

IMPLEMENTATION OF FOLKSONOMY BASED TAG CLOUD MODEL FOR INFORMATION RETRIEVAL FROM DOCUMENT REPOSITORY IN AN INDIAN UNIVERSITY

Sohil D. Pandya¹ Paresh V. Virparia² and Rinku Chavda³

¹Sardar Vallabhbhai Patel Inst. of Technology (SVIT), Vasad, India

²G H Patel PG Dept of Computer Science & Technology, Sardar Patel University, Vallabh Vidyanagar, India

³Sardar Vallabhbhai Patel Inst. of Technology (SVIT), Vasad, India

ABSTRACT

In the magnitude of internet one need to devote extra time to investigate anticipated resource, especially when one need to search information from documents. For the higher range internet there is serious need to demand the essentiality to discover the reserved resources. One of the solutions for information retrieval from document repository is to attach tags to documents. Numerous online social bookmarking services permit users to attach tags with resources which are eventually meta-data, frequently stated as folksonomy. In current paper, authors implemented this model for information retrieval by utilizing these tags, after retrieving by using delicious API and synthesize tag cloud in an Indian University to search and retrieve information from document repository.

KEYWORDS

Folksonomy, metadata, social bookmarking services, tag cloud

1. INTRODUCTION

As the magnitude of Internet is growing in recent years, we need to spend more and more time on Internet to search for specific resource. Even to search within an organization document repository is a difficult and time-consuming task. Because they are unstructured in nature, unorganized in storage, naming conventions are different, but they are needed to be retrieved as and when required. In an organization like University, various notifications, circulars, notes are published regularly in general. To become specific, authors have taken an example of Gujarat Technological University (GTU), which regularly updates its website contents.

GTU publishes all its circulars, notification, tenders, recruitments notices, academic calendars, examination schedules, results, etc. on its website (<http://www.gtu.ac.in>) regularly. The GTU has driven a paradigm shift among various educational institutes, faculty members and students in access and reach ability. Today more than 10,000 circulars or notifications are there on the website. GTU nicely manages the repositories of circulars with respect to time frame. Anyone can find any circular on the website, if there is an idea of date or month and year when it is published.

In this paper, authors tried to implement folksonomy based tag cloud model for information retrieval. One of the reasons is of that people identify the same resource with different words.

Popular social bookmarking services like Delicious, Digg, etc. allow us to add our own words, i.e. tags, to specific resources while we are doing online bookmarking of the resources. By adding extra words to the resources builds up pool of words chosen by people called Folksonomy. These tags are ultimately meta-data, but they are labeled by users. This folksonomy can be used for information retrieval by generating tag cloud. Tag cloud is visual representation of folksonomy based on frequency of words.

Here, authors have attached tags to various circulars which will be used later to retrieve circulars. Users will have a tag cloud as an interface where they are allowed to see tag related circulars. Authors also have given an interface where user can move generic to specific by clicking tags.

In this paper, authors have proposed a model for information retrieval by generating folksonomy based tag cloud. Authors wish to study other social tagging based platforms in future..

2. RELATED WORK

A.W. Rivadeneira et al [1] and Kim et al [4] studied tag clouds and its evolution. Trant et al [11] have proposed a framework for tagging system. Hearst et al [2] considered tagging as social signals which needs to be identified and can be used for various situations.

James Sinclair et al [3] and Knautz [5] compared traditional search box and tag cloud approaches to identify users' behavior. They have simulated 10 questions for 89 users and tried to see in which type of questions users were selecting different approaches. And found users' tendency towards tag cloud based approach.

Satoshi Niwa et al [8] proposed a web page recommender system using tag generated by users. J Trant et al [11] have studied folksonomy and found that this data can be further analyzed using various tools like vocabulary analysis, classification, user interaction theory and social network theory.

Kuo et al [6] found that tag cloud offer not only an overview of knowledge but also an interface that enables users to navigate to potentially relevant information hidden.

M A Hearst et al [7] found that tag clouds are social signal or mark and are more suggestive device than as a precise depiction of the underlying phenomenon.

