

Analysis of User Behavior on Multilingual Tagging of Learning Resources

Riina Vuorikari¹, Xavier Ochoa², and Erik Duval¹

¹ Dept. Computerwetenschappen,
Katholieke Universiteit Leuven,
Celestijnenlaan 200A, B-3001, Heverlee, Belgium
{Riina.Vuorikari,[Erik.Duval](mailto:Erik.Duval@cs.kuleuven.be)}@cs.kuleuven.be?

² Information Technology Center, Escuela Superior Politecnica del Litoral,
Vía Perimetral Km. 30.5, Guayaquil – Ecuador
xavier@cti.espol.edu.ec

Abstract. Although social, collaborative classification through tagging has been the focus of recent research, the effect of multilingual tags is often overlooked. This work presents an early exploratory study of the production and consumption of multilingual tags in a European educational K-12 context. The data, produced by teachers bookmarking and tagging learning resources during three month period, was analysed. Thereafter, a focus group of teachers evaluated a sample of learning resources with metadata records containing both thesaurus terms and multilingual tags. The results of this early study suggest that some tags are found as useful as thesaurus terms and that users are divided about the benefits of multilingual tags. As some tags are useful for some users, “hiding all but the right tags” becomes crucial for the success of a multilingual collaborative tagging system.

Keywords: Collaborative tagging, multilinguality, learning resources.

1 Introduction

The use of social, collaborative classification systems has gone through a continuous growth in the latest years [1]. An example of this is a multitude of sites that provide some type of social annotation of digital artefacts and a social navigation system (Flickr, del.icio.us , CiteULike, Last.fm, among others). Social tagging, i.e. allowing individuals to apply free text keywords to digital objects, potentially offers advantages in terms of personal knowledge management, serendipitous access to objects through tags, and enhanced possibilities to share content with emerging social networks.

Several studies have been undertaken to better understand the behaviour and evolution of social tagging systems. Early research has been conducted by Mathes [2] where the term “folksonomy” is used to compare the emerging socially generated vocabulary with the more formal ontology concept. Golder and Huberman [3] first

looked at user patterns of collaborative tagging systems. Recent studies focus on the navigability of such social systems [4] and on understanding the network properties [5].

A prevailing aspect among current studies concerning tagging is that they assume that tags are represented in a common language [6], understandable by all the members of the user community. Guy suggests that it is not always the case [7], but does not offer insight on how to deal with tags in multiple languages.

Lately, multilingual tags have started emerging on popular social tagging systems as their user-base grows on the international level. Roughly, two different ways to deal with multiple languages can be observed; ones taken care of by users (i.e. crowd-sourcing¹) and others where the system supports multiple languages to certain extent.

Examples of crowd-sourcing include del.icio.us and LibraryThing. It can be observed that del.icio.us users add tags in different languages (e.g. achat, shopping), and even in some occasions, add language identification of the source language as a tag (e.g. lang:fi). There is no system level support that allows users to see tags, say, only in French or Finnish. In LibraryThing experienced users can combine tags under one tag. In some occasions, tags in different languages have been grouped together. On the other hand, there is Yahoo!'s MyWeb, that offers tags and tag clouds in different languages in their localised parts of the portal (e.g. .fr, .es, ...), thus some language identification of tags takes place on the system level.

Our work, still at its early stage, attempts to shed light on a community of users who shares a common educational interest to use a social tagging system across country and language borders, but does not necessarily share one common language, as the users are free to choose the language(s) in which they apply tags. This exploration takes place in the context of two European Community founded projects, CALIBRATE¹ and MELT², both focusing on sharing and re-using of digital learning resources for K-12.

European education, especially that of K-12 education, is inherently multilingual and multicultural. Offering educational resources and services in native languages is deemed important, but equally important is the exposure to other languages. One way to promote this is to make learning resources available across national and linguistic borders. This puts constraints on semantic interoperability, i.e. how well content and its metadata can be understood by other systems and users.

