Journal of Universal Computer Science, vol. 15, no. 3 (2009), 678-691
submitted: 17/11/08, accepted: 15/1/09, appeared: 1/2/09 © *J.UCS*

Providing Multi Source Tag Recommendations in a Social Resource Sharing Platform

Martin Memmel, Michael Kockler and Rafael Schirru
(Knowledge Management Department
DFKI GmbH, Kaiserslautern, Germany
{memmel, schirru}@dfki.de
kockler@googlemail.com)

Abstract: In today's information environments, tagging is widely used to provide information about arbitrary types of digital resources. This information is usually created by end users with different motivations and for different kinds of purposes. When aiming to support users in the tagging process, these differences play an important role. In this paper several approaches to generate tag recommendations are discussed, and a prototypical recommender system for the social resource sharing platform ALOE is presented. This interactive system allows users to control the generation of the recommendations by selecting the sources to be used as well as their impact. The component was introduced at DFKI, and a first evaluation showed that the recommender component was considered as helpful by a majority of users.

Key Words: classification, collaborative tagging, digital resources, knowledge management, knowledge sharing, metadata, recommender, tagging, web 2.0
Category: H.3.0, H.3.2, H.3.3, H.3.4, H.3.5, H.5.0, H.5.1, H.5.2

1 Introduction

With the advent of the Web 2.0 various social resource sharing platforms arose which allow their users to easily organize and share digital resources. Well-known examples of such platforms are Delicious[1] for bookmarks and Flickr[2] focusing on photos. In order to organize content for future search, navigation and filtering the users of such platforms can usually assign tags (i.e., freely chosen keywords) to the resources in the system. In contrast to systems where information about resources is only provided by a small set of experts, collaborative tagging systems take into account that the way individuals conceive the information contained in a resource differs a lot. It depends on a variety of factors such as their knowledge, experience and the current task. As Maron states in [Maron 1965][p.9]:

> 'information is not a stuff contained in books as marbles might be contained in a bag – even though we sometimes speak of it in that way. It is, rather a relationship. The impact of a given message on an individual is relative to what he already knows, and of course, the same message could

[1] http://delicious.com/
[2] http://www.flickr.com/

convey different amounts of information to different receivers, depending on each one's internal model or map.'

A folksonomy ([VanderWal 2007]), i.e., the result of the collborative tagging efforts, can reflect this diversity. As pointed out in [Golder and Huberman 2005], collaborative tagging is most useful when there is nobody in the "librarian" role or if there is just too much content for a single authority to classify. For the web, where collaborative tagging has grown popular in the recent years, both cases apply. Providing tag recommendations in such scenarios not only has the potential to supporting users in the tagging process, it can also help to reduce undesired noise in the folksonomy.

In this paper, we present an interactive approach to generate tag recommendations using different kinds of sources. The approach allows users to influence the tag recommendation process by offering them to select and combine different sources and services. Thus, it provides the possibility to adapt the recommendation outcome to the users' current needs.

The paper is structured as follows: Section 2 describes available sources for tag recommendations. In Section 3, we describe the functions of tags, motivations for users to annotate tags as well as the resulting implications for tag recommenders. In Section 4 we describe criteria to measure the quality of tag recommendations. A prototypical implementation of a multi source tag recommender for the ALOE system as well as a short evaluation will be depicted in Section 5. We then summarize our results and give an outlook on future work.

2 Available sources for tag recommendations

As shown in Figure 1, one can distinguish four main sources that can be exploited for the generation of tag recommendations: the tagging *user*, the *system* providing the tagging facility, the *resource* to be tagged and *background knowledge*. In the following, each of them as well as further sources will be examined briefly.

2.1 Main sources

User: First, information from and about the user who wants to tag a resource can be taken into account. This includes the tags already used, tags used by the user's contacts, information that can be drawn from the user's profile (e.g., containing information about interests), and information about the current context of the user (e.g., gathered with user observation components).

