

# Semantic Analysis of Tags to Renovate Folksonomies in Social Tagging System

J.Annie Jeba Dafney, A.Lourdes Mary

**Abstract** - Grouping resources into set of classes allows easy access to the resources we use in our day-to-day lives. This classification makes the search faster and easier. The process of classifying the resources manually becomes expensive. This cost effectiveness switch to automated classification of resources which depends on the content of the data. This paper deals with the semantic analysis of tags. Collaborative tagging system allows any user to annotate the web resources. The user annotations can be useful to discover the aboutness of resources and also helps to ascertain classification. The social tagging system focuses on using support vector machine (SVM) as a state-of-the-art classification algorithm. We have two large scale social tagging data sets to interpret the characteristics of social classification. This System supports user defined annotations which immensely renovate folksonomies.

**Index Terms**— Collaborative tagging system, Social Tagging System, social classification, Support Vector Machine, user annotations, semantic analysis, Folksonomies.

---

## 1 INTRODUCTION

Grouping resources that we use in our day to day lives such as web pages, books, songs, etc. will make our search faster. The classification of such resources by manually is a cost effective job. Hence the arrival of automated classifiers can make the categorization easier. The social tagging is designed for the effective information retrieval. As the data is growing very huge in web the manual categorization of data is not at all an easy task. So it is not possible for an expert or an author to classify all the documents based on their categories. Hence the emergence of Social Tagging System paved a new way for the classification of data by user provided tags. The STS allows the user to annotate the images, web pages, songs, videos, books, etc. The user annotations can enhance the classification task at free of cost just by the tags given by the user.

This tagging will improve search and navigation facilities. Tags are the means to collect, store and retrieve the resources based on the keywords provided. For instance it is possible for

a librarian to categorize nearly 1000 books in a library. But it is not easy in case of web, as the data is growing each and every day. Hence it is not possible to classify all these resources by single person. Tagging enables such classification based on the keywords provided by the user. The Social Tagging Systems supports free form tagging or uncontrolled vocabulary. The free form tagging allows any user to tag on any resource as they wish while uncontrolled vocabulary represents the freedom to tag, which means any keyword to any content without any restrictions. The popular tags form an organization called folksonomy which is a portmanteau of the words folk and taxonomy defined by Vanderwal. This implies that the organization of the contents available on the web by the users. Folksonomy is a free form tagging user supplied categorization system. It can enhance the search, information discovery, and retrieval. As there is no control over tagging any user can tag any keyword to any resource even if the tag is not relevant to that content. This uncontrolled vocabulary and free-form tagging leads to semantic ambiguities such as synonymy, polysemy, base form variation, specificity, relatedness, spam etc.

In this paper we have explored the semantic ambiguities such as synonymy and polysemy in social tagging systems such as delicious and LibraryThing. Synonymy refers to different words with same meaning, for instance the tags "image" and "picture" may be used by the users will be treated as different tags by the system but they have same meaning. The search for image will not retrieve the document tagged picture thus

- 
- J.Annie Jeba Dafney is currently pursuing masters degree program in Computer Science and Engineering in Scad college of Engineering and Technology, Tirunelveli, Tamilnadu, India.  
E-mail: [anniejdafney@gmail.com](mailto:anniejdafney@gmail.com)
  - A.Lourdes Mary is currently working as an Associate Professor in Department of Computer Science and Engineering in Scad college of Engineering and Technology, Tirunelveli, Tamilnadu India.  
E-mail: [lourdesmaryamulu@gmail.com](mailto:lourdesmaryamulu@gmail.com)

yielding incomplete results. Polysemy refers to a word has two or more similar meanings, for example the word "bank" may be referred to a financial institutions as well as the river bank. Eliminating such semantic ambiguities will make the user to understand the role of each tag and make the retrieval more effective and accurate. Semantic analysis of tags in social tagging system improves the search by understanding the users query and also makes it more users friendly.

