

# Could Social Tags Enrich the Library Subject Index?

Constantia Kakali<sup>1,2,\*</sup>, Christos Papatheodorou<sup>2,3</sup>

<sup>1</sup>Library and Information Service, Panteion University, Athens, Greece

<sup>2</sup>Department of Archive and Library Sciences, Ionian University, Corfu, Greece

<sup>3</sup>Digital Curation Unit, Institute for the Management of Information Systems, 'Athena'  
Research Centre, Athens, Greece

## Abstract

Social tagging aims to generate folksonomies through the users' collaboration and activation. This paper is motivated by the trend of several libraries to adopt social tagging functionalities and presents a tag analysis study aiming to exploit a social tag collection for the benefit of the subject description of an academic library material. In this context, cataloguers are interviewed to assess the semantic value of the concepts expressed by the set of social tags and discuss the possibility of their incorporation in the well-formed and structured library authority file.

## 1. Introduction

Web 2.0 enables users to index resources and organize information according to their own background and needs. While up to now information resources were described, organized and classified, either by experts, or by their creators, nowadays users produce new sets of metadata by adding to the resources uncontrolled keywords, which are referred as "tags". The association of tags to resources is named social tagging and generates folksonomies. A folksonomy is a flat classification system using as descriptors the tags created by the users collaboratively.

The phenomenon of social tagging appeared in blogs and social bookmarking sites, but nowadays has been expanded to information organizations such as libraries, museums and archives, which use tags along with their knowledge organization systems (KOS). This trend affects the scholar communication and information seeking behavior and this is evident by the increasing number of social bookmarking sites for scientific material as well as the number of academic libraries that use such technologies. Recently many pioneer libraries are inspired by the Web 2.0 technologies and their catalogues (OPAC) provide a set of key features, promoting digital scholarship and encouraging their user communities to collaborate. Given that an increasing number of libraries develop social tagging systems (Penn Tags, SOPAC, MTagger, etc.) in parallel to their traditional services, a key issue concerns the impact of social tags to the subject indexing process of an information organization.

This paper introduces OPACIAL (<http://library.panteion.gr/opacial/>), an OPAC system enhanced by Web 2.0 features, developed by the Panteion University Library, Athens, Greece and aims to investigate (a) whether the social tags correct, enhance or refine the subject description of an academic library collection and (b) how the social tags could improve the library's local authority file. For this purpose it presents the results of a study, which analyzes the tags associated with a representative sample of 30 bibliographic records. The tags belong to OPACIAL and LibraryThing (<http://www.librarything.com/>), a social cataloguing web-based service that supports the tagging of bibliographic records. In the study the library's subject cataloguers were interviewed to (a) assess the semantic value of the tags with respect to the corresponding subject headings that describe thematically the selected bibliographic records, and (b) suggest alternatives to exploit the tags for the benefit of subject indexing. The next section presents the research that focus on the correlation of folksonomies and subject

indexing, while section 3 describes OPACIAL. In section 4 the experimental setting of this study is presented and its findings are described in section 5. The conclusions derived by this experiment are presented in section 6.

## **2. Related work**

Many scientists foresee possibilities and dynamics in tagging (Quintarelli, 2005; Mathes, 2004; McCulloch & Macgregor, 2006) but at the same time they recognize significant limitations. The current state of the art on the semantic correlation between folksonomies and knowledge organization systems (KOS) centers upon the analysis and integration of user, creator, expert and machine generated vocabularies.

Kipp (2006) compared the vocabularies of users, authors and cataloguers analyzing tags on CiteULike, a social bookmarking site specialized on academic articles. She correlated the tags associated to specific articles with author keywords and thesaurus descriptors and she showed that user tags are related to the author keywords and cataloguers subjects, and the majority of tags were broader or new terms. Lin, Beaudoin, Bui, and Desai (2006) compared social tags with controlled vocabularies and title-based automatic indexing in three empirical studies. They observed overlaps among the three approaches and investigated how tags could be categorized to improve the searching and browsing effectiveness. Moreover the study of Al-Khalifa and Davis (2007) showed that the folksonomy tags overlap significantly with the human generated keywords in contrast to the automatically generated.

Heckner, Mühlbacher and Wolff (2008) created a document classifier for a collection of articles from Connotea, based on the linguistic and functional aspects of tag usage, as well as on the relationship between the tags and document's full text. The classifier was applied to approximately 500 randomly selected tagged articles from the information and computer technology domain and the findings demonstrated a great overlap between the tags and text. Voss (2006) explored the similarities and differences between Wikipedia, folksonomies and traditional hierarchical classification systems (e.g. Dewey Decimal Classification) and he concluded that Wikipedia's category system constitutes a thesaurus based on a special combination of social tagging and hierarchical subject indexing.

