

Exploring the Interaction between Users and Tagging Systems on the Web

Chuttur M Yasser

Indiana University, Bloomington, USA

mchuttur@indiana.edu

Abstract

With tagging becoming an increasingly important social phenomenon, some researchers view the tags supplied by users as a useful tool for helping individuals find relevant information on the web. Other researchers, however, still question the utility of tags in information retrieval systems, contending that tags are only relevant for the individual user supplying the tags. In this paper, we argue that tags are, in fact, the result of an interaction between the user, the tagging system, and the resource being tagged. Using Web Content Analysis, we analyze 33 websites that support tagging and identify the structural features (interfaces) that allow a user to tag a resource. Findings show that depending on the features supported by tagging systems, tags may either represent the information content of a resource as viewed by individual users only or they may be the result of a collaborative effort from a group of users.

Keywords: *World wide web, Tagging system, Web Content Analysis*

1. Introduction

Finding relevant information on the web is an ongoing problem. Search engines like Google rely on sophisticated algorithms to index huge collection of web pages to make them accessible to user queries. Users, however, are still frequently overloaded with irrelevant results. Web directories like Yahoo Directory¹ partially solve the problem of irrelevancy by categorizing information and allowing users to filter undesired information by selecting only those categories that address their needs. For example, by having different categories for information about *Paris* (see Figure 1), Yahoo Directory makes it easy for an individual to select the determine where to look for information that will meet her needs. Directories usually employ indexers who manually visit the web and collect information that they classify under different categories. The process is slow, highly human intensive, and cannot cover the majority of pages that are actually available on the web. Indexers and users may not agree on the same vocabulary to use for representing the same information such that users often find themselves alienated from the language used by indexers [2] [5]. In consequence, users may have different expectations regarding where certain information should be categorized and this can make it even more difficult for users to find relevant information.

¹ <http://dir.yahoo.com/>

[France > Paris > Local Travel Guides](#)
[France > Paris > Lodging > Reservation Services](#)
[France > Paris > Vacation Rentals](#)
[France > Paris > Hotels](#)
[France > Paris > Hotel Directories](#)
[France > Paris > Magazines](#)
[France > Paris > Shopping > Travel](#)

Fig.1.: Snapshot of Categories Available in Yahoo Directory for Information on Paris.

Jacob [8] and Olson [12] have pointed out the short comings of rigid structures such as classification schemes created by professional indexers. The authors argue that a knowledge domain, through language evolution, may undergo changes and so a more universal approach to the representation of information is needed. They have also highlighted the fact that a knowledge domain may also undergo changes in the ways in which related queries are expressed, implying the need for a dynamic approach to knowledge representation.

Tagging offers an inexpensive, scalable, dynamic, and fast alternative to traditional methods of indexing information available on the web: By using descriptive labels, known as tags, users can easily provide an indication of the information content of a web resource (e.g., web page, image, video, music) without the intervention of professional indexers. This process is particularly useful for non-textual materials available over the web like images and audio files, which cannot be easily processed by search engines as the underlying algorithms require the presence of textual description. While the index terms used by search engines view the content of web pages from a system's point of view, tags represent the information content as interpreted by users. As pointed out by Voß, there is increasing use of tagging systems to annotate web resources using tags [14]. Some researchers, view tagging as useful for information retrieval, but others still question the utility of tags assigned to resources.

Heyman, Koutrika, and Garcia-Molina, for instance, contended that tags can effectively improve web searching results [7]. As Golder and Huberman observed, users have a tendency to reach agreement on the tags used to describe a given resource [3]. Lin, Beaudoin, and Desai concluded that, with time, tags converge on a few popular tags that are used to describe similar or same resources [9]. Munk and Mork also showed that most users have a tendency to use similar tags to describe similar or same resources [10]. By extension, the results from these studies suggest that tags offer general consensus about the content of a resource as viewed by users in their own language and would therefore be useful in information retrieval systems: for example, if most users agree on a set of tags to represent a given resource, then the same set of tags would be very appropriate for matching user queries and discovering the tagged resource.

