INTERNATIONAL JOURNAL OF
RESEARCH IN COMPUTER
APPLICATIONS AND ROBOTICS

SOCIAL TAGGING IN SOCIAL MEDIA A REVIEW

Surabhi D Thorat[1], C.Namrata Mahender[2]

[1] Assistant Professor, MCA Department, MIT (E), Aurangabad (M.S), India
[2] Assistant Professor, Computer Science & I. T Department, Dr.B.A.M. University, Aurangabad

Email: thorat.surabhi@gmail.com[1], nam.mah@gmail.com [2]

Abstract: - Social tagging initiates an intuitive and easy method to organize resources in information systems. These social tagging services have been among the stars of the Web 2.0 era. They allow their users to label diverse resources with freely chosen their own vocabulary in form of tags. In this paper, we tried to high lighten the evolution process of social tagging, categories of tagging approaches, different types of tags available in social media, linguistic tools available now a days and major problematic issues in social tagging.

Keywords: Tagging evolution, Tagging categories, Tagging, Types of tagging, Linguistic tools, Major issues.

1. Introduction

The trend of social tagging that allows the user to create content in virtual environment is growing tremendously in recent years by the launch of media-based social networking sites. These virtual contents can be Blog entries, reviews, online votes, message board posts, videos, audio, images, user profiles, bookmarks etc. These websites provide support for both social interaction, and the presence of rich media objects which can be tagged during such interactions. In the early days of the web keywords meta-tags were used by web page designers to tell search engines what the web page was about.

The social networking sites facilitates the user with social bookmarking services .A social bookmarking service is a centralized online service which enables users to add, annotate, edit, and share bookmarks of web documents[1]. Since 1996 a number of online bookmark management services have launched .Tagging is one of the significant features of social bookmarking systems that enable users to organize their bookmarks in flexible ways and develop shared vocabularies known as folksonomies.

In a social bookmarking system
- Users save links to web pages that they want to remember for later reference or to share with users.
- Collaboratively created knowledge or social knowledge.
- Contains rich semantic information about the described web objects.
- Provide a window for information providers to learn a user's information interests and preferences.
- Web feeds for their lists of bookmarks, including lists organized by tags. This allows subscribers to become aware of new bookmarks as they are saved, shared, and tagged by other users. It also helps to promote your sites by networking with other social book markers and collaborating with each other.

2. Social Tagging Evolution

The advent of social news and bookmarking sites in the mid-2000's brought about a whole new way of see what's going on in the world and discovering interesting content. These sites allow users to create and manage labels (or "tags") that categorize content using simple keywords. The use of keywords as part of an identification and classification system long predates computers. Today's tagging takes the meta keywords concept and re-uses it. The users add the tags. The tags are clearly visible, and are themselves links to other items that share that keyword tag.

Table 1: Tag Evolution History

Origin Year	Site	Founder
2003	Delicious	Joshua Schechter
2004	Digg	Kevin Rose, Ron Gorodetzky, Jay Adelson, and Owen Byrne
2004	Facebook	Mark Zuckerberg , Eduardo Saverin.
2005	Reddit	Steve Huffman and Alexis Ohanian
2006	Twitter	Jack Dorsey, Evan Williams, Biz Stone and Noah Glass

3. Categories of Social Tagging

A folksonomy is a system of classification derived from the practice and method of collaboratively creating and translating tags to annotate and categorize content[2][3]. This practice is also known as collaborative tagging [3], social classification, social indexing, and social tagging. Social tagging is broadly divided into 2 categories [4].

3.1 Broad Folksonomy
In broad folksonomy multiple people tag the same object by using their own vocabulary. It allows multiple people to tag the page with the same tag. Like delicious site.
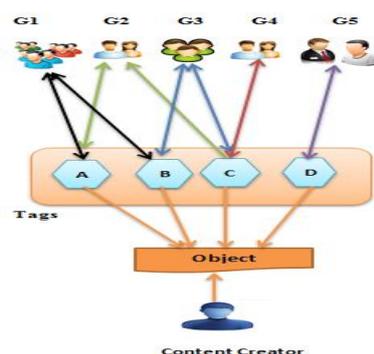

Figure 1: Broad Folksonomy

In this content creator creates the content and makes it accessible to others. Other people or groups of people with the same vocabulary tag the object with their own terms. The people also find the information based on the tags. Arrows to the tag indicate that Group applies that tag to content and the arrows from tag to Group indicate that Tag is used to find the content.