## 2 RELATED WORK

As the social tagging system grows exponentially the quality of huge amount of data can be maintained only when the data is organized properly. The organization of data available on the web is not at all an easy task. Due to the vast available content on the web and changing nature of the data needs some automatic methods for classification [5]. The existing system already has the problem of noisiness and lack of effective information retrieval. The automatic classification method needs to separate the noisy part from noiseless part. The social tag extracts the content with less noisy content, high speed and stability. Tagging is the process of marking content with descriptive terms called keywords or tags which can improve search and navigation facilities. Collaborative tagging allows anyone to freely attach any keyword to the content available. The social tagging sites allow users to tag and share content which not only categorize and also search the content which is tagged by others. Collaborative Tagging and social annotations provided by authors, publishers, expert editors and end users to find out the kind of document which is annotated the most and how the folksonomy is maintained by classification. This work has the limitation of disambiguation of keywords. Golder and Humberman [2] work shows the dynamic aspects of collaborative tagging. This in turn retrieves the content with the union of tags not the intersection of tags.

Recommender System requires users' explicit rating and implicit rating information. In real time the explicit rating is not always present [8]. Implicit ratings are provided by the user generated tags to make the classification. As there is no restriction on tagging the tags used by users are free formed tags and hence it contains semantic ambiguity which means same tag name has different meaning for different users and tags called synonyms. Most of the tags that is nearly 60% of tags used in many sites are personal tags which are used by a single user. The recommendation system provides multiple relationships among users, items, and tags. It will further helps to find the related tags that an individual user tags as well as find each user most interested items.

Social annotations provided by the readers of documents, hyperlink anchor text provided by authors and search queries of users to find the documents. These three different metadata used in delicious and Flickr dataset investigated a number of characteristics such as length, novelty, diversity and similarity. This work does not study more about the search queries. Noll and Meinel explore the information provided by the authors and publishers of web documents compared with the metadata provided by the users of the same content [21], [23]. The user provided tags do not provide any meaningful information about the content. The web document which is not bookmarked will not be allowed to tag in the site delicious. The content based approaches do not able to retrieve any meaningful information. Shipitsen, Gemmill, Mobasher, and Bruke proposed a personalization algorithm for recommendation in folksonomies [2]. The work of Symeonidis, Nanopoulos, and Manolopoulos shows the 3-order or tensor semantic analysis which focuses on tags, items and users with common interest [9]. Social annotations provided by the end users on social classification of resources. It focuses on the Support Vector Machines (SVMs) as the state-of-the-art classification algorithm. It has the limitation of semantic ambiguity.

## 3 TAGGING SYSTEM

Tagging is also called as collaborative tagging, social classification, folksonomy, etc. The Tagging websites has increased since 2004. The word *folksonomy* is a blend of the words folk and taxonomy where folk refers to people and taxonomy refers to classification. Hence folksonomy is literally means a classification done by users (people). Folksonomy based systems enable the user to categorize the resources with tags. It is useful when there is no one in the role of librarian or a huge amount of content is classified by a single authority. Tagging system allows the user to publicly share and tag the content. This Tagging not only classify the information but also permits the user to browse the information which is classified by others. Some of the well known examples are delicious, furl, reddit, flickr. Tagging services that allow the user to manage the bibliographic metadata on the web are BibSonomy, Connotea, CiteULike, and LibraryThing. Metadata is a information about the webpages, books, songs, videos, documents, etc.

[2009 2010 2011 2013 Author](#)  
[general books challenge Challenge](#)  
[Group Challenges](#)[Christianity discussion](#) [fantasy fiction gam](#)  
[es genres](#) [Group](#)

[Group authors book-](#)  
[Reads Harry](#)

Potter help **history** horror Identity  
grouplib~~rary~~thing lists literature location-  
based mystery philosophy politics reading **religion** reviews **sc**  
**ience** **fictionsite** **related spam** translation vampires Yearly  
Group young adult

Fig.1. Tag cloud of LibraryThing

A folksonomy is a community based taxonomy where the classification is non-hierarchical. It allows all the users to tag the resource, not only by authors. Tags are usually defined for a resource, as a result of tagging a resource by many users, tagging the same item for a long time will generate a considerable set of tags available for each resource. For example delicious is a social bookmarking site where each URL (resource) can be tagged by as many as users who are interested in.

Social tagging sites have a tag cloud as a result of tagging. The tag cloud is a collection of most popular tags on that site while some tags with larger in size have more resources in it. Fig.1 shows the tag cloud of LibraryThing dataset. In social tagging system, there is a set of users (U) who add bookmarks (B) for a resource (R) which is annotated by tags (T). In delicious and LibraryThing the URL and books are the resources. Tagging will enhance search, navigation, retrieval, discovery, spam detection, identifying emerging topics globally within a community.

