

# Leveraging Social Web Functionalities in Tele-Teaching Platforms

Franka Moritz, Maria Siebert, Christoph Meinel  
*Hasso-Plattner-Institute  
University of Potsdam  
Potsdam, Germany*

## Abstract

*A lot of content for tele-teaching portals was produced in the last decade. But metadata to filter and search through the content was not generated adequately. That is why it is a new challenge to find solutions how to filter the large amount of data with a small base of metadata available. To engage the user community to generate metadata is one option. Due to the small size of tele-teaching user communities this approach needs to be enhanced with automatic methods. This paper describes community rating and tagging as two widely used social web functionalities and motivates their usage in the tele-teaching context. Furthermore approaches to extend and enhance this user-generated metadata are explained. The potential of other social web features is explained afterwards. Finally the need to activate more users and the opportunity to connect the social with semantic web features in tele-teaching are motivated.*

## 1 Introduction

The learners' main challenges in nowadays society concerning learning are the limited time and the huge amount of resources and data available. Tele-teaching was introduced as one solution where people can learn independent from time and place according to their interests and learning speed. But even in those closed environments the amount of data increases rapidly due to recording technology like tele-TASK [1] becoming cheaper as well as easier and faster to use.

But still the problem of searching the data is not solved. In tele-teaching it is in fact even more present, due to the nature of the e-learning content. The lecture recordings are multimedia content, which cannot easily be indexed or searched. Manual effort is mostly required in order to create sufficient administrative metadata for finding the appropriate content. Several solutions can be thought of for this problem. First of all the automatic metadata harvesting. This can be done by using optical character recognition of the lecture video slides or audio transcription of the lecturer's speech. But then suddenly a large amount of unpriori-

tized metadata is available where the problem is to find the most essential information of the content.

A second opportunity turned up with the era of Web 2.0, the idea of which was introduced by Tim O'Reilly [2]. The idea describes that users join together to communities and participate in the creation and sharing of web and media content. The benefit is that the joint force of a lot of users is able to generate more data and provide it to the group again than individuals or administrative personnel can. Furthermore it was observable through user statistics, that fun communities made a huge impact on people and grew tremendously. This potential should be leveraged for the tele-teaching context as well. Challenges and approaches to do so will be discussed in this paper.

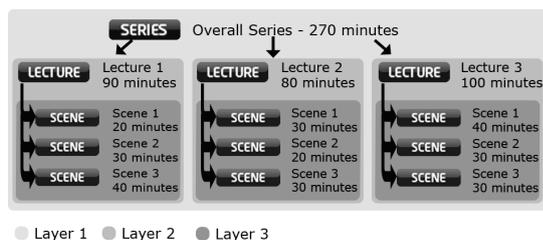
Before the utility and challenges of those social web features in the tele-teaching context will be described, some specifics of tele-teaching itself as well as the sample tele-teaching application tele-TASK will be briefly introduced in the next paragraphs.

### 1.1 Characteristics of Tele-Teaching Portals

Usually tele-teaching data has two characteristics that distinguish it from other e-learning content. First of all this is the nature of the content itself - it is multimedia content instead of single media content or even just plain text. This characteristic itself makes it more difficult to find the matching content to a search keyword, because the multimedia data itself is not searchable. Additional aids, like administrative metadata or user-generated metadata is needed to enable a proper search function, as explained in the introduction.

Second, mostly such a tele-teaching-portal consists of several layers of content, like visualized in figure 1. The primary content layer are lecture recordings that are held by lecturers. The lectures itself are mostly embedded in a larger context, for example the course which runs a whole semester, here called series. Furthermore the lectures are often subdivided into smaller pieces, called scene in this paper. This is done in order to facilitate the usage of mobile players where the content needs to be downloaded, for podcasting and also to simplify a more precise metadata collec-

tion and search.



**Figure 1. Several content layers in tele-lecturing portals**

Sample social web functionalities were implemented at the research tele-teaching portal tele-TASK<sup>1</sup> of the Hasso-Plattner-Institut (HPI). As the tele-TASK project includes a recording system as well as a portal for distributing e-lectures, some details of the project will be explained in the next paragraph.

## 1.2 Tele-Teaching with tele-TASK

The tele-Teaching Anywhere Solution Kit [1], 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. The goal of the project is the recording and distribution of lectures, seminars, reports and other presentations with as little as possible effort of material and resources.