Controlled vocabularies, such as multilingual LRE Thesaurus³, can be used to overcome some hurdles of semantic interoperability. However, the gap between the terms used by experts and practitioners in the field is also problematic. For that reason, the current research looks into co-existence of taxonomies and end-user generated tags, which is also the case of the MELT project.

A federation of learning resource repositories in a multilingual context needs to support multiple languages at the system level in order to support each repository and its national user-base, but at the same time, there is a need to allow people (i.e. user

¹ <http://calibrate.eun.org>

² <http://info.melt-project.eu/>

³ http://insight.eun.org/ww/en/pub/insight/interoperability/learning_resource_exchange/metadata.htm

information and preferences), resources and tags to “travel” across national and linguistic borders.

This paper is structured as follows: first, in section 2, we analyse the early stage of the bookmarking and tagging behavior of our community in order to better understand how teachers bookmark and tag resources in a multilingual context; what types of tags are provided and in which languages. Then, in the section 3, an experiment is presented that measures the effect of multilingual tags on the descriptiveness, usefulness and overall quality of the metadata. Finally, the findings are discussed and applied to design decisions for multilingual tagging systems.

2 Analysis of tagging behaviour in multilingual context

The CALIBRATE project makes K-12 digital learning resources available to its pilot schools (78 schools) in Hungary, Austria, Estonia, Czech Republic, Lithuania and Poland in their different curriculum areas. Schools can access material in different languages through a portal that is connected to a federation of learning resource repositories [8] in the pilot countries.

As part of the project's multilingual search interface⁴, a personal bookmarking and tagging tool has been available since the beginning of 2007. This tool allows a user to create personal collections of learning resources by bookmarking interesting resources found through the portal. To facilitate the management of these personal collections (also called favourites in the project), the user can also add keywords to resources to make it easier to “keep found things found”. These keywords are free for the user to choose and can be expressed in any language(s). The collections and keywords are kept private to the user, and at this stage of the experiment, they cannot be shared among users.

The data for this analysis is from a period of about three months (January 24 to April 21 2007). There were 77 teachers who made 459 bookmarks with 417 multilingual tags on 320 different learning resources. It is intended to have regular analysis of this data within the projects lifespan (-2008).

2.1 Quasi-Experimental Set-up

A total of 173 subjects used the portal during the time of the experiment, however, the subjects of this dataset comprise a group of 77 teachers. These teachers had done at least one bookmark, thus, it was a self-selected group formed based on the bookmarking behavior. The group represents 45% of all the pilot participants. As there was no overall methodology to introduce bookmarking and tagging to the subjects, more than half of the participants had not shown interest in using this feature of the portal.

The bookmarking habits, at this very early stage, varied a lot in terms of what languages to use, how many tags to add, how to add multiple tags (with comma separated or without commas), etc. Also, hardly any of the participants had previous

⁴ <http://calibrate.eun.org/merlin/>

experience on tagging, so not one single tagging convention emerged, rather many different ways to use tags in multiple languages. As there is very little research done on the multilingual context, we think it is important to study the early stage of tagging behaviour to better anticipate the effect of multilinguality on the system to improve its design.

It is noteworthy to mention that the bookmarking and tagging system, at this stage of the pilot, offers very little social features or support. Oftentimes in social bookmarking sites, social cues are made available (e.g. most bookmarked items, tag clouds, tags are recommended based on previous tags, etc). At the time of the experiment, the system had hardly any tags attached to resources, and in the case where a resource already had related tags, they were only shown to users who used the interface in the same language (e.g., a user using the interface in Polish will only see tags in Polish, if any are available. No tags in any other languages are exposed).

2.2 Results

In the part we present the results of the analysis, which will be discussed further in conjunction with the other results in the discussion section.