## 2.1 Dataset

The two social tagging systems such as delicious and LibraryThing are taken. *Delicious* allows the free and easy access to save, organize and discover interesting links. If anyone wants to add link delicious requires login using an email id. We can add link by typing the link while adding a link it ask for a tag to that link. It also suggests some tags and displays the details about the link. The link we have typed will be added to the user account and we can share, edit and even delete when needed. The user can also connect to the link whenever they want. Using delicious we can also search a specific tag, specific user, specific keywords, tag for a specific user, keywords for a specific user, a specific web page. In delicious we can find our friends in the network and start discovering the world around us.

*LibraryThing* is a home for books (library quality catalog). It connects us to the people who read. It also requires login before we have access. LibraryThing is a friend finder, connect us to the twitter and facebook account. Here we can add books to our library, wish list, currently reading, to read, read but not owned and favorites. LibraryThing allows the user to

search the books in the Library of Congress, Amazon.com, and over 690 libraries around the world. We can also add tags while we add books. Overcat searching allowed in LibraryThing which means searching can be done using author name, title, series, publication date, publisher and ISBN. It has nearly 102,854,851Tags added till now. It can recommend nearly 10-25 books for a user based on their interest. It also has the member recommendation facilities, reviews, statistics, authors and tag clouds. Both delicious and LibraryThing allows any user to tag the URL and books but there is a few differences between these sites. The LibraryThing does not suggest any tags will the user annotate the resource while the delicious suggest the tags about the URL. The LibraryThing will recommend books while delicious will not recommend anything to the user. LibraryThing has many facilities for the book readers. We can even read the books from the Library of Congress. It also allows the user to import the books they need. Generally LibraryThing is an ocean of books which connect the book readers by a community. These bookmarking sites will be more useful for the user for their later reference and search.

## 3 SEMANTIC ANALYSIS

Semantic analysis of tags is useful to understand the role of each tag and also enhance the search and navigation capabilities. The two datasets delicious and LibraryThing is taken for the research. Synonymy and polysemy, base form variation, spam detection, sparsity are some of the semantic ambiguities. The presence of such ambiguities in web documents results in incomplete search. Hence eliminating such ambiguities is more essential to make the retrieval more exact and accurate. The collected dataset delicious contains 193938 URLs are being tagged by many users with more than 4 tags for each resource. In the available resource 41204 URLs are having their appropriate web content. The unwanted tags and resource can be eliminated by manual validation. Nearly 776752 tags are being used in the appropriate resource. While in LibraryThing there is nearly 39576 books are being tagged. 114598 web content is collected. More than 791520 tags are used to tag the books. The resources (URL, books) are manually validated before preprocessing to eliminate the resource which is not tagged. When semantic analysis is applied to the dataset the number of tags obtained will be reduced by eliminating synonymy and polysemy. The architecture of semantic analysis is given in the Fig.2 The delicious and LibraryThing dataset including the tags posted by the user along the web content are being collected. The collected dataset is manually validated for any stop words removal and eliminating the resources without tags.