Earlier, the access of these tags was limited. But due to availability of open source social bookmarking systems, authors wish to use these meta-data and generate folksonomy based tag cloud for information retrieval.

3. DELICIOUS API, FOLKSONOMY AND TAG CLOUD

<http://www.delicious.com> is one of the popular website for social bookmarking over the Internet; it is also called web based tagging system. It not only allows adding URL as a bookmark, but it also allows adding some extra information related to the URL like title, keywords and remark [16]. The website also provide to search among all the bookmarks added by other users also (provided the users have added bookmarks as public).

The keywords added during bookmarking act as tags to content or resource, which is also a way the users identify or recognize or categorize the content. A folksonomy is a system of classification derived from the practice and method of collaboratively creating and translating

tags to annotate and categorize the content, which is also known as collaborative tagging, social classification, social indexing, and social tagging.

The Fig. 1 shows a conceptual model for social tagging systems. In the model, users assign tags to a specific resource; tags are represented as edges connecting users and resources. Resources may also be connected to each other (as links between web pages) and users may be associated by a social network, or set of affiliation (like users works for the same organizations).

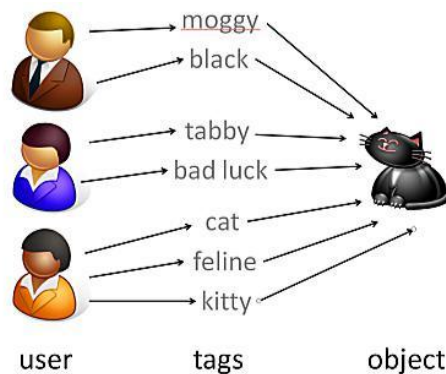


Figure 1. Conceptual Model for Social Tagging

Now this database of folksonomy is a great pool of knowledge as they are generated collectively by users themselves. It is having lots of potential for satisfying users' information need. It can be represented in many ways like tag cloud, tag clusters, etc.

A tag cloud (word cloud, or weighted list in visual design) is a visual representation for text data, typically used to depict keyword metadata (tags) on websites, or to visualize free form text [16]. Fig. 2 is an example of tag cloud (which has having tags sized as per population of the countries). A tag cloud can be of two types – based on frequency or category. In this paper we emphasize on frequency based tag cloud. In the later sections we have described how it can be generated.

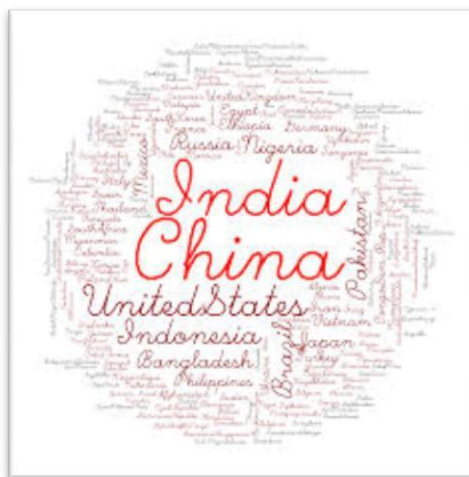


Figure 2. Example of tag cloud of Countries with reference to population

4. PROPOSED MODEL

4.1. Function of Generating Tag Cloud

```
$(function(){ $("#div-  
id").jqCloud(word_list);  
});
```

Where,

- jqCloud [13] is a method of jqCloud (an extension of jQuery [14, 15] for generating tag cloud, which can be download from www.github.com)
- word_list is list of words with weight (weight mentioning font size), which can be computed using following formula [16]:

$$S_i = \left[\frac{f_{max} * (t_i - t_{min})}{(t_{max} - t_{min})} \right]$$

Where,

S_i = display font size
 f_{max} = max. font size
 t_i = count
 t_{min} = min. count
 t_{max} = max. count

For larger values logarithmic representation may make sense.

- #div-id is the portion of a page where you want to place tag cloud.

4.2. Function of Retrieving Tags from Delicious

```
$(function(){  
  
String word_list[]=getDeliciousTags(String Delicious_Username,Delicious_Password, String t);  
});
```

Where,

getDeliciousTags() is method which using Delicious API [12] to get tags with parameters like username, password, search string, if any. The function returns tags with frequencies.

