and Cooperative Technologies
Universidad de las Américas Puebla
{j.alfredo.sanchez,omar.valdiviezo,rosa.gpe.paredes}@gmail.com,
aquinoperezemm@hotmail.com

Abstract. REC is a software environment designed to generate recommendations of potentially useful web resources based on tags assigned by a community of users. Recommendations prompt users to access resources that otherwise would go unnoticed, thus promoting improved data exploitation. REC implements the notion of “induced tagging”, a technique devised to improve the effectiveness of social bookmarking for digital libraries by having a group of specialists tag resources as part of their job, in addition to tagging performed by the community. This paper describes the initial experiences with the introduction of induced tagging into an actual learning community, which has motivated an implementation of REC that is embedded in a popular social networking environment. We have gathered initial evidence that indicates that REC mediates an effective collaboration that promotes resource discovery and results in an improved utilization of digital library collections to support academic activities.

Keywords: Tagging, social networks, recommendations, digital libraries.

1 Introduction

Tagging clearly has become a ubiquitous mechanism that supports both personalized and collaborative organization of varied information spaces. Also known as social bookmarking, tagging is present in an ever increasing variety of contexts: photo and video sharing environments such as Flickr and YouTube, citation databases such as CiteULike, many electronic newspapers and popular blog services, to name a few cases, provide handy mechanisms for users to assign keywords, or tags, that function as metadata that describe documents or multimedia objects. When tags are assigned by a sufficiently large number of users, they can be used to effectively support browsing or searching very large information spaces. Also, tag collections can be regarded as determining dynamic classifications of digital objects that depend only on the keywords chosen by all kinds of users, rather than the more conventional controlled vocabularies used by specialists. In that sense, tags are said to originate “folksonomies”.

© A. Gelbukh, M. Adiba. (Eds.)
Advances in Computer Science and Artificial Intelligence.
Research in Computing Science 39, 2008, pp. 83-93

Received 01/07/08
Accepted 01/08/08
Final version 12/09/08

In addition to proprietary tagging mechanisms that are embedded in systems such as those mentioned above, a significant number of services have been developed that help users assign tags to any web resources and provide them with tools to manage their tag collections and to take advantage of tags assigned by all users in a community. The most popular of such services is arguably *del.icio.us*, which only a few months ago boasted three million registered users and about 100 million bookmarked unique URLs¹. Other services that specialize in social bookmarking include *Mister Wong* (*mister-wong.com*) and *Simpy* (*simpy.com*).

The success of social bookmarking has prompted a lot of work aimed to investigate, for example, how tagging occurs, how tag collections are structured, how close folksonomies are to formal classifications generated by specialists, or how tags can be used to enhance information retrieval. Our group has devised a technique termed induced tagging, which is based on the idea that a group of specialized users within a community can be in charge of tagging as part of their jobs. We posit that this will have a positive impact on the size and quality of the tags collection and, as a result, on the usefulness of social bookmarking to support the discovery of digital resources and to improve the exploitation of large information spaces.

We have developed REC, a software environment that implements induced tagging and we report in this paper on our experiences with its introduction into an actual community, the challenges we have faced for its adoption and how these experiences have influenced our redesign of REC.

In what follows, we discuss work that is related to our research. Then, we introduce induced tagging with more detail, as well as its implementation via REC. Our initial experiences with the use of REC are presented next, followed by a discussion of the changes to our initial version of the software. Finally, we present an overview of our ongoing and future work as well as the conclusions that can be drawn from the work we have conducted so far.

2 Related Work

Among the studies of the issues and impact of tagging and its potential for generating useful recommendations or to improve the effectiveness of information retrieval, the following are salient.

An early but still current study of the structure of collaborative tagging was conducted by Golder and Huberman [3]. Among other findings, they reported regularities in user activity, tag frequencies, kinds of tags used and bursts of popularity in bookmarking. Stable patterns for tags assigned to certain URLs are attributed to imitation and shared knowledge. As our work progresses, we plan to use this work as a reference for analyzing the dynamics of our data sets.