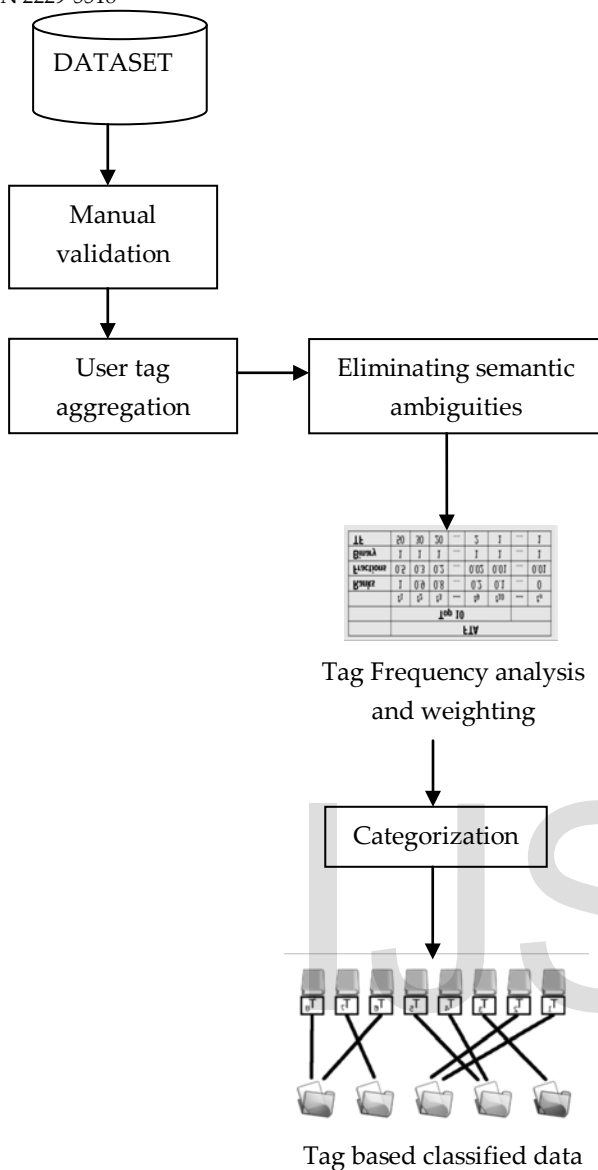


Fig.2. Semantic Analysis of Tags

The tags are aggregated into a unified model where the semantic ambiguities such as synonymy and polysemy are being eliminated. Synonymy is defined as the same word has different meaning while polysemy is defined as different words will have the same meaning. Elimination of such ambiguities will make the websites more and more user friendly, by understanding the user query and retrieves the exact content, what is needed by the end user. For instance the word image and picture infers same meaning even though if the search query is image it will not retrieve the content which is tagged as picture this is what the synonymy infers. Likewise the word bank refers to financial institution as well as the river bank. But if the search query is given as bank it retrieves the content related to the financial institution. This is the example

of polysemy. Our system provides a user friendly environment which retrieves all the content related to the query.

Tag Frequency analysis and ranking is done on the unified model. The tag frequency analysis represents the number of times the keyword is being tagged by the user. We analyze the number of tags that are used more, equal or less frequently in an item (i.e., resources, users or bookmarks) than in another. This reflects the large number of tags that users utilize just once on these sites. The weight is computed according to the fraction of users who utilize a tag, i.e., the number of users utilizing a tag on a resource, divided by the total number of users who annotated the resource. The only feature considered for this representation is the occurrence or nonoccurrence of a tag in the annotations of a resource. Frequency based tag weighting (term frequency (TF)). It considers the number of users assigning the tag (wt) as a weight. The weight for each of the tags of a resource is considered as it is in this approach. Term frequency-inverse document frequency (TF-IDF) is an inverse weighting function, which has been widely applied to text collections, and has proven to be beneficial for a large number of tasks. Classification is done based on the TF-IDF value. The finally the tag based classified data is obtained.

### 3.1 Tag Distribution

A tag which is posted by the user is checked with the web content available. The web content describes what the resource is actually composed of. The term frequency-inverse document frequency (tf-idf) is a weighting approach for ranking the tags. The tf-idf value gives the frequency of the tag by comparing it with the web content and returns how the tag is related to the web content. The tf-idf actually refers to how important the tag is in the document. The tf-idf value increases proportionally when the tag occurrence increases. In the collected dataset the most of the tags provided by the user are not related to the web content. This infers the user provided tags are mostly of their wish and it is not more useful for the other user for their search and retrieval.

The TF is:

$$tf(t, b) = 0.5 + \frac{0.5 \times f(t, b)}{\max\{f(t, b)\}}$$

The inverse document frequency is:

$$\log \frac{|D|}{\{|b \in D: t \in b\}}$$

Then tf-idf is:



$$tfidf(t, b, D) = tf(t, b) \times idf(t, D)$$

$$tf - idf_{ij} = tf_{ij} \times \log \frac{|D|}{|\{b:t_i \in b\}|}$$

Where  $|D|$  refers to the available number of documents,

$b$  refers to bookmark given,

$t$  refers to tags supplied.

The tf-idf value can be used for classification of tags.

