Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 00 (2011) 1–8

Procedia Engineering

GCSE 2011: 28-30 December 2011, Dubai, UAE

Community Tagging in Tele-Teaching Environments

Franka Gruenewald, Maria Siebert, Christoph Meinel

Hasso-Plattner-Institute
University of Potsdam
Potsdam, Germany

Abstract

Collaborative tagging as such is introduced and the usage of it in the tele-teaching context is motivated. Several use cases for community tagging are suggested. Those include the improvement of the learning process through tagging, the extension of the search possibilities and the determination of similarities between content through tag comparison. The importance of a certain tag for a content item together with the intersection of the tags for two content items are used to determine the similarity of content items for the use in recommender systems. A detailed formula is suggested. The activation of users, the automatic creation of tags and the exploitation of semantics to enhance the utilization of tags are described as outlook for future work.

Keywords: Tagging, Community, E-Learning, Tele-Teaching, Web 2.0, Recommendation

1. Introduction

Several new concepts have revolutionized the world wide web in the last decade. Among those are the community driven features that enhance the static and administration driven web to a rapidly changing as well as collaboratively created and improved world wide web. This concept is called web 2.0 [1]. Whereas user participation is the key to this new concept, leveraging the possibilities of creating new knowledge and information with the help of computation is the second concept. The technology this is based on is called semantic web [2]. Due to the increasing percentage of the population that is interested in the world wide web and actively taking part in it, the amount of data available on the web is rapidly growing.

The same issue can be observed in tele-teaching. As more and more people have access to tele-teaching systems and the recording technology, like tele-TASK [3], becomes easier to use and therefore enables content producers to create more content a lot easier, the amount of lecture recordings increases rapidly.

With the help of the aforementioned new concepts, new possibilities arise to improve the search, filtering and personalization opportunities amongst the large amount of content and therefore support learners in still finding the needle in the haystack. Good search and filtering techniques, especially within media learning content, are currently

Email address: franka.gruenewald|maria.siebert|christoph.meinel@hpi.uni-potsdam.de (Franka Gruenewald, Maria Siebert, Christoph Meinel)

within the research focus [4, 5]. Creating appropriate metadata to those content items is the key to a successful search. Some manual metadata creation techniques, where not administrators but the users of the content themselves are involved, derived from the web 2.0. An introduction into web 2.0 concepts and the combination of those methods with tele-lecturing is introduced in the next paragraph.

1.1. Community and Social Web Functionalities in Tele-Lecturing Scenarios

Since the beginning of the web 2.0 era [1] numerous social web portals, whose main motivation is fostered around the user participation, have evolved and grew very quickly.

A number of social web and community features have been found to be useful to the users. These include blogging, the collaborate creation of wikis, social annotating and tagging, evaluating (eg. rating and commenting), recommending, content sharing and linking of content items [1]. The key success factor for these features is the interaction amongst users [6].

That community functionalities are not only useful for networking, but also for learning contexts was found out at the beginning of the e-learning era around 2000 already [7, 8]. But only recently research started on joining tele-lecturing with community functionalities. During the workshop *eLectures 2009* at the conference DeLFI 2009 [9] an approach of integrating tele-lecturing applications into facebook, a combination of wikis and tele-lecturing and other social e-learning approaches were shown.

This paper deals with community tagging as one web 2.0 feature with a lot of potential for e-learning. The tagging functionality will be introduced in the next section.

1.2. Collaborative Tagging

Tagging is the assignment of keywords, the so called tags, to a resource. In collaborative tagging, where people are taking over the role of assigning tags to resources, this keyword can be chosen according to the experiences with the resource and the perception of the person choosing the keyword, the tagger. The new aspect thereby is that not a librarian or an administrator chooses those tags, but all users of the community are allowed to participate in the tagging activity.

Each keyword is meant to be one aspect of the resource. Tags therefore help to summarize and classify the content of a resource. [10, 11] In collaborative tagging, the users can tag freely without limitations. That means that they can choose any vocabulary they can think of including compound words and a combination of several words. This enables to user to categorize content according to his personal strategy. The compound of all tags collected via collaborative tagging is finally called folksonomy, although there is some discussion about the correctness of the term [11]. A folksonomy is formally defined as a quadruple $F := (U, T, R, Y)$ whereby U, T and R are finite sets whose elements are users, tags and resources. Y represents the tag assignment, a relation between the three other sets, which can formally be described as $Y <= UxTxR$ [12].