When we look at the distribution of bookmarks per users, we can find that on the average, each user had 6 bookmarks (Fig.1). However, the distribution was very wide; 10% of the users had more than the average amount of bookmarks, which leaves 90% under the average. Eight of the users could be called “super users”, as they had more than 20 bookmarks, and 12 users had between 20 and 6 bookmarks. About 30% of the users seem to have only experimented with the bookmarking system, as they only have one single bookmarked item in their favorites folder.

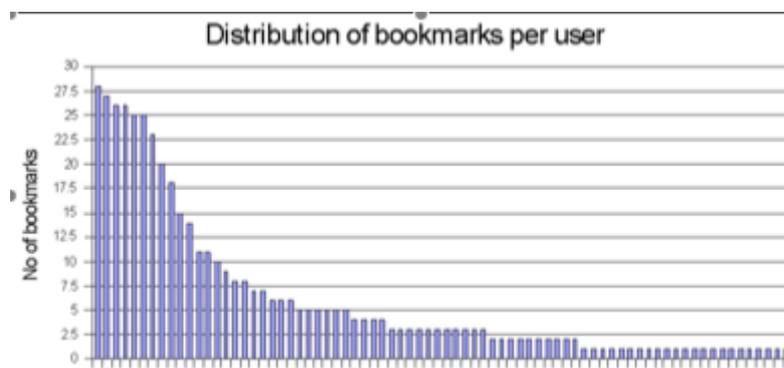


Fig. 1: Distribution of bookmarks on the CALIBRATE portal

We had recorded 418 tags in the system. During the semantic analysis of tags we found that many tags actually contained multiple terms, i.e. they were bundles of terms without comma separation. This was due to a technical feature of the tool that treated terms without comma separation as one tag. When broken down, they resulted in 585 terms. One third of these tags were in Hungarian, another third in German and

Polish and 26% in English, even though none of the users were native English speakers.

Tags were translated into English and a semantic analysis was performed to better understand the types of tags. We used the classification from Sen [9] that is also based on the categories of Golder et al. [3], which are Factual tags (Golder: item topics, kinds of item, category refinements); Subjective tags (Golder: item qualities) and Personal tags (Golder: item ownership, self-reference, tasks organisation)

The vast majority of the tags at this early stage (Table 1) are of the factual type. From the factual tags, 79% were put into a rough category of topic and 14% of the category refinement with richer information. The rest of the tags were subjective in their nature and could be used to describe the quality of the resources or how the person felt about them. None of the tags fell into the category of personal tags as Golder describes them (e.g. tags related to item ownership, self-reference or personal tasks organisation). When we analysed how these tags were used and re-used among users, we found that 80% of tags related to bookmarks were factual and 20% of tags subjective tags. In a MovieLens study [9], for comparison, the distribution was 63% factual, 29% subjective, 3% personal and 5% other.

Table 1. Types analysis of each tags (no re-use)

| | | |
|---------------------|-----|-----|
| Factual | 340 | 93% |
| Topic | 288 | 79% |
| Category refinement | 52 | 14% |
| Subjective | 24 | 7% |
| Personal | 0 | 0% |

After categorising the tags, we further studied their nature. Some trends seemed to emerge. First, about 13% of tags contain a general term, a name, place, e.g. EU, Euroopa, Euroopa, Europa, europe, geograafia, Pythagoras, etc . We hypothesise that this type of “travel well” tags, even if not translated, could be found useful for other users for their close similarity in spelling in many languages. We think it could be of interest to test automatic filtering methods (e.g., matching them against existing multilingual vocabulary lists available on the Internet) for this type of terms to single them out from the main pool of all multilingual tags. This would help presenting them to users, instead of exposing them to all available tags.

When we looked at the number of tags that users related to bookmarked resources, we found that some of the tags were re-used. There was an average of 1.92 tags/resource. More than half (56%) of the tags were entered as a bundle of terms, i.e. most teachers had added 2 to 6 terms without a comma separation. In quite a few cases these terms were comprised of the terms in the title of the resource (Fig.2). In 28% of the cases only one term was entered as one tag.