Tagging System: Secondly, the system in which the recommendations shall be provided can play an important role. In case such a system uses certain tagging conventions (e.g., multi-word terms can be used as tags using quotation

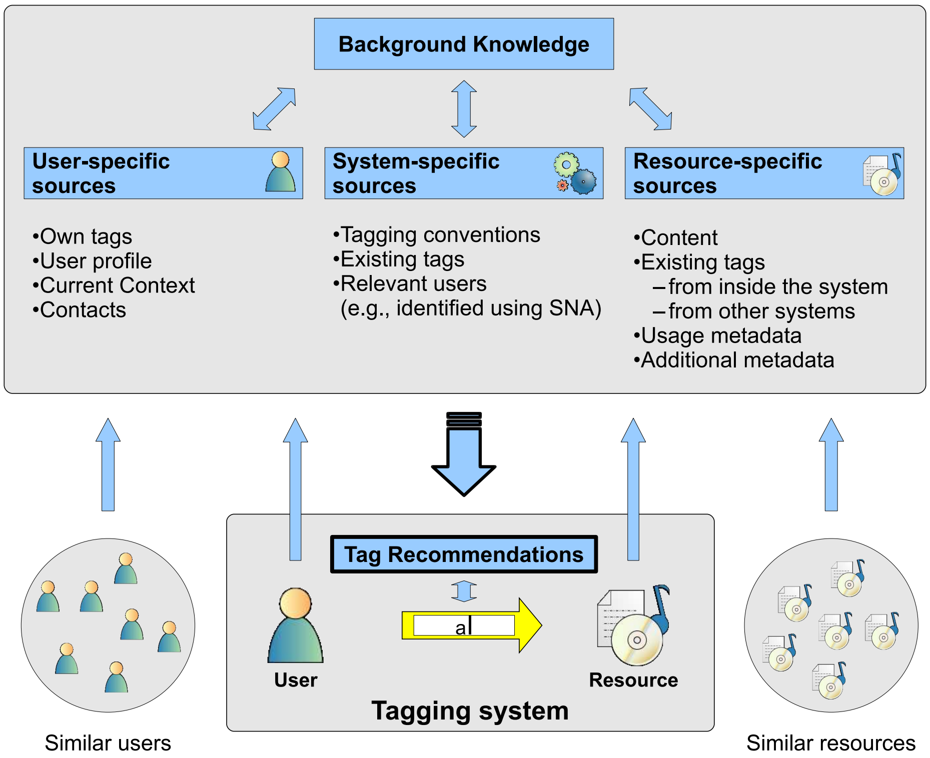

Figure 1: Available sources for tag recommendations (adapted from [Kockler 2008])

mechanisms), this has to be taken into account. Furthermore, information about popular tags can be used as a source, and social network analysis can provide additional information about the users of the system (e.g., identifying them as hubs or authorities) that can be used to estimate the quality of existing contributions.

Resource: Thirdly, the resource itself can be examined. Using techniques from the fields of Information Extraction and Natural Language Processing, information about the content of the resource (e.g., concepts such as topics, people or organizations) can be derived. Furthermore, existing tags and additional metadata about the resource (e.g., containing bibliographic information or information about the context of use) can be taken into account.

Background knowledge: Finally, a tag recommender can of course make use of any kind of available background knowledge to achieve a better understanding and representation of the information at hand. This especially concerns structures such as dictionaries, gazetteers and thesauri. They can provide important information about the meaning of terms, about similarity and other relations between terms, about categories a term might belong to, or they can be used to extract entities from a given input (e.g., using tools such as OpenCalais[3]).

2.2 Further sources

Apart from the sources mentioned above, we can also take into account information about users that are similar to the tagging user, resources that are similar to the resource to be tagged, and inital tag input provided by the tagging user.

Information about users that are similar to the active user have to be considered when applying collaborative filtering approaches for recommendations ([Konstan et al. 1997]). In [Marinho and Schmidt-Thieme 2008] user-based collaborative filtering is applied for tag recommendations. The method yielded good results when applied on the user-tag matrix. It could be shown that users with a similar tag vocabulary tend to tag alike.

Resources that are similar to the resource to be tagged can also be analyzed to provide tag recommendation approaches. E.g., the AutoTag system suggests tags for weblog posts by using the tags that are associated with posts considered as similar ([Mishne 2006]).