As sample the tagging functionality was implemented at the tele-teaching portal tele-TASK[1] of the Hasso-Plattner-Institut (HPI). Some details of this project will be explained in the next paragraph, as the tele-TASK project includes a recording system as well as a portal for distributing e-lectures.

1.3. Tele-Teaching with tele-TASK

The tele-teaching Anywhere Solution Kit [3], short tele-TASK, is an e-learning project at the chair Internet-Technologies and -Systems at the HPI. The tele-TASK project was started in 2002 at the university of Trier by developing a hardware system for lecture recording. An all-in-one solution was developed including hard- and software for lecture recording. It is a plug-and-play set-up. Two video steams (a video of the lecturer and screen capturing of his laptop or a smart-board) and one audio stream can be recorded at once. More than 2000 lectures and 4000 podcasts of the tele-TASK archive can be accessed free of charge via web-browser or portable device. The large video archive and the web-platform tele-TASK are the basis for further research and development at the HPI.

The next paragraph will give an introduction into use cases of tagging in tele-teaching environments and show a sample implementation within the tele-TASK portal.

[1] http://www.tele-task.de

2. Community Tagging in a Tele-Teaching Portal

In a tele-teaching environment three different types of tags from the tag categories defined by Golder et al. [11] can be imagined. Those are task organising tags and content-related tags. The other categories of tags are not expected in tele-teaching environments. Tags to identify the resource are not needed, as the content type is defined by the surrounding framework. The ownership and self-reference are also represented by the framework. Subjective opinions about the content should rather be expressed via content rating than tagging.

The task organising tags serve the individual users for the administration of their learning schedule. 'Revise for the exam', 'Essential for the next homework', 'Consume before the next lecture' and other tags relating to the time when the learning content shall be consumed or for which event in the course schedule the learning content is needed can be thought of in this connection. The administrative tags are only sensible for individual learners or a learning group of closely connected learners.

Content-related tags include the two tag functions to identify the content of the tagged resources and refining the categories created by existing tags [11] or an existing categorization scheme. Those tags are most useful for the whole community using the tele-teaching portal, because those can support search, filtering and personalization within the media content of the portal.

This paragraph will give an introduction into the use of tagging in a tele-teaching environment. Several sources can be used to create content tags and diverse concrete use cases can be named for tags in tele-teaching environments. This paragraph will further deal with different sources from which tags can be delivered and explain use cases for tags as well as a data model, which fully supports all possible capabilities of the tag functionality.

2.1. Sources for Tags in Tele-Teaching Environments

Tags are a form of metadata that can either be created manually or harvested automatically. Sources for the automatic metadata harvesting can for example be the video or audio of the lecture recording. Optical character recognition is used to extract keywords from the presentation slides, that means the video of the lecture recording [13]. Audio transcription is used to gather keywords from the speech of the lecturer, that means from the audio recording of the lecture [14]. In both cases algorithms like word stemming, word clustering and frequency calculations have to be used to only accumulate the most relevant keywords from the files and convert them into tags[14].

One can distinguish between administrative and collaboratively created manual tags. Administrative metadata is usally created by media technology engineers or recording personnel. It is imaginable to include administrative tagging into the post processing workflow of the lecture recording as well. But it was found out, that tags created by administrative personnel quite often do not match tags allocated by users [10].

Finally tags can be collaboratively created by the community of a website. All users are invited to participate in it and enter their own keywords freely through a text field. The motivation for learners to participate in collaborative tagging can be diverse. Therefore use cases and benefits will be explained in the next section.

2.2. Use of Tagging in Tele-Teaching Environments

In summary the use of tagging includes the enhancement of content finding capabilities and the improvement of the cognitive processes in the students learning workflow. The single use cases will be described more detailed in the following paragraphs.

2.2.1. Uses for Content Providers

It was found out that web 2.0 features support the user's loyalty to the website and result in a rise of the mean browsing time as well as the intensity of the site visit [6]. Also the inviting friends paradigma of web 2.0 will eventually lead to learners inviting fellow learners to join in and register for the tele-teaching plattform as well. For the content provider that means that more users can be attracted to the site for a longer time.