The work by Stoilova and colleagues [8] developed a similarity measure for generating recommendations based on the personal bookmark files of users who donate them to help improve search for a web community. Their method takes advantage both of the hierarchical structure of the bookmark files of individual users,

¹ <http://www.techcrunch.com/2007/09/06/exclusive-screen-shots-and-feature-overview-of-delicious-20-preview/>, last accessed on August, 2008.

and of collaborative filtering across users. Though this work is not developed in the context of a tag management system, it provides leads to ways in which recommendations can be improved. In our current stage, the similarity measures we are applying are based only on the number of tags, the assigned ratings and the types of users.

For a survey of tagging systems and a discussion of their potential for knowledge organization and discovery, an interesting source is [4].

What motivates users to actively participate in social bookmarking is among the most important issues that are being studied by researchers. Marlow and colleagues [5] suggest five main motivators for tagging: future retrieval, contributing to the visibility of a resource, attracting attention, expressing opinions, competing with other users and leaving persistent marks. Out of these motivations, the first three (or a combination of them) appear as having particular importance for generating useful tags, possibly to be used as the basis for recommendations. The other two (competing and leaving marks) are features that could be promoted to increase user motivation. A related taxonomy of incentives, only oriented more specifically to image tagging, has been suggested by Ames and Naaman [1], whereas the various roles played by users when tagging has been explored by Thom-Santelli and colleagues [10].

A taxonomy of tagging styles that is particularly relevant for our work was proposed by J. Cañada [2]. After analyzing patterns in del.icio.us and Flickr, he found four such styles: (1) Selfish tagging, when users assign tags that are related only to their personal context and most likely are not meaningful to others; (2) friend-oriented, when tags are familiar only to a closed circle of people; (3) altruistic, when the assigned tags are consciously selected as the most descriptive and generally accepted; and (4) populist, when the assigned tags are enticing but intentionally deceive other users. Evidently, the greatest benefit for the community occurs if users tag altruistically, but it is also true that incentives need to be devised in order for this to happen, as most well-intentioned users tend to tag for themselves or only for a few people. Our approach aims to address precisely this need for incentives, as explained in the following section.

3 Rationale for Induced Tagging

We defined induced tagging [7] as a kind of 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 that group's regular responsibilities as a reference team.

The concept of induced tagging has been proposed to take advantage of the shift that is occurring in the role of information professionals, particularly reference librarians. Personnel at the reference desk generally become knowledgeable and capable of locating and recommending resources from vast and dynamic collections in a timely manner. Increasingly, those resources are part of web-accessible digital collections. In the process of helping users, staff and users often discover resources that might be useful for supporting current or future tasks. It should be possible for information experts to bookmark those resources and share their findings with their

colleagues and the entire community they serve. Since these information experts are also familiar with the needs of the community and the terms most commonly used, altruistic tagging should happen naturally.

In induced tagging, there is a service component that involves mandatory tagging from information experts. Policies would be needed to require information experts in an organization to participate in collaborative tagging. Although all users are encouraged to tag, having a specialized group that does altruistic tagging continuously and applies tags consistently for extended time periods as part of their job, addresses concerns on the advantages of controlled vocabularies as well as incentive issues. Moreover, given the familiarity of the information experts with the needs of their user community, schemes can also be devised to facilitate the generation of personalized recommendations that are based on the tags assigned to relevant resources.

A critique to induced tagging is that it may contravene a popular view that is often used to explain the success of social bookmarking, which is best represented by the so-called wisdom of crowds [9]. According to this view, the collective intelligence that results from aggregating imperfect judgments (tags assigned by regular users) typically outperforms the judgments of experts (tags assigned by information professionals).

In contrast, recent experiments by Razikin and colleagues [6] show that the wisdom of crowds theory is not consistently supported. Their experimental evidence indicates that tags assigned by experts tend to be better descriptors for resource sharing and information retrieval than those assigned by the general public.