Last but not least, if the tagging user already provided an initial tag input, this information can be used to recommend only tags with the same prefix. If already one or more complete tags have been entered, further tags can be recommended, e.g., by using a co-coccurence matrix that provides information about how often tags co-occured in a system.

3 Why are people using tags?

Tagging is used in a variety of scenarios, and people tag for very different reasons. To understand how tag recommendations can help users in the tagging process, an understanding is required about what kinds of tags exist and what motivates people to tag.

3.1 Functions of tags

Tags can convey information about potentially any facet of a resource. This concerns information about the content and creation of a resource, about the way it should be or was used, etc. In [Golder and Huberman 2005], the following kinds of tags for resources are identified:

[3] http://www.opencalais.com/

- *identifying what (or who) it is about*,
- *identifying what it is*,
- *identifying who owns it*,
- *identifying qualities or characteristics*,
- *self reference*, and
- *task organizing*.

Each of these purposes requires different kinds of tag recommendations, generated by using different kinds of sources. E.g., the content of a resource can be used to identify entities in a resource and thus to suggest keywords that describe what or who it is about, whereas it makes almost no sense to use such information for task organizing or to identify qualities or characteristics.

3.2 What motivates people to tag?

People use tags for different purposes. But not only the purposes of tags, but also the motivation of users to tag resources has to be considered. Marlow et al. identified the following criteria ([Marlow et al. 2006]):

- *future retrieval*,
- *contribution and sharing*,
- *attract attention*,
- *play and competition*,
- *self presentation*, and
- *opinion expression*.

It is obvious that different kinds of motivations require different kinds of tag recommendations. Whereas tag recommendations derived from the analysis of the resource content will be useful in most cases, they won't be helpful when users want to express their opinions. The way existing tags should be taken into account also depends on the users' motivations. E.g., support for users that want to attract attention should take into account existing, popular tags from the whole system, whereas mainly the tagger's existing tags should be taken into account to support future retrieval or self presentation.

3.3 Implications for tag recommenders

We have seen that the way tag recommendations should be generated strongly depends on what kind of tag shall be used and for which reason. Thus, a one-size-fits-all solution is not adequate. It should be possible to adapt a recommender system to a variety of usage scenarios. As it is rather unlikely that the recommender knows in advance about what characterizes the current scenario, the user should have the possibility *to decide which sources to use for recommendations*, *to determine the impact the different sources should have*, and *to decide whether the recommended tags should be restricted to a certain vocabulary*, e.g., the user's own tags or the tags already existing in a system.

As a recommender component with that many features might become very complex and requires some knowledge and expertise about how recommendations are generated, it might not necessarily be useful to provide it to arbitrary end users in any kind of tagging system. Yet, it might be very useful to provide it to a restricted group of users, or to a community manager or admin to specify values that best correspond to the characteristics of the community and content in the according system. These settings can then be used for a reduced version of the tag recommender with less complex interaction possibilities that can be made available for all end users.

4 Quality of tag recommendations

As argued in Section 3, the reason tags are used strongly depends on a user's motivation and the context of use. Furthermore, depending on the focus of a system some of the motivations to tag resources (e.g., *attract attention*) might be considered as undesirable from a system point of view.

Thus, the quality of tag recommendations also depends on these factors. Whereas a recommendation might be very useful to describe the topic of a resource, it might be at the same time almost useless when users want to express their relations to the resource. In this case, existing tags already used in such a context would be helpful. Nevertheless, we would always consider a tag recommendation as bad when it does not provide valuable information for any kind of usage scenario, e.g., if a tag was recommended that is not related to the resource.

When existing tags will be recommended, one also has to consider what is sometimes denoted as the *echo chamber effect* ([Jamieson and Cappella 2008]). This effect turns up when popular tags from the head of the long tail of tags (see [Anderson 2006, Barnett 2006]) will be recommended very often, thus causing that the respective tags become even more popular. Sen et al. examined community influence in tagging systems and stated that their results suggest that 'users would tend to follow the pre-seeded tag distribution' ([Sen et al. 2006]).