4.3. Steps and Visualization

The entire model is divided in to two clear portions: (i) fetching tags from Delicious database using APIs and (ii) Generating tag clouds. For both of tasks functions have been defined earlier in this section. To have independent model and view mechanisms, both are kept separate.

Step 1: Initial tag cloud is generated by retrieving tags.

Step 2: User click on any tag and corresponding tag cloud is generated by retrieving tags (as shown in Figure. 3).

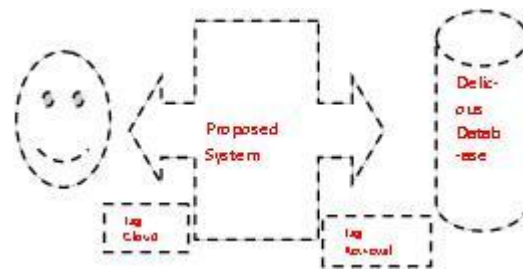


Figure 3. Visualization of Steps

4.4. Demonstrating Expected Outcomes

Authors are intended to implement the proposed model and expected outcome will be tag cloud based information retrieval interface of circulars of the said University. Authors experimented this by letting a normal user to add circulars over Delicious with keywords chosen by him. After adding of around 2000 circulars with tags, both the above mentioned phases implemented. Its demonstration can be seen below:

Interaction 1:



Figure 4. Initial Tag Cloud for Course Options Available after XII Sci.

Interaction 2:

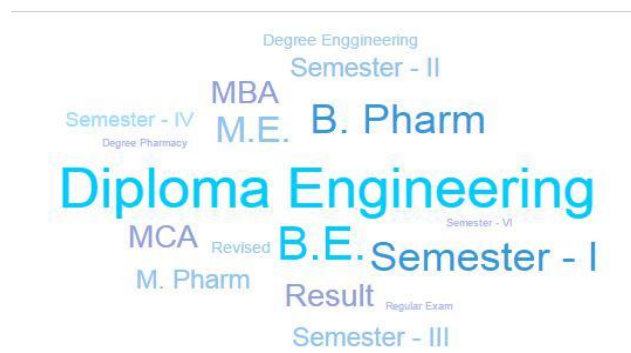


Figure 5. Tag Cloud Generated after Selection of Engineering

Interaction 3:

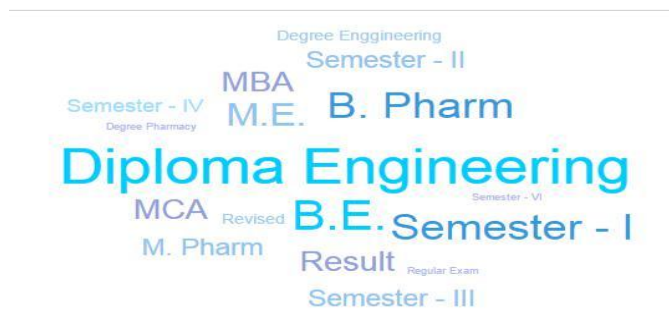


Figure 6. Tag Cloud Generated after Selection of Diploma Engineering

4.5. Applications

- Information retrieval
- Quick visual search
- Anywhere where generic to specific path approach is followed
- Studying trends
- Developing environment to learn local/region/industry specific Folksonomies

4.6. Limitations

Resources restricted by people because of private mode may not be accessible to API.

Resources are identified and named by people, hence if people may wrongly / useless /extra / little tagging may lead to wrongly generated tag cloud.

As there are millions of resources available to API, the tag cloud is limited to top 50 tags, which can be further specific in next level (sub) to top 50 tags.

5. CONCLUSIONS

The above proposed model will not only improve information retrieval but can be implemented for alternative mechanism over conventional query based search approach. Authors wish to implement the above model and explore experimental results to prove its efficiency. In this paper, we have used Delicious API, but one can explore any other open source social tagging services to generate the tag clouds.

