DESIDOC Journal of Library & Information Technology, Vol. 35, No. 1, January 2015, pp. 41-46 DOI: 10.14429/djlit.35.1.8036
© 2015, DESIDOC

Tag Cloud Application and Information Retrieval System: Visualisation to Create Information Literacy

Shri Ram

Jaypee University of Information Technology, Waknaghat, Solan-173 234
E-mail: shriram2576@gmail.com

ABSTRACT

The purpose of this paper is to discuss the application of tag clouding for information visualisation and information retrieval over social cataloguing tool such as Shelfari (http://shelfari.com). This is a case study where tag cloud is being used as a tool for information visualisation and information retrieval to create awareness about new arrivals in the library. For the purpose of creating awareness amongst the user of the library and promote the use of newly catalogued resources, an online social cataloguing system 'Shelfari' (http://shelfari.com) has been used. The new arrivals are catalogued through social cataloguing system. The Library of Congress subject heading (LCSH) and Medical Subject Headings (MeSH) are being used as tag words for each resource. The tag cloud created in this way has been evaluated for the effectiveness of the social cataloguing. The results suggest that tag clouding helps to create a visual effect and help users to learn about availability of the resources in the library along with the collaborative activities such as reviews and feedbacks, to choose right kind of reading materials. Social cataloguing activities are now getting momentum in online social network system. The use of internet and social network activities are increasing day-by-day so as the use of reviews and feedback for choosing right reading material. This study shall be valuable for those who are interested in choosing right reading materials based on popularity.

Keywords: Social cataloguing, tagging, tag clouds, information search and retrieval, new arrivals

1. INTRODUCTION

Recent years have seen an explosion in online collaborative information sharing. Social cataloguing is one of the collaborative network activity which is characterised, firstly by cataloging of resources and interaction with others based on shared item and, secondly by the enrichment of cataloguing descriptions through either explicit cooperation of cataloguing metadata or through analysis of implicit data. Social cataloguing helps user to catalogue their own resources on web owned by them or liked by them. In this process, user assigns some tags (freely chosen keywords) for their convenience based on the principle of mnemonics and follow social networking sites commonly used for video-sharing (YouTube), photo-sharing (Flickr), book-marking (Citeceer, CiteULike, Del.cio.us), and cataloguing of personal books (LibraryThings, Shelfari). It is found that over these social networking websites, the users are freely using certain keywords to demonstrate the theme of the content. This has given the origin of social tagging as one of the key web 2.0 application that helps users to identify the resources tagged by them. Such tagging-based systems enable users to categorise web resources by means of tags, in order to re-find these resources later. Mendis *et al.,* debated such activities as collaborative tagging, social indexing, and mob indexing[1]. Shiri discussed the thesauri development on web and further, in 2009 used the terms communal categorisation, ethno-classification and free-text tagging system[2, 3]. Thus, in the process of choosing a set of popular tags populated in a particular context is known as 'Folksonomy'. According to Vander Wal, who coined the term 'folksonomy', social tagging had its origins in the 1990s, but it remains one of today's major successes through websites such as 'Delicious'(tagging webpages), 'Flickr' (tagging images) and YouTube (tagging videos)[4]. Within the library world, much attention has been paid to 'LibraryThing', a social tagging based website that allows users to upload records of their entire book collection[5]. Folksonomy is the combination of taxonomy and folk. At the time of tagging, the keywords popularised as an interface model for visual information retrieval known as 'Tag-Cloud'. Description of resources over internet and to ensure effective retrieval from plethora of information is a challenging task. It is essential to describe resources in an effective ways.

Received 29 September 2014, revised 7 November 2014, online published 22 January 2015

Metadata, a structured information which is used to describe, explain, locate an information. It adds value to description of information element and makes it easier to retrieve[6]. Metadata aids in the identification, description, management and location of information in both digital and non-digital environment. Digital environments finds its usability in enhancing resource discovery and disclosing sufficient description about the resource, to enable users to discriminate between what is relevant and what is not relevant to a specific query. To facilitate retrieval of information or subject, the resources are assigned subject headings, based on the content, in the process of cataloguing. Lancaster had given the processes of indexing system in order to achieve high precision in information retrieval. To ensure the effective indexing and to maintain the overall efficacy of the retrieval system, it is necessary to apply some of the degree of control of the indexing process. By controlling the indexing process using a so-called controlled vocabulary, index terms were standardised and similar or related resources were collocated for ease of the discovery by user[7]. The preeminence of controlled vocabularies has recently been challenged by the appearance of "Collaborative tagging" in a variety of social cataloguing and bookmarking applications[8].

