

Folksonomy in Social Question & Answer Platform: A Case study of TCS

M Nandeeshha

Abstract

In the Web 2.0 environment, users are empowered to collaborate and share information. This has made the Web more participatory in nature thereby increasing the content creation at a rapid pace. With this, the ever challenging problem of categorizing content over the Web has increased tremendously. To combat this, Web 2.0 has one feature known as “Folksonomy”. This allows users to attach tags to the content so that they can use these tags to retrieve the content in the future. Folksonomy has pros & cons compared to traditional subject indexing. This paper describes the fundamental problems inherited to Folksonomy with a case study of user tagging in Social Question & Answer Platform used within Tata Consultancy Services (TCS).

Keywords: Folksonomy, Tagging, Web 2.0, Tata Consultancy Services

1. Introduction

It's always difficult to find the content the user needs. This is the problem inherited from the day human started generating information more specifically online information. To combat this, various mechanisms were found. One such is the classification of the content using various classification systems. Though various classification system exists, a system or an organization uses one such classification (with or without appropriate local changes) to meet their requirement. This was successful to some extent in retrieving the relevant content by the users.

With World Wide Web (WWW) moving towards next generation, known as Web 2.0, more and more applications are encouraging users not only to consume content but also to generate content with ease. This is resulting in generating enormous content on the web. Organizing content on the web is still a challenge for everyone. Web 2.0 also brings

a mechanism called “Folksonomy”, which will help in organizing the content by the users. Folksonomy allows users to tag or assign keywords to the content generated from their perspective so that these tags can be used in the future to retrieve them.

2. What is Folksonomy?

The jargon “Folksonomy” is a blend of two words “Folk” and Taxonomy”. It stands for the conceptual tags assigned to the content by the users. As per Vander Wal, Thomas ^[1] who coined this word, “Folksonomy is the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval. The tagging is done in a social environment (usually shared and open to others). Folksonomy is created from the act of tagging by the person consuming the information.”

A system which allows for “folksonomy”, users are free to add tags to a piece of content (picture, information, etc) from their perspective so that they will find it easy to retrieve it later. Since, there are no predefined categories; users are free to categorize their resources. This makes the task of categorizing or tagging simple to the user.



3. About Tata Consultancy Services (TCS)

Tata Consultancy Services Limited ^[2] is an IT services, business solutions and outsourcing organization that delivers real results to global businesses, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled services delivered through its unique Global Network Delivery Model, recognized as the benchmark of excellence in software development.

A part of the Tata Group, India's largest industrial conglomerate, TCS has over 120,000 of the world's best trained IT consultants in 42 countries. The company generated consolidated revenues of US \$5.7 billion for fiscal year ended 31 March 2008 and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

3.1 Social Question & Answer (Q&A) platform in TCS

TCS has developed a Social Q&A platform for the enterprise. It allows users to Ask, Answer and Discover questions and answers in a social and participatory ecosystem. All associates in TCS can post and respond to queries. The system has a categorization part & tagging part. Under the categorization part, user has to categorize the question under one of the predefined categories. These categories are based on the organizational business and hierarchy. Under tagging part, users can provide their own tags to the questions. The system also provides a view to see the frequently used tags in the tag clouds and the bigger and bolder tags indicate that they are used more.

4. The Experiment

The goal of this experiment is to study the selection of category and tags assigned by the users in TCS Social Question & Answer (Q&A) platform ^[3]

(internal). When a group of people start working on a task without proper guidelines or procedures, the expected result will not be achieved. This can be analyzed with special reference to folksonomy. Though folksonomy allows users to tag their content, many systems employing this feature do not give any guidelines or procedures to the user in terms of tagging the content. This leads to inconsistent tagging of the content and in turn the system which is meant to help users with folksonomy approach, tends to become chaotic. By educating the user with proper guidelines or procedures for tagging the content, the system can be more reliable and less chaotic.

TCS Social Question & Answer (Q&A) platform is used as the data source for this experiment with a special focus on the category "Library". This system has a categorization (designed based on the Organization Structure) which needs to be selected and then tags can be assigned to the questions being asked by the users. TCS Social Q&A platform is browsed to find all questions categorized under Library and the tags assigned to these questions form the core set of data for this experiment.

A search was made with the category "Library" and 39 questions were retrieved. Those questions which are not related to TCS Library were removed from the data set. Finally, 33 questions related to TCS Library were considered as the raw data for further study and analysis.

The 33 questions had 50 tags distributed among them. Out of these, three questions have been assigned with tags like semicolon (;), colon (:) and "a". Since these tags don't have any meaning, they have been excluded from the data set. Finally 30 questions with 47 tags have been considered. This set of tags constituted raw data for the present experiment.

4.1 Results

The findings of the experiment are as follows –

4.1.1 Average number of tags

The present study considered 30 questions found in TCS Social Q&A platform under the category Library. The below table gives the details of the average number of tags

Title	User Tags
Total number of tags found	47
Average tags per question	1.566

The average number of tags per question is fairly good considering the fact that a categorization system is in place to supplement the tagging system. In addition to the categorization, users have used additional tags which show that users are interested in multiple approaches.