### 3.2 Classification Algorithm

The dataset collected is first undergone manual validation and preprocessing. The words which considered unwanted and the document without tags are being eliminated. The tags are then aggregated along with the annotations provided. The classification is done based on the TF-IDF value. The tags are classified based on the Term frequency. Table I shows the tags with the term frequency of the delicious dataset. In the above table the sql, report, storage having the higher term frequency. This shows that the tag is more relevant to the web content. The tf-idf, cosine similarity are computed according to the relatedness of the tag to the web content. Finally based on the tf-idf the tags are classified to obtain the tag based classified data with the elimination of semantic ambiguity. The tags with higher frequency in the dataset are being classified. For instance in our dataset of LibraryThing dataset the 21 different types are given to the classification algorithm it classifies into only 8 tags. These tags are more relevant to the web content. The tags are classified as if they have higher frequency. The higher term frequency is obtained only if the tag is relevant to the web content. While collecting the dataset the tags, records along with the web content is obtained. The classified tags contain the tags with higher term frequency, inverse document frequency is 0 and cosine similarity is 0. The classification algorithm classifies depending on the term frequency of each tags. This classifies only the tags with higher term frequency other tags are being eliminated.

TABLE 1

Tags showing TF value of delicious dataset

Tags	Term Frequency
Sql, report, storage	0.938928
Dovecot, linux	0.198423

Suprememo, mind, train	0.341693
Free, download, book	0.308707
Opensource, virtualization, sun	0.193656
Internet, security	0.088276
Carlin, comedy, interview	0.011163
Mar, science, phoenix	0.021309
Rack, ruby	0.129412
Highlight, program, tool	4.69E-04

The cosine similarity between two vectors in machine learning is

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|}$$

where  $a, b$  are the two vectors or two documents on the vector space which calculates the cosine of the angle between them.

Cosine similarity calculates the comparison between the documents. It is a metric that measures the relation between two documents by using the angles. The cosine similarity is related to the tf-idf value of the machine learning.

The tf-idf term frequency-inverse document frequency is a numerical value that shows how the tag is important to the web content. This weighting approach is used in many social sites for scoring and ranking the relevancy of the documents. If the tag is related to the web content the term frequency will be increased this in turn makes the Inverse Document Frequency to 0. This will make the cosine similarity value to 0. The term frequency increases when the occurrence of the word in the document increases. TF-IDF value is used in the classification for effective information retrieval and text mining. In existing the Inverse Bookmark Frequency, Inverse Resource Frequency, Inverse User Frequency are being considered in order to get the relativity of the web content. But here Inverse Document Frequency (IDF) is used to identify the relatedness of the tags to the web content.

TABLE 2

Tags with TF-IDF and Cosine Similarity of Delicious Dataset

Tags	Term Frequency	IDF	Cosine Similarity
Sql, report, storage	0.938928	0	0

Dovecot, linux	0.198423	0	0
Supermemo, mind, train	0.341693	0	0
Free, download, book	0.308707	0	0
Opensource, virtualization, sun	0.193656	0	0
Internet, security	0.088276	0	0
Carlin, comedy, interview	0.011163	0	0
Mar, science, phoenix	0.021309	Infinity	NaN
Rack, ruby, serv	0.129412	Infinity	NaN
Highlight, program, tool	4.69E-04	Infinity	NaN

The cosine similarity value of 0 represents that the tag is more relevant to the content. If the cosine similarity is 0 only when the term frequency is high. This will make the Inverse document frequency to 0 which shows that the tag describes more about the content of the resource. The table II shows the Term frequency, inverse document frequency and cosine similarity. The IDF value of 0 and cosine similarity of 0 shows the relevancy of the tags to the web content. While the IDF, cosine similarity value of Infinity and NaN (not a number) shows that the tag is not relevant to the web content.

### 5 EXPERIMENTAL RESULTS

The experimental result of semantic analysis shows that the most the tags posted by the user does not describe about the content of the data. Mostly the people tag the word as they wish which may not relate to the content available.

Fig.3. Classification of delicious dataset using Term Frequency

The classification of the delicious dataset is shown in the Fig.3. Among 52 datasets 39 are being classified in the result set based upon the term frequency of the each tags. The tags with higher frequency or the tags which is more related to the content are being classified in the result set.