Vocabulary Trends

Based on the figure we see that Tags A and B are the popular terms used to describe the content. A Long Tail exists with tags C and D that is created by people of similar vocabulary mind-sets that makes the searching of specific tags more effective. The benefits of the long tail and power curve come from the richness provided by many people openly tagging the same object.
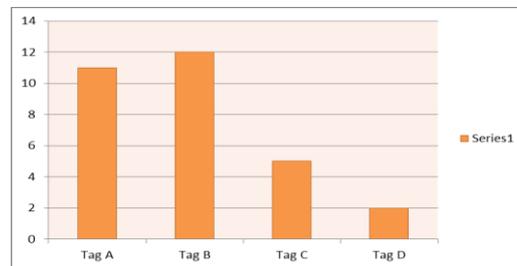

Figure 2: Power Curve and Long Tail

3.2 Narrow Folksonomy

In narrow folksonomy one or few people provides tags that the person uses to get back information. The tags are singular in nature only one tag per term. An object is created and tagged by the creator for reasons unknown to the other users .A good example of narrow tagging tool is Flickr represents, provides benefit in tagging objects that are not easily searchable or have no other means of using text to describe or find the object.
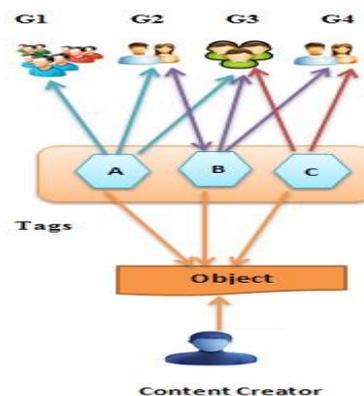

Figure 3: Narrow Folksonomy

4. Tag

Tag is a non-hierarchical keyword or term assigned to a piece of information .Like an Internet bookmark, digital image, or computer file. Tags are kind of metadata helps describe an item and allows it to be found again by browsing or searching. Tags are generally chosen informally and personally by the item's creator or by its viewer, depending on the system [5]. Tagging gives the road map by which a user can classify, indicate ownership rights, and indicating online identity. They may can be in the form of text, images, or other identifying marks.

In the early websites the web keywords meta tags were used by web page designers to describe the web page. Now days along the raise in growth of social media the term tagging takes the Meta keywords concept and re-uses it. The users add the tags that are clearly visible, and are themselves links to other items that share that keyword tag. Tagging was popularized by websites associated with Web 2.0 and is an important feature of many Web 2.0 services. A Web 2.0 site may allow users to interact and collaborate with each other in a social media dialogue as creators of user-generated content in a virtual community. Web 2.0 offers all users the same freedom to contribute. While this opens the possibility for serious debate and collaboration, it also increases the incidence of "spamming" and "trolling" by unscrupulous or misanthropic users. The impossibility of excluding group members who don't contribute to the provision of goods from sharing profits gives rise to the possibility

that serious members will prefer to withhold their contribution of effort and free ride on the contribution of others[6].

The key features of Web 2.0 [7] include:
1. **Folksonomy-** free classification of information; allows users to collectively classify and find information (e.g. Tagging)
2. **Rich User Experience-** dynamic content, responsive to user input.
3. **User Participation -** information flows two ways between site owner and site user by means of evaluation, review, and commenting. Site users add content for others to see.
4. **Long tail-** services offered on demand basis; profit is realized through monthly service subscriptions more than one-time purchases of goods over the network.
5. **Software as a service -** Web 2.0 sites developed API to allow automated usage, such as by an app or mash-up.
6. **Mass Participation -** Universal web access leads to differentiation of concerns from the traditional internet user base.

Like Web 2.0 website include: Social networking sites, Blogs, Wikis, Folksonomies, Video sharing sites, Hosted services, Web applications, Mash-ups.

5. Types of Tags

1. General tags

Content-Based Tags

Identify the actual content of the resource. Like Physical entity (Person, animal, plants etc.) and Non-Physical entity(Organization). In Figure 5. Honda models are taken as the content-based tags.
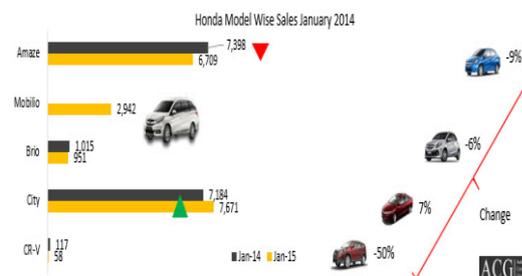
Content Based Tag on Honda Models[12]

Context-Based Tags

Provide the contextual information of an object in which the object was created or saved. Like the time and location

Context Based Tag on Facebook

Subjective Tags

These tags express user's opinion and emotion, e.g. exited, funny, sad. They can be used to help evaluate an object reviews, status. They provide a way of self-representation of individual expression.

Subjective Tag expressing feelings

Organizational Tags

Tags that identify personal usage, tasks for further reference as remainder or review This type of tags is usually not useful for global tag aggregation with other users' tags. These tags are specifically used by the user based on their personal interest or task.