For that reason it is furthermore worthwhile to investigate if those web 2.0 features may also positively influence learners to spend more time on the e-learning website and to intensify the learning experience.

2.2.2. Improvement of the Learning Process Through Tagging

Tagging is the action of making sense of the content that is consumed [11]. In order to perform tagging learners have to spend time consuming the learning object and reason about its content. With this process they actively participate in the learning instead of just leaning back and consuming the learning content without analysing and actively processing the knowledge. Formally speaking the learner ascends one step in the Taxonomy of Learning, from the level 'knowledge and comprehension' to the level of 'application and analysis' [10, 15]. When further utilizing the tags for choosing adjacent topics and browsing similar topics even the highest level of 'synthesis and evaluation' can be reached.

The great opportunity in collaborative online learning is also that the students may learn from each other with the help of sharing and organizing the knowledge [11]. The creation of compositions of content [16] is one aspect of this opportunity, tagging is another. How can students learn from each other when using tags? When learners organize themselves in online learning groups, they can follow the learning process of their team colleagues. This concept of following is known from social networking websites like Facebook. As students in one learning group usually have the same learning goal, learners can follow the same learning path they have seen from a team colleague who started learning this topic. They know about this learning path, because they are able to see actions (like the creation of playlists, rating and tagging of content) this team colleague performs. Tagging is especially useful in this context, because reflection about the content is required to find appropriate tags and other learners can benefit from that process as they receive a more accurate description of the content [10].

2.2.3. Extending Search Possibilities with the Help of Tags

Several new search possibilities arise when tagging is enabled for a tele-teaching environment. This is especially necessary, because the search amongst video files is not a trivial issue. Without tagging the following search strategies were realizable. Usually the learning videos are categorized in a simple top-down structure, either according to different semesters or according to a category of subjects. The students could simply use this structural browsing via the menu to click through the categorization until they find the desired content.

The second opportunity is the search function, which queries the database according to keywords that are looked for in the title or description in different languages. Furthermore it was possible for students to open up their own organisational structure of the content by using compositions, so called playlists [16]. The playlist titles, descriptions and groups can be searched for keywords as well.

The last possible method is pivot browsing [10]. This type of search is enabled by clickable headlines, names and links in the text. When using this browsing method for searching, it is a rather explorative search.

Including tags into a tele-teaching environment opens up a new world of search for the users. Basically using a tag system can be considered as one kind of keyword-based search [11]. But when using this tagging system, it is visible to the learner how important a certain keyword is for a specific content, because the more users use this tag the more important this tag is. The whole folksonomy can be visualized in a tag cloud. The importance of the single tags is thereby visualized with the help of the font-size - the larger the font size the more important the tag. The tag-cloud of the tele-teaching portal tele-TASK is shown as example in figure 1.

Furthermore a tagging system is a form of alternative categorization system that the user can utilize to retrieve the information he organized in that way more easily [17]. The tag cloud as well as a standard search interface can serve as entry point for browsing in those categories.

With the help of tags it is furthermore possible to provide personalized content for the users [6]. On the one hand interesting topics for them can be determined by processing the tags, on the other hand a personalized tag-cloud can provide an easier access to topics the learner is currently interested in.

2.2.4. Utilizing Tags to Determine Similarities Between Content

A very important feature for learners is the determination of related content. If the learner has not understood a certain topic, he might want to look into further material about this topic. If however he comprehended everything, he may want to proceed to the next topic within the learning domain. This topic belongs to the field of recommender systems. Those are used to propose content to the user in order to support him in finding the content he is looking for.

The importance of a tag for a content item can be determined through two factors. The first is the frequency with which the tag is used for the content item. The second is the relevance of the tag for the content item. The question

Figure 1: Tagcloud of the tele-teaching portal tele-TASK

is, how this relevance can be computed. Existing approaches work the other way around and suggest appropriate tags to the content items [18]. Up to now the data set is not large enough to also consider a meaningful inclusion of the users that assigned the tags in our calculation, although those are part of the definition (as was described in paragraph 1.2). Therefore only the tags and tag assignments will be considered in our proposed calculation for the similarity of content items with respect to their tags.

