The Journal of Academic Librarianship 40 (2014) 565–573

Contents lists available at ScienceDirect

The Journal of Academic Librarianship

A Utilization Model of Users' Metadata in Libraries

Constantia Kakali

Panteion University of Social and Political Sciences, Library and Information Centre, Athens, Greece

ARTICLE INFO

Article history:
Received 22 January 2014
Accepted 28 August 2014
Available online 6 November 2014

Keywords:
Subject indexing
Social tagging
Folksonomy
Classification
Users' metadata
Cataloguers

ABSTRACT

The purpose of this paper is to define a utilization model of meaningful users' tags in subject indexing work in libraries. The research work was originally performed with a quantitative method; a large number of relations (tag–bibliographic record) were examined and analyzed, resulting in a definition of the classes of the model. This model was attempted to be verified by a survey addressed to cataloguers in Greek libraries. This paper is based on the principle that the users' collaboration and their vocabulary provide useful feedback for the enhancement of the subject description of the documents.

INTRODUCTION

Various knowledge organization systems have been developed over the centuries, and they still do without losing their significance in the new digital environment; on the contrary they have reinforced it with new tools of information management. These systems were created by experts/scientists and handled by professionals: cataloguers, indexers or classifiers who have historically created order in the universe of knowledge. The wide spreading of social media has limited the exclusivity of the traditional model, since the appearance and use of folksonomies allows users to contribute to the retrieval process without the involvement of professionals' interpretation.

The implementation of Web 2.0 technologies has resulted in the transformation of the information environment, especially in libraries. The appearance of new generation catalogs in libraries is one of the most important innovations. They have replaced the classic catalogs since they are enhanced with significant features such as social tagging and annotations. These applications are used by library's users for the purpose of management of their personal information space; in fact, they are being operated as a way of inserting metadata, forming the "personomy" of each user. The reason why they were developed and expanded has more to do with the influence of dominant paradigms of the social web in library, despite the fact that "very few legacy or classic ILSs are capable of this function" (Yi Lee & Yang, 2012).

A prominent tool for these functionalities is the cloud of social tags, which allows users to navigate to information through an immature way, a flat and unregulated structure. The "repository of the terms" has been created by users, and it is the dynamic visualization of the total of the tags called "folksonomy", a new compound word deriving from the words folk and taxonomy. The folksonomies, functioning in parallel with long-standing and robust knowledge organization systems, have been proved incredibly simplistic, therefore they are treated with skepticism as to whether they can be identified as mechanisms for information retrieval. The flaws of folksonomy systems are inherent to their structure and development. The possibility of users inserting terms immediately and freely overweighs the lack of precision and control of these terms. In folksonomies there is absence of synonyms or homonyms, there is the parallel use of plural and singular forms, as well as a major number of "sloppy tags" which are probably generated as being of "single-use" (Guy & Tonkin, 2006) without more meaning and value except their creators.

Recent research (Jiang, 2013) shows that those libraries which have developed social web features are bound to become a diverse and dynamic information seeking environment. In a survey conducted in 100 academic libraries in the USA included in the Association of Research Libraries' membership list, Mahmood and Richardson (2011) found that 55 of them already had social tagging in OPACs or social bookmarking features in their website. Earlier researches found out that social tagging had been less widely implemented than other Web 2.0 tools (Kim & Abbas, 2010), and a reason for that is that the technical services of the academic library (e.g. cataloging) "have yet to explore the value of tagging", as Xu, Ouyang, and Chu (2009) explained.

The outcome of users' involvement in tagging systems is the creation of their metadata, which concerns several scientists for different reasons: the analysis of user needs or the comparison between users' description language and knowledge organization systems can lead them to suggestions for improvements of these social tagging systems. However, the total or partial utilization of the terms—concepts included

E-mail address: nkakal@panteion.gr.

http://dx.doi.org/10.1016/j.acalib.2014.08.004
0099-1333/

in folksonomies from the known knowledge organization systems is of little concern.

This paper aims to probe the utilization of social tags for the subject indexing enrichment of bibliographic records. The selected bibliographic records related with tags constitute the pairs that are rated highly by a process of evaluation that was presented in previous research work. The present work demonstrates that the utilization of tags from libraries is not deterministic and that the role of professionals is important when more people participate in the organization and representation of information and knowledge. The degree of their evaluation and, furthermore, their utilization both vary in different libraries and in indexers. Even if optimized techniques for more useful tags were revealed, their incorporation in the existing knowledge systems would depend on the policies of subject indexing of libraries and their indexers.

RESEARCH BACKGROUND

The potential of enrichment of the subject description from folksonomies has been studied by several researchers. The recent literature is divided in three groups of studies, which are related to this article. The three different groups reflect the evolution steps of our research. Having as a starting point, the comparison of folksonomies with traditional knowledge organization systems and in particular the subject headings, we continued by presenting some research works that examine similarity measures in-between the two systems. The third group included research papers relevant to the categorization of tags.

