

Towards Collaborative Metadata Enrichment for Adaptive Web-Based Learning

Róbert Móra, Ivan Srba, Maroš Unčík, Mária Bielíková, Marián Šimko

Institute of Informatics and Software Engineering, Faculty of Informatics and Information Technologies,
Slovak University of Technology, Ilkovičova 3, 842 47 Bratislava, Slovakia
{xmoror, xsrba, xuncik, maria.bielikova, marian.simko}@stuba.sk

Abstract—In recent years we have witnessed expansion of Web 2.0. Its main feature is allowing users' collaboration in content creation using various means, e.g. annotations, discussions, wikis, blogs or tags. This approach has influenced also web-based learning, for which the term "Learning 2.0" has been introduced. In this paper we explore using tags in such systems. Tags can be used for improving of searching, categorization of web-documents, creating folksonomies and ontologies or enhancing the user-model. Another aspect of tags is that they act as a bridge between resources and users to create a social network. We integrated tags in a learning framework ALEF and experimentally evaluated their usage in education process.

Keywords—Web collaboration; folksonomy; tag; learning framework

I. INTRODUCTION AND RELATED WORKS

An unprecedented expansion of web applications can be observed since September 2005 when official definition of the term Web 2.0 was published by O'Reilly [10]. During this time, the main principles of Web 2.0 proved to be very successful. Similar changes occurred also in the field of e-learning. In the same year Downes specified the term E-learning 2.0 [5] but progress of technologies in e-learning systems is not as fast as in the Web 2.0.

We research how to extend such systems effectively with *Users Add Value* rule, which is one of the key principles of Web 2.0. Users can be involved into the content creation in many ways [11] and participate directly in the content creation, i.e., in wiki systems, or can be involved in adding potentially valuable metadata such as annotations or tags.

In this paper we focus on collaborative tagging in learning systems and the characteristics of the social network created by tags. We provide several recommendations how to effectively integrate a social network acquired from tags into learning systems. We also present an evaluation of our experiment in real educational settings.

A. Collaborative Tagging

Although the idea of tagging is not new, it has increased in popularity with the arrival of social bookmarking service Delicious and photo sharing service Flickr. These services popularized using of tags by adding the social aspect to the process of tagging. Much research has been done in this domain in recent years. Golder and Huberman [7] studied the social bookmarking service Delicious, identified several motives why users use tags and discovered stable patterns in

tag usage. They also discussed linguistic problems of tags such as polysemy, synonymy and basic level variations.

The term folksonomy is strongly related to collaborative tagging. It has been coined by Thomas Vander Wal in a discussion on an information architecture mailing list, in order to name the system of organization that emerges when collaborative tagging is applied. This term originated as a combination of folk and taxonomy and reflects the main principle of collaborative tagging – people (users) use their own vocabulary to categorize web documents. Mathes [8] argues that despite folksonomy's problems, namely ambiguity and synonymy, it works, mainly due to the ability to directly reflect the vocabulary of users and it is also much cheaper – both in terms of time and effort – than building the complex classification hierarchy.

Collaborative tagging is one of the characteristics of the Web 2.0 and also e-learning 2.0, which tries to take advantage of the principles of the social web. This feature has been recently integrated to learning frameworks such as Moodle, Elgg or Blackboard (in its variation of social bookmarking, Bd Scholar).

Torniai et al. designed a method for leveraging folksonomies for learning [13]. Their LOCO-Analyst tool provides teachers with means for visualization of a domain model associated with a selected learning object as well as a tag cloud. Using Context-Based Relatedness Measure it suggests user-provided tags related to a selected concept to teachers, which can be used to update an existing ontology. Through tags, teachers can also find out that students' perception of the course content differs from the course conceptualization encoded in domain ontology.

