Doreen Böhnstedt, Philipp Scholl, Christoph Rensing, Ralf Steinmetz: *Collaborative Semantic Tagging of Web Resources on the Basis of Individual Knowledge Networks.* In: Houben, G.-J.; McCalla, G.; Pianesi, F.; Zancanaro, M.: Proceedings of First and Seventeenth International Conference on User Modeling, Adaptation, and Personalization UMAP 2009, vol. Lecture Notes in Computer Science , no. Vol. 5535, p. 379-384, Springer-Verlag Berlin Heidelberg 2009, June 2009. ISBN 978-3-642-02246-3.

Collaborative Semantic Tagging of Web Resources on the Basis of Individual Knowledge Networks

Doreen Böhnstedt[1], Philipp Scholl[1], Christoph Rensing[1], Ralf Steinmetz[1],

[1] Multimedia Communications Lab, Technische Universität Darmstadt,
Merckstraße 25, 64283 Darmstadt, Germany
{Doreen.Boehnstedt, Philipp.Scholl, Christoph.Rensing, Ralf.Steinmetz}@
KOM.TU-Darmstadt.de

Abstract. The web is increasingly used as an information source to gain new knowledge but the management of found web pages can be a challenging task. Often social tagging systems are used for resource management. Besides the obvious use of tags – organizing a collection of web resources – they support functionalities like sharing resources with other users and recommendation of further possibly relevant web pages. This paper describes a novel application based on an extended tagging concept that can improve resource management and recommendation. Adding semantic information to tags and tagging fragments of web pages instead of whole web pages enhance the possibilities of well-known tagging applications. Individual knowledge networks are the basis of this tagging concept. A first prototype is developed as proof of concept.

Keywords:

1 Introduction

In this time of fast changing circumstances and new challenges in job and life it is often necessary to learn continuously. It is not feasible anymore to learn for the whole life in advance or attending many advanced trainings. Permanently changing tasks require learning when it is needed. This form of “learning on demand” is a kind of learning where the learning process is self-directed and resources are searched autonomously. Increasingly, the web is the source for new information but the web pages are seldom worked up for learning purposes. There is no teacher who structures the learning process and provides relevant resources. Thus a big challenge is the resource management.

Collaborative tagging systems are one possibility for this sort of knowledge organization [1]. We developed an extended semantic tagging concept which is explained and compared to popular tagging applications in the next section. Our extension of the well-known tagging concept is based on individual knowledge networks as described in section 3 and has some benefits that are described in section 4 e.g. with regard to reflection of the learning process, collaboration, filtering and recommendation. Furthermore we developed a prototype with basic functionalities which is shown in section 5. This paper concludes with a summary and further steps.

2 Concept of Extended Semantic Tagging

This section shows that collaborative tagging systems are useful to support resource management. Based on evaluating existing applications we developed a concept that enhances the well-known tagging concept towards an extended semantic tagging where not only whole web pages can be tagged but also fragments of web resources. These tags can be specified through a type and linked with other tags.

2.1 Related Systems

In collaborative tagging systems users can label resources with tags that make sense to them to sort the web pages into their own personal resource organization. By aggregating the tags of all users, folksonomies emerge that everyone can profit from. Either the tags are predefined in a controlled vocabulary or they are arbitrary. [2] provides an overview of the collaborative tagging phenomenon/applications and discusses potentials and problems of uncontrolled vocabularies. Unlike organizing web pages into browser bookmark folders or saving in folders on hard disks where users have to decide which folder to choose, in tagging systems they can add more than one keyword. For example, a web page about a comparison between the programming languages Java and Python, could be tagged with the keywords “java”, “python”, “comparison”, “interesting” and can later be easily found again using each one of these tags, whereas in a hierarchical organization the user would have to decide which folder to use.