Therefore an all-in-one solution was developed including hard- and software for lecture recording. Two video streams (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 3500 lectures and 8800 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.

A topic that is within the research focus currently is the utilization of community and social web functionalities to enhance the tele-lecturing for students. The following paragraph will briefly give an introduction into the topic of social web and motivate the combination of tele-lecturing with concepts derived from the social web.

## 1.3 Community and Social Web Functionalities in Tele-Lecturing Scenarios

Since the beginning of the Web 2.0 era numerous social web portals have evolved and grew very quickly. Their main motivation is fostered around the user participation. A number of social web and community features have been found

<sup>1</sup><http://www.tele-task.de>

to be useful to the users. These include blogging, the collaborate creation of wikis, social annotation and tagging, evaluating (eg. rating and commenting), recommending, content sharing and linking of content items [2].

That community functionalities are not only useful for networking, but also for the learning context was already found out at the beginning of the e-learning era around 2000 [3]. But only recently research focused on joining tele-lecturing with community functionalities. During the workshop *eLectures 2009* at the conference DeLFI 2009 [4] an approach of integrating tele-lecturing applications into facebook and other social e-learning approaches were shown.

## 1.4 Utility of Social Web in E-Learning

There are two points of view that should be taken into consideration when talking about the utility of social web features in e-learning in general and tele-teaching in the specific context of this paper. First of all the point of view of the single user and his benefits from the social web functionalities and second the community of the e-learning portal that is interlinked with the providers of it.

The learners themselves can benefit, when they learn easier or faster with the help of social web functionalities. The uses cannot be explained directly with the old learning theories behaviourism, cognitivism and constructivism, because those theories do not consider the technology supported side of learning [5]. The more recent learning theory connectivism [5], that is adapted to the digital age, supports the hypothesis that social web can be beneficial for learners though. This is the case, because connectivism describes learning as the creation of connections between information. Social web features support the creation of those connections through functionalities like tagging and exchange of knowledge in groups. Furthermore connectivism also describes the "cycle of knowledge development" where individuals provide knowledge to a community and may also gain knowledge from the community, which is the core of the Web 2.0 philosophy.

When the single user becomes more active, the benefits for the community can be deducted from the connectivism theory as well. Knowledge sharing and connection is one of the great benefits for the community as well. Furthermore a lot more metadata is created, which can help to improve search, recommendation and filtering functionalities within the portal. A more user friendly interface will be advantageous for the community and also the providers, as more users can be attracted. But still there are some challenges that need to be considered, which will be briefly addressed in the next paragraph.

## 1.5 Challenges of Social Web in E-Learning

At present specifically e-learning communities often suffer from several problems. First of all their member base is mostly not as large as in the huge private and free-time-oriented Web 2.0 platforms. That means that the knowl-

edge connection and sharing may only take place amongst a smaller number of people and may thus be not so effective, because the key success factor for the community features is the interaction amongst users [6]. Additionally, individuals in e-learning environments are less active than people in leisure and private communities, so the participation is generally low. A study about the Web 2.0 video service YouTube [7] and also experience with the example portal showed this.

Those two issues together lead to the main challenge of social web initiatives in e-learning: to engage users and to reach a critical mass of user-generated metadata. It is necessary to have a minimum base of the metadata that it makes sense to provide additional functionality for this metadata. In order to improve the metadata base there are therefore two options. First of all, one can try to engage the users more into the community and thus animate them to create more metadata. Second, one can try to enhance and extend the user-generated data that is already in the system with the help of automatic processes.

The following chapters will introduce community rating and tagging, their specific utility for the tele-teaching context as well as one option how the user-generated data may be enhanced and more data generated on the basis of this user-generated metadata.

## 2 Community Rating in Tele-Lecturing

In this chapter a definition of rating, an easy community feature, is given and the application of rating for the tele-lecturing context explained. Furthermore a new algorithm to calculate more precise ratings is suggested and its advantages and disadvantages balanced.

### 2.1 Introduction to Collaborative Rating

Rating is the classification of content according to order or grade it. In the context of the rating of media items, rating is the quantification of the personally perceived quality of an item. It belongs to community functionalities which originate from Web 2.0 platforms.