Yi and Chan (2009) investigated the relation of the LCSH and social tags selected from Delicious. The study of the tags distribution over LCSH concluded that LCSH "may greatly enhance the collaborative tagging systems information control process" and "it is possible to connect collaborative tagging systems with OPACs or digital libraries". Thomas, Caudle and Schmitz (2009) performed also a comparison of social tags with LCSH. They report an effort of the librarians of the Cataloging Department, Auburn that compares the social tags and LCSH assigned to a sample of ten books in problematic subject areas across a sample of libraries. The analysis followed a combination of tag classification criteria mentioned by Golder and Huberman (2006) and Kipp (2006).

LibraryThing (<http://www.librarything.com/>) is a social cataloguing web application permitting the tagging of the bibliographic records. The inserted tags are used for organizing personal book collections, recommending related books, linking editions and translations of a work, etc. According to (Mendes, Quinonez-Skiner & Skaggs, 2008) the usage of LibraryThing tags might transform an OPAC from static to an open, interactive and usable site. LibraryThing content has been used by several tag analysis experiments and innovative systems. For instance EnLibS system, which expands users' queries with tags from LibraryThing in order to reduce the high percentage of the failed queries (Pera, Lund, & Ng, 2009). Smith (2007) explored the relationship between

folksonomy and subject analysis in a study of LibraryThing tags and Library of Congress Subject Headings (LCSH) associated with the same documents. Her results showed that the tags identified latent subjects. Finally Lawson (2009) compared the 31 top-level subject divisions and the tags from Amazon.com and LibraryThing associated with a sample of 155 books and she claimed that social tagging enables librarians to partner with users to enhance subject access.

### 3. The OPACIAL system

The presented experimental study is based upon the tags inserted to OPACIAL, an OPAC system enhanced with Web 2.0 features, developed by the Panteion University Library, Athens, Greece. The added-value features of OPACIAL include tagging functionalities, folksonomy-based navigation to the library material, as well as tag searching (Figure 1 shows a small part of the tag cloud of OPACIAL). Successive tag selections operate as faceted information retrieval and narrow down the retrieved records. Furthermore OPACIAL provides user annotations, ranking functionalities and use of reference tools. The users are able to annotate and rank each resource (on a 1 to 5 scale) and to export a record to external social networking sites by using a social networking site aggregator, like Socializer.

A significant feature of OPACIAL is the integration of OPAC records with the ones of the University's digital repository, named Pandemos and also deployed by the Library. Thus, for each OPAC record the user is capable to retrieve similar digital objects.



Figure 1: Part of OPACIAL's tag cloud

OPACIAL has been evaluated by a technology acceptance experiment (Gavrilis, Kakali & Papatheodorou, 2008), in which twenty users (post graduate students and faculty members) used all its functionalities for a week, inserted more than 500 tags and finally were interviewed to assess the system usability and usefulness. The results of this experiment were encouraging since the users declared that they are satisfied by the offered service and consider useful and reliable the information searching using tags. Moreover they prefer to use both the tags and library subject index in information seeking. Given these findings the tags inserted by these users were analyzed in order to explore and categorize their tagging behavior. The results emerged that users insert tags to either correct, or complement weak subject descriptions.

These promising results triggered the design of the presented experiment, which aims to survey the subject cataloguers' opinion concerning the impact of the user community

vocabularies to the local authority file evolution and the definition of a policy to converge the user-based and the expert-based subject indexing approaches.

#### **4. Experimental setup**

The Panteion University library's authority file consists of up to 100,000 entries – most of them translated in Greek language from LCSH – and the majority is interlinked by references of broader, narrower and non-preferred terms. The authority entries correspond to a collection of 80,000 titles of books, serials, video, and grey literature specialized on social and human sciences. For the investigation of the mentioned issues a representative random sample of 30 socially tagged bibliographic records was selected, which carried 72 subject headings, 66 being unique. The 18 records correspond to books written in English, 1 in French and 11 in Greek language.

Totally 540 tags were gathered, 120 being from OPACIAL and 420 from LibraryThing. The distribution of the tags over the bibliographic records is as follows: 12 records has been annotated exclusively by the OPACIAL tags, 8 records has been annotated only by LibraryThing tags and the rest 10 carry tags from both folksonomies.