In the following, four free collaborative tagging applications are described representatively in regard to tagging features: Delicious[1], Faviki[2], Zigtag[3] and CiteULike[4]. They allow users to tag, save, manage and share web pages. CiteULike is specialized in managing and discovering scholarly references. In tag clouds, tags are visualized differently, e.g. in varying sizes based on frequency of use. They serve to browse the own collection or the resources and tags of other users. To be up to date, tags can be subscribed to via RSS feeds or so-called watch lists. If users save a resource, most systems recommend tags, e.g. on the basis of a combination of tags of the user and the folksonomy. In tagging applications it is often possible to build up networks or groups with other users, in order to share bookmarks and to follow group tagging activities. While saving web pages, the Delicious plug-in fills in the title of the web page and text snippets in the page if something is marked. CiteULike offers BibTeX import services for selected publishers while saving articles and users can indicate the reading priority of articles. Faviki und Zigtag use semantic tags. In Faviki, tags must correspond to Wikipedia concepts i.e. tags refer to Wikipedia[5] pages. When tagging a page in Zigtag, users can define a meaning for their tags so that the tags are more than just simple keywords. Semantic tagging can solve the problems of synonyms and homonyms. Based on these tagging features we developed a concept of extended semantic tagging.

[1] Delicious, http://www.del.icio.us, Online 2009-01-17
[2] Faviki, http://www.faviki.com, Online 2009-03-15
[3] Zigtag, http://www.zigtag.com, Online 2009-03-15
[4] CiteULike, http://www.citeulike.org, Online 2009-01-17
[5] Wikipedia, http://en.wikipedia.org/wiki/Main_Page, Online 2009-03-15

2.2 Extended Tagging of Web Page Fragments

Tagging is described in [3] as “a way of making sense of many discrete, varied items according to their meaning”. Tags which are added to a resource have a semantic meaning for the user who saves the resource – sometimes tags make only sense for the individual depending on the personal situation. Users have diverse motivations to tag. Thus tags can have different functions [4], e.g. expressing opinions like “interesting” or “relevant”. Moreover they can be used for identifying persons and topics like “java” or “python”. They can be used for goal management as well, for example “writing a paper for UMAP09”. These different motivations and functions show that users have a particular concept in mind while tagging, like shown above it can be e.g. a topic. We propose to let users add this concept type while tagging. Therefore we examined possible tag types which are relevant for personal resource management and introduced an extendable type set that contains topic, person, goal, event, location and miscellaneous. In section 5 we show with our prototype that users can add the type easily with little effort.

Furthermore, users should be able to connect tags with tags in order to express relatedness between them. For example, a user discussed an article of a web page with an expert whom he met at a conference. Then the user could tag the article with this person and the person with the name of the conference and he could label the relation between the tags with e.g. “met at”. Probably the user will remember the person or the conference later instead of the article title. Because a web page can contain different articles with dissimilar topics, the user is interested in a particular fragment of the page. Thus we propose to link tags to the fragment instead of the URL in contrast to known tagging systems, where tags are bound only to URLs. Moreover, users should not be forced to comply with a predefined typology or structure, but should be able to add information to tags or resources that are considered relevant by them and to decide themselves which and how much information, e.g. to add a deadline to a goal.

3 Individual Knowledge Networks as Basis

Knowledge networks are a good technical basis to realize the concept described in the last section. A knowledge network – also called semantic network – is defined in [5] as “a graphical notation for representing knowledge in patterns of interconnected nodes and arcs”. We use the concept *knowledge networks* in this paper to avoid association with formal ontologies. As demanded, each user can configure his knowledge network individually. There are no experts upfront who model the semantic networks. Each user builds up his individual semantic network gradually through tagging resources and adding information. Web page fragments and tags are nodes in the knowledge network and the action of tagging connects the nodes. Relations between tags can have properties as well, e.g. the relation between two goals can have the label “is sub-goal of”. Besides managing nodes and relations, knowledge networks can be searched semantically; this means that the knowledge network can use the relations between nodes to answer the search query (see figure 1 right for a more detailed example). The whole concept of our knowledge networks and their usage in learning scenarios is described in [6].