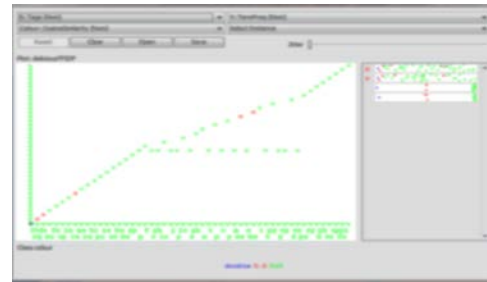


Fig.4. shows the tags versus term frequency in the delicious dataset

The graph shows the tags along with the term frequency in the delicious dataset.

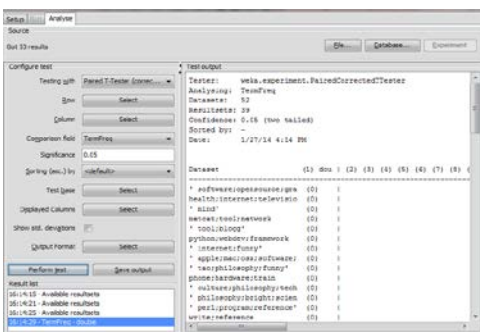


Fig.5. shows the tags versus term frequency in the LibraryThing dataset

The graph shows the tags along with the term frequency in the LibraryThing dataset.

The red spots show that the IDF value is infinity and the cosine similarity is NaN. While the green spots show that the IDF and cosine similarity is 0.

As the social tagging system allows free form tagging where any user can tag any keyword to the content that is the user can even tag the word as they wish without knowing the content of the web resources. This reduces the relevancy of the tags towards the content. The tf-idf value shows that the classified data based on the user tags does not provide useful information about the search. The tf-idf value of the tags related to the web content describes how the user provided tag is relevant to the document.



## 5.1 Analysis Graph

The Term Frequency of delicious is higher when compared to that of the LibraryThing. This is because the delicious suggest tags to the user while we are tagging a resource while LibraryThing will not suggest. Delicious is social tagging site will allows the user to tag any URL while LibraryThing allows the user to tag the books.



Fig.5. Graph showing Term Frequency of delicious and LibraryThing dataset

The result shows that if a social tagging system recommends something then the relevancy of the tag to the content is naturally increased. Hence the delicious has the higher tag frequency. In the LibraryThing the site does not suggest any tags but recommend books for the readers based on their interest and hence LibraryThing has lower term frequency when compared to the delicious.

## 6 CONCLUSION AND FUTURE WORK

Our experimental results of delicious and LibraryThing dataset shows that the user supplied tags does not infer more about the available content on the web. Most of the tags are unrelated to the web content. But the elimination of semantic ambiguities such as synonymy and polysemy enhance the search and navigation facilities a little more. As a future work we can do further semantic analysis of tags in other social tagging datasets in order to improve the search and navigation facilities. The semantic analysis of tags allows the user to understand the role of each tag and even improve the precision of recommendation in social tagging system.

## REFERENCES

[1] Arkaitz Zubiaga, Víctor Fresno, Raquel Martí'nez, and Alberto Pérez Garcí'a-Plaza, "Harnessing Folksonomies to

Produce a Social Classification of Resources", IEEE Transactions on knowledge and data engineering, vol.25 no.8, Aug 2013.

[2] S.Golder and B.A. Huberman, "The Structure of Collaborative Tagging Systems" J.Information Science, vol. 32, no. 2, pp. 198-208, 2006.

[3] A. Shepitsen, J. Gemmell, B. Mobasher, and R. Burke, "Personalized Recommendation in Social Tagging Systems Using Hierarchical Clustering," RecSys '08: Proc. ACM Conf. Recommender Systems, pp. 259-266, 2008.

[4] R. Angelova, M. Lipczak, E. Milios, and P. Pralat, "Characterizing a Social Bookmarking and Tagging Network," Proc. European Conf. Artificial Intelligence Mining Social Data Workshop, pp. 21-25, 2008.

[5] S. Aliakbary, H. Abolhassani, H. Rahmani, and B. Nobakht, "Web Page Classification Using Social Tags," Proc. IEEE Int'l Conf. Computational Science and Eng., vol. 4, pp. 588-593, 2009.