The bibliographic records along with the corresponding subject headings and the associated tags were presented in a tabular form (Table 1 presents a part of the data) and given to the 9 subject cataloguers of the Panteion University Library. Regarding their profile, they all hold a BSc on librarianship, 1 holds a second BSc on History and 4 of them hold a MSc on Information Science. All of them are women and 3 of them have more than 15 years professional experience, 2 have an experience between 6-10 years and 3 are junior cataloguers. Most of them acquainted with social tagging due to OPACIAL, while 4 of them are familiar with social networking applications.

The cataloguers had to study the thematic description and the tags associated with each record in a period of a week; then each cataloguer was interviewed. The interview was structured in three axes:

(A1): Comparison of the expressiveness and usefulness of the OPACIAL and LibraryThing tags.

(A2): Assessment of the semantic value of both the OPACIAL and LibraryThing tags, with respect to the corresponding subject headings of the selected records. The focus of the discussion was on whether the tags (a) are identical to a part of the subject description of the selected documents, (b) are identical to some subject descriptors of the authority file but disjoint to the description of the selected documents (c) correct the subject description of the selected documents, (d) enhance the subject description of the selected documents. The enhancement of the subject description by a tag is defined as either

(i) the tag introduces a new descriptor; it might be synonym, broader or narrower to the terms of the subject description, or

(ii) the tag is identical to existing descriptors and disjoint to the terms of the subject description; it might to be correlated to the terms of the subject description with equivalence, hierarchical or association relations.

(A3): Exploitation of the social tagging in subject indexing and the development of a social tagging policy by the library. Indicative questions of this axis are:

(i) does social tagging upgrade the library information access services,

(ii) whether the library should encourage its users to insert social tags and why,

(iii) how a library could incorporate social tags in its authorities; by creating new subject headings; by correlating existing subject headings; by correcting/modifying the translation of the subject headings to the local language according to the user communities vocabulary evolution.

The first two were closed questions aiming to disambiguate the participants' opinion about the usefulness of social tagging. These attitudes were recorded using 7-point Likert scale, with 1 being in the negative side and 7 being the positive. Finally each interview had an average duration 30 – 40 minutes.

## 5. Results

Concerning the first axis (A1) the interviews proved that OPACIAL has more representative and accurate tags than LibraryThing. The librarians found a number of pointless expressions and misunderstood some of the tags of the LibraryThing's clouds. In particular they "vote" for the 60% of OPACIAL tags are useful and more precise and 40% for LibraryThing. However the OPACIAL percentage increased after the exhaustive enumeration and examination of the tags.

This finding is explained by the fact that OPACIAL hosts material focusing on social sciences and serves a scholar community that uses a specialized vocabulary. On the other hand LibraryThing is a general-purpose collaborative cataloguing service, thus carries several tags that do not add value to the subject description. For instance 20 (out of the 420) LibraryThing tags represent the taggers instead of the subject of the documents (e.g. Norton, wanted), or are too general terms (e.g. book). Furthermore some tags are identical to the last names of authors, editors, or personal names as subject. Finally a couple of them are annotations instead of tags since they are phrases, comments or definitions.

Regarding axis (A2), all librarians confirmed that in general the tags enrich the subject description of the documents. The tags supplement the thematic description of most documents, while a librarian said, "they complement the information" of the books. However they hardly discerned a trend of the users' tags to correct the subject description of some documents but they admitted that they have already corrected some bibliographic records due the appearance of more accurate tags. Most of them they found a significant number of tags that are identical to authority records, and certainly enrich the subjects of the bibliographic records. This opinion is confirmed by the fact that only 21 tags are the same with the subject description of the selected documents, while the majority of the tags, 355 out of 540, are identical to the subject descriptors of the library authorities.

Indicative examples of this analysis are given in Table 1. In the record of the first row the 2 subject headings are included in the tag cloud. The tag cloud consists of 34 tags and 28 of them belong to the local authority. The evaluation of the tag cloud revealed that 11 of the tags could be used in the subject description of the record, while 2 of them are new terms. In the second row there are 3 subject headings and 2 tags are part of them. Totally 21 tags are associated with this record, while 12 of them are terms in the authority file. The evaluation emerged that 5 tags enhance the subject description of the record, while just the tag "digital humanities" is a new term.

Moreover librarians opined that the majority of OPACIAL tags are narrower terms as compared with the subjects of the documents and sometimes too specific. Contrary LibraryThing tags are diverse; some of them are broader terms (e.g. Philosophy, Sociology, Economics, Culture) and do not add value to information retrieval.

Finally the librarians found that several tags constitute either new concepts or neologisms, or alternative translations of terms to the Greek language. However most of them hesitate to create new authority records but are eager to consider the tags non-preferred terms and add relations to them.