But sometimes tags do not accurately reflect the content of web resources. Some tags are personal and have no meaning to a wider community. They also claim that users lack sufficient training and are not skilled to describe resources correctly. Peterson [13] explained that tags allow both true and false statements to co-exist. She contends that although a group of users may agree on some tags assigned to a web resource, another group may disagree on those same tags. Golder and Huberman [4] raise the question of specificity when tagging

resources. They note that users may use tags that are only useful for a specific group of individuals. Noll and Meinel pointed out that some users may tag resources only for their own use, whereas other users may understand the implications of tags to other users and therefore would use tags that are more useful for a wider community [11].

Tagging is in fact system mediated as shown in Figure 2. Users are exposed to resources, e.g., a web page, and they assign tags to the resource through the interface provided by a system. In this paper, we are interested in the influence that the system can have on the tagging behavior of users. Given that researchers criticize the fact that some users tag resources only for their own use, hence limiting the utility of tags in information retrieval systems as claimed by other researchers, we investigate the following research question.

RQ1: Do all tagging systems offer the same features to users?

RQ2: How do the features provided by tagging system influence the tagging behavior of users?

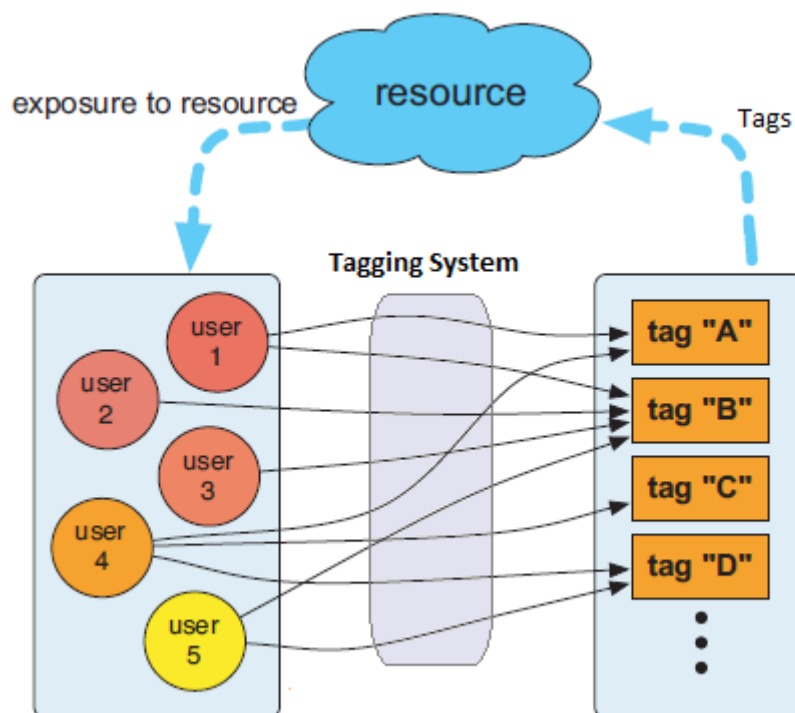


Fig. 2: Schematic Description of the Process of Tagging
(Adapted from Cattuto, Loreto, Pietroneno, 2007, p. 1464 [1])

2. Methodology²

An initial list of 210 tagging systems was created after visiting blogs, news posts, search engine results and Wikipedia entries during the period from March 2008 to May 2008. By using a random generator, 33 from the 210 identified tagging systems were shortlisted for

² It should be noted that due to the very dynamic nature of the Web, some links referred to in subsequent sections may not be active anymore, may be broken or may have completely changed their contents.

further analysis. To obtain a list of features that assist users in assigning tags to resources, each of the 33 systems was manually analyzed using web content analysis (see [6]). Every feature identified was recorded into a spreadsheet. To obtain the distribution of features across tagging systems, the number of times a feature appeared in the set of 33 systems was counted and this process was repeated for all the features that have been identified.