Tags can be categorized as a special type of annotations. In this paper, by annotations we mean various remarks that are related to a resource. Many web-based systems support various types of annotations. Much work has been devoted to educational domain. Tags and other types of annotations, such as comments, bug reports or external sources, are often used to enrich educational texts. Current educational systems use automatic annotation frequently and at the same time they allow users to add content by themselves. The common way of automatic annotation is to use techniques for information extraction and semantics discovery [4]. We are, however, interested in the enrichment of content by users, which is a different approach. New content is added by the users and relies only on the users. This feature is employed in several existing web-based educational systems, such as AnnotatEd [6].

B. Social Networks

Online social networks are already a very important part of many popular web sites. Social networking sites often offer a possibility of tagging, which allows users to interactively annotate resources using descriptive tags. Thus, the tag as a kind of social annotation can be used to create social network. Although there are some research works on this topic, a little attention is paid to investigate this area in the domain of learning.

Wu and Zhou consider tags in social bookmarking sites as a bridge in three dimensions: tag connects users and resources, tag connects resources and tag connects users [15]. In this tag-centric approach the social network is formed by users linked together by collaborative tagging through sharing of resources. This way it is relatively straightforward to find users' interests via shared tags. Such a social network of tags has the feature of small world and a scale-free network. Their results also showed that connected tags had relatively strong semantic relatedness.

Mika formulated a model of semantic-social networks in the form of tripartite graph of persons, concepts and instance associations, extending the traditional concept of ontologies with social dimension [9]. This representation allowed him to derive a social network of persons based on overlapping sets of tagged objects and a semantic network based on community relationships.

In this paper, we focus on collaborative tagging in learning, because though it is an active area of research and tagging proved to be very successful in other domains, their full potential in the domain of learning is yet to be explored. An important difference with a general purpose tagging system is that tags in these systems reflect not only user interest in a subject, but also his or her knowledge. Thus, we believe that tags can be used to derive domain model, or to find users (students) with similar knowledge (and interests), which is important for creating virtual study groups.

II. TAGS IN ADAPTIVE LEARNING FRAMEWORK ALEF

In order to provide students with a truly interactive environment where they could participate in content creation and collaborate, an adaptive learning framework ALEF has been designed and implemented [12]. ALEF adapts main principles of adaptive web-based learning. It is designed with flexibility to develop loosely coupled components as one of primary goals. ALEF represents a full-featured e-learning portal that supports learning and collaboration/creation flow [2] currently used for learning programming (C, lisp and prolog languages) and principles of software engineering.

The collaboration/creation flow allows adding various types of content by using collaborative content creator tools family. Currently, three collaborative content creators are implemented – commentator, external resources inserter and questions creator (see Fig. 1, which shows two of them). We have extended this family of tools by implementing the tagger as is also shown in Fig. 1.

Domain model of ALEF is divided into two layers: content and metadata. The content consists of documents, which can be accessed via an URI and can be presented to the users. In order to describe the characteristics of learning objects, the metadata layer is used. Metadata includes annotations and concepts – knowledge domain elements. ALEF's data model is sufficiently general to ease the adding of new components such as collaborative content creators, which was with advantage used in previous studies [14].

We have taken this advantage also for implementing the tagger as a component in ALEF. From data model design view we consider the tags as a special kind of annotations. ALEF data model had already contained data entity for annotations, as well as for special kinds of annotations (i.e. comments, bug reports); we have easily extended the data model using inheritance by creating a new special data entity for tags (see Fig. 2).

The screenshot displays the ALEF web interface. At the top, there is a navigation bar with the ALEF logo and a user profile for 'Maros Unck (študent) | Odlhásť'. The main content area is titled '2.1 Způsob zpracování programu' and contains text about C programming, a diagram of the compilation process, and definitions for 'Editor' and 'Preprocessor'. The diagram shows the flow from 'Editor' to 'Preprocessor/Compiler' (which takes '.H' files and produces '.OBJ' and '.LIS' files), then to 'Linker' (which produces '.EXE' files), and finally to 'Debugger'. A sidebar on the left shows a list of lessons, with '2.1 Způsob zpracování programu' selected. On the right, there is a 'Tvoje skóre' section showing a score of 7.5 and 3, and a 'Tagy' section with options for 'moje' and 'populárne' tags, and a 'Pridaj' button. A 'Filter' box is located at the top right of the main content area.