Table 1. A sample of tagged records

Bibliographic Record	Subject Headings	Tags
<p><b>Author:</b> Weber, Max (1864-1920), Roth, Guenther (Editor), Wittich, Claus (Editor).</p> <p><b>Title:</b> Economy and society : an outline of interpretive sociology / Max Weber; edited by Guenther Roth and Claus Wittich</p> <p><b>Publication:</b> Berkeley, Calif. : University of California Press, c1978</p>	<p>Sociology</p> <p>Economics</p>	<p><u>19th century</u> <u>20th century</u> <u>Europe</u>  <u>Germany</u> <u>Verstehen</u> <u>Weber</u>  <u>bureaucracy</u> <u>class</u> <u>structure</u> <u>economic</u>  <u>sociology</u> <u>economics</u> <u>economy</u> <u>german</u>  <u>history</u> <u>interpretation</u> <u>knowledge</u>  <u>philosophy</u> <u>political</u> <u>economy</u> <u>political</u>  <u>science</u> <u>political theory</u> <u>politics</u> <u>religion</u>  <u>social theory</u> <u>society</u> <u>sociological</u>  <u>theory</u> <u>sociology</u> <u>state</u> <u>the state</u> <u>theory</u>  <u>world history</u> <u>Αξιολογική Ελευθερία</u>  <u>Γερμανοί Φιλόσοφοι</u> <u>Κατανόηση</u>  <u>Κοινωνιολογία</u></p>
<p><b>Author :</b> Janes, Joseph ( 1962- )</p> <p><b>Title:</b> Introduction to reference work in the digital age / Joseph Janes</p> <p><b>Publication:</b> New York: Neal-Schuman Publishers, c2003</p>	<p>Reference services (Libraries)</p> <p>Internet in library reference services</p> <p>Electronic reference services (Libraries)</p>	<p><u>Digital Reference</u> <u>LIS</u> <u>Professional</u>  <u>Books</u> <u>Reference/Business</u> <u>best</u>  <u>practices</u> <u>biblioteques</u> <u>computers</u> <u>digital</u>  <u>humanities</u> <u>guidelines</u> <u>info</u> <u>science</u>  <u>information management</u> <u>information</u>  <u>science</u> <u>internet</u> <u>librarians</u> <u>librarianship</u>  <u>library</u> <u>library science</u> <u>reference</u>  <u>reference service</u> <u>textbook</u></p>

Summarizing the third axis (A3) of the interview, the cataloguers are positive towards the adoption of social tagging by the Panteion University Library and they find it useful. Some of them argued, “we have the chance to discover our weakness in subject indexing, especially the non-subject librarians”. In addition they agreed that social tagging could help them to approach the user’s way of thinking and help them more effectively as well as to observe the communities terminology evolution. As cataloguers, they believe that tags provide new terms for the benefit of subject indexing – even the subject description precedes the insertion of tags. It is worthy to mention that metadata cataloguers proposed the implementation of tagging system in the digital library of the University (Pandemos, <http://library.panteion.gr/pandemos>), which hosts specialized material such as theses and digitized scientific journals.

An issue revealed by the interviews is the quality assurance of the tagging process. The cataloguers expressed their reservation about the tags being inserted by undergraduate students and external users, while some of them stated, “we must encourage faculty and post-graduate students to tag”. Nevertheless an open question is to “buy” LibraryThing’s tags or to encourage the users to enrich the local folksonomy.

A relevant issue concerns the process of the social tags exploitation. The discussion among the others concerned the frequency and the criteria of the tag assessment process. Some librarians suggested the refinement of the inserted tags by searching, in pre-defined time periods, the LCSH to identify overlapping terms and keep in the folksonomy only the non-overlapping tags, while the overlapping to be inserted in the local authority file. However this proposal characterized as a “luxury” for an academic library with a limited number of personnel that serves almost 3,000 users per day and inserts more than 4,000 records in the catalog.

Two librarians proposed the creation of a wiki to enhance the collaboration of subject cataloguers and the faculty members for the disambiguation of the inserted tags, the apodosis of subject descriptors in the Greek language and in general the improvement of the library authorities. Moreover two other cataloguers suggested the creation of new authority records “out of the LCSH frontiers”, especially in cases the tags are identical to Wikipedia entries.

Concerning the closed questions, the librarians believe that OPACIAL upgrade the library services giving an average grade 5.66 in the 7-point Likert scale. Regarding the reasons for which a library must encourage its users to tag records, the results are as follows: (i) to activate the user participation gained an average grade 6.22, (ii) to develop an user – friendly alternative for information retrieval, gained an average 6.44, (iii) to receive feedback for the users needs, gained an average 6.44, and finally (iv) to develop a direct way for subject indexing according the users' vocabulary, gained an average 5.66. These results confirm the hypothesis that the cataloguers consider useful the OPACIAL tagging functionality and thus the cooperation between the local index and folksonomy.