Organizational Tag on LinkedIn

Table 2: Comparison of purpose-based categorization of social tags, adapted from [8]

Our categories	Xu et al. [9]	Sen et al. [10]	Golder et al. [11]	Bischoff et al. [8]
Content Based	Content-based	Factual	What or who is about	Topic
	Attribute		What it is	Type
Context Based	Context Based		Who own it	Author/owner
			Refining other categories	Time/Location
Subjective	Subjective	Subjective	Qualities/Characteristics	Opinions/qualities
Organizational	Organizational	Personal	Task organization	Usage context
			Self-reference	Self-reference

2. Linguistic Classification of Tags [13]

- **Functional:** Tags that describe the function of an object. (e.g., weapon)
- **Functional collocation:** These are defined by function but in addition, they have to be collected in a place (and/or time). (e.g., furniture, tableware)
- **Origin collocation:** Tags that describe why things are together? (e.g., garbage, contents, dishes (as in "dirty dishes" after a meal)).
- **Function and origin:** Tags that describe why an object is present, what is the purpose, or where did it come from. (e.g., "Michelangelo" and "medieval" on an image of a painting by Michelangelo)
- **Taxonomic:** They are words that can help in classifying the object into an appropriate category. (e.g., "Animalia" or "Chordata" tag to an image of a heron)
- **Adjective:** They describe the object that denotes the resource. (e.g.,"red", "great", "beautiful", "funny")
- **Verb:** These are action words. (e.g., "explore", "todo", "jumping")
- **Proper name:** Most of the tags are of this category. (e.g., "New Zealand","Manhattan bridge")

Special Tags

1. Triple Tags- `It Define extra semantic information about the tag .Triple tags comprise three parts: [namespace]:[key]=[value]

Example:

If I wanted to sell my bike, I may add a photo of it to Flickr and add the tags:

selltagged
sell:price=79.99
sell:currency=dollar

2. Hash Tags-hash tag is metadata tag marked by the prefix #, sometimes known as a "hash" symbol. This form of tagging is used on micro blogging and social networking services.

Example-Twitter, Facebook, Google+, VK and Instagram.

3. *Knowledge Tag-* Knowledge tags are an extension of keyword tags. They were first used by Jumper 2.0, an open source Web 2.0 software platform released by Jumper Networks on 29 September 2008. Jumper 2.0 was the first collaborative search engine platform to use a method of expanded tagging for knowledge capture. It's a metadata that captures knowledge in the form of descriptions, categorizations, classifications, semantics, comments, notes, annotations, hyper-data , hyperlinks , or references that are collected in tag profiles. These tag profiles reference an information resource that resides in a distributed, and often heterogeneous, storage repository.

Knowledge capturing in tags takes many different forms:
- Factual Knowledge
- Conceptual Knowledge
- Expectational Knowledge
- Methodological Knowledge
-

4. *HTML Meta Tag--* Meta elements are tags used in HTML or XHTML documents to provide structured metadata about a Web page. *They are part of a web page's head section.* Meta elements can be used to specify page description, keywords and any other metadata not provided through the other head elements and attributes.
Like: Page creator, how often updated, about page, Keywords that represent page content.

6. Linguistic Tools
Now a days there are many linguistic tools are available in the market, they may vary in terms of features provided by them.

Table 3: Linguistic tools features

Tool	Comparative Study of Feature							
	User friendly GUI	Dimension Reduction	Flexible	Graphs & Visualization	Multi Language Support	Clustering	Decision Tree	Data Partitioning
SASText /miner	√	√	√	√	√	√		√
IBM Text Analytics	√	√	√		√		√	
SAP Text Analytics	√	√	√	√	√	√		√
Angoss	√		√	√	√			√
Provalis Research	√		√	√	√	√		√
Rapid Miner	√	√	√	√	√	√	√	√

7. Major Issues in Tagging

1. Spamming

They are misleading tags that are generated by the spammers in order to increase the visibility of some resources of their own interest or simply to confuse users. The web is susceptible to search engine spam, that is to content that is created to mislead search engines into giving some pages a higher ranking than they deserve [14].

Example
- Like, in a system that annotates web pages, one mobile company may annotate many pages (except the page of its competitor) with the string "buy mobiles", so that users looking to buy mobiles will not easily find the competitor's page.

2. Polysemy

Since words often have multiple meanings, which meaning is supposed to be associated with a tag?

Example
- Does the tag orange refer to the fruit or to a color?
- The Tag Bank means financial institution or the building offering financial services. If a tag *I'm your friend, you can bank on me* comes. It's difficult to get the sense of this tag as here bank used as means of "rely upon". It is different, but related, as it derives from the theme of security initiated by word bank.
- Zeugma is the typical rhetorical device used to pun on the different senses of polysemous words. Cruse (1986: 73) gives the example Arthur and his driver's licence expired on Thursday where the verb expire has two distinct, but related, senses (i.e. 'died' and 'no longer valid'). If zeugma is possible, the relevant word is ambiguous and therefore polysemous.