2. OBJECTIVES

The objective of the paper is to study the tagging process which can help in information retrieval in social cataloguing environment.

3. METHODOLOGY

A profile of library was created by the name of 'Learning Resource Center, Jaypee University of Information Technology, New Arrivals'on Shelfari (http://www.shelfari.com/lrcjuit/shelf) (Fig. 1). Each month new books catalogued in the library were being updated on this virtual shelf for displaying new arrivals globally. After this process, a feedback on impact on the users of this facility was obtained through a survey. The faculty, student and staff members were asked to give their feedback on this service. The result of the survey is presented in analysis section.

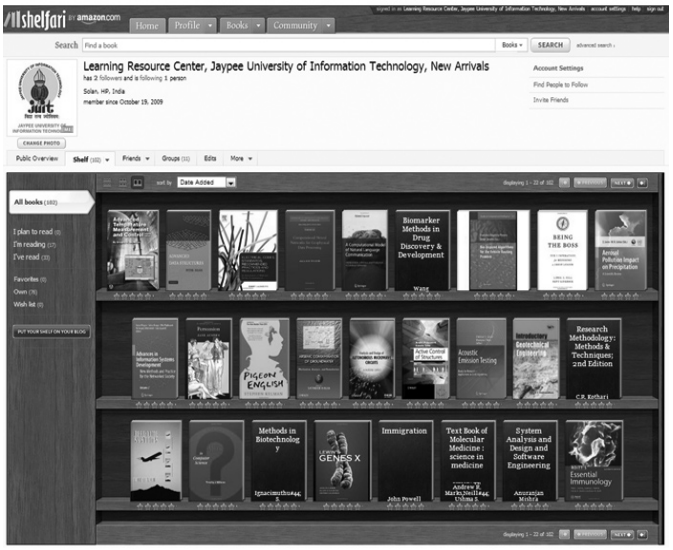

Figure 1. Virtual shelve of library on Shelfari.

Shelfari, based in Seattle, introduces readers to a global community of book lovers and encourages them to share their literary inclinations and passions with peers, and friends. The aim of Shelfari is to gather authors, publishers and readers at one place. Shelfari has many tools and features that help authors, publishers and readers connected each other. These tools are indexing of books, reviews, suggestions, and sharing of information about a book, they write, publish and read (http://shelfari.com). In August 2008, Shelfari was purchased by Amazon[27].

4. SOCIAL CATALOGUING

A Social cataloging application is a web application designed to help users to catalogue reading materials owned or otherwise of interest to them. The terminology is characterised by two peculiar characteristics arising due to the multi-user cataloguing environment, firstly, the ability to share catalogue and interact with others based upon shared item and secondly, the enrichment or improvement of cataloguing description through either explicit co-operation in the production of cataloguing metadata or through analysis of important information architecture associated with the resources under cataloguing activity. Steele described such activity as a social phenomenon and its use in cataloguing activity, metadata description, discussing folksonomies, social bookmarking, tagging and use of controlled vocabulary such as Library of Congress Subject Heading (LCSH) or Medical Subject Heading (MeSH) for information retrieval[9]. In recent times, the discussion is on the information architecture by community about the classification of information as 'social classification', 'ethnoclassification' and 'falkosonomies'[10]. The folksonomies allows user to describe documents (in web based environment) with subject headings called 'tags', without regard for conventional rules[11]. The tags acquire their size according to the amount of use, and are organised in a 'cloud' of tags. Folsosonomies act as an Internet-based information retrieval methodology and involves collaboratively generated, open-ended labels which categorise any content, web resources, and photographs over internet, etc. There are so many websites that exists today which has become increasingly popular in collaborative tagging and appears in a variety of prominent web-based services. The example includes CiteULike (www.citeulike.org/), Del.icio.us (http://del.icio.us/), Flickr (www.flickr.

com/), etc. Similarly, there are also various social networking website which are dedicated to describing books by various user groups. These websites are LibraryThings (www.librarythings.com), and Shelfari (www.shelfari.com). While describing a document over social networking website, the user often marks a label to describe his/her documents. The labels are commonly known as tags and the labeling process is called tagging. Tags are helpful in improving the efficiency of search-engine, because the content is categorised using a familiar, accessible, and shared vocabulary. Tags also have an additional benefit of creating a group of related web resource[12]. This process of collaborative tagging emerged in this way is still developing as a means of organising information resources on the web.