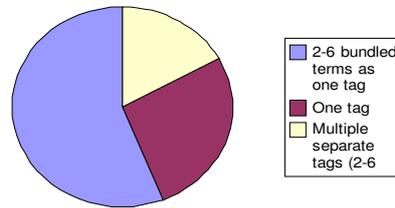


Fig. 2: One vs. multiple tags per resource

The rest had used multiple separate tags (2-6 tags). In the latter case the terms were not necessarily related to the title alone, but carried other types of information (e.g. title: Umweltkids and tags: Oekologie, Artenschutz, Regenwald, Tierschutz, Skisport).

Contrary to our expectations, the users took liberties to add tags in multiple languages and to use the portal interface in different languages than that of their mother tongue (interface was made available in the languages of the pilot and in English). This made the identification of the language of tags more difficult, as we had expected to be able to identify the language of the tag from the language of the interface that the user used when inserting the tag. In about 70% of the cases we were able to identify the language of the tag correctly using this method, which leaves us with a 30% error rate on language identification. This error in identifying the language of the tag correctly would make it hard, for example, to display tags and tag clouds in one single language, an issue that is related to the usability of the portal, and the one of which the second experience was set up to find more evidence.

We found the following scenarios for tagging, however, due to our logging, we can't give percentages for these use cases:

- Interface and tags in mother tongue
- Interface was used in mother tongue, but tags in other language
- Interface was used in a language that is other than the mother tongue, but tags were entered in mother tongue
- The tagging language was other than the interface language and the mother tongue

These scenarios were found through comparing the real language of the tags to that of identified language by using the interface language. In this early stage of the experiment it is impossible to draw firm conclusions, but it seems that users are likely to use, or at least try, the interface in different languages. We found, for example, that the tags entered through the English interface were in English only in 50% of the cases, which means that users added tags in languages within their areas of competences. On the other hand, we also found that there were many more tags in English than we expected from the choice of the interface language and mother tongue of the participants. These users had chosen to tag in English, even if they used the interface in some other language, most likely to be able to share tags with users from other countries.

3 Experiment with Multilingual Tags

An exploratory experiment was set up in order to measure the perceived descriptiveness, usefulness and quality [10] of traditional metadata, expert classification keywords and multilingual tags. We were also interested in how users reacted when they were confronted with tags in multiple languages that they did not have knowledge of. The experiment subjects were shown a list of learning resources metadata with keywords in multiple languages. This list was like a sample of the search result list of the portal. The results of this experiment will be useful to guide design decision in the development of retrieval tools for learning objects in a multilingual environment.

3.1 Experimental Set-up

Thirteen teachers, who belong to the MELT focus group, were selected to participate in the experiment. They were confronted with metadata regarding five learning resources in different areas of primary and secondary education curricula, namely in health education, social science, physics, mathematics and biology. An online form was used for the experiment⁵.

Each learning resource had a metadata description, but the number of elements varied. However, they all had the following metadata: title, description, age range (all in English) and keywords. The keywords were comprised of multilingual tags and thesaurus terms. They were mixed together and displayed in an alphabetical order to participants.

The number of thesaurus terms and tags varied for resources. Twenty of these keywords were English classification terms from a thesaurus by an expert. The rest (39) were multilingual tags provided by pilot teachers during the three first months of the CALIBRATE pilot. These tags were both in commonly used languages and in less used languages as listed below:

- 11 in Hungarian
- 7 in German
- 7 in English
- 6 in Polish
- 4 in Estonian
- 1 in Finnish

The participants were asked to look at each learning resource at the time and go through the metadata related to it. Then, they were exposed to two different task related questions: first, to select the keywords that they found helped them to learn about the resource (i.e. descriptiveness), and secondly, a question related to decision support (i.e. help using the learning resources in teaching) was asked. Finally, they were also asked to rate the perceived overall quality of all the metadata displayed (traditional metadata plus keywords). This procedure was repeated for each one of the five learning resources.