COMPARISON OF FOLKSONOMIES WITH LCSH

A significant number of information scientists have studied the relations between folksonomies and traditional indexing and classification systems, some of which focus their research on the library environment. Using techniques such as the content analysis, they examine the relation and the overlapping of the tags with Library of Congress Subject Headings (LCSH). Thomas, Caudle, and Schmitz (2009) present a comparison between social tags and headings (LCSH), from a sample of 10 books from different libraries and thematic areas. According to a study performed by Lu, J-r, and Hu (2010) social tagging can be an excellent tool for improving access to the library collection, despite the problematic issue of non-subject related tags. Comparing the LCSH with Delicious' folksonomy tags, Yi and Chan (2009) found that 61% of the tags which were used twice or more matched the subject headings. Additionally, the authors represent the subject headings in tree form, along with their synonyms, the narrower and broader terms. They propose a methodology with which a subject heading is associated to a tag with the ultimate goal of improving the recall of documents. In a similar direction, Heymann and Garcia-Molina (2009) compared term to term and made semantic matches between LCSH and tags assigned to books in LibraryThing. They found that about 50% of the keywords were the same in both vocabularies. Moreover, they also underlined that the expert and users differ in the usage of the same terms.

Many studies focused on a particular subject field to demonstrate the similarities and differences in the systems' vocabulary (LCSH, folksonomy). M. Adler (2009) highlighted the contrast between the language of the people (terminology) using tags and the LCSH mentioned in books on transgender issues, while Iyer and Bungo (2011) examined the semantic relationship between user tags and the assigned subject headings of popular literature in the domain of complementary and alternative medicine. Similarly, DeZalar-Tiedman (2011) explored and evaluated the usefulness of LibraryThing tags in a set of bibliographic records for literature works which lacked subject headings.

Comparing tags with traditional subject access, Lawson (2009) describes and evaluates their intersection and how their integration might be useful. Furthermore, Pirmann (2012), examining the utility of tags with a usability test, indicated that while tags can be a useful mechanism for finding materials in library catalogues, they cannot replace the traditional subject headings. In the same vein Rolla (2009) compared the subjects of 45 bibliographic records with the tags assigned by the LT users to the same documents and found that the tags reflect mainly topical information, though a large proportion of them are personal and without value to information retrieval. Another attempt, defining a methodology for the utilization of social tagging in subject analysis (Kakali & Papatheodorou, 2010a), was focused on the characteristics of users' tagging behavior, something which enhances the subject description of documents.

REFINEMENT OF TAG QUALITY BY SIMILARITY METRICS

The second category consists of researches which present tag recommendation techniques by using various metrics applied for the evaluation of tags. Identifying tags of high quality, Krestel and Chen (2008) proposed an algorithm to measure the quality of tags through a graph (Tag–Resource Pair Rank). They then evaluated their approach on a BibSonomy data set, where tags generated by suspicious spammers were manually labeled. Continuing the research on the relations between LCSH and tags, Yi (2010) experiments and compares five similarity measures (tf–idf, Cosine and Jaccard similarity etc.), aiming at predicting subject headings from social tags assigned to the corresponded resources. Recently, Yi (2012) evaluated the degree of Collective Intelligence embedded in social tagging, and utilized the five different metrics for assessing the similarity between ranking lists: overlapList, overlapRank, Footrule, Fagin's measure, and the Inverse Rank measure. The work by Szomszor et al. (2007) is focused on the combination of multiple tag clouds for movies through a measure of tag cloud similarity, done in order to construct user profiles that reflect their interests in different kinds of movies and predict their rating for an unseen movie. Beyond the use of content analysis, Goh, Chua, Lee, and Razikin (2009) adopted tf–idf values in order to indicate the weights of terms. Their findings showed that tags whose semantic meanings are more specific perform better classification. Moreover, tags vary in their effectiveness as navigational aids to resources.

In order to face the problem of users' query failure in a library's catalog, Pera, Lund, and Ng (2009) developed an enhanced library system, in which the results from queries depend on the degree of similarity between social tags assigned to each item and the words of the users' queries. One of the latest publications in this area, Wu, He, Qiu, Lin, and Liu (2013), studied the relationship between social tagging and controlled vocabulary in the domain of information science in two different languages: English and Chinese. They used a strong similarity (measured by Jaccard's coefficient) method and their results showed that the shared keywords between social tags and subject terms are still relatively low, even among the most frequently used tags and subject terms.

MODELS OF TAG CATEGORIZATION

Finally, an important group of publications refers to different models of categorization of tags. These models were used as tools for understanding users' behavior and revealing their needs.

A first attempt to create a categorization of tags was the research study of Kipp (2005), where she exposed the results of the comparison among users' tags, author's keywords and library index terms. That categorization is particularly the transformation of a taxonomy of relationships presented by Voorbij (1998), describing the relations between title words and subject descriptors in a set of monographs of a library. Shortly later, Golder and Huberman's (2006) categorization was based on different functions that tags perform: an extrinsic group for describing a resource and an intrinsic one for personal meaning. The previous typology was used by Thomas et al. (2009) to select the tags that were related with the meaning of the resource. Then they continued by applying the Voorbij/Kipp typology and designed their model by