5. CONCEPT OF TAG CLOUDING

Many changes have happened to the web in the past 10-15 years in terms of increased sophistication of the technology supporting internet and have led to an online paradigm shift commonly label web 2.0. The concept as described by Tim O'Reilly as web platform with collective intelligence, rich user experience, and light weight programming. These concepts of web 2.0 have applicability of remixibility, convergence, and participation amongst the user of internet[13]. These phenomenon's are taking place since the time of Carl (Carolus) Linnaeus, the father of taxonomy, to classify the living organism in to Kingdom-Phylum-Sub Phylum-Class-Family-Genus. On the same principle library resources are also being classified according to their subject matter based on the same or other principle devised by Melville Dewey (The Dewey Decimal Classification) and S. R. Ranganathan (Colon Classification). Dewey system organises books on library shelves in a specific and repeatable order that makes it easy to find any book and return it to its proper place. On the other hand, Colon classification is a system based on hierarchical classification, applying the principle of development from the general to the specific in disciplinary and subject relationships, under the terminology Faceted classification scheme[14].

Every school of library and information science has incorporated taxonomy into their classification and cataloguing curriculum to teach structure of the subject. This taxonomy is the 'science of finding, describing, classifying, and naming organisms (here in this case, books and other reading materials). Over the time, the concept of taxonomy has come to include categorisation of not just 'organism' but anything that can be categorised, given the phrase the 'Hierarchical system of classification'. The National Library of Medicine Medical Subject Heading (MeSH) and Library of Congress (LCSH) are using this system. With the convergence of online information system and development in computer programming, a new social service, known as bookmarking, has given user's a flexibility to comment on the subject in their own way. During this process, reader assigns keywords or tags for which they can search it at later stage. Tagging is implicitly a social indexing process; in which users share their tags and resources, constructing a social tag index, so-called 'folksonomy'. Del.icio.us is one of the example website which provides this facility. This tagging system has been defined by information scientist in various ways depend upon its use in the context. Congregation of tag words results in Tag Clouding. A tag cloud is a text based visual representation of set of tags which usually depicts tag importance by font size and/or font color. Recent trends in social and collaborative software, social cataloging, and web based library and information management environment have greatly increased popularity of tag representation form. A tag cloud provides users with comprehensible overview on the content of large tagged repositories. Each of the individual tag links to relevant subsets of repository content and tag clouds act as an instrument for topical browsing of the content[15]. Tag clouds occupy a peculiar niche in the domain of visualisation.

Many authors have studied the purpose of tag clouding for various purposes. Kuo *et al.,* evaluated the pubcloud interface by comparing it to the PubMed list layout[16]. The authors found that the response time was less when using lists but the quality of the answer was improved by the cloud layout. Halvey & Keane evaluated the influence of the different tag cloud properties on search time[17]. They compared vertical and horizontal list of tag cloud, all of them either sorted by tag's relevance alphabetically or using larger font size of the text in tags. They conclude that alphabetisation can aid users to find information more easily and quickly and also the font size helps in identification of information more quickly than smaller size. Flicker tags for finding landmark photos[18], use of social tags for classification of web pages[19, 20], application of falkosonomies for social classifications of resources[21] were some of the highlighted applications demonstrated by these authors. Gupta *et al.,* surveyed and summarised different techniques employed to study various aspects of tagging[22] and again Gupta *et al.,* discussed the various properties of tags in terms of tag streams, tagging models, tag semantics, generating recommendations using tags, visualisations of tags, applications of tags and problems associated with tag usage[23].

As the technology is changing, libraries are also adopting new techniques for getting to person's personal interest. These changes in new technologies are the application of web technology such as social networks, syndication techniques, and information retrieval technologies. Information communication

through OPAC enhancement is one of the important aspects that libraries should consider as media for information literacy[24]. Another study suggested the specific applications of the tags as a lightweight way of enhancing descriptions of on-line information resources and improving access to resources through broader indexing. Searching tags can enable the discovery of relevant resources and developing social relationship among taggers and act as a means of information discovery[25,26].