On the one hand, this can be considered as positive, as the usage of the same tags for resources fosters that more potentially relevant resources can be found when searching for the respective tag (positive impact on recall). On the other hand, the usage of a diversity of tags allows for more subtle distinctions (positive impact on precision). Marlow et al. therefore distinguish *blind tagging* where a user does not see the tags which have been assigned to the same resource by other users, *viewable tagging* where users can see these tags, and *suggestive tagging*, where the system suggests possible tags to the user ([Marlow et al. 2006]). Finding a balance between independence and guidance can be considered as one of the most important tasks when providing tag recommendations.

5 Multi source tag recommendations in ALOE

Before presenting the prototypical realization of the multi source tag recommender, we will first introduce the ALOE system in which it was embedded. We will conclude this chapter with the results of a first evaluation.

5.1 The ALOE platform

ALOE[4] is a social media sharing platform developed at the Knowledge Management Department of DFKI[5]. ALOE offers possibilities to share and organize digital resources and arbitrary information about them. Users can use ALOE either as a repository (i.e., uploading and kind of multimedia files) or as a referatory (i.e., just refering to a URL). ALOE provides a rich user interface (see Figure 2 and Figure 3) with a wide range of social media functionalities. Among others, the following features are offered:

- upload and sharing of arbitrary types of digital resources,
- sharing and organizing of bookmarks,
- tagging, rating, and commenting on resources and bookmarks,
- initiating groups and communicating with other users,
- publishing as private, public, or only for certain groups,
- finding resources with different types of search filters,
- ranking search results according to different criteria,
- associating arbitrary metadata sets with resources.