4 Benefits

The extended tagging concept based on individual knowledge networks has some benefits e.g. regarding to reflection, collaboration, filtering and recommendation.

Reflection. Tagging with goals can serve as a simple task management. Users can structure their search and learning process if they specify tags as goals. Users can observe their search progress based on the web resources they have saved and identify knowledge gaps. Thus, a visualization of resources connected to goals can support the reflection of the self-directed learning process.

Collaboration. Users can integrate subsets of other knowledge networks in their own network, so they don't have to switch between the view of the own collection and the whole network – as it is often the case in tagging applications. The integration could be supported with the aid of (semi-)automatic merging algorithms. For example, if a user merges a topic tag with the topic tag of someone else he can be informed if the other user adds new resources. This subscription mechanism is comparable to RSS feeds or watch lists.

Filtering and Recommendation. Often tags fit only to particular contents and recommendation algorithms can provide more precise results if the algorithms work on fragments instead of comparing URLs or whole pages. Folksonomies are increasingly used to suggest further relevant web resources. As described in section 2 tags can have different functions. Recommending could be refined if tag types (like the topic type) are included in the algorithms. Similarity matching algorithms on schema and structure level (like graph matching algorithms) could profit from this additional information. Tag types which make only sense to single individuals like "interesting" could be left out. Semantic searches enhance the possibilities of full-text searches because it can search along relations, e.g. a resource can be found with a location tag – even if the resource doesn't contain the location. While navigating through the knowledge networks it can be helpful to filter particular tag types because filtering reduces the amount of displayed data.

5 Implementation

We implemented an application with basic functionalities of extended tagging as a proof-of-concept. Our prototype system is composed of three parts: a platform for managing the knowledge networks, a client and a web service that handles the access between them. The client is embedded as a plug-in into the sidebar of the web browser Firefox[6] because mostly a web browser is used to search for information in the web. Users can save the current opened web resource in their knowledge network via button or drag and drop. The URL, the title and marked fragments of the web resource are extracted and a screenshot is saved automatically. The client offers a dialog with different input fields – one for each tag type. The users have only to

[6] Mozilla Firefox, http://www.firefox.de, Online 2009-01-10

choose the respective text field to specify a tag with a type. By tagging, the according nodes (i.e. tags, tag types and resources) in the knowledge network are linked. Tags are arbitrary and are created automatically, if they haven't existed yet. The tag types are individually extendable; this means that the users can configure a tag type with new properties and thereby new input fields are created in the dialog form.

The sidebar (see figure 1 left) presents additional information when a new web page is opened in the browser, e.g. the user is notified that the web page has already been saved in the own knowledge network or in the network of someone else. In addition, further web resources are recommended that can be relevant for the current search process. Thumbnails of saved resources in the bottom part of the sidebar visualize the current research management progress so that the users can reflect their search process.