6. TAGGING AND TAG CLOUD WITH SHELFARI

Library is making huge investment for acquiring different kinds of books every month on various subject areas. Each book, which is being received in the library, metadata elements set is prepared for title, author/s, publisher, subject heading, purchased date, department group, and placed in the category 'currently reading'. These element sets are also updated at the new arrival lounge of Learning Resource Center on Shelfari. The tag/s or the words, which describe the central theme of the books, are assigned in Tag section on Shelfari. This tag word is used as "subject heading" of the book. The subject heading is based on Library of Congress Subject Headings (LCSH) or Medical Subject Heading (MeSH). Being an engineering institution the subject heading have its implications in co-relating subjects of the title. Jill O'Neill describes subject heading as a 'language that is more in-keeping with users approach to information seeking rather than more traditional classification schemes'[28]. The subject headings assigned to each book automatically converted into 'tag clouds'. Different members tag differently for a particular book. Most commonly used tags are represented by larger font size.

The tagging system contributes an exciting feature on Shelfari to create a visual information retrieval[29]. The tag cloud is a simple visual interface model, which may have some limitation as it depend upon the frequency of use of tag words. Each items catalogued on Shelfari can be rated, reviewed and flagged as all time favorite items.

7. RESULTS AND DISCUSSION

The distribution and creation of the tag cloud depends upon the fonts of the texts in cloud. In principle, the font size of a tag in a tag cloud is determined by its incidence. For a word cloud of categories like weblogs, the frequency of use for example, corresponds to the number of weblog entries that are assigned to a category. For small frequencies it is sufficient to indicate directly for any number from one to a maximum font size. For larger values, a scaling should be made. In a linear normalisation, the weight t_i of a descriptor is mapped to a size scale of 1 through f, where t_{min} and t_{max} are specifying the range of available

$$s_i = \left\lceil \frac{f_{\max} \cdot (t_i - t_{\min})}{t_{\max} - t_{\min}} \right\rceil \text{ for } t_i > t_{min}; \text{ else } s_i = 1$$

s_i: display font size; f_{max}: max. fontsize; t_i: count; t_{min}: min. count; t_{max}: max. count

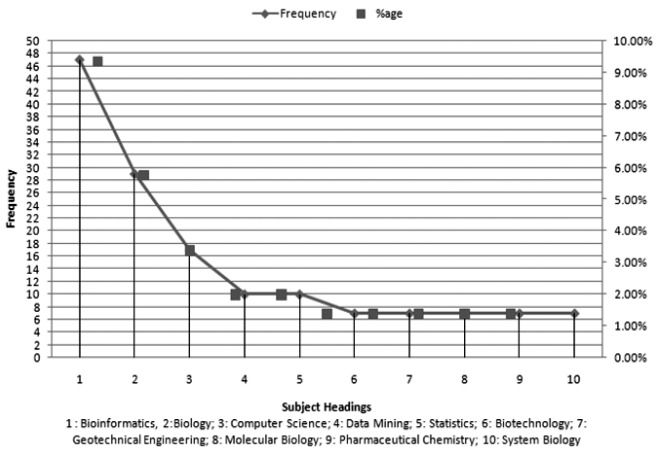

Figure 2. Picture of collection on virtual shelve.

weights[30].

After careful observation of the tag cloud, it can be visualised that, there are certain words which have larger font size and appear bigger in size. The increase in font size indicates the number of resources available or catalogued in the virtual shelves. Fig. 2 gives the overall pictures of the top ten subjects heading, used as tag, on virtual shelve of Learning Resource Center, JUIT. Out of 503 tag words used on this virtual shelve 'Bioinformatics (47, 9.37 %), Biology (29, 5.7 %), Computer Science (17, 3.38 %), Data Mining and Statistics (10, 1.99 %), Geotechnical Engineering, Molecular biology, Pharmaceutical Chemistry (Each 7, 1.39 %) are the some of the larger font words and represents larger collection.