Our view is that induced tagging should get the best of both worlds: by overcoming the lack of incentives for altruistic tagging, the process should get a significant boost resulting in an accelerated growth of the collection of tags. Information experts know the vocabulary most commonly used by the user community, which should result in tags that are helpful for information retrieval and discovery, which in turn should attract community members to use existing tags and assign new ones. In the long run, the tags universe will reflect the wisdom of the community.

We believe 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. With appropriate tools, tagging can become part of the regular support offered to users for their information retrieval needs.

We have developed REC, an environment for collaborative tagging that supports induced tagging and generates resource recommendations based on tags. REC is described in the following section.

4 REC

REC is an Ajax-based platform developed to explore induced tagging. REC provides a toolbar that can be added to a web browser so users may label resources in a minimally disruptive manner while they navigate around the web. The dialog box that pops up when the user is browsing the web and selects the "tag" option from the REC toolbar is illustrated in Figure 1. In this case, the user "waldo" is using the tags

“ENC”, “SMCC” and “Computer Science” to describe a website, and is assigning a five-star rating to this site.

[illegible]


Figure 1. Tagging a website in REC.

We are particularly interested in studying how induced tagging works in the context of digital libraries, but REC actually can be used on any web-accessible resources. Additionally, users may manage tags and request recommendations using the main REC interface, which is illustrated in Figure 2 (a complete interface in English is still in progress at the time of this writing).

As illustrated, various tabs in REC allow the user ("waldo", in this case) to access recommendations that result from specific queries, as well as his tagged resources and the tags he has assigned. Additionally, at the bottom of this interface, REC displays a list of the most popular web resources that have been tagged by the user community. In the figure, as the user types a keyword, REC suggests tags that start with the characters being typed. The total number of resources that have been assigned each tag is indicated in the list. The user is also indicating that resources with at least a two-star ranking are preferred. After selecting "consumer behavior" from the list of suggestions, the user is presented with a list of resources that have been tagged and rated by information experts or other users. This is illustrated in Figure 3.

In Figure 3, the four resources displayed (out of the seven that were indicated in the list of suggestions in Figure 2), in this case scholarly papers in subscribed digital libraries, have been tagged as related to "consumer behavior" and rated with two or three stars. The title of each paper is displayed along with other tags that also have been used as descriptors. What is important here is that each of these tags can also be clicked upon, resulting immediately in a new list of resources, possibly with new tags that can lead to other resources, and so on. In Figure 3, if the user clicks on "social

behavior”, which is a tag assigned to the first document in the list, he obtains a new set of three resources, as illustrated in Figure 4. At any moment during the interaction with REC, the user may choose to peruse the suggested documents or to continue to explore the related tags.



Recommendations based on collaborative tagging BETA

User **waldo**

Interested in: [Get recommendations](#) ★ ★ ☆ ☆ ☆

consumer behavior	7
consumo de sustancias	1
consumo dietético	1
consumo energético	1
consumo de drogas	1

<< My resources(13 resources found) **waldo**


<< My tags

[OAI recommendation tagging recommender systems](#) [hci](#) [congresos ihc](#) [riesgo pais](#) [OAI-PMH](#) [server](#)
[generation user interfaces](#) [personalization](#) [Dublin](#)
[Core tags](#) [folksonomies](#) [libros](#) [must-read](#) [SMCC](#) [ENC](#) [computer science](#)

Popular

[Yahoo en español - Mexico](#) etiquetado por 6 usuarios ★ ★ ★ ★ ★
[Google](#) etiquetado por 6 usuarios ★ ★ ★ ★ ★
[CIRIA Centro Interactivo de Recursos de Información y Aprendizaje](#) etiquetado por 7 usuarios ★ ★ ★ ★ ☆

Figure 2. Main interface of REC.