Figure 1. Example screenshot of ALEF (in Slovak). Recommender with recommendations is situated over the content navigation (1). Collaborative tools are visualized on the left – motivation tool (3), external resources inserter (4) and tagger (5) allow students to collaborate. Content can be enriched by adding different types of annotations and is accessed by pop-up menu over the highlighted sections of text (2).

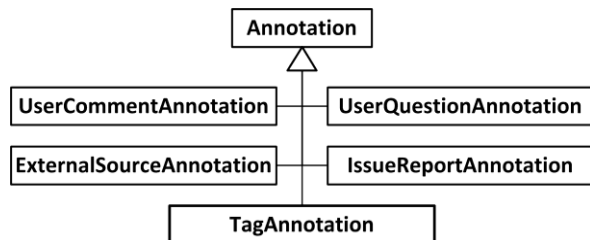


Figure 2. Fragment of ALEF data model relevant to tags (our extension of model is at the bottom).

Using the tagger users can add tags to learning objects, such as explanatory texts, questions or exercises. Tag can be a keyword or a multiword phrase, which is limited only by maximal character length. A user chooses, whether he wants a tag to be *private* (other students will not see the tag) or *anonymous* (tag will be public, but others will not see by whom it was added). Users can view either tags they added themselves (in *my tags view*) or *popular tags* (in *popular tags view*). A tag is considered popular in a learning object, if it was added by at least three users as an anonymous tag to the learning object. Users can also delete tags, but only those they added themselves.

Finally, tags serve as a means of navigation in learning objects space – when a user selects a tag, a list of all learning objects tagged with the tag selected is shown. List of learning objects shown depends on the tag view too. In *my tags view*, only learning objects tagged by the user are shown, while in popular tags view, all learning objects tagged with popular tags are shown.

III. EVALUATION

We have evaluated tagging behavior during an uncontrolled experiment. We have assessed how students use tags in ALEF and the capability of these tags to cover important concepts of learning objects. Students were asked and motivated to use ALEF and to tag learning objects to prepare themselves for final exam in programming course. We consider a social network of users created implicitly by their collaboration and explicitly by their grouping into courses, setups or virtual study groups.

The experiment lasted for 15 days. During this period we gathered 2 272 tags, out of which 947 were unique. In order to be able to evaluate these tags, we further normalized them as follows: we removed diacritic marks and some other characters, e.g. brackets that occur together with function names (for example *malloc()* to *malloc*), converted all tags to lowercase, lemmatized them and finally we translated some frequently occurring tags (such as string or structure) from English to their Slovak equivalents (*retazec*, *struktura* resp.). After normalization, 755 unique tags were left.

From all the tags gathered, only 2% were deleted by students and only 15% were private. It shows, that students used tags to describe the content of the learning objects and not as a means for personal notes (for example, only two *ToDo* tags were added). Each unique tag has 2.4 occurrences on average; the overall distribution of tag occurrences follows the power law distribution as expected (see Fig. 3), when only one tag occurred 96 times, but 461 tags occurred

one time. Tab. 1 shows the number of top 5 tags' occurrences (after normalization). We can see that these tags well reflect important concepts in the domain of C programming language, which was taught in the course.

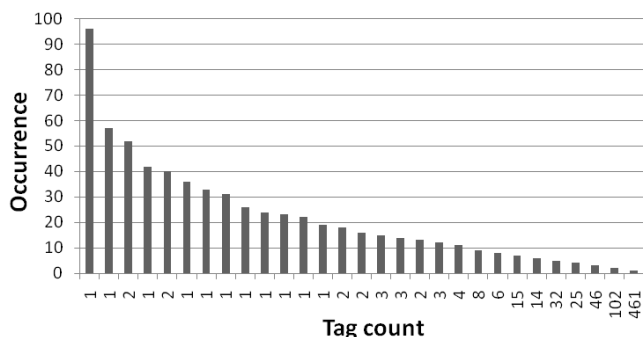


Figure 3. Distribution of tag occurrences.