⁵ <http://www.zoomerang.com/survey.zgi?p=WEB226K9M7EHWB>

Once the review of all the resources was concluded, the users were asked to identify their language competencies, and to indicate their comfort level when keywords were presented in languages that they did not understand. All these questions were mandatory to answer. The subjects commented later that in some cases they did not feel that any of the keywords was descriptive or useful, but they had to choose one to conclude the web-survey. This might have skewed the results to some extent. Finally, participants had a choice to leave free comments about their experience during the experiment.

3.2 Results

In this part we present the results of the experiment, which will be discussed further in the following section.

On average, only 35% of the presented keywords, both thesaurus and tags, were found descriptive for the learning resource. The thesaurus terms were found descriptive in 58% of the cases, while the tags in 25% (Fig.3). When we look at the two top terms for each resource, we find that thesaurus terms were somewhat more popular (60%) than tags (40%) (Table.2). All but one of the most popular tags were in English, which was also the most spoken language among the focus group. There were a lot of variations, by resource and by language groups, on how users perceived the keywords.

For example, for the first resource in the Fig.3, there was only one thesaurus term and nine tags, which were in English and German, the languages widely spoken by participants. In this case two of the tags were chosen almost as often as the thesaurus term. As for the second resources in Fig3, there was almost equal amount of tags and thesaurus terms; two tags, both generic terms (EU, Europa) were chosen more often than thesaurus terms.

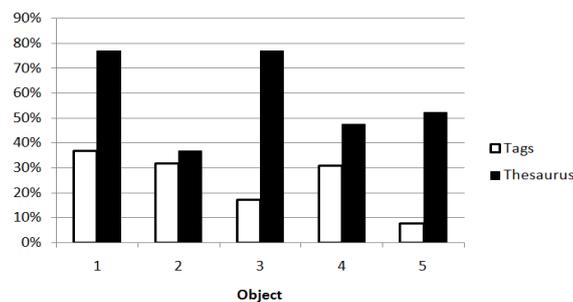


Fig. 3. Percentage of tags and thesaurus terms found descriptive

The no:3 in the same figure represents a case of multilingual tags in less spoken languages in which the users did not have competences. In this case two thesaurus terms were most chosen, however, two “travel well” tags (JavaApplets, Applets) were very high on the list. As for the resource no:4, there was an equal number of tags and thesaurus terms which was also displayed in the results, top two positions were held

by both. In the last case the tags were in less spoken languages, in Hungarian and Estonian; one Hungarian tag was found useful by all with Hungarian skills.

| Type | Lang | Keyword | % of votes |
|-----------|------|------------------|------------|
| Tag | En | healthy meal | 100% |
| Thesaurus | En | health education | 77% |
| Tag | En | EU | 77% |
| Tag | Pl | Europa | 62% |
| Thesaurus | En | physics | 85% |
| Thesaurus | En | mathematics | 92% |
| Thesaurus | En | mathematics | 92% |
| Tag | En | GeoGebra-program | 62% |
| Thesaurus | En | Geography | 92% |
| Thesaurus | En | Island | 69% |

Table 2. The two most popular keywords for each resource

From the total number of keywords, 54% were in a language within users competencies; however 87% of the keywords found descriptive were in a language that the user had skills in (Fig. 4). The remaining 13% of tags that were found useful, but not in the languages that users had competences, seem to comprise of terms of the generic type, the “travel well” tags, as described previously.