## 6. Conclusion

This study signified the opening of an in depth discussion between the library staff about the power of Web 2.0 opportunities. The library cataloguers recognized the role of “long tail” and the importance of the development of social networks through the activation of their users to create metadata.

The findings of this study provide the opportunity to the library staff to evaluate the library subject index, and to renew its content by new terms or relations. In particular the study addresses that the tags express directly the evolution of a scientific domain and the library should (a) create new subject descriptors, (b) substitute the current subject headings with more appropriate ones and (c) create references between the subject descriptors of the local authority file.

Essentially, the development of a policy for the exploitation of social tagging is equivalent to the establishment of a Library 2.0 environment in an information organization grounded on the concept of user collaboration and the design of collective information services.

## References

- Al-Khalifa, H.S., & Davis, H.C. (2007). Exploring the value of folksonomies for creating semantic metadata. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 3 (1), 13-39.
- Gavrilis, D., Kakali, C., & Papatheodorou, C. (2008). Enhancing library services with Web 2.0 functionalities. In *Proceedings of the 12th European Conference on Research and Advanced Technology for Digital Libraries (Aarhus, Denmark, September 14 - 19, 2008)*. B. Christensen-Dalsgaard, D. Castelli, B. Ammitzbøll Jurik, and J. Lippincott, Eds. Lecture Notes In Computer Science, vol. 5173. Springer-Verlag, Berlin, Heidelberg, 148-159. Doi: 10.1007/978-3-540-87599-4\_16
- Golder, S., & Huberman, B. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2), 198-208.
- Heckner, M., Mühlbacher, S. & Wolff, C. (2008). Tagging tagging: Analysing user keywords in scientific bibliography management systems'. *Journal of Digital Information*, 9(27).
- Kipp, M.E.I. (2006). Complementary or discrete contexts in online indexing: a comparison of user, creator, and intermediary keywords. *Canadian Journal of Information and Library Science* 30(3). Retrieved October 26th, 2009 from <http://dlist.sir.arizona.edu/1533/01/mkipp-caispaper.pdf>
- Lawson, K. (2009). Mining social tagging data for enhanced subject access for readers and researchers. *The Journal of Academic Librarianship* 35(6), 574-82.

- Lin, X., Beaudoin, J. E., Bui, Y., & Desai, K. (2006). Exploring characteristics of social classification. In *Proceedings of the 17th ASIS&T Classification Research Workshop*, Austin, Texas, USA. J. Furner & J. T. Tennis (Eds.). Retrieved November 20th, 2009 from <http://dlist.sir.arizona.edu/1790/01/lin.pdf>.
- Mathes, A. (2004). Folksonomies - cooperative classification and communication through shared metadata. *Report, Graduate School of Library and Information Science, Illinois Urbana-Champaign*. Retrieved November 20th, 2009 from <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- McCulloch, E. & MacGregor, G. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review* 55(5), 291-300.
- Mendes, L.H. Quinonez-Skinner J., & Skaggs D. (2008). Subjecting the catalog to tagging. *Library Hi Tech* 27(1), 30-41.
- Pera, M.S., Lund, W., & Ng, Y-K., (2009). A sophisticated library search strategy using folksonomies and similarity matching. *Journal of the American Society for Information Science and Technology*, 60 (7), 1392-1406.
- Quintarelli, E. (2005) Folksonomies: power to the people. In *ISKO Italy-UniMIB meeting. Milan, Italy*. Retrieved December 20th, 2009 from <http://www.iskoi.org/doc/folksonomies.htm>.
- Smith, T. (2007). Cataloguing and you: Measuring the efficacy of a folksonomy for subject analysis. In 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research, Milwaukee, Wisconsin, USA. J. Lussky (Ed.), Retrieved July 18th, 2009, from <http://dlist.sir.arizona.edu/2061/01/Smith%5FUpdated.doc>.
- Thomas, M., Caudle D. & Schmitz C. (2009). To tag or not to tag? *Library Hi Tech* 27(3), 411-334.
- Voss, J. (2006). Collaborative thesaurus tagging the wikipedia way. Retrieved July 18th, 2009 from <http://arxiv.org/abs/cs.IR/0604036>.
- Yi, K. & Chan, L. M. (2009). Linking folksonomy to Library of Congress subject headings: an exploratory study. *Journal of Documentation* 65(6), 872-900.