Rating is the user-generated enhancement to standard metadata that is easiest for the users. It is usually a small set of integers where the user chooses one of the values. The evaluation of content in this manner is therefore an easy and quick process for the user, which he might be more willing to go through than a more time intense process, like writing comments or annotations. Facilitating the engagement of users is an important issue in this context as the user participation is usually not very high. The uses of the rating functionality in the tele-teaching context is explained next.

### 2.2 Utility of Rating in Tele-Teaching Environments

Once the rating is accepted by the users, it will facilitate the information retrieval process for users in a tele-teaching

portal.

First of all a content filtering system may utilize the rating results. When a user searches amongst the content of the portal via a keyword, a large amount of content items in all layers of the content hierarchy may be found. The choice, which of the search results is the most useful for the student in that moment, will be more difficult the more results are found. Therefore it should be possible to filter the results and thereby reduce the total number of them. One filtering option is the rating. The best rated items will be shown first, worst rated items last. As shown by a study with a search engine [8], users only scan through the first few search results and only thoroughly look at the first two to five results. The best rated (and therefore most relevant) results will be on those first pages when using the filtering with the rating results.

Second, a recommendation engine can make use of rating results. When similarities between the current content item and all other items are measured it will often be the case that several relevant items for the same topic or keyword are determined. But neither is there space for a large number of suggestions on the content page nor will the users scan through large numbers of recommended similar content items. When including the rating results into the recommendation algorithm, the number of similar content items per related keyword can be reduced by only selecting the best rated item.

The idea of the rating over several layers of content items is briefly introduced in the next chapter and aspects of the implementation described.

### 2.3 Enhancing the Rating Data - Calculating the Rating Results Over Several Layers

In the tele-teaching context there are several content layers as visualized in figure 1. As all the three layers include tele-teaching content, all of them should be rateable individually. As the rating across several layers is used, a way of reliably calculating a result that reflects the ratings across the different layers needs to be thought of. The following paragraph shows a calculation that combines the ratings across several layers.

As rating uses a pre-set interval of values that the user can choose and which are used to calculate the mean afterwards, deviant values need not be considered and average calculations, like median or truncated mean, need not be used to ensure a valid result. As the rating is furthermore not a mean value that is calculated with a factor that includes a relative reference to another unit and no changing rate is required, the arithmetic mean is the mean calculation of choice for ratings. Because the rating shall be calculated across several layers, a weighting of the subset ratings is required. The weighted mean (WM) rating of a content item will be calculated by combining and weighting the means (M) of all ratings for the content item and the ratings for its connected content items of the layers underneath and above.

Equation (1) shows how the arithmetic mean of all ratings for one content item is calculated. This equation is the basis for all further calculations of the mean rating that consider a weighting.

$$M_{CSin} = \frac{\sum_{i=1}^p R_p}{p} \quad (1)$$

The calculation of the weighted mean for one layer of connected content items (for example all segments that belong to one lecture or all lectures that belong to one series) is shown in equation (2). The factor for weighting the different arithmetic means that were calculated in (1) is the length of the content items. One example: a lecture which is 30 minutes long consists of 3 segments, the first is 5, the second 10 and the third 15 minutes long. The mean rating for the longest segment should have most influence on the weighted mean calculation for the lecture and the other two have lower priority. Equation (2) calculates the combined mean of one layer of content items (as for example the before mentioned three segments) by weighting the means of the single segments with their length.

$$WM_{CLay} = \frac{\sum_{i=1}^n L_{CSin_i} \cdot M_{CSin_i}}{\sum_{i=1}^n L_{CSin_i}} \quad (2)$$

The overall calculation of the weighted mean for one content item, considering all connected layers underneath and on top, is shown in equation (3). It follows the same principles as equation (2), but it uses the means of all layers that were calculated with equation (2) and combines them to a weighted mean. The factor for weighting is also a different one now. As one content item has the same length as the sum of all connected items in the layer underneath, the length is no proper weighting factor in this case. The number of ratings is the factor that determines, which mean ratings have which prioritization. But as all the segments (which are used for podcasting) together will most certainly receive more ratings than the single lecture they belong to, the ratio of prioritizing only by number of ratings would minimize the effect of the mean rating of the single content item. Therefore the ratio of the number of ratings to the number of content items of the layer is used as weighting factor to combine the means of the different layers.

$$WM_{CSin} = \frac{\sum_{i=1}^m \frac{NoR_{CLay_i}}{NoC_{CLay_i}} \cdot WM_{CLay_i}}{\sum_{i=1}^m \frac{NoR_{CLay_i}}{NoC_{CLay_i}}} \quad (3)$$