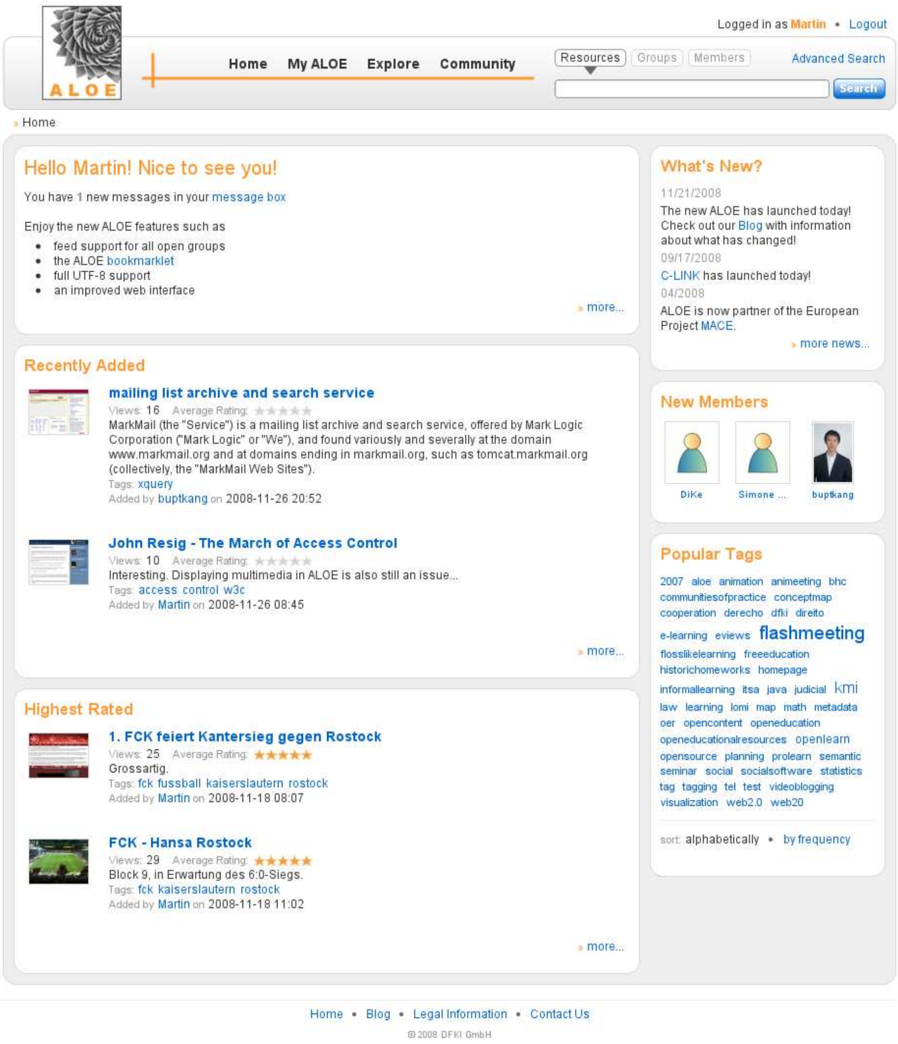

Figure 2: Screenshot of the ALOE start page

ALOE offers access to data and the complete range of functionalities via a Web service API (SOAP) and thus allows to introduce social media paradigms in existing (heterogeneous) infrastructures. The system is used in several projects,

[4] see `http://aloe-project.de/`

[5] funding for the development within the project CoMet (see `http://www.dfki.uni-kl.de/comet/`) was provided by the "Stiftung Rheinland-Pfalz für Innovation"

e.g., the European project MACE[6] (Metadata for Architectural Contents in Europe) and C-LINK[7] (a web based tool to support conference attendees).

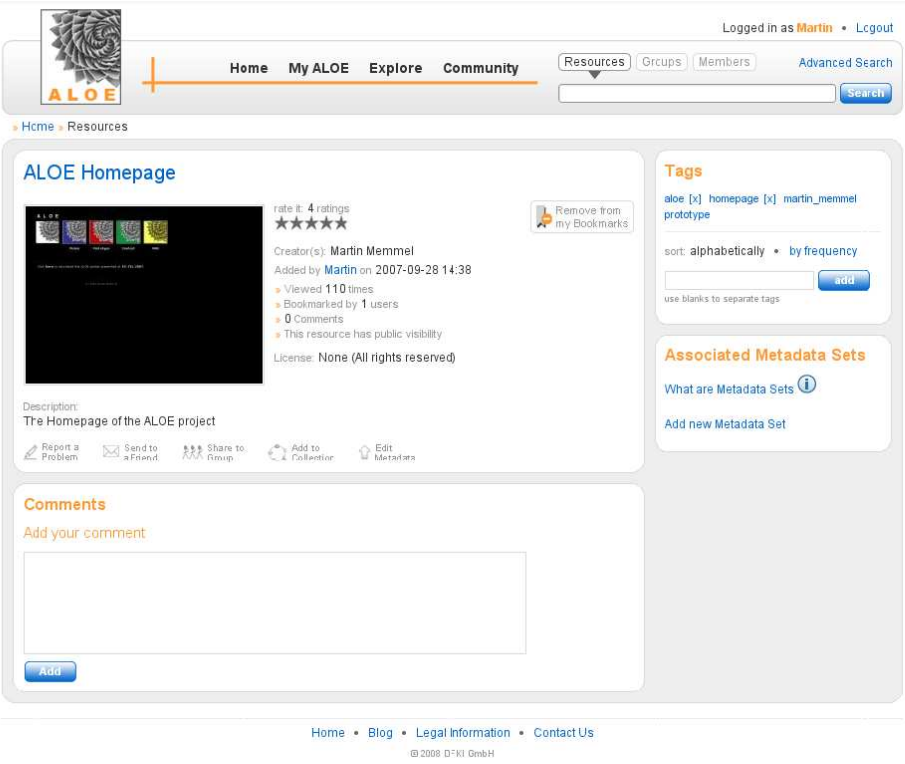

Figure 3: Screenshots of a detail view page for a resource in ALOE

See [Memmel and Schirru 2007] and [Memmel and Schirru 2008] for more details about functionalities, use cases and the system architecture.