One example should show how the relevance of a tag for a content item can be determined. Considering that one lecture has 100 tags amongst which one is 'internet'. A second lecture only has 5 tags where one of them is 'internet'. In the case of the second lecture the tag 'internet' is more relevant, because it is one amongst a very limited number of tags whereas the first lecture has such a large number of tags in total that the single tag is less relevant.

Figure 2 shows how tag relations can be visualized with two tags connected to four different objects. Each arrow represents one connection. Because a tag can be used multiple times with the same object when different users use the same tags, multiple connections between the tags and the objects are possible.

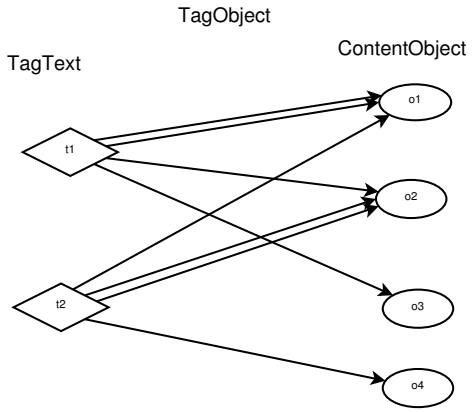

Figure 2: Relations between tags and content objects

When calculating the similarity between two of these objects, at first the weight of each tag has to be considered. Therefore let T be the set of all tags and $c_i(o,t)$ the i-th connection of object o with tag t. Furthermore we define:

$$T_{o_i} = \{\text{tags of object } o_i\} \tag{1}$$

$$w_{\{M\}}(t_1) = \#\{c_i(o,t_1) | o \in M\} \tag{2}$$

When comparing two objects o_i and o_k, the easiest approach is to look at the number of tags available in both objects and compare it with the number of tags available in at least one object, as shown in formula 3.

$$sim_{o_i,o_k} = 100 \cdot \frac{\#(T_{o_i} \cap T_{o_k})}{\#(T_{o_i} \cup T_{o_k})} \tag{3}$$

Because of the possibilities of more than one connection between a tag and an object, this approach is a little too easy. The number of usages of a tag should be considered. The following formula considers that fact as well.

$$sim_{o_i,o_k} = 100 \cdot \frac{\sum\limits_{t \in (T_{o_i} \cap T_{o_k})} w_{\{o_i,o_k\}}(t)}{\sum\limits_{t \in (T_{o_i} \cup T_{o_k})} w_{\{o_i,o_k\}}(t)} \tag{4}$$

The resulting combined formula of the formulas 3 and 4 is shown in formula 7.

$$p_1 = \frac{\sum\limits_{t \in (T_{o_i} \cap T_{o_k})} w_{\{o_i,o_k\}}(t)}{\sum\limits_{t \in (T_{o_i} \cup T_{o_k})} w_{\{o_i,o_k\}}(t)} \tag{5}$$

$$p_2 = \frac{1}{\max\left(1, \sum\limits_{t \in (T_{o_i} \cup T_{o_k}) \setminus (T_{o_i} \cap T_{o_k})} w_{\{o_i,o_k\}}(t)\right)} \tag{6}$$

$$sim_{o_i,o_k} = 100 \cdot p_1 \cdot p_2 \tag{7}$$

The problem is that this formula will have small values in most cases, because the number of equal tags between two objects is small in most cases. That is why it is better to use a smaller decreasement than linear, for example a logarithmic one (see exchange of the second part of the formula in formula 8).

$$p_2 = \frac{1}{\max\left(1, \sum_{t \in (T_{o_i} \cup T_{o_k}) \setminus (T_{o_i} \cap T_{o_k})} \log(w_{\{o_i,o_k\}}(t))\right)} \tag{8}$$

This calculation can be combined with similarity calculations of other meta data using the plug-in architecture [19] of the portal. Because the calculation can take long when many tags are used, it is useful to save the results of the comparison in an extra table and to update the table each time the tags are changed.

3. Evaluation of Tagging

Up until now more than 440 tag assignments have been made for which about 200 different tags were chosen. But only 6 users were involved into the creation of those tags. This means that the user participation is very low. This fact could already be observed with other web 2.0 features in the tele-TASK portal and was shown in a study about the web 2.0 video service YouTube [20] as well. Of course this is a big problem, as it was found out that these collaborative processes gain popularity as more users participate. That means it is crucial that a minimum number of users participate before the functionality is widely accepted [10]. Some possible solutions to this problem will be explained in the outlook section.