CSin = Single content item  
CLay = All content items in one layer  
p = Number of ratings per content item  
n = Number of content items per layer  
m = Number of layers

R = Rating  
L = Length of the content item  
M = Arithmetic mean of all ratings for one content item  
WM = Weighted mean  
NoR = Number of ratings  
NoC = Number of content items in this layer

## 2.4 The Rating Algorithm Implementation

The first implementation for rating is a simple one. It focuses on the forms for generating rates by the user and provides only simple evaluation of the results. This implementation uses the average function of the underlying database (see listing 1) to calculate the score of one object. This implementation can even be used inside the search function, but it does not regard the influences between the different data layers.

```
1 votes = Vote.objects.filter(contentType =
    ctype.id, objectId = data.id).aggregate(
    average=Avg('vote'))
```

### Listing 1. Simple rating calculation

Therefore the implementation of the described calculation method was done parallel. We used a plug-in architecture as described in [9] to add the rating functionality. This allowed us to implement the second version on the base of the same database tables as the first one using the data gained so far. Both versions use the same database and can be displayed together on the page. Using the plug-in architecture we can switch the functions on and off, so it is easy to compare them and provide the better one in the live version of the portal.

For the calculation of the results of one lecture, two nested loops were required. To calculate each included lecture, the average of the containing segments has to be calculated as well. The first idea of doing this calculation each time when displaying the results would consume a big amount of calculation time.

So we were forced to think about a better way to implement it. The requirements are less time consumption and ideally the possibility to use the rating results for search result ordering. When thinking of the rating it became obvious, that saving a new rating results happens less often, than displaying a result. Therefore we decided to do the calculation of the results at saving time instead of displaying time.

For saving the results we had to create an extra database table which contained only the calculation results for every object. To reduce the number of calculations, we use these saved intermediate results for calculation as well. This results in the following workflow:

1. If the object which is rated is a segment, then the average rating result of the segment is determined. Afterwards the new rating result of the parent lecture is calculated as described in (2). At last the rating result of the corresponding series is calculated as described in (3).

- If the rated object is a lecture, the average rating results of all segments are retrieved from the database, together with the number of ratings. From the database the metadata of the segments, like the duration, are also fetched. Each rating of a segment is now weighted with the ratio of the duration of this segment and the overall duration of all segments. These values from all segments are added together and represent the average rating of all segments.

Similar to the calculation of the average of a segment, the average value of the lecture is calculated using the average function of the database. With both values the overall average rating of the lecture is calculated using the number of rates concerning the lecture and the segments.

Afterwards the average of the series is recalculated like described in (3).

- If the rated object is a series, then the average rating results of all lectures of this series are retrieved from the database, as well as the number of rates and the lecture metadata. Now the same calculation process starts, which is described in (2), using the data of the lectures like the segment data and the series data like the lecture data.

Because the data of the segment is included in the rating result of the lecture, it does not have to be calculated again.

Our calculation approach therefore brings the effect, that it minimizes the number of calculations, because saved results can be used. It also helps the usage of the rating data for ordering the search results, because the order criteria is directly inside the database and therefore available for the database request itself.

## 2.5 Evaluation of the Rating Functionality

Several advantages are expected by using the new rating algorithm. First more exact results and second the automatic calculation of ratings for superordinate content items in the different content layers and therefore more rating results are expected. This is necessary, because although the rating functionality has been online for about a year, only 613 of the about 12600 content items (about 8850 segments, 3500 lectures, 250 series) have been rated and only 43 of those have more than two ratings and can therefore be utilized for rating filtering (average ratings with less than three ratings by different people will not be taken in consideration for fairness reasons). In this paragraph both of these theses are evaluated. Advantages and disadvantages of both calculation methods are compared.

### 2.5.1 Advantages of the new rating algorithm

When comparing the number of rating results shown, the algorithm calculating the rating over several layers shows

a lot more content items than the standard algorithm. This means that, with the help of the calculation over several layers, content items that have not been rated by users so far, or not enough ratings have been collected to make up an objective results, receive a rating calculated via the new algorithm (which can be seen in figure 2). As consequence more rating results will be available as metadata base for further processing within features that use rating to improve the usability and searchability of tele-teaching portals.