3. Synonymy

When multiple words can have the same or similar meaning, which tag should you use, and how do you find all the tags that mean the same? **Example:** Are "Independence Day" in the India and "Aug 15th" the same?

4. Word inflections

This problem arises because words are modified for specific grammatical contexts, which variation do we use for a tag? **Example:** We might see mouse and mice.

5. Syntactic constraints

Many syntactic constraint involves with tags .
How to create tags out of phrases when spaces are not allowed?
How to deal with punctuation?
How to deal with non-ASCII words?

6. Sparsity, no consensus and search inefficiency

Sparsity is related to the annotation coverage of the data . Certain pages may not be tagged at all. Users do not generally associate tags to newly emerging web pages or web pages that can be accessed easily from hub pages, or uninteresting web pages.

8. Conclusion

In this paper we focus on the evolution of social tagging with the raise of web 2.0 services, categories to tagging chosen by a person of group of peoples all together, various types of tags used by the tagger and the content creator and the major issues that leads to less effective diagnosis of tags. This paper gives the good understanding of tags and its different aspects.

9. REFERENCES

[1] Noll, Michael G.; Meinel Christoph (2007). "Web Search Personalization   Via Social Bookmarking And Tagging". Lecture Notes in Computer Science. Lecture Notes in Computer Science 4825: 367–380. doi:10.1007/978-3- 540-76298-0_27. ISBN 978-3-540-76297-3.

[2]Pink, Daniel H. (December 11, 2005). "Folksonomy". New York Times. Retrieved14 July 2009.

[3]Jump up Lambiotte, R, and M Ausloos.2005 Collaborative tagging as a tripartite network . http://arxiv.org/abs/cs.DS/0512090

[4]ww.vanderwal.net/random/entrysel.php?blog=1635

[5] en.wikipedia.org/wiki/Tag_(metadata)

[6] Gerald Marwell and Ruth E. Ames: "Experiments on the Provision of Public Goods. I. Resources, Interest, Group Size, and the Free-Rider Problem". The American Journal of Sociology, Vol. 84, No. 6 (May, 1979), pp. 1335–1360.

[7] "7 Key features of web 2.0". www.webApprater.com.2012-09-17. Retrieved2012-09-17.

[8] Bischoff, K., Firan, C. S., Nejdl, W., Paiu, R. 2008. Can All Tags Be Used for Search? In Proceeding of the 17th ACM Conference on Information and Knowledge Management (CIKM'08),pp. 203-212.

[9] Xu, Z., Fu, Y., Mao, J., Su, D. 2006. Towards the Semantic Web: Collaborative Tag Suggestions. In Proceedings of the WWW'06 Collaborative Web Tagging Workshop.

[10] Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, M. F., Riedl, J. 2006. Tagging, Communities, Vocabulary, Evolution. In Proceedings of the 20th ACM Conference on Computer Supported Cooperative Work (CSCW'06), pp. 181-190.

[11] Golder, S. A., Huberman, B. A. 2006. Usage Patterns of Collaborative Tagging Systems. Journal of Information Science 32(2), pp. 198-208.

[12] http://www.autobei.com/blog/honda-car-registered-highest-monthly-sales/

[13] Csaba Veres. The language of folksonomies: What tags reveal about user classification. In Natural Language Processing and Information Systems, volume 3999/2006 of Lecture Notes in Computer Science, pages 58–69, Berlin / Heidelberg, July 2006. Springer.

[14] Z. Gy¨ongyi and H. Garcia-Molina. Web spam taxonomy. In 1st Intl. Workshop on Adversarial Information Retrieval on the Web (AIRWeb), pages 39–47, 2005.

BIOGRAPHIES

1. **Surabhi D Thorat** is Assistant Professor of Master of Computer Applications at Marathwada Institute of Technology, Aurangabad, India. Her research areas are Data mining and data ware housing He has published widely in several journals including the International Journal of IT, Engineering and Applied Sciences Research, International Journal of Computer Science and Information Technologies, International Journal of Research in Computer and Communication Technology.

2. **C. Namrata Mahender**(Ph.D.) is Assistant Professor of Computer Science & Information Technology at Dr Babasaheb Ambedkar Marathwada University, Aurangabad, India. Her research areas are Operating System, Image Processing, Natural Language Processing, Artificial Neural Network, Fuzzy logic, Pattern Recognition, Brails Script, Software Engineering, Concurrent programming, Multi core System. She is the author of many papers in these areas in National and International Journals. She has taught post graduate courses in the area since 2006.