TABLE I. TOP 5 TAGS OCCURRENCES.

	Tag (in Slovak, normalized)	Tag (translated to English)	Occurrence count
1	pointer	pointer	96
2	makro	macro	57
3	retazec	string	52
4	struktura	structure	52
5	operator	operator	42

Next, we have explored the distribution of tags regarding the learning objects. Students managed to tag 613 learning objects (more than 77%) with 3.63 tags in average (considering only learning objects tagged by at least one tag). The overall distribution again follows the power law (see Fig. 4), when one learning object was tagged with 28 tags, while 131 of them were tagged only by one tag. 182 learning objects (less than 23%) were not tagged; these are not considered in our evaluation.

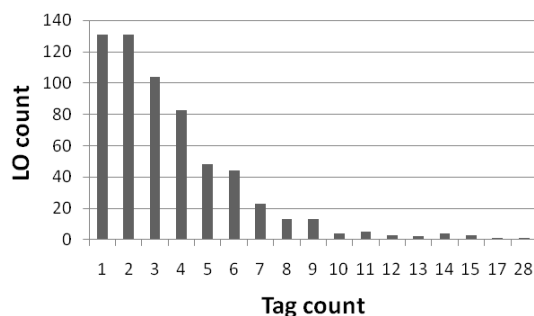


Figure 4. Distribution of tags in regard to learning objects (LO).

We have also evaluated students' behavior when adding the tags. However, this particular evaluation has been influenced by the fact that only 35 out of 82 students using the system at the time of our experiment added tags. Therefore, the average number of tags per student is 63.6. Number of tags varies significantly from student to student and the distribution does not follow the power law. Users were allowed to add phrases instead of a keyword as a tag

(up to 30 characters). However, almost 78% of tags consisted of only one word and only 16% of two words. Average number of words per tag was 1.28 and no tag consisted of more than five words.

Finally, we were interested in the capability of tags to cover important concepts defined by experts in the domain model of the course. To calculate the overlay ratio φ , we compared normalized tags to normalized concepts associated with the same learning object using (1), where $|LO|$ is a number of tagged learning objects, $Tags(lo)$ and $Concepts(lo)$ are sets of tags and concepts associated with the learning object lo respectively.

$$\varphi = \frac{\sum_{lo=1}^{|LO|} |Tags(lo) \cap Concepts(lo)|}{\sum_{lo=1}^{|LO|} |Concepts(lo)|} \quad (1)$$

We have found out that the overlay ratio φ is 27.76%. In the explored domain model, 263 unique concepts were identified by the experts. From these, 49.8% were covered by the folksonomy, i.e. about half of the domain concepts have an exact match (after normalization) with one of the tags. This measure does not reflect relations between concepts and learning objects, but rather indicates the capability of folksonomy to identify important concepts in the domain. Interestingly, only 17% of unique tags added by users were covered by the concepts. This can be partially explained by the imperfection of our comparison measure or by the use of semantically similar words, which were not fully covered by our normalization method. However, many of these tags were actually concepts which were not originally contained in the domain model.

IV. CONCLUSIONS

In this paper, we have explored the characteristics of personalized web-based learning 2.0 systems and focused on the role of collaborative tagging in these systems. Our main contribution lies in demonstrating how the tagging can easily be introduced to an e-learning framework as an extension to existing domain model. We have designed and implemented tagger component in ALEF and experimentally evaluated tagging during programming course in a real world setting.

Findings of our experiment suggest that it is possible and beneficial to use folksonomies for domain modeling, either to enrich an existing model or even to build it from the bottom up. This approach is collaborative and social, because it relies on users of the system to find consensus on the concepts and their relationships are derived by co-occurrence of tags. Thus, semantic network of the domain is derived by social network of users. However, this process is bidirectional and semantic network emerging from usage of tags can enrich the social network of users (students).