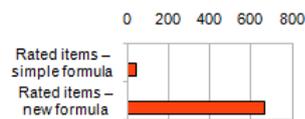


Figure 2. Number of rating results when using the new rating algorithm compared to the old

Second, the rating results will be more precise. This is the case, because a weighted mean calculation consists of more detailed ratings of the content layers underneath in connection with ratings to that specific content item. An evaluation of the current metadata base did not show a lot of difference in the results of the two rating calculations. Only 53 out of the 613 currently rated objects have differing results with a nearly equal amount of more positive and negative results (see figure 3).

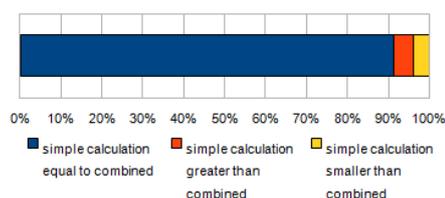


Figure 3. Rating results when using the new rating algorithm compared to the old

This may be the case because of the small interval. When calculating a mean of very deviant values a middle value will be the result and if a lot of people rate in the middle sector the same effect will occur. Therefore most of the mean ratings will be around two to four and the differences between the rating calculations might often only be marginal. Another reason is the ongoing low participation and the therefore small metadata base. There have been 809 ratings for 613 different objects. Only when more connected content items have been rated a proper evaluation on the influence of the new calculation algorithm on the real data can be started.

Until then a sample calculation can show the advantages of the method. Following ratings were inserted for a series with 3 lectures belonging to it:

content item	rating	duration
series	1,1,2,2,1,4,4,3	160
lecture 1	1,1,2,1,1	10
lecture 2	4,4,5,4,5,5,5	90
lecture 3	3,3,3,3,2,4,3,3,2,3	60

**Table 1. Sample ratings for connected content items over several layers**

In this example the first lecture was only an introduction and therefore not rated very interesting. The last lecture was rated only moderately interesting. But the second one was rated very good. Most of the people who saw the first lecture rated the series as well, later on most students only rated the lectures not the series. Therefore the mean rating of the series itself is not very good (2,25).

Calculation step	result
Simple rating mean calculation	Series = 2,25 Lecture 1 = 1,2 Lecture 2 = 4,57 Lecture 3 = 3
Mean of all lecture ratings weighted by duration	3,77
New series rating calculated over several layers	2,96

**Table 2. Sample calculation of rating over several layers for connected content items**

Looking at the weighted mean of all the lectures one will get a more positive result of 3,77. When using the calculation over several layers, the final rating result of the series combines the ratings of all lectures and calculates a weighted mean together with the standard mean of the series ratings. As result the series will now be rated more positive and the very positive ratings of the second lecture are reflected in the overall rating of the series, as can be seen in figure 2.

**2.5.2 Disadvantages of the new rating algorithm**

The major disadvantage of the new algorithm is the increased computing time. At the beginning the new calculation for the rating was performed when opening up each page. This resulted in a huge overhead as a page which included a list of for example 20 content items had to process the rating calculation 20 times and thereby consider all connected content items as well. As solution to reduce the increased computing time, the rating result for the content item and the items connected in the layers above is calculated and stored in the database when a new rating is saved.

**2.5.3 Tradeoff of advantages and disadvantages**

When storing the combined rating results for affected content items in the database each time a new rating is entered by a users, the calculation overhead for the new algorithm is minimized. Additionally the results can easily be retrieved each time a content item is shown. This bit of more computing time stands against a large number of more rating results and more precise results. Considering that the user participation for the generation of metadata will not grow exponentially very soon, the required metadata needs to be generated another way. Therefore the usage of the rating calculation over several layers is advisable.

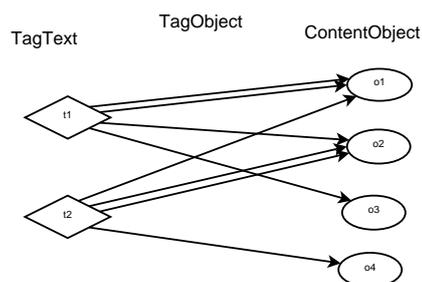
The next chapter will introduce and evaluate community tagging as second social web feature in tele-teaching applications.