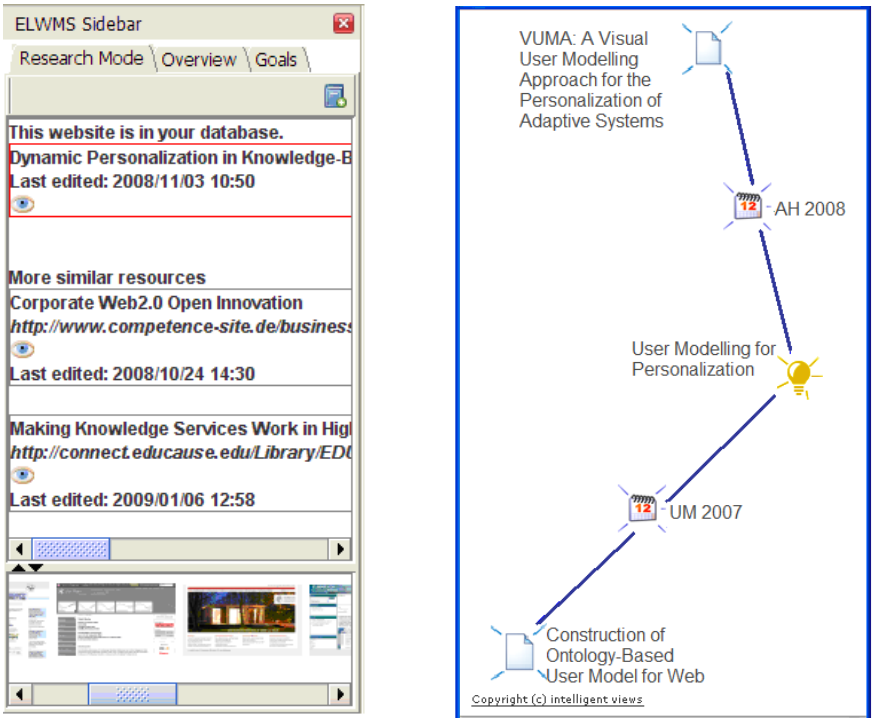

Fig. 1. (left) Screenshot of the sidebar. (right) Screenshot of the Net Navigator showing a subset of a knowledge network. A connection between two resources is inferred over the relations between two conferences covering the same topic.

We use K-Infinity[7] for storing, managing and searching the knowledge networks. This platform has a Knowledge Portal with special features. For example, the users can graphically browse the knowledge network in the Net Navigator, where tag types are marked with different icons. In figure 1 (right), the connection between two resources (page icon) inferred over relations between two conferences (calendar icon) covering the same topic (bulb icon) is shown.

6 Conclusion

This paper introduces a novel extended collaborative tagging concept based on individual knowledge networks where semantic information can be added easily to tagged web pages or fragments of web pages while specifying a type to a tag, linking tags with tags and the possibility to extend tags if needed. This concept enhances the possibilities of collaborative tagging systems with regard to resource management, and despite of its simplicity it can have effects on reflection of learning processes, collaboration, filtering and recommendation. The prototype shows that this extended semantic tagging concept is not only a theoretic construct and the simplicity of its usage will be evaluated soon in a user study. Further steps are improvements of the prototype, for example implementing buddy lists in order to realize different access and visibility permissions in the knowledge network.

Acknowledgments. The project is funded by Hessian State Ministry of Higher Education, Research and the Arts.

References

1. Müller-Prove, M.: Modell und Anwendungsperspektive des Social Tagging. In: Gaiser, B., Hampel, T., Panke, S. (eds) Good Tags - Bad Tags. Social Tagging in der Wissensorganisation. Medien in der Wissenschaft, Band 47, pp. 15--22. Waxmann, Münster (2008)
2. Macgregor, G., McCulloch, E.: Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. Library Review, vol. 55 no. 5, pp. 291--300 (2006)
3. Golder, S., Huberman, B.: Usage patterns of collaborative tagging systems. Journal of Information Science, vol. 32 no. 2, pp. 198--208 (2006)
4. Marlow, C., Naaman, M., Boyd, D., Davis, M.: HT06, tagging paper, taxonomy, Flickr, academic article, to read. In: Proceedings of the Seventeenth Conference on Hypertext and Hypermedia, pp. 31--40, ACM Press, New York (2006)
5. Sowa, J. F.: Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks Cole Publishing Co., Pacific Grove, CA (1999)
6. Böhnstedt, D., Scholl, P., Benz, B., Rensing, C., Steinmetz, R., Schmitz, B.: Einsatz persönlicher Wissensnetze im Ressourcen-basierten Lernen. In: Seehusen, S., Lucke, U., Fischer, S.: DeLFI 2008: 6. e-Learning Fachtagung Informatik, Lecture Notes in Informatics P-132, pp. 113--124, GI (2008)

[7] Intelligent Views: http://www.i-views.de, Online 2009-01-17