3. Findings

Table 1 lists the features that were observed during the analysis of the 33 tagging sites. These features allowed users to add tags to resources and thus acted as the interface between users and the system.

Table 1: Features of Tagging Sites

| Feature | Description |
|--------------------|---|
| Tagging Permission | This feature place constraints on who has permission to tag which resource. |
| Tag Visibility | This feature determines whether a user's tag is made visible to other users. |
| Tag Recommendation | This feature may provide suggestion on the tag to use for a given resource. |
| Tag Duplication | This feature controls whether duplicate tags are allowed for the same resource. |

3.1 Tagging Permission

Different sites offer different tagging permission to its users. Three categories of permission setting observed are: own, peer, and anyone. In the 'own' tagging permission category, the site allows only the user who uploads the resource to the site to tag the actual resource. Most sites fall under this category (see Figure 3). This is typical of social bookmarking sites in which users who provide the link to a resource (e.g., web page) are also allowed to add tags to describe the contents pointed by the link. Examples of web sites in this category are <http://www.delicious.com/>, <http://www.jumptags.com/>, and <http://slashdot.org/>.

In the 'peer' tagging permission category, users are allowed to choose who they want to tag the resources they upload. In our sample, only 1 site offered such feature: the popular photo sharing website Flickr (<http://www.flickr.com/>) allows users to upload and tag their own resources and it also gives them the ability to extend permission to tag the same resources to other users.

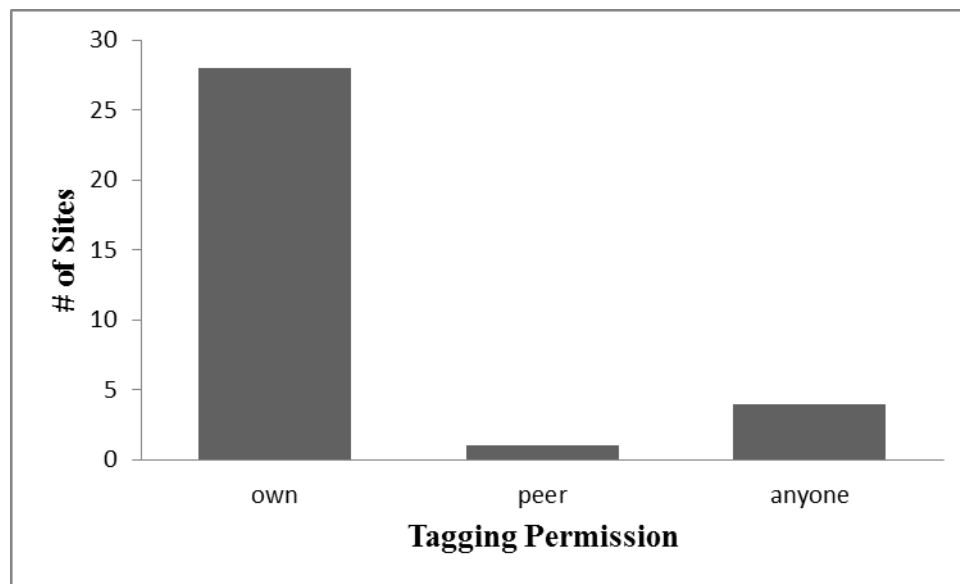


Fig. 3. Distribution of Tagging Permission Feature Across Sample Data (N=33)

In the 'anyone' tagging permission category, web sites allow anybody who visit their sites to add tags to any resources. More sites allowed this feature compared to sites falling in the peer permission category. Examples of websites falling under this category are <http://odeo.com>, <http://www.stylehive.com/>, and <http://imeem.com>.

3.2 Tag Visibility

Tags that have already been assigned to resources may be visible when users are tagging similar or same resources. Figure 4 shows that most tagging sites do not make tags visible and thus prevent users from knowing which tags have been already been assigned to a given resource. Fewer sites allowed users to see the tags previously added to existing or similar resources.