**3 Community Tagging in Tele-Lecturing**

First, collaborative tagging will be introduced. Afterwards its utility for the field of application tele-teaching will be evaluated and several use cases explained. The usage of the functionality and the experiences with it will finally be evaluated.

**3.1 Introduction to 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.



**Figure 4. Relations between tags and content objects**

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 the 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  $Y \subseteq U \times T \times R$  [12]. This also means that one content object might have several tags assigned to it and one tag may be allocated to multiple content objects (see figure 4).

### 3.2 Community Tagging in a Tele-Teaching Portal

In a tele-teaching environment two 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. 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.

The next paragraph will give an introduction into the use of tagging in a tele-teaching environment. A data model, which fully supports all possible capabilities of the tag functionality, is also explained.

### 3.3 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.

#### 3.3.1 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 paradigm of Web 2.0 will eventually lead

to learners inviting fellow learners to join in and register for the tele-teaching platform 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.

#### 3.3.2 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]. 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 [13] 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].

#### 3.3.3 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. 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.

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 [14]. 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.

### 3.3.4 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 is, how this relevance can be computed. Existing approaches work the other way around and suggest appropriate tags to the content items [15]. 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 3.1). 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.

Figure 4 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.

We suggested an equation that considers the importance of each tag by using weights and thereby calculates the similarity of two content items using all its tags. [16] This calculation can be combined with similarity calculations of other metadata using the plug-in architecture [9] 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.

## 3.4 Evaluation of Tagging

Up until now nearly 600 tag assignments have been made for which nearly 300 different tags were chosen. But only 15 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 [7] 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.

Because the tagging functionality in the e-learning platform suffers from a lack of user participation, automatic means to enhance the tagging data should be thought of. The Semantic Web offers one potential to retrieve more relevant metadata from the user-generated tags. One solution will be briefly described in the next paragraph.

## 3.5 Enhancing the Tagging Data - Implementing a Semantic Extension to Community-Tags

The vision of the Semantic Web is that all the knowledge that exists throughout the whole world can be inter-linked and new knowledge can be created by reasoning over all the united data. Semantic means, that the context of a source is known through the connections that are created with the source and other sources. In order to push the spreading of the Semantic Web vision amongst researchers, initial data sets for research were required. Therefore the Linked Data initiative was founded by a joint initiative of several research groups in order to build one initial data set.

Linked Data is a technique to publish resources and their descriptions on the web. It is based on two main technologies: the resource description framework (RDF) and uniform resource identifiers (URIs). The URI is the address under which the descriptions of the resource, that is identified by that URI, can be found. The description itself is presented in the RDF format and contains information about the resource as well as links to related resources. [17]

One project within the Linked Data initiative is DBpedia. Their goal is the conversion of all content from the very popular user-generated online-encyclopaedia Wikipedia into structured knowledge. Ultimately all this data should be accessible with Semantic Web methods and already web interfaces are offered that allow to retrieve Wikipedia data in CSV, RDF or OData formats.

The idea of a semantic information retrieval module in the tele-TASK portal is based on two things. First, the knowledge, that there are about 2 million resources described in the DBpedia dataset, each of which is identified by a unique URI. Second, the observation that we do regularly have new tags that can be semantically enhanced. Once a new tag was saved in the system, the keyword will

trigger the extraction of related metadata from DBpedia. By parsing the RDF for this keyword, further information about it as well as connections to other entities in the DBpedia database can be extracted. Relations between tags can finally be derived from the connections between the DBpedia entities.

There are multiple use cases for the enhanced tagging data. First of all an alternative navigation can be suggested to users which visualizes the connections between tags and thereby animates browsing behaviour. Second, the metadata base for search algorithms can be enhanced. When a user searches for a specific tag, there is first of all the problem that the number of tags in tele-teaching environments is still very limited. When the issue of synonyms and connections between terms is taken into consideration, the user can be provided with a lot more satisfying search results. This can be realized by using the semantically enhanced tagging data. The same foundation can be used to improve the before mentioned algorithm of the detection of content similarity (see 3.3.4) which is used for recommendation systems. All in all this enhancement can improve search and filtering amongst the lecture videos.

The next section will give an overview about potentials of other community features to improve the user experience in tele-teaching environments.

## 4 Other Community Functionalities and Their Potential

We learned that there are two core tasks that determine the success of community functionalities in the tele-teaching context. First of all, the user needs to gain benefit for his learning experience and second the community and the platform itself should profit through a larger amount of metadata generated. Two more community functionalities that can fulfil those two requirements will briefly be described in this chapter.