As future work we intend to utilize folksonomy for this purpose and explore other possible usages of collaborative tagging, i.e. enriching the user model following term-based user modeling approach [1] or providing personalized search or automatic text summaries in the learning environment.

ACKNOWLEDGMENT

This work was partially supported by the grant KEGA 028-025STU-4/2010 and it is the partial result of the Research & Development Operational Programme for the project Research of methods for acquisition, analysis and personalized conveying of information and knowledge, ITMS 26240220039, co-funded by the ERDF.

The authors wish to thank all members of the PeWe group (pewe.fiit.stuba.sk) for their invaluable contribution to discussions, development of the ALEF framework and support in experiments. The most current state of ongoing projects within the group is reported in [3].

REFERENCES

- [1] M. Barla, "Towards Social-based User Modeling and Personalization," Information Sciences and Technologies Bulletin of the ACM Slovakia, vol. 3, no. 1, 2011, pp. 52-60.
- [2] M. Bieliková, M. Šimko, M. Barla, "Personalized Web-Based Learning 2.0," Proc. of 8th Int. Conf. on Emerging eLearning Technologies and Applications, 2010, pp. 5-10.
- [3] M. Bieliková, P. Návrat, M. Barla, J. Tvarožek, M. Tvarožek (Eds.), "Workshop on the Personalized Web – Science, Technologies and Engineering," 2011. ISBN 978-80-227-3485-1. Available online at pewe.fiit.stuba.sk/ontoparty-2010-2011-spring/.
- [4] M. Brut, D. Kukhun, F. Sedes, "Ensuring Semantic Annotation and Retrieval within Pervasive E-Learning Systems," Proc. of the Int. Conf. on Complex, Intelligent and Software Intensive Systems. IEEE Computer Society, Washington DC, USA, 2008, pp. 959-964.
- [5] S. Downes, "E-learning 2.0," eLearn, vol. 2005, no. 10, 2005.
- [6] R. Farzan, P. Brusilovsky, "AnnotatEd: A Social Navigation and Annotation Service for Web-based Educational Resources," Proc. of E-Learn '06, 2006, pp. 2794-2802.
- [7] S.A. Golder, B.A. Huberman, "Usage patterns of collaborative tagging systems," Journal of Inf. Science, vol. 32, no. 2, 2006, pp. 198-208.
- [8] A. Mathes, "Folksonomies - Cooperative Classification and Communication through Shared Metadata," Computer Mediated Communication - LIS590CMC, 2004.
- [9] P. Mika, "Ontologies Are Us: A Unified Model of Social Networks and Semantics," Proc. of ISWC, LNCS, Vol. 3729, Springer-Verlag 2005, pp. 522-536.
- [10] T. O'Reilly, "What is Web 2.0 – design patterns and business models for the next generation of software," oreilly.com/web2/archive/what-is-web-20.html (accessed 06/01/2011).
- [11] H. Rollett, M. Lux, M. Strohmaier, G. Dosinger, K. Tochtermann, "The Web 2.0 way of learning with technologies," Int. Journal of Learning Technology, vol. 3, no.1, 2007, pp. 87-107.
- [12] M. Šimko, M. Barla, M. Bieliková, "ALEF: A Framework for Adaptive Web-based Learning 2.0," Proc. of IFIP Advances in Inf. and Comm. Tech., vol. 324. Springer-Verlag, 2010, pp. 367-378.
- [13] C. Tormiai, J. Jolanović, S. Bateman, D. Gašević, M. Hatala, "Leveraging Folksonomies for Ontology Evolution in E-learning Environments," Proc. of the IEEE Int. Conf. on Semantic Computing, 2008, pp. 206-213.
- [14] M. Unčík, M. Bieliková, "Annotating Educational Content by Questions Created by Learners," Proc. of 5th Int. Workshop on Semantic Media Adaptation and Personalization, IEEE, pp.13-18.
- [15] Ch. Wu, B. Zhou, "Analysis of tag within online social networks," Proc. of the ACM Int. Conf. on Supporting Group Work, ACM, New York, NY, USA, 2009, pp. 21-30.