The above data analysis gives an idea about the content of the tags available over Shelfari. But, the real challenge was whether this service of the library, actually had any impact on the users in terms of retrieval of information through tag clouds. In order to assess the impact; a feedback survey was conducted amongst the users of the library. A total of 352 responses were received. Table 1 presents the result of the survey. When asked about profile of library on Shelfari for display of new arrivals, 38.15 % of student and 46.67 % of faculty responded that they are not aware about this, in contrast, 61.85 % student and 53.33 % faculty accepted that they are not aware of this service.

The result of impact of this service is quite satisfactory in terms that 146 users who were aware of the services also responded on the impact. Out

Table 1. Awareness about the profile of library over Shelfari

Category of respondent	Total no. of respondent (N=352)	Aware of virtual shelves on shelfari	Not aware of virtual shelves on shelfari
Student	173	38.15 %	61.85 %
Faculty	90	46.67 %	53.33 %
Others	89	42.50 %	57.50 %

of these 51.52 % student and 64.29 % of faculty feel that the tag cloud has helped them in knowing about the new books in library (Table 2). There were negative feedbacks also, which may be useful for the understanding about specific reason and may

Table 2. Impact of tag cloud on awareness about new arrivals in the library through Shelfari

Category of respondent	Total no. of respondent (146)	Awareness of new books	
		No	Yes
Student	66	48.48 %	51.52 %
Faculty	42	35.71 %	64.29 %
Others (Non-Teaching and Alumni)	38	31.58 %	68.42 %

be helpful in improving the service in future.

8. CONCLUSIONS

This paper has presented a practical approach of creating information literacy about arrivals of newly catalogued books in the library. While using this web-based technology the library has experienced one of the novel ideas of visual information retrieval and information literacy mechanism to create awareness of new arrivals. Since the data and information are very small, still it requires a comprehensive study for future development and testing of the usefulness of this service. Some of the earlier works in this direction have already demonstrated the application of social cataloguing activity in the library and information centers of the global community[31, 32] and these trends could be possible more elaborative ways as the information and communication technology is having great potential and impact over all subjects including library.

REFERENCES

1. Mendes, L.H.; *et al.,* Subjecting the catalog to tagging. *Library Hi Tech*, 2009, **27**(1), 30-41.
2. Shiri, A. & Revie, C. Thesauri on the web: Current developments and trends. *Online Inf. Rev,* 2000, **24**(4), 273–79.
3. Shiri, A. An examination of social tagging interface features and functionalities: An analytical comparison, *Online Inf. Rev.*, 2009, **33**(5), 901–19.
4. Vander Wal, T. Folksonomy, 2007. http://vanderwal.net/folksonomy.html. (accessed on 16 February 2014).
5. Santolaria, A.M. Librarything as a library service. Assessment report ((2009). http://e-collection.library.ethz.ch/eserv/eth:777/eth-777-01.pdf. (accessed on 16 February 2014).
6. Understanding metadata. http://www.niso.org/publications/press/UnderstandingMetadata.pdf. (accessed on 14 February 2014)
7. Lancaster, F.W. Information retrieval systems: Characteristics, testing and evaluation, Ed.2. John Wiley & Sons, Chichester, 1979.
8. Steele, L. Collaborative tagging as distributed cognition. *Pragmatics & Cognition*, 2006, **14**(2), 287-92.
9. Steele, T. The new cooperative cataloging, *Lib. Hi Tech*, 2009, **27**(1), 68-77.
10. Speller, E. Collaborative tagging, folksonomies, distributed classification or ethnoclassification: A literature review. *Lib. Stu. J.*, 2007. http://www.librarystudentjournal.org/index.php/lsj/article/view/45/58 (accessed on 16 February 2014).
11. Peters, I. & Stock, W.G. Folksonomy and information retrieval. Proceedings of the American Society for Information Science and Technology, 2007, **44**(1), 1-28.
12. Shirky. C. Ontology is overrated: Categories, links and tags, 2005. http://www.shirky.com/writings/ontology_overrated.html (accessed on 16 February 2014).
13. O'Reilly, T. What is web 2.0: Design patterns and business models for the next generation of software, 2007. http://mpra.ub.uni-muenchen.de/4580/ (accessed on 16 February 2014).
14. Heiner-Freiling, M. & Landry, P. The use of the Dewey Decimal Classification (DDC) for the organisation of national bibliographics: Switzerland and Germany/Austria. IFLA General Conference and Council, Oslo, Norway, 2005.
15. Seifert C; *et al.* On the beauty and usability of tag clouds. *In* Proceedings of the International Conference on Information Visualisation, London, 2008, pp.17-25.
16. Kuo, B.Y. L, *et al.* Tag clouds for summarising web search results. *In* Advances in information retrieval (LNCS: 5478), Springer, Germany, 2009, pp. 654-61.
17. Halvey, M. & Keane, M.T. An assessment of tag presentation techniques. *In* Proceedings of the 16th International World Wide Web Conference, Banff, Alberta, Canada, 2007.
18. Abbasi, R,*et al.* Exploiting flickr tags and groups for finding landmark photos. *In* ECIR '09: Proceedings of the 31st European Conference on IR Research. Springer-Verlag ,Berlin, Heidelberg, 2009,pp. 654–661