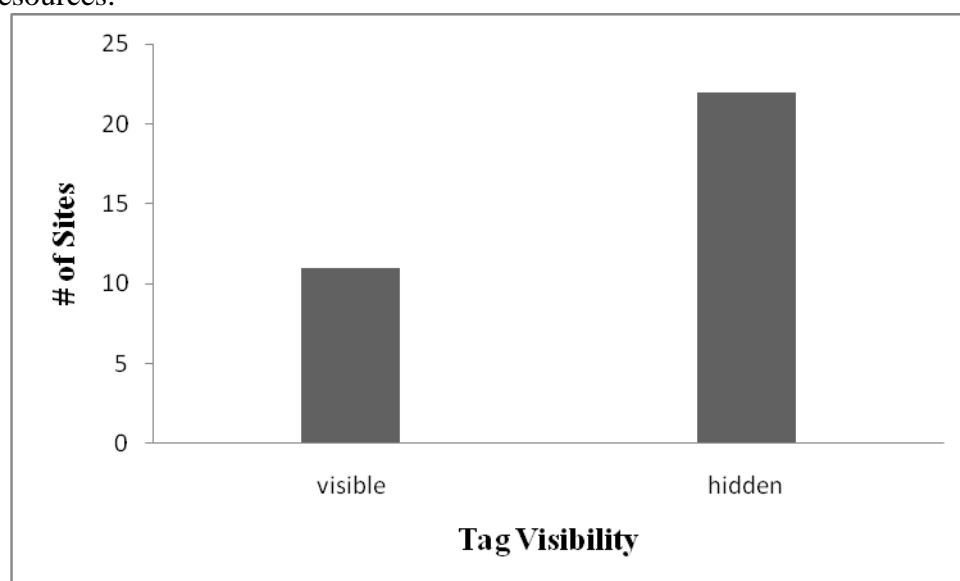


Fig. 4. Distribution of Tagging Visibility Feature Across Sample Data (N=33)

The social bookmarking site <http://www.stumbleupon.com/> allow users to provide links for websites considered interesting but will not let users see if any tags have been submitted to the same or similar sites. The web site <http://www.delicious.com/>, in contrast, lets users bookmark a link that has already been recommended by other users and it also shows all the tags that have been assigned to the same link.

3.3 Tag Recommendation

Some sites suggest tags that users can use for a given resource. The websites www.stylehive.com , www.delicious.com and www.clipmarks.com are examples of sites that offer this feature to their users. However, there is no standardization regarding the suggestion mechanism used by the different sites. Some sites would suggest:

- the user's own set of previously used tag;
- a dictionary-like alphabetical list;
- keywords that have been pulled from the resource being tagged; or
- a subset of most popular tags assigned by other users for same resource.

As shown in Figure 5, fewer sites provide tag suggestions to users compared to those sites that do not have this feature.