### 4.1 Personal Video Playlists

In well-known online audio or video services, like iTunes and Youtube, users have the possibility to create personal lists of their favourite content. This very popular feature is known as playlist and it became favoured when the audio compression codec mp3 allowed to store large amounts of audio content. In the context of tele-teaching this functionality has to be altered, though. The content types from the different layers of content connected with media content, lecture and scene, which were explained in paragraph 1.1, can both be combined in playlists. In order to allow the user to do his own content clustering, the content items are subdivided into groups that have their own title and description.

The benefit for the users in terms of their learning is that they may create and manage their own content combinations. They may even subscribe to their own playlist feeds and thus download their collections on their mobile players for mobile learning. This option can be used for exam

preparation or collections of special topics, that means it enables the learner to manage his own independent learning process. [13] Profitable for the community and the portal are additional metadata for search, that is created through description and group titles. Also the newly created connections between content, that are generated by clustering and grouping content items together, can be beneficial for the recommendation process. This is the case, because those can be integrated into the similarity calculation that was explained in paragraph 3.3.4.

### 4.2 Video Annotations as Digital Manuscripts

Analogue lecture notes were widespread for centuries. The conversion of those to the digital world of lecture videos uncovers a lot of new potential though. First of all, the digital annotation is allocated to a specific point on the timeline, which enables the user to browse the video content accordingly. Also it is possible not only to use descriptive free-text metadata as annotation, but also links and other media formats, like images or audio, might serve as annotation [18].

Beneficial for the users is not only the additional metadata they can utilize, but also the process of annotation itself. Because it includes interpretation, reflection and weighting of the content [18] digital annotation leads to a deepened understanding of the topic [19]. The students may print out their collected annotation as manuscript, but also use it as basis for time-based navigation and revision. The community gains profit from the additional metadata, because time-constraint keywords are generated that facilitate to find the specific point in time where the lecturer actually talks about a topic. Also it was found out, that time-based annotation serves as anchored discussion as opposed to forums and thus encourages discussions amongst students [19].

## 5 Conclusions and Future Work

In this paper we motivated the usage of social web functionalities in the tele-teaching context. In detail we explained the utility of the two community functionalities, rating and tagging. Both features were implemented and evaluated in our sample tele-teaching application tele-TASK. We discussed sample use cases and addressed challenges that need to be considered. One of the most essential issues is the lack of participation and activity from the user side. Until a critical mass of active users and user-generated metadata is created, those community functionalities cannot be fully leveraged.

One approach to tackle this lack of user-generated metadata that was introduced in this paper is the extension of the metadata with automatic processes. For the rating application, the rating-calculation over several content layers was introduced to do so. In the tagging context, the extension of the tagging data with semantic data from the Linked

Data dataset was introduced. Of course this generated data should be used more extensively in the future. The search- and recommendation algorithms will for example benefit from this additional data.

A task for the future and the second approach to the lack of user-generated metadata is to activate more users to actively participate in the metadata creation. Facilitating the usage of the provided functionality is one way to achieve this goal. Usability tests and user studies can support this approach. Also many Web 2.0 applications give incentives to their users to promote user activity. The easiest type is the usage of ranks, like it is done in different forums. A user receives a special rank, when he writes a specific number of messages in the forum. Other incentives, like special prices or more rights, are known as well. But, user incentives may also produce useless data. Many forums have problems with user spamming the forum threads with short texts to get a higher rank. This will result in a lower quality of the data and more damage than the additional participation is worth. Therefore an evaluation needs to be done, which incentives will have a positive effect on the user activity without producing spamming.

In the future the implementation of a group functionality throughout the portal is planned as additional feature. It was shown, that learning in a group is positive for supporting the individuals motivation and eagerness to engage into academic activities [20]. Once the group functionality is available, we hope to improve the user participation by supporting more collaboration amongst the students. The more interesting and diverse features can be provided, the easier it will hopefully be to attract more users and engage them in an active learning process and an intense usage of the community features. Because only with active participation more metadata will be generated and only then those data will provide real benefits for the users, the community and also the research perspective of the tele-teaching environment.