[6] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and S. Gerd, "Evaluating Similarity Measures for Emergent Semantics of Social Tagging," Proc. 18th Int'l Conf. World Wide Web, pp. 641-650, 2009.

[7] A. Zubiaga, "Enhancing Navigation on Wikipedia with Social Tags," Wikimania '09: Proc. Fourth Ann. Conf. Wikimedia Community, Aug. 2009.

[8] H. Liang, Y. Xu, Y. Li, R. Nayak, and X. Tao, "Connecting Users and Items with Weighted Tags for Personalized item Recommendations," HT '10: Proc. 21st ACM Conf. Hypertext and Hypermedia, pp. 51-60, 2010.

[9] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos, "A Unified Framework for Providing Recommendations in Social Tagging Systems Based on Ternary Semantic Analysis," IEEE Trans. Knowledge and Data Eng., vol. 22, no. 2, pp.179-192, Feb.2010.

[10] C. Lu, J.-R. Park, and X. Hu, "User Tags versus Expert-Assigned Subject Terms: A Comparison of Librarything Tags and Library of Congress Subject Headings," J. Information Science, vol.36, no.6, pp.763-779, 2010.

[11] del.icio.us, <http://del.icio.us>.

[12] LibraryThing, <http://www.librarything.com>.

[13] Golder, Scott and Bernardo A. Huberman. (2006). "Usage Patterns of Collaborative Tagging Systems."Journal of Information Science, 32(2). 198-208.

[14] Harris Wu, Mohammad Zubair, and Kurt Maly. Harvesting social knowledge from folksonomies. In *Proceedings of the 17th conference on Hypertext and hypermedia*, 2006.

[15] S. Niwa, Takuo Doi, and S. Honiden, "Web page recommender system based on folksonomy mining for itng'06 submissions", In *ITNG 2006. Third International Conference on Information Technology: New Generations*, pages 388{393, 2006.

[16] K'orner, C., Benz, D., Hotho, A., Strohmaier, M., Gerd, S, "Stop thinking, start tagging: tag semantics emerge from collaborative verbosity", in: *Proceedings of WWW 2010*, pp. 521-530. ACM, NY (2010).

[17] M. Gupta, R. Li, Z. Yin, and J. Han, "Survey on Social Tagging Techniques," SIGKDD Explorations, vol. 12, no. 1, pp. 58-72, 2010.

[18] P. Heymann, G. Koutrika, and H. Garcia- Molina, "Can Social Bookmarking Improve Web Search?" WSDM '08: Proc. Int'l Conf. Web Search and Web Data Mining, pp. 195-206, 2008.

[19] R. Abbasi, S. Chernov, W. Nejdl, R. Paiu, and S. Staab, "Exploiting Flickr Tags and Groups for Finding Landmark

- Photos," ECIR '09: Proc. 31th European Conf. IR Research, pp. 654-661, 2009.
- [20] D. Ramage, P. Heymann, C.D. Manning, and H. Garcia-Molina, "Clustering the Tagged Web," Proc. Second ACM Int'l Conf. Web Search and Data Mining, pp. 54-63, 2009.
- [21] M.G. Noll and C. Meinel, "Authors versus. Readers: A Comparative Study of Document Metadata and Content in the WWW," Proc. ACM Symp. Document Eng., pp. 177-186, 2007.
- [22] H. Halpin, V. Robu, and H. Shepherd, 'The complex dynamics of collaborative tagging', in World Wide Web (WWW) Conference, pp. 211-220, Banff, Alberta, (May 8-12 2007).
- [23] M.G. Noll and C. Meinel, "Exploring Social Annotations for Web Document Classification," Proc. ACM Symp. Applied Computing, pp. 2315-2320, 2008.
- [24] Robert Wetzker, Carsten Zimmermann, and Christian Bauchhage. Analyzing social bookmarking systems: A del.icio.us cookbook. In *Proceedings of the ECAI 2008 Mining Social Data Workshop*, pages 26-30. IOS Press, 2008.
- [25] David R. Millen and Jonathan Feinberg. Using social tagging to improve social navigation. In *Workshop on the Social Navigation and Community based Adaptation Technologies*, 2006.

IJSER