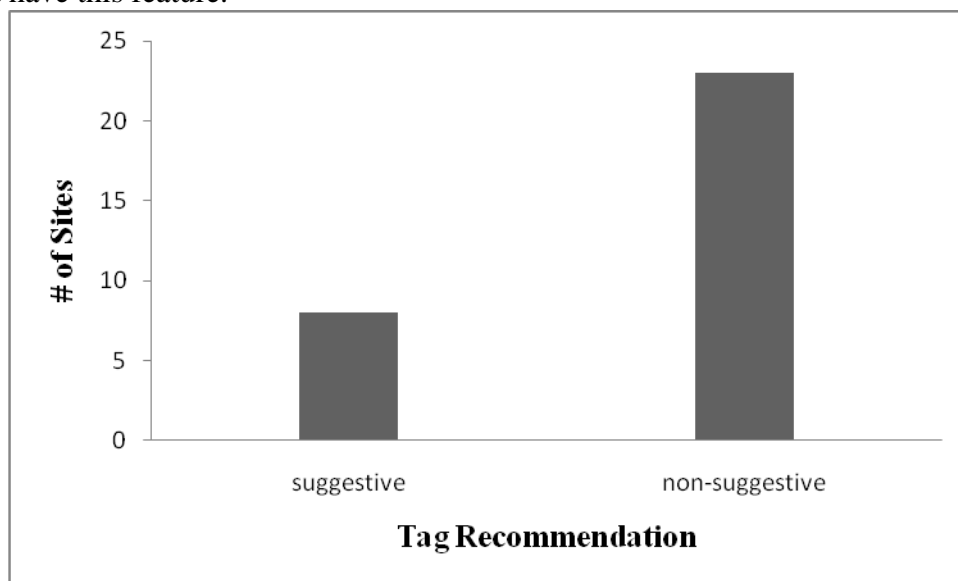


Fig. 5. Distribution of Tagging Recommendation Feature Across Sample Data (N=33)

3.4 Tag Duplication

Figure 6 shows the distribution of tagging sites based on their feature for allowing duplicate tags for a resource. Most tagging sites allow users to duplicate tags for the same resources. This is particularly observed in tagging sites that allow users to create their own account to bookmark web links: A user can bookmark web sites that have already been bookmarked and tagged by others, and she can also use those similar tags for the links that

are added to her personal account. Examples of such websites include <http://www.delicious.com>, <http://www.clipmarks.com> and <http://www.furl.com>.

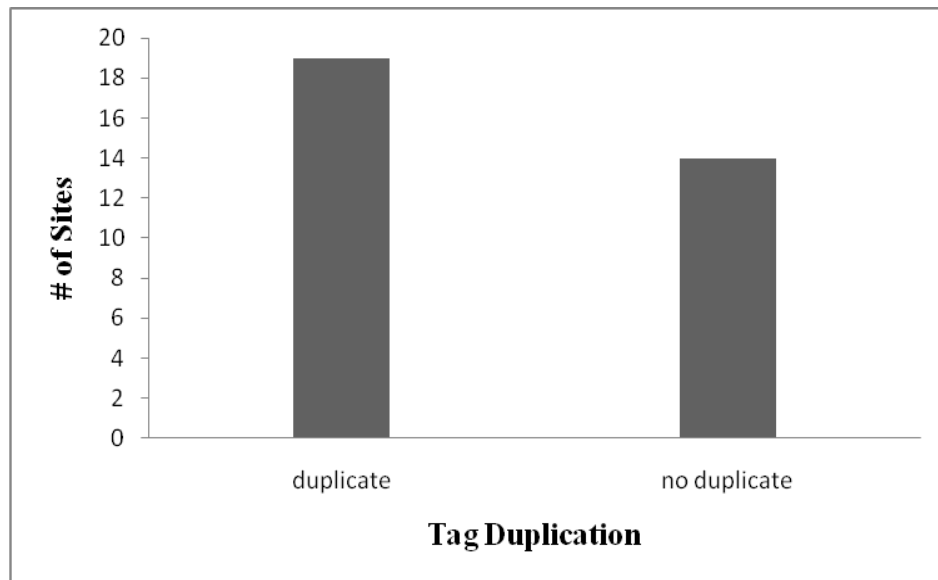


Figure 6. Distribution of tagging duplication feature across sample data (N=33)

Certain sites do not allow duplicate links to be posted: Users cannot upload duplicate resources and only one occurrence of a resource is possible. These sites will only let the user who uploaded the resource add tag to it. Any duplication of tags for the same resource is, thus, not possible. Examples of tagging sites falling under this category are <http://www.kaboom.it.com>, <http://www.barksbookmarks.com>, <http://www.actualtopics.com>.

4. Discussion

Following the analysis of the 33 websites in our sample, it is found that tagging sites offer a number of features that mediate the interaction between users, resources and tags. For each feature, a tagging site can fall in different category depending on the support that the site provides to the user. Table 2 summarizes the different categories under which different sites may fall. It is found that not all of the features are common to all tagging sites. The two features that were offered by all tagging sites are ‘tagging permission’ and ‘tag duplication’. The other features, ‘tag visibility’ and ‘tag recommendation’ were relevant to certain sites only.