5.2 The multi source tag recommender

The interface of the prototypical implementation of the multi source tag recommender is shown in Figure 4. First, users have to enter the URL of the resource they want to tag. Using the Recommender Mixing Desk, users can choose the source from which the recommendations shall be derived. Available sources are

[6] http://www.mace-project.eu
[7] http://c-link.dfki.uni-kl.de

the content of the resource, *metadata about the resource* that has been published in ALOE (this concerns the usual ALOE metadata as well as external metadata sets such as FlashMeeting-XML that is associated with public FlashMeeting replays[8] published in ALOE), *the usage context of the resource* that has been gathered with selected user observation components developed at DFKI (see [Kiesel et al. 2008] for more details), and *existing tags* for the resource from different sources.

By using the sliders, users can decide which impact each source will have in the tag recommendation process. In the box below, the information extraction services which shall be used to extract the entities that appear in the considered resource can be selected. Services that can be used here are

- Semager[9],
- Yahoo![10],
- TagTheNet[11] and
- ALOA[12].

Stemming can optionally be applied to the determined keywords. If existing tags shall be used, users can select tags of the resource in ALOE, tags that the users' contacts have used, and the tags which have been assigned to the resource in Delicious. The third box allows to specify for each selected source whether related tags shall be taken into account to generate the tag recommendations. If this is the case then also the services which have to be used to determine related tags have to be specified. Available services are:

- Semager,
- Yahoo!,
- WordNet[13],
- MobyThesaurus[14] and
- Watson[15].

[8] see http://flashmeeting.open.ac.uk/ for more details about the FlashMeeting project
[9] http://www.semager.de/api/textcloud
[10] http://developer.yahoo.com/search/content/V1/termExtraction.html
[11] http://www.tagthe.net/
[12] http://eiche.informatik.rwth-aachen.de:3333/ALOAInterface/index.jsp
[13] http://wordnet.princeton.edu/
[14] http://www.mobysaurus.com/
[15] http://watson.kmi.open.ac.uk/WS_and_API.html