Recommendations based on collaborative tagging BETA

User **waldo**

Interested In: [Get recommendations](#) ★ ★ ☆ ☆ ☆

<< Recommendations(4 resources found)

2007-05-02 13:31:50.0	Norms, Preferences, Identities and retention decisions
	consumer behavior social behavior preferences
2007-05-02 13:28:30.0	Status as a value resource
	consumer behavior status
2007-05-02 12:20:52.0	Social cues in the store environment and their impact on stores' image
	shopping consumer behavior customer satisfaction
2007-05-03 11:08:54.0	Social class influences on purchase evaluation criteria
	gender social class consumer behavior

<< Common tags

[gender](#) [social class](#) [shopping](#) [customer satisfaction](#) [status](#) [social behavior](#) [preferences](#)

<< My resources(13 resources found) **waldo**

<< My tags

[OAI recommendation tagging recommender systems](#) [hci](#) [congresos ihc](#) [riesgo pais](#) [OAI-PMH](#) [server](#)
[generation user interfaces](#) [personalization](#) [Dublin](#)
[Core tags](#) [folksonomies](#) [libros](#) [must-read](#) [SMCC](#) [ENC](#) [computer science](#)

Figure 3. Obtaining recommendations in REC.

Resources tagged by a particular user may also be selected. A convenient list of all related tags in the list of resources is provided as an additional tab, and so is a list of all the resources tagged by the user. Lists of tags 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).

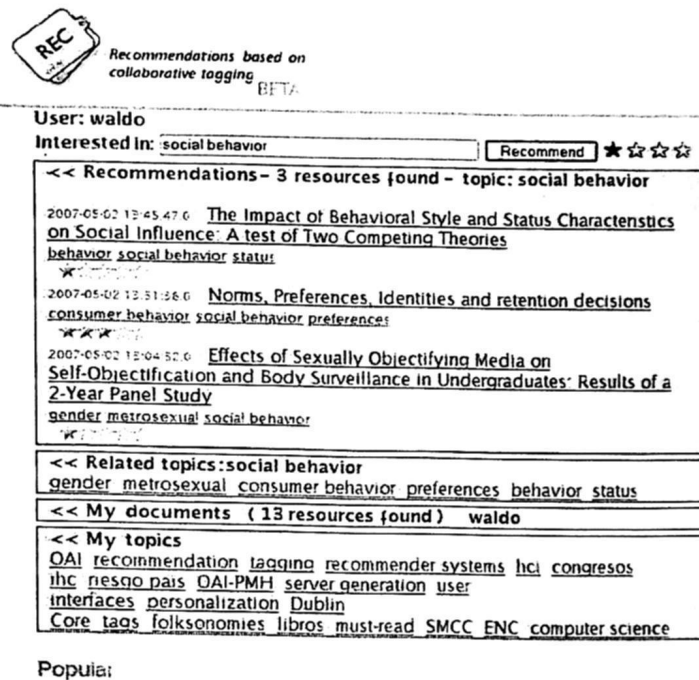


Figure 4. Exploring the tag space in REC.

5 Experiences with the Adoption of REC

REC is open to the public for evaluation and feedback at <http://ict.udlap.mx>. In order to assess its potential, we presented REC to our information experts and provided basic training for them to start labeling resources in the digital collections built or subscribed to by our institution.

Over the period of about six months, six information experts have been tagging resources using REC in addition to their regular duties, which include assisting users at the reference desk and via a virtual reference environment. We also have had students use REC as part of project assignments in a first-year college course called "Information Culture", which is taken by students from all majors and is intended to

provide orientation for them to learn about the information services and resources provided by the university. Although students also have tagged resources, they mostly have been users of existing tags. Usability studies of REC's interface have been conducted by students of courses on human-computer interaction. We also have a number of additional users who have registered to explore the tagging environment.

Collections provided by different vendors typically are available through highly heterogeneous interfaces. Finding related documents on those collections as part of a student or faculty research project often becomes a time-consuming and frustrating task. When tags are assigned to those documents by the information experts (or other users), REC becomes a uniform interface that establishes tacit links that are used to discover resources comprised by diverse collections that otherwise would go unnoticed. Participating users have been able to fetch documents carefully selected by information experts, by just following links recommended by REC, and remaining essentially unaware of the specific collections that contain the documents. Obtaining such documents normally would entail formulating complex queries. Tag clouds have resulted also in a discovery mechanism, as they invite users to explore relevant tags and their associated resources.