Going back to the first research question investigated in this study (i.e., do all tagging systems offer the same features to users?), it can be concluded that all tagging sites do not offer the same features for the tagging of resources. Our analysis showed that different tagging sites had different system interfaces and these sites handled tags differently providing different tagging support to their users. In regards to the second research question, (i.e., how do the features provided by tagging system influence the tagging behavior of users?), we discuss the implications of each tagging feature on the tagging behavior of users.

Table 2.: Categorizing the Features of Tagging Sites

| Feature | Categories |
|--------------------|-----------------------------|
| Tagging Permission | Own Peer Anyone |
| Tag Visibility | Visible Not-Supported |
| Tag Recommendation | Suggestive Not-Supported |
| Tag Duplication | Allowed Not-Allowed |

Regarding the tagging permission feature, it is seen that most users are allowed to tag only those resources they upload. Fewer systems allowed users to collectively tag the same resource and even lesser systems would allow anyone to tag any resource. Because most systems allow users to tag only their own resources, the tags assigned will represent the interpretation of only one individual. These findings seem to support the concerns for the second group of researchers who consider that tags associated with a resource mostly represent the personal information needs of one individual user. Eventually these tags will not be of much value to a wider community of users. To support the claim of the first group of researchers who consider that users, over time, will assign similar tags to similar or same resources, tagging systems should allow users to tag the same resources. Different users tagging the same resource based on their own interpretations will create a more universal representation of the resources that can positively contribute to information retrieval.

When a user is tagging a new resource, most of the tagging systems analyzed did not show tags that have already been used by other users. Users rely on their own interpretation of the resource and they can only assign tags to resources without the influence of any other external factors, such as imitation. This finding is interesting because it suggests that if different users, without prior knowledge of the tags other individuals have used, agree on similar tags for the same resource, then tagging could provide a broader view of a resource from a user's perspective. This contrasts with the situation when the vocabulary used by professional indexers is different from the one used by users.

Most systems also did not suggest any tags to users. By recommending which tags to use, a system can in effect encourage users to use similar set of vocabularies. Sometimes, it may be easier for users to pick up a tag, suggested by the system, rather than thinking over which tags to use for a resource. The benefit of such an approach is that it reduces the cognitive load on users to determine the tags most suitable for a resource, but the downside of it is that the tags assigned to a resource may mostly represent the system's view and less of the user's own interpretation of the resource. For systems that offer tag suggestion, the probability that, over time, similar tags will be used by different users will be higher than for systems in which users have to formulate their own tags. Researchers claiming that tags assigned by users converge over time should make sure that it is not, in fact, the system which is encouraging users to use similar tags.

Users are sometimes not allowed to post the same resources to a tagging site. Eventually, users cannot tag the same resource multiple times. In consequence, tags assigned in sites that do not support duplicates will only represent an individual's perspective of a

resource. When duplicate resources are allowed, the tags assigned would represent a broader view of the resources and may be of more use to a wider community. In our analysis, more systems allowed the presence of duplicate resources and multiple users were able to tag the same resource. This is because most tagging sites in our sample were social bookmarking sites that let users create their personal list of bookmarks while sharing the same list to others. If tags are to be aggregated and used for representing a resource, systems should allow users to post and add tags to duplicate resources; otherwise, as some researchers have argued, tags will only be relevant to the person who assigns the tags to a resource.

It follows that features provided by tagging sites have a direct influence on the tagging behavior of users. The mixed feelings that researchers express about the utility of tags for information retrieval are in fact the result of looking at tags only. As discussed above, depending on the features offered by a tagging site, the resulting tags may be either unique to one individual personal information need or they can be aggregated from multiple users to represent a broader view of a resource. It is in the latter case that tags may be more useful in delivering relevant information to users.