4.1.2 Most frequently used tags

The below table gives the top most frequently used tags by the users

User Tags	Number of times used	Percentage
Library	19	40.425
Learning	03	06.382
Books	03	06.382
Library and Information Centre	02	04.255
Others < 2 times	20	42.533

It is evident from the above table that the name of the category stands first in the ranking of the user tags. Of the 47 tags collected, 19 (4.425%) tags contained the name of the category “Library”. For question retrieval, the name of the category is the most important approach. However, the same can

be retrieved through the categorization (by browsing) without assigning the category name as a tag. Also, the tendency to give other conceptual tags to describe the work is found to be very less.

4.1.3 Single word V/s multiword terms (Terms V/s String)

TCS Social Q&A platform does have an option to use multiword terms for tags like most of the subject heading list. The below table shows the findings from the experiments with regards to number of words used for constructing the tags.

Number of Words	Tags created by users	Percentage
One Word	36	76.595
Two Words	5	10.638
>= Three Words	6	12.765

It is evident from the above table that the users are interested in assigning the single words as tags which constitutes to 76.595% of the tags assigned. Where as 23.405% of the tags are of Two Words or more.

4.1.4 Plural & Singular forms of Tags

The below table shows the distribution of the noun forms in the data sample.

Parameter	Tags created by the users	Percentage
Singular Noun	41	87.234
Plural Noun	6	12.765

There is a strong preference of the users for singular form of tags than the plural form. Most of the search engines give different hits for singular & plural form of the subject. This will result in lack of precision.

4.1.5 Inappropriate Tags

The tags assigned by the users sometimes are inapplicable or irrelevant to the questions asked in the system. Tags like “Colon (:), “Semicolon (;), “a”, etc were found initially and were removed from the data set for the experiment. Also, tag like “Top” was found which doesn’t have any relevance to the question asked. So, all users are not completely aware as to how to assign a proper tag.

4.1.6 Reuse of the category name

Though the system, by default, indexes the category name for retrieval purpose, most of the users have provided the category name “Library” (40.425%) as the tag. Instead of this they should have given some other appropriate tag

which would have helped in describing the question better. This shows that the most of the users are not aware of the indexing mechanism used by the system and also to make best use of the tagging feature.

4.1.7 Inappropriate selection of category

The below table gives relevant & irrelevant questions categorized under “Library” –

Questions under Library Category	Total Number	Percentage
Relevant	33	84.615
Irrelevant	06	15.384

Out of the 39 questions categorized under Library, 6 questions (15.384%) are not related to the Library. In addition to this, there are some questions which are related to Library but are categorized under different categories. This shows that users are not completely aware of how to categorize the questions depending on the respective system.

5. Learning

When we compare the traditional & the emerging approach to content organization, we realized that there was a lot of thought processing went in to organization of content in traditional approach and rightly so considering the manual dependencies. Though emerging approach like folksonomy has eliminated manual efforts from a group of experts, it does not make the traditional principles of content organisation redundant. In fact, they can be the guiding forces for web content organization as well.

In the above experiment, we have seen the following problems associated with users tagging the content –

- ◆ Reuse of categorization names as tags
- ◆ Conceptual tags to describe the content is found to be very less
- ◆ Terms frequently used against Strings
- ◆ Inappropriate use of plural & singular forms of tags
- ◆ Inappropriate use of category, and so on.

These are some of the common problems of folksonomy approach to content categorization. And here in lies the opportunity for Library professionals – to use traditional skills to make search in the emerging platforms non-chaotic. Librarians need not accept Web 2.0 the way it is growing, but librarians need to give it a direction and make it better. (E.g. on the social network sites, librarians can include a small user tip to providing tags).

6. Conclusion

Though folksonomy approach empowers users to categorize the content, it has its strengths &

weaknesses. On the positive side, it allows users to categorize the content based on their perceptions, have multidimensional approach, etc. But, on the negative side, it faces the problem of consistency, inappropriate tags, and others. To overcome some of these problems, some applications are coming out with top level categorization and also allowing users to tag the content. Even this has some limitations as the users may not properly categorize the content and may not tag appropriately as exhibited in the above study. This calls for user education. Unless & until the user is properly aware of the system being used and its features, it's not possible to get the maximum benefit out of it. So, the organizations which are building these kinds of applications have to keep these things in mind before releasing and expecting maximum benefit out of it. Though, the folksonomy approach will

involve more users' participation, appropriate measures are to be taken to ensure the application is reliable.

References

1. **Vander Wal, Thomas. Folksonomy Coinage and Definition.** Available at <http://www.vanderwal.net/folksonomy.html> (Accessed on 24-Dec-08)
2. **TCS. About TCS.** Available at <http://www.tcs.com/about/Pages/default.aspx> (Accessed on 30-Dec-08)
3. **TCS Internal. TCS Social Question & Answer Platform.** (Accessed on 30-Dec-08)

About Author

Mr. M Nandeesh, Knowledge Officer, Tata Consultancy Services Ltd.