REFERENCES

- [1] A.W. Rivadeneira, D.M. Gruen, M.J. Muller, and D.R. Millen, "Getting our head in the clouds: Toward evaluation studies of tagclouds", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2007, pp. 995–998.
- [2] Hearst, Marti A., and Daniela Rosner. "Tag clouds: Data analysis tool or social signaller?" In Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, pp. 160-160. IEEE, 2008.
- [3] James Sinclair, and Michael Cardew-Hall, "The folksonomy tag cloud: when is it useful?" in Journal of Information Science, Vol 34 No 1, pp. 15-29, May 2007.
- [4] Kim, Hak-Lae, John G. Breslin, Sung-Kwon Yang, and Hong-Gee Kim. "Social semantic cloud of tag: Semantic model for social tagging." In Agent and Multi-Agent Systems: Technologies and Applications, pp. 83-92. Springer Berlin Heidelberg, 2008.

- [5] Knautz, Kathrin, Simone Soubusta, and Wolfgang G. Stock. "Tag clusters as information retrieval interfaces." In System Sciences (HICSS), 2010 43rd Hawaii International Conference on, pp. 1-10. IEEE, 2010.
- [6] Kuo, Byron YL, Thomas Hentrich, Benjamin M. Good, and Mark D. Wilkinson. "Tag clouds for summarizing web search results." In Proceedings of the 16th international conference on World Wide Web, pp. 1203-1204. ACM, 2007.
- [7] M.A. Hearst, and D. Rosner, "Tag clouds: Data analysis tool or social signaller?“, Proceedings of 41st Hawaii International Conference on System Sciences (HICSS 2008), Social Spaces minitrack, 2008.
- [8] Satoshi Niwa, Takuo Doi, and Shinichi Honiden, "Web Page Recommender System based on Folksonomy Mining” in Proc. of the Third International Conference on Information Technology: New Generation (ITNG'06), pp. 388-393, April 2006.
- [9] Seifert, Christin, Barbara Kump, Wolfgang Kienreich, Gisela Granitzer, and Michael Granitzer. "On the beauty and usability of tag clouds." In Information Visualisation, 2008. IV'08. 12th International Conference, pp. 17-25. IEEE, 2008.
- [10] T. Russell, "Cloudalicious: Folksonomy over time“, Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries, 2006, p.364.
- [11] Trant, Jennifer. "Studying social tagging and folksonomy: A review and framework." Journal of Digital Information 10, no. 1, 2009.
- [12] <https://delicious.com/developers>
- [13] <https://github.com/lucaong/jQCloud>
- [14] <http://jquery.com>
- [15] <http://www.w3schools.com/jquery>
- [16] <http://en.wikipedia.org>

AUTHORS

Paresh V Virparia joined the Department of Computer Science of Sardar Patel University, Vallabh Vidyanagar in 1989 and currently working as a Director and Professor. He completed his MCA in 1989 from Sardar Patel University and Ph. D. in 2002 from Sardar Patel University. SEVEN research scholars have completed their Ph.D. (Computer Science) under his guidance. Currently, SEVEN students are doing their Ph. D. under the guidance of him. Also, three students have completed their M.Phil. (Comp. Sc.) under his supervision. His publications include 40 papers in International Journal, 16 papers in National Journals and 47 papers in national conferences/seminars. His research interests include the areas of Computer Simulation & Modeling, Data Mining, Networking and IT enabled services. He is an editor and editorial review board member in several journals/magazines.



Sohil D. Pandya is working as an Asst. Professor at Sardar Vallabhbhai Patel Institute of Technology (SVIT), Vasad. He has 10+ years experience of teaching in MCA. He has published more than 6 Papers in International/ National Journals and presented more than 6 papers in International/ National conferences. His research interest includes data mining and information retrieval. He is an editor and editorial review board member in some journals.



Rinku Chavda is working as an Asst. Professor at Sardar Vallabhbhai Patel Institute of Technology (SVIT), Vasad. She has 4+ years experience of teaching in MCA. She has published 2 Papers in International/ National Journals. She has expertise in Android App development and has guided several Android App projects.