5. Conclusions

This paper has presented the different features offered by tagging systems to tag resources. The implications of each of these features on the tagging behavior of users have also been discussed. Despite the small sample size analyzed in this study, it can be concluded that tags are in effect the result of the interaction between the user, the tagging system and the resource. By providing different features, tagging systems have a direct influence on the tagging behavior of users and the resulting tags.

Acknowledgement

The author extends his sincere thanks to Professor Susan Herring and Professor Elin Jacob from Indiana University, Bloomington, USA, for reviewing earlier versions of this paper.

About the Author

Chuttur M Yasser is a doctoral candidate and lecturer at Indiana University, Bloomington. He has a Masters degree in Internet Computing from the University of Surrey in UK and his research interests include knowledge representation on the Web.

References

- [1] Cattuto, C., Loreto, L., & Pietronero, L. (2007). Semiotic Dynamics and Collaborative Tagging. *Proceedings of the National Academy of Sciences of the United States of America*. 104 (5). 1461-1464.
- [2] Furnas, G. W., Landauer, T. K., Gomez, L. M. & Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communication ACM*, 30(11), 964-971.
- [3] Golder, S. A., & Huberman, B. A. (2005). The structure of collaborative tagging systems. Information Dynamics Lab: HP Labs, Palo Alto, CA. Retrieved March 30 2008, from: <http://arxiv.org/abs/cs.DL/0508082>
- [4] Golder, S., & Huberman, B. (2006). Usage patterns of collaborative tagging Systems. *Journal of Information Science*, 32(2), 198-208.
- [5] Hammond, T., Hannay, T., Lund B., & Scott, J. (2005). Social Bookmarking Tools- A General Overview. *DLib Magazine* 11, (4). Retrieved March 30 2008, from <http://www.dlib.org/dlib/april05/hammond/04hammond.html>

- [6] Herring, S. C. (2007). Web Content Analysis: Expanding the paradigm. In J. Hunsinger, M. Allen, & L. Kladstrup (Eds.), *The International Handbook of Internet Research*. Springer Verlag.
- [7] Heymann P., Koutrika, G., & Garcia-Molina, H. (2008). Can social bookmarking improve web search? Proceedings of the international conference on Web search and web data mining, (pp. 195-206). Palo Alto, California, USA.
- [8] Jacob, E. K. (1991). Classification and categorization: drawing the line. In B. H. Kwasnik and R. Fidel (Eds.), *Advances in classification research*, vol. 2 (p. 67-83). Washington D.C.: American Society for Information Science.
- [9] Lin, X., Beaudoin, Y., B., & Desai, K., (2006). Exploring characteristics of social classification. In proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop, Austin, Texas (US). Retrieved March 30, 2008 from <http://dlist.sir.arizona.edu/1790/>
- [10] Munk, T.B. & Mork, K. (2007). Folksonomies, tagging communities and tagging strategies – An empirical study. *Knowledge Organization*, 34(3), 115.
- [11] Noll, M. G. & Meinel, C. 2008. Exploring social annotations for web document classification. In Proceedings of the 2008 ACM Symposium on Applied Computing (Fortaleza, Ceara, Brazil, March 16 - 20, 2008). SAC '08. ACM, New York, NY, 2315-2320. DOI= <http://doi.acm.org/10.1145/1363686.1364235>
- [12] Olson, H. (1994). Universal models: a history of the organization of knowledge. In H. Albrechtsen and S. Oernager (Eds.), *Knowledge organization and quality management: Advances in knowledge organization*, vol. 4 (p. 72-80). Frankfurt/Main: Indeks Verlag.
- [13] Peterson, E. (2006). Beneath the Metadata: Some philosophical problems with Folksonomy. *Dlib Magazine*. 12(11). Retrieved October 12, 2010 from <http://www.dlib.org/dlib/november06/peterson/11peterson.html>
- [14] Voß, J. (2007). Tagging, folksonomy & co. – renaissance of manual indexing? Retrieved March 30 2008, from <http://arxiv.org/abs/cs/0701072>