Although the size of our tags collection is still modest, our observations have helped shape new features and requirements for our tagging environment to provide support for improved exploitation of vast digital collections. Out of over 11,000 tags that describe more than 4,150 web resources, about 76% have been assigned by information experts. This portion of the tag collection has a high descriptive value and is expressed in the terms most likely used when requesting recommendations, as our staff has been consciously considering the users' needs in the tagging process. This shows that the role of designated personnel is fundamental for describing the collections. Given that each information professional has an area of expertise, the available collections are being covered rapidly, but there is still little overlap among the tags being assigned. Thus, the current average number of tags per resource is only 2.17 (standard deviation of 1.79). We expect these figures to change significantly as REC becomes more popular in our community.

One of the requirements derived from our initial observations was that participants wanted to be able to recommend resources to specific users. They also wanted to be able to define groups of users and suggested interface features that were similar to other familiar environments, such as those in their social networking systems. This suggested the version of REC described next.

6 REC on Facebook

We believe the combination of social networks and induced tagging can provide an even more powerful environment for exploiting information spaces. Not only are our users familiar with various social networking systems, but they also are keen on extending their network of contacts, which could just as well include information experts that could help them accomplish academic endeavors without having to leave their social environment of choice.

We developed a version of REC that can be added as an application of Facebook, one of the most popular social networking systems. Users of Facebook who install REC can obtain or give recommendations within the same familiar environment. More specific registration as a user of REC is needed only to enable tagging and is required only the first time a user tags resources. Both versions of REC share the same databases, so tags and recommendations managed by either of them are equivalent.

Profile edit Friends ▾ Inbox ▾ home account privacy logout

Main Search Tag REC-ommend Idioma Invite About

REC

consumer behavior

Recommend

Recommendations

Norms, Preferences, Identities and retention decisions
consumer behavior, social behavior, preferences

Status as a value resource
consumer behavior, status

Social cues in the store environment and their impact on stores image
shopping, consumer behavior, customer satisfaction

Personal values' influence on e-shopping attitudes and behavior
shopping, consumer behavior, e-shopping, social values

Web advertising: gender differences in beliefs, attitudes and behavior
gender, advertising, consumer behavior

Gender identity and consumption of advertising
gender, advertising, consumer behavior

Social class influences on purchase evaluation criteria
gender, social class, consumer behavior

Page built by REC (report) About Find Friends Advertising Developers Terms Privacy

Figure 5. REC as a Facebook application.

Profile edit Friends ▾ Inbox ▾ home account privacy logout

Main Search Tag REC-ommend Idioma Invite About

Type your friend's names and the URL to recommend

Friends

Rosa Paredes × Emmanuel Aquino ×

URL: http://www.insp.mx/rsp/articulos/articulo.php?id=0001

Recommend or Cancel

Figure 6. Recommending resources to friends.

Figure 5 illustrates the interface of REC as a Facebook application. In this case, the recommendations displayed correspond to those in Figure 3; only they have not been filtered by rating. As illustrated in the figure, other functionality offered by REC is available via tabs that include options for tagging resources, recommending resources to friends, switching to another language (Spanish or English in the current version), and inviting friends to become users of REC. By clicking on the small icon that appears next to each document in the list, users can generate recommendations for their friends. Figure 6 is an example of how recommendations of resources can be sent to Facebook friends.

7 Ongoing and Future Work

At the time of this writing we are deploying REC for its use by our entire learning community, which comprises about 7,000 students and 500 faculty members. We also are sharing the software for installation by member institutions of the Open Network of Digital Libraries (ONeDL, <http://www.onedl.org.mx>) and the Institutional Network for Library Cooperation (also known as the Amigos Group, <http://ciria.udlap.mx/amigos>). Whereas REC is, as mentioned earlier, open to the general public, we particularly are interested in conducting formal experimentation and further observation of the impact of induced tagging in the context of digital libraries and academic communities. We will conduct longitudinal studies aided by surveys, questionnaires and activity logs.