Hitherto experiences have shown that the tagging functionality as it was implemented is intuitively usable. This can be seen, because the few participants have created large amounts of tags in a very short timespan. That would not have been the case if the functionality was unusable. In order to verify this rather rough deduction, usability tests can be conducted in the future. Further activities are planned as well to improve the tagging functionality for the tele-teaching purpose. The next paragraph will give a conclusion and outlook.

4. Conclusion and future work

This paper motivated the usage of collaborative tagging for e-learning. Several use cases for tags in the tele-lecturing context were proposed, all of which were implemented at the tele-lecturing portal tele-TASK. However some major issues still need to be solved and further research and development is still required.

The most important issue that needs to be solved is the activation of users to participate in the collaborative creation of tags. Up until now the user participation is very low, which is why all functions derived from tagging cannot be utilized to the full extend as the underlying data has not been created. This issue has been observed by others as well. It was found out that a critical mass of content has to be created before the community starts participating more actively. One suggestion in this context is to automatically create tags from the content until the critical mass is reached [10].

There exist several possibilities to automatically generate tags by processing the lecture recordings. Of course tags could be derived from the title or description of the lecture, but this would not bring any additional value, as this data is already searchable itself, because it is in text form. Added value is only created by information retrieval from the audiovisual content. The two main strategies existing in conjunction with this topic are the processing of videos with the help of optical character recognition [13] as well as the transliteration of the audio channel [14]. In both cases tags can be generated by determining the most important keywords for a given time span.

If tags were created automatically it would be interesting to compare the qualitative differences of the automatically created tags versus user generated tags. Other research has shown that the intersection of user generated tags with tags generated by a metadata harvesting software was quite significant [10]. If the same conclusion can be drawn when comparing collaboratively created tags to automatically retrieved tags, a combination of both can be thought of for a successful tagging system.

It is also intended to further refine the tags. First of all it should be possible for the users to create public as well as private tags. The default will obviously be set to public, because the intention is that the whole community can benefit from the tags of the individual users. But private tags should help the users to organize the learning content for himself only. The private tags can for example be organizational tags as described in a previous section.

The next step to fully leverage the tags for knowledge management and search within large amount of data is the connection of tags with regards to content. One approach how the similarity of tags can be determined by how frequently they were used together by the same user is shown in [6].

A more advanced method is the utilization of ontologies. In [21] is described how tags can be used to maintain domain ontologies in e-learning. However the goal should be to re-use existing ontologies and map the user tags to them in order to provide added value to the tags, like similar tags, parent or child tags and their belonging content items. A method of manual mapping was described in [22], but was assessed not to be user friendly and feasible [23]. Therefore the ambition is to automatically map existing tags from users and also automatically created tags to existing ontologies and enable automatic reasoning over the tags with this approach.