## References

- [1] V. Schillings and C. Meinel, "Tele-TASK – tele-teaching anywhere solution kit," in *Proceedings of ACM SIGUCCS*, Providence, USA, 2002.
- [2] T. O. Reilly and O. R. Media, "What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software," *Communications & Strategies*, vol. No. 1, pp. 17–37, 2007.
- [3] R. M. Palloff and K. Pratt, *Building Learning Communities in Cyberspace: Effective Strategies for the Online Classroom*, 1st ed. Jossey-Bass, 1999.
- [4] S. Trahasch, S. Linckels, and W. Hürst, "Vorlesungsaufzeichnungen - Anwendungen, Erfahrungen und Forschungsperspektiven. Beobachtungen vom GI-Workshop 'eLectures 2009'," *i-com*, vol. 8, pp. 62–64, 2009.
- [5] G. Siemens, "Connectivism: A Learning Theory for the Digital Age," *International Journal of Instructional Technology & Distance Learning*, vol. 2, no. 1, 2005.
- [6] D. Mican and N. Tomai, "Web 2.0 and Collaborative Tagging," *2010 Fifth International Conference on Internet and Web Applications and Services*, pp. 519–524, May 2010.
- [7] M. Cha, H. Kwak, P. Rodriguez, Y.-y. Ahn, and 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.
- [8] L. Granka, H. Hembrooke, B. Pan, T. Joachims, and G. Gay, "In google We Trust: Users' Decision on Rank, Position and Relevancy," *Journal of Computer-Mediated Communication Special Issue on Search Engines 1*, vol. 12, no. 3, pp. 1–36, 2005.
- [9] M. Siebert, F. Moritz, and C. Meinel, "Establishing an Expandable Architecture for a Tele-Teaching Portal," in *2010 Ninth IEEE/ACIS International Conference on Computer and Information Science Article*. Yamagata, Japan: IEEE Computer Society, 2010.
- [10] S. Bateman, C. Brooks, G. McCalla, and P. Brusilovsky, "Applying Collaborative Tagging to E-Learning," in *Proceedings of the Workshop on Tagging and Metadata for Social Information Organization (WWW'07)*. Banff, Canada: ACM, 2007.
- [11] S. A. Golder and B. A. Huberman, "The Structure of Collaborative Tagging Systems," *Journal of Information Science*, vol. 32(2), pp. 198–208, 2005.
- [12] M. G. Noll, "Understanding and Leveraging the Social Web for Information Retrieval Dissertation," PhD thesis, Hasso-Plattner-Institut für Softwaresystemtechnik, 2010.
- [13] M. Siebert, F. Moritz, and C. Meinel, "Enriching E-Learning Meta Data with User Generated Playlists," in *5th International Conference for Internet Technology and Secured Transactions (to appear)*. London, UK: IEEE Computer Society, 2010.
- [14] M. D. O. C. Ames, S. University), and M. Y. R. B. Naaman, "Why We Tag : Motivations for Annotation in Mobile and Online Media," in *CHI 2007 Proceedings*. San Jose, CA, USA: ACM New York, USA, 2007, pp. 971–980.
- [15] D. Liu, X.-S. Hua, L. Yang, M. Wang, and H.-J. Zhang, "Tag Ranking," in *18th International World Wide Web Conference*, Madrid, Italy, 2009, pp. 351—351.
- [16] F. Moritz, M. Siebert, and C. Meinel, "Community Tagging in Tele-Teaching Environments," in *2nd International Conference on e-Education, e-Business, e-Management and E-Learning (to appear)*. Mumbai, India: IEEE Computer Society, 2011.
- [17] C. Bizer, G. Kobilarov, J. Lehmann, and Z. Ives, "DBpedia : A Nucleus for a Web of Open Data," in *6th International Semantic Web Conference (ISWC 2007)*, Busan, Korea, 2007.
- [18] C. Hofmann, N. Hollender, D. W. Fellner, I. Graphics, S. Group, and T. U. Darmstadt, *Workflow-Based Architecture for Collaborative Video Annotation*, 5621st ed. Heidelberg: Springer Berlin / Heidelberg, 2009, pp. 33–42.
- [19] B. Zupancic, "Vorlesungsaufzeichnungen und digitale Annotationen Einsatz und Nutzen in der Lehre," Dissertation, Albert-Ludwigs-Universität Freiburg, 2006.
- [20] H.-J. So and T. A. Brush, "Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors," *Computers & Education*, vol. 51, no. 1, 2008.