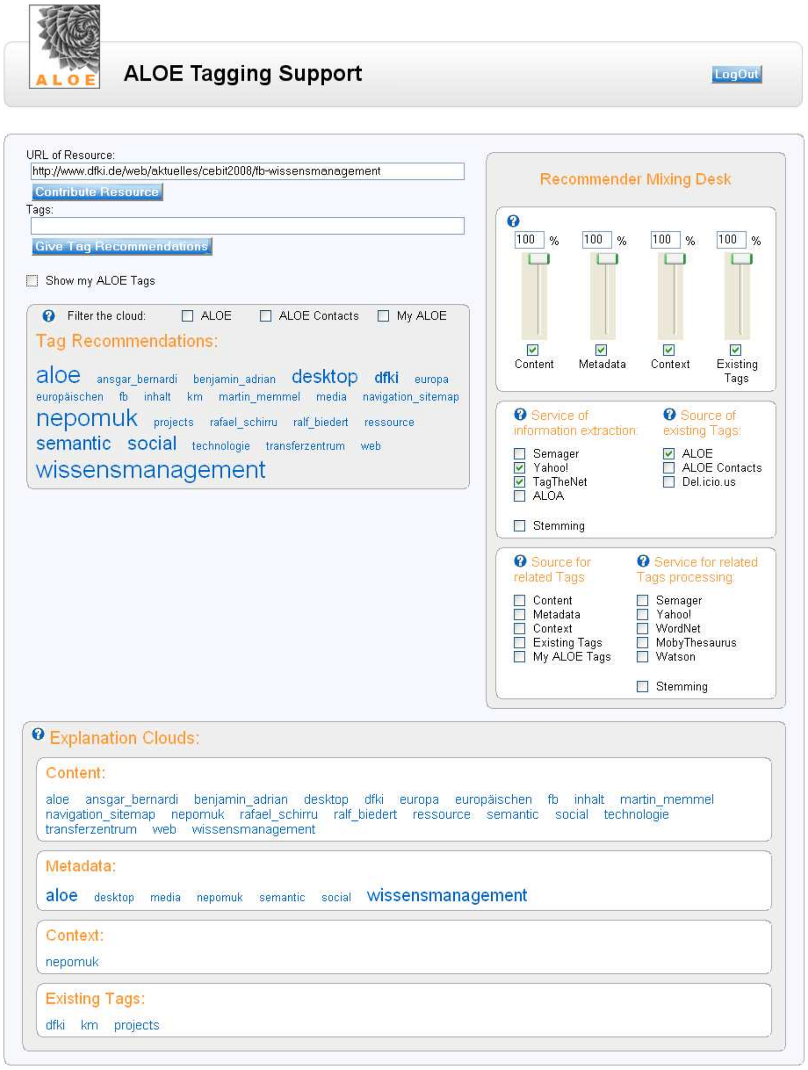

Figure 4: A screenshot of the multi source tag recommender interface

After choosing sources and services, the generation of a recommendation cloud can be initiated. The size of the tags in the cloud corresponds to the relevance that has been determined for the respective tag. Tags in the recommendation cloud can be restricted to tags from the users themselves, their contacts, or tags in ALOE by applying different filters. At the bottom, the results for each different source are shown to provide information about the impact they had.

5.3 A first evaluation and results

In a first evaluation at DFKI, 15 users were asked to try out the recommender with three specified resources as well as with arbitrary resources they could choose freely. They then had to fill out a questionnaire to provide feedback. 10 users were already familiar with tagging, mainly using it for "future retrieval" and "contribution and sharing". Figure 5 shows that the vast majority of users

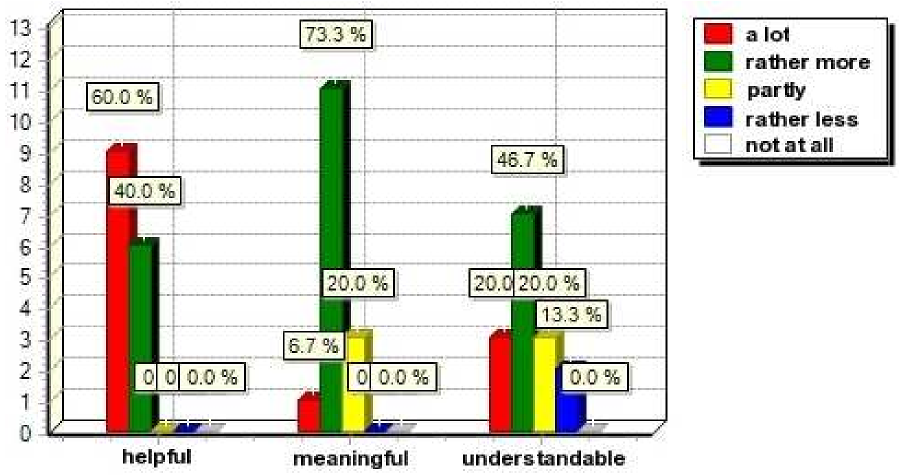

Figure 5: Evaluation results for the quality of recommended tags

considered the recommendations as meaningful and helpful. Furthermore, the evaluation showed that the selection of sources as well as information extraction services has a huge impact on the user's satisfaction with the recommended tags.

6 Summary and future work

Supporting users in the tag generation process requires an understanding of the nature of tags as well as the users' motivations to tag a resource. Thus, a respective recommender component should be adaptable to different usage scenarios. We therefore suggested a tag recommender that allows users to select the sources as well as the impact of these sources for the recommendation process. A tag recommendation protoype was developed as part of the ALOE system, and a first evaluation showed that the recommendations were considered as meaningful and helpful by a majority of users. In the future, we plan to use the full version with all interaction possibilities mainly as a tool for experts within ALOE. Meaningful settings that correspond to the charateristics of users, resources, tags, and additional metadata can thus be determined and used for a reduced version of the tag recommender with less complex interaction possibilities that will be made available for all end users of the system. This reduced version will also be offered in a bookmarklet that will allow users to tag web sites in a comfortable way without the need to enter a URL manually. Furthermore, stopword lists and lemmatizing will be applied to filter out unwanted recommendations.