We also are working on functionality enhancements such as improved visualization of the tagging spaces, so users can discover relationships among the resources in the digital collections by relying on graphical representations. This visualization will provide filtering mechanisms so users may include or leave out tags by date, tagger or Internet domains, among other criteria.

Both on the web-based and Facebook versions of REC, we plan to provide options for users other than the “official” information professionals to participate as domain experts. In the realm of social networks, we plan to produce an implementation of REC that takes advantage of OpenSocial², a specification that defines a common API for social applications across multiple websites. OpenSocial is being implemented by other popular social networking service providers, such as Friendster, hi5, LinkedIn, MySpace, and orkut. Our goal is to make REC accessible from as many social networking platforms as possible.

8 Conclusions

We have made progress in exploring the notion of induced tagging, a technique for accelerating the construction of tag collections that may effectively support the exploitation of vast digital collections. Progress includes the implementation and refinement of a social bookmarking environment called REC, which has been used by

² <http://www.opensocial.org>, last accessed on August, 2008.

a group of information experts to assign tags to scholarly web resources. Their tagging style is altruistic as they are familiar with the vocabulary and needs of the target user community and tagging has been included in their regular duties. Our initial experiences with REC motivated the development of a version of this software that can be installed as an application of Facebook, the popular social networking environment. By combining collaborative induced tagging and social bookmarking we are contributing to the development of tools that will support users in at least four areas: (1) exploiting the information resources they have available, (2) organizing their information spaces in a personalized manner; (3) discovering digital resources that otherwise would remain hidden, and (4) participating in the social construction of a tag collection that reflects the wisdom of the community.

Acknowledgments. We would like to thank the group that participated at the April 2008 Seminar Series of the Information Systems and Interactive Systems Group at the Manchester Business School, organized by Víctor M. González, for valuable feedback on the notion of induced tagging.

References

1. Ames, M. and Naaman, M.: Why we tag: motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA, April 28 - May 03), pp. 971-980. ACM, New York, NY (2007)
2. Cañada, J.: Tipologías y estilos en el etiquetado social. Blog entry, available at <http://www.terremoto.net/tipologias-y-estilos-en-el-etiquetado-social/>, (2006). (In Spanish).
3. Golder, S., Huberman, B. A.: The structure of collaborative tagging systems. HP Labs Tech. Report. (2005). (<http://hplabs.com>)
4. Macgregor, G., McCulloch, E.: Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, 55, 5 (2006)
5. Marlow, C., Naaman, M., Boyd, D., and Davis, M.: HT06, tagging paper, taxonomy, Flickr, academic article, to read. In: *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia* (Odense, Denmark), pp. 31-40. HYPERTEXT '06. ACM, New York, NY (2006)
6. Razikin, K., Chua, A. Y. K., Lee, C. S., Goh, D. H.-L.: Can social tags help you find what you want? In: *Proceedings of the European Conference on Digital Libraries*, (ECDL 2008, Aarhus, Denmark). *In print* (2008)
7. Sánchez, J. A., Arzamendi-Pérez, A., Valdiviezo, O.: Induced tagging: Promoting resource discovery and recommendation in digital libraries. In: *Proceedings of the Joint Conference on Digital Libraries* (JCDL 2007, Vancouver), pp. 396-397 (2007)
8. Stoilova, L., Holloway, T., Markines, B., Maguitman, A. G., and Menczer, F.: GiveALink: mining a semantic network of bookmarks for web search and recommendation. In *Proceedings of the 3rd international Workshop on Link Discovery* (Chicago, Illinois, August 21 - 25, 2005). LinkKDD '05, pp. 66-73. ACM, New York, NY (2005).
9. Surowiecki, J.: *The Wisdom of Crowds*, Doubleday, New York (2004)
10. Thom-Santelli, J., Muller, M. J., and Millen, D. R.: Social tagging roles: publishers, evangelists, leaders. In *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy, April 05-10), pp. 1041-1044, ACM, New York, NY (2008)