19. Aliakbary, S, *et al.* Webpage classification using social tags. *In* Proceedings of IEEE International Conference on Computational Science and Engineering, 2009, **4**, 588-93.
20. Castano ,S; Ferrara ,A; Montanelli, S. Mining topic clouds from social data. *In* Proceedings of the 5th International Conference on Management of Emergent Digital Eco-Systems, 2013, pp.108-12.
21. Zubiaga, A, *et al.* Harnessing folksonomies to produce a social classification of resources, knowledge and data engineering. IEEE transactions, 2012, **5**(8), 1801-813.
22. Gupta, M., *et al.* Survey on social tagging techniques, ACM SIGKDD explorations newsletter, 2010, **12**(1), 58-72.
23. Gupta Manish; *et al.,* An overview of social tagging and applications. *edited by* Charu C. Aggarwal Social Network Data Analytics.Springer, Netherlands: 2011, pp. 447-497.
24. Ram,S. Information Literacy literacy through web 2.0 integrated Webopac: An experiment at Jaypee Group of institution. *DESIDOC J. Lib. Inf. Technol.*, 2010, **30**(3), 43-50.
25. Marlow, C., *et al.* Tagging paper, taxonomy, Flickr, academic article, to read. *In* Proceedings of the 17th Conference on Hypertext and Hypermedia. ACM, New York, 2006, 31-40.
26. Trant J. Studying social tagging and folksonomy: A review and framework. *J. Dig. Inf.*, 2009, **10**(1). https://journals.tdl.org/jodi/index.php/jodi/article/view/269/278
27. Shelfari joins the Amazon.com family (2008), Shelfari.com. http://shelfari.typepad.com/my_weblog/2008/08/shelfari-joins-the-amazoncom-family.html (accessed on 16 February 2014).
28. O'Neill, J. LibraryThing: Cataloguing for the (Social) masses. *Information Today*, 2007, **24** (8), 23.
29. Hassan-Monteroa, Y. & Herrero-Solana, V. Improving tag-clouds as visual information retrieval interfaces. International Conference on Multidisciplinary Information Sciences and Technologies, Mérida, Spain, 2006.
30. Lohmann, S.; Ziegler, J. & Tetzlaff, L. Comparison of Tag Cloud Layouts: Task-related performance and visual exploration. *edited by* T. Gross, *et al.* INTERACT 2009, Part-I, LNCS, 2009, **57**(26), 392–404.
31. Jeffries, S. Social cataloging tools: A comparison and application for librarians. *Library Hi Tech News*, 2008, **25**(10), 1-4.
32. Hvass, A. Cataloguing with LibraryThing: As easy as 1,2,3!, *Library Hi Tech News*, 2007, **25**(10), 5–7.

About the Author

Dr Shri Ram is working as Deputy Librarian at Jaypee University of Information Technology (JUIT), Solan since 2002. He holds PhD degree in Library and Information Science ,along with MSc in (Anthropology). He has also qualified UGC-NET in Library and Information Science. Prior to joining JUIT, he has worked with Indian Council of Medical Research, New Delhi and Commonwealth Educational Media Center for Asia in various capacities. He has been awarded with *Commonwealth Professional Fellowship 2013* and went to University of East London, London, UK during September–December 2013. He has published 19 papers in various national and international journals, 12 papers in national and international conferences, and also edited two conference volumes. His research areas include: Implementation of information and communication technology, multimedia-based information literacy, user education, library automation, and mobile application in libraries.