[1] T. O. Reilly, What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, Communications & Strategies No. 1 (65) (2007) 17–37.
[2] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, Scientific American Magazine 284 (5) (2001) 34–43.
[3] V. Schillings, C. Meinel, Tele-TASK – tele-teaching anywhere solution kit, in: Proceedings of ACM SIGUCCS, Providence, USA, 2002.
[4] M. Siebert, C. Meinel, Realization of an Expandable Search Function for an E-LearningWeb Portal, in: Workshop on e-Activity at the Ninth IEEE/ACIS International Conference on Computer and Information Science Article, Yamagata/Japan, 2010, p. 6.
[5] J. Waitelonis, H. Sack, J. Hercher, Z. Kramer, Semantically Enabled Exploratory Video Search, in: Proc. of Semantic Search Workshop at the 19th Int. World Wide Web Conference, Raleigh, NC, USA, 2010.
URL `http://km.aifb.kit.edu/ws/semsearch10/Files/video.pdf`
[6] D. Mican, N. Tomai, Web 2.0 and Collaborative Tagging, 2010 Fifth International Conference on Internet and Web Applications and Services (2010) 519–524doi:10.1109/ICIW.2010.83.
[7] R. M. Palloff, K. Pratt, Building Learning Communities in Cyberspace: Effective Strategies for the Online Classroom (The Jossey-Bass Higher and Adult Education Series) Share your own customer images Search inside this book Building Learning Communities in Cyberspace: Effective St, 1st Edition, Jossey-Bass, 1999.
[8] M. J. Rosenberg, E-learning: strategies for delivering knowledge in the digital age, McGraw-Hill, 2001.
[9] S. Trahasch, S. Linckels, W. Hürst, Vorlesungsaufzeichnungen - Anwendungen, Erfahrungen und Forschungsperspektiven. Beobachtungen vom GI-Workshop "eLectures 2009", i-com 8 (3 Social Semantic Web) (2009) 62–64. doi:10.1524/icom.2009.0040.
[10] S. Bateman, C. Brooks, G. Mccalla, P. Brusilovsky, Applying Collaborative Tagging to E-Learning, in: Proceedings of the Workshop on Tagging and Metadata for Social Information Organization (WWW'07), ACM, Banff, Canada, 2007.
[11] S. A. Golder, B. A. Huberman, The Structure of Collaborative Tagging Systems, Journal of Information Science 32(2) (2005) 198–208.
[12] M. G. Noll, Understanding and Leveraging the Social Web for Information Retrieval Dissertation, Phd thesis, Hasso-Plattner-Institut für Softwaresystemtechnik (2010).
[13] H. Sack, Automated Annotation of Synchronized Multimedia Presentations, in: In Workshop on Mastering the Gap: From Information Extraction to Semantic Representation, CEUR Workshop Proceedings, Berkeley, CA, USA, 2006.
[14] S. Repp, C. Meinel, Automatic Extraction of Semantic Descriptions from the Lecturer's Speech, in: I. Press (Ed.), Proc. 3rd ICSC, Berkeley, CA, USA, 2009, pp. 513–520.
[15] B. S. Bloom, Handbook 1: Cognitive Domain, 2nd Edition, Addison-Wesley Publishing Company, 1956. doi:ISBN 978-0582280106.
[16] M. Siebert, F. Moritz, C. Meinel, Enriching E-Learning Meta Data with User Generated Playlists, in: 5th International Conference for Internet Technology and Secured Transactions (to appear), IEEE Computer Society, London, UK, 2010.
[17] M. D. O. C. Ames, S. University), M. Y. R. B. Naaman, Why We Tag : Motivations for Annotation in Mobile and Online Media, in: CHI 2007 Proceedings, ACM New York, USA, San Jose, CA, USA, 2007, pp. 971–980.
[18] D. Liu, X.-S. Hua, L. Yang, M. Wang, H.-J. Zhang, Tag Ranking, in: 18th International World Wide Web Conference, Madrid, Italy, 2009, pp. 351—-351.
[19] M. Siebert, F. Moritz, C. Meinel, Establishing an Expandable Architecture for a tele-Teaching Platform, in: 2010 Ninth IEEE/ACIS International Conference on Computer and Information Science Article, IEEE Computer Society, Yamagata, Japan, 2010.
[20] M. Cha, H. Kwak, P. Rodriguez, Y.-y. Ahn, S. Moon, I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System, in: ICM, San Diego, California, USA, 2007, pp. 1–13.
[21] C. Torniai, J. Jovanovic, S. Bateman, D. Gasevic, M. Hatala, Leveraging Folksonomies for Ontology Evolution in E-learning Environments, in: 2008 IEEE International Conference on Semantic Computing, IEEE Computer Society, Los Alamitos, CA, USA, 2008, pp. 206–213. doi:10.1109/ICSC.2008.15.
[22] S. Bateman, C. Brooks, G. Mccalla, Collaborative Tagging Approaches for Ontological Metadata in Adaptive E-Learning Systems, in: Proc. Workshop on Applications of Semantic Web Technologies for e-Learning, at the 4th International Conference on Adaptive Hypermedia and Adaptive Web-based Systems, Dublin, Ireland, 2006, pp. 3–12.
[23] C. Brooks, S. Bateman, J. Greer, G. Mccalla, Lessons Learned using Social and Semantic Web Technologies for E-Learning, IOS Press, 2009, Ch. 14, pp. 260–278.