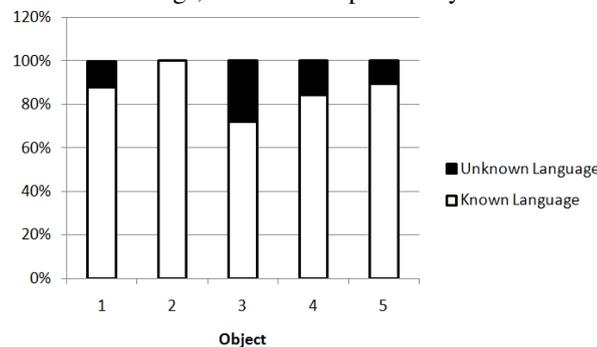


Fig. 4. Percentage of keywords in a known and unknown language that were found descriptive

When we asked about how well the keywords would help to use the resource, in average, only 27% of the presented keywords were found useful to indicate possible uses of the learning resource. In this case, the thesaurus terms were found useful 50% of the time, against only 18% for tags. When we look at the languages in which the participants had skills in, we find that in 83% of the time they mark those terms useful, again leaving more about 15% tags in other languages found useful.

We can say that the issue of multilingual tags evokes sentiments and also splits users. From the thirteen users, two “love” being able to see multilingual tags and four

found them useful, whereas six found them confusing and one “hates” to see keywords in languages that he/she does not understand (Fig. 5).

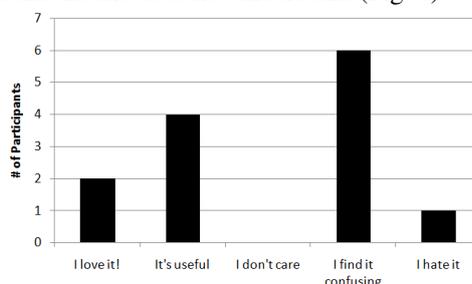


Fig. 5. Answer of the participants to the question: “What do you think when you see the keywords in many languages?”

Lastly, we were also interested in how users evaluated the overall quality of the metadata record [10]. The quality assigned to the metadata record correlate in a statistical significant way with the amount of words in the description (.909) and with how descriptive (.944) and useful (.994) the user found the keywords for that learning object. The first correlation was already found in a previous study [10].

4 Discussion on the results

The main argument that comes out of this early experimental research is that certain multilingual tags seem to be useful for some users – the challenge is how to make the other tags invisible? Moreover, the results can lead us to discuss the multilingualism of tags and indexing keywords from different perspectives; what are the user needs and requirements in a multilingual Europe, how can they be supported at the system level, what are the ramifications on the design and usability, and how is the overall quality of the portal enhanced through multilingual tags?

In the spirit of “how to hide all but the right tags for each user” this research has identified two topics that need further investigation: one is that of identifying “travel well” tags and the other that of how to correctly identify the language of each entered tag. After tackling these two issues, hiding all but the right tags becomes a much more manageable task.

Solving those two issues would greatly enhance the usability of the portal that offers multilingual tags: as shown in the experiment with the focus group, being exposed to tags in many languages has a dividing effect. One half of the subject expressed that they liked to see multilingual tags, whereas the other half found them rather irritating, especially when they were in languages that they did not recognise. It was also mentioned that multilingual tags make it harder and slower to pick the useful terms out of all the tags.

Two possible ways to further advance the cause could be envisaged: to automate the recognition of “travel well” tags and the identification of languages of all tags by using already existing vocabulary and dictionary lists on the Internet, or by crowd-

sourcing” it to users. The latter involves asking the end-users to identify “travel well” tags, and allow them to translate and correct the language of tags. A co-existence of both could also be envisaged.

Another interesting outcome of the study is that keywords in general received a rather low appreciation rate among the subjects: 35% of the keywords were found descriptive and 27% were found helpful to the use of the resource. Overall, the thesaurus terms performed better than the tags, however, it can be argued that tags, after all being produced with no outlay, showed an overall encouraging and potential gain in overall usefulness. Thus, social tagging could add value to keywords in terms of sharing the accumulated knowledge of actual use of resources in teaching and learning. More design level effort is needed in guiding and encouraging users in using tags for such purposes.