Up to now the impact of tag recommendations on the folksonomy of a tagging system is not fully understood. On the one hand tag recommendations support the users in the tagging process by reducing the cognitive costs for the annotation of tags, but on the other hand also negative effects cannot be excluded. Research so far indicates that users will then predominatly annotate recommended tags that might not fully reflect their individual view on a resource. This can reduce positive features of tagging systems such as the provision of resource descriptions that allow users to 'tap into the long tail' and to find the niches that are relevant for them. Further research will be required in this area.

References

[Anderson 2006] Anderson, C.: "The Long Tail", Random House Business Books, 2006.

[Barnett 2006] Barnett, A.: "The Long Tail of Tags"; Electronic document. Date of publication: September 16, 2006. Retrieved April 27, 2008, from http://alexbarnett.net/blog/archive/2006/09/16/The-Long-Tail-of-Tags.aspx.

[Golder and Huberman 2005] Golder, S.; Huberman, B.A.: "The Structure of Collaborative Tagging Systems"; HP Labs, 2005.

[Jamieson and Cappella 2008] Jamieson, K. H.; Cappella, J. N.: "Echo Chamber: Rush Limbaugh and the Conservative Media Establishment"; Oxford Univ. Press, 2008.

[Kiesel et al. 2008] Kiesel, M.; Schwarz, S.; van Elst, L.; Buscher, G.: "Mymory: Enhancing a Semantic Wiki with Context Annotations"; Proceedings of the 5th European Semantic Web Conference (ESWC), 2008, pages 817–821.

[Kockler 2008] Kockler, M.: "Conception & Implementation of Tagging Support Mechanisms in a Web 2.0 Social Media Sharing Platform"; Diploma Thesis, Technische Universität Kaiserslautern, DFKI, 2008.

[Konstan et al. 1997] Konstan, J. A.; Miller, B. N.; Maltz, D.; Herlocker, J. L.; Gordon, L. R.; Riedl, J.: "GroupLens: applying collaborative filtering to Usenet news"; Commun. ACM, ACM Press, 1997, 40, pages 77–87.

[Marinho and Schmidt-Thieme 2008] Marinho, L. B.; Schmidt-Thieme, L.: "Collaborative Tag Recommendations"; Data Analysis, Machine Learning and Applications, 2008, pages 533–540.

[Marlow et al. 2006] Marlow, C.; Naaman, M.; Boyd, D.; Davis, M.: "HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, ToRead."; Proceedings of Hypertext 2006; New York: ACM Press, 2006.

[Maron 1965] Maron, M. E.: "Mechanized documentation: The logic behind a probabilistic interpretation"; in M.E. Stevens et al., editors, Statistical Association Methods for Mechanized Documentation Symposium Proceedings 1964, volume 269, National Bureau of Standards, Washington, 1965, pages 9–13.

[Mishne 2006] Mishne, G.: "AutoTag: A Collaborative Approach to Automated Tag Assignment for Weblog Posts"; WWW '06: Proceedings of the 15th International Conference on World Wide Web, ACM Press, 2006, pages 953–954.

[Memmel and Schirru 2007] Memmel, M.; Schirru, R.: "Sharing Digital Resources and Metadata for Open and Flexible Knowledge Management Systems"; in Klaus Tochtermann and Herrmann Maurer, editors, Proceedings of I-KNOW 2007, Journal of Universal Computer Science, Graz, Austria, 2007, pages 41–48.

[Memmel and Schirru 2008] Memmel, M.; Schirru, R.: "ALOE White Paper"; Technical report, DFKI GmbH, 2008.

[Sen et al. 2006] Sen, S.; Lam, S. K.; Rashid, A. M.; Cosley, D.; Frankowski, D.; Osterhouse, J.; Harper, F. M.; Riedl, J.: "tagging, communities, vocabulary, evolution"; CSCW '06: Proceedings of the 20th anniversary conference on Computer supported cooperative work, ACM, 2006, pages 181–190.

[VanderWal 2007] Vander Wal, T.: "Folksonomy Coinage and Definition"; Electronic document. Date of publication: February 02, 2007. Retrieved October 15, 2008, from http://vanderwal.net/folksonomy.html.