5 Conclusions

This early study contributes to the understanding of tags in multiple languages. Despite the small sample size and early tagging behaviour of the participants, we can assume that tags in a multi-cultural and lingual context offer potential advantages to the collaborative tagging system and its multilingual user communities (e.g. Europe). However, there are challenges and research questions that need further attention. As it becomes clear that some tags are useful for some users, the design challenge becomes “hiding all but the right tags”. This implies for both entering and viewing the tags, e.g. what tags and in what languages to show/recommend to users when they are about to add a tag and what kind of tags to show for retrieval and social navigation.

First, it seems important that the system has a capacity to infer and identify tags entered in multiple languages, so that users can be shown or exposed to tags only in languages that they desire. Second, it appears that there are tags that “travel well”, i.e. tags that are easily understood by many users despite the lingual barriers. It appears important that those terms are identified, either automatically or by users, so that they could be better taken advantages of. The two above findings seem to further indicate that tags in different languages should not be kept as separate silos, but interaction between languages should be used for connecting like-minded people across country and linguistic borders.

The issue of multilingual tags is intriguing and offers interesting possibilities for both the learning resources repository managers and administrators, as well as for end users. In a multilingual environment such as Europe, where making learning resources available in languages others than in mother tongue is becoming more mainstream, mixing tagging with top-down expert classification system seem to offer interesting possibilities for accessing resources and for other novel educational applications that leverage the social network aspects of a given community. From this early experiment it becomes clear that further research into the topic of multilingualism is needed to better understand its complexity, but also to be able to design more adaptable applications.

Acknowledgments. We would like to thank Sylvia Hartinger from European Schoolet for making the tags available for analysis and Jim Ayre from Multimedia Ventures Europe Ltd. for valuable comments. Acknowledgment also goes to Helsingin Sanomain 100-vuotissäätiö for the research grant that made this research possible.

References

1. Marlow, C., Naaman, M., Boyd, D., Davis, M.: Position paper, tagging, taxonomy, flickr, article, toread. In: Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland. (2006).
2. Mathes, A.: Folksonomies-cooperative classification and communication through shared metadata. In: Computer Mediated Communication, Graduate School of Library and Information Science, University of Illinois Urbana-Champaign. (2004)
3. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. *Journal of Information Science* 32(2), pp. 198—208. (2006).
4. Chi, E., Mytkowicz, T.: Understanding navigability of social tagging systems. In: Proceedings of CHI. Volume 7. (2007)
5. Catutto, C., Schmitz, C., Baldassarri, B., Servedio, V.D.P., Loreto, V., Hotho, A., Grahl, M., Stumme, G. Network Properties of Folksonomies. *AI Communications Journal*, Special Issue on "Network Analysis in Natural Sciences and Engineering", 2007.
6. Hammond, T., Hannay, T., Lund, B., Scott, J.: Social bookmarking tools (i). *D-Lib Magazine* 11(4) (2005)
7. Guy, M., Tonkin, E.: Tidying up tags. *D-Lib Magazine* 12(1) (2006)
8. J.-N. Colin and D. Massart. LIMBS: Open source, open standards, and open content to foster learning resource exchanges. In Kinshuk, R. Koper, P. Kommers, P. Kirschner, D. Sampson, and W. Didderen, editors, Proc. of The Sixth IEEE International Conference on Advanced Learning Technologies, ICALT'06, pp. 682-686, Kerkrade, The Netherlands, July 2006.
9. Sen, S., Lam, S.K., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F.M., Riedl, J.: tagging, communities, vocabulary, evolution. In: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work. pp. 181–190 (2006)
10. Ochoa, X., Duval, E.: Towards automatic evaluation of learning object metadata quality. In: Advances in Conceptual Modeling - Theory and Practice, ER 2006 Workshops BP-UML, CoMoGIS, COSS, ECDM, OIS, QoIS, SemWAT. pp. 372–381. Lecture Notes in Computer Science, Tucson, AZ, USA, Springer (November 2006)