Social Tagging as a Knowledge Organization and Resource Discovery Tool

Hesham Allam

Abstract: The purpose of this paper is to give an overview of the social tagging phenomenon, including how it evolved and the debate surrounding its benefits and limitations. Further, social tagging's potential as a new tool for knowledge organization and resources discovery will be considered. Finally, some questions concerning social tagging will be presented for future research. The paper reflects an interdisciplinary overview of the technical aspect of social tagging along with a behavioural look at why and how it is used as a tool to discover and organize knowledge.

About the Author(s): Hesham Allam is a second year student in the Interdisciplinary PhD program at Dalhousie University. His interdisciplinary interests are business administration and computer science. He got his Master degree in Management Information Systems (MIS) from California State University Sacramento in 2003. He also has a Master degree in Developmental Psychology from Egypt in 1994. He worked as an IT manager in the State of California for two years after graduating from his Master degree. He lectured MIS for six years. This paper was originally researched and written for CSC 6606, Human Factor online Information System Course at the school of Computer Science.
Background

Metadata is known as data about data. Metadata is often a highly organized description of information resources, which may include documents, books, articles, photographs, multimedia or other items. The goal of metadata design is to facilitate the organization and the access to these information resources (Taylor, 1999). With its standard schemes and conventions that help produce accurate and consistent results, metadata plays an important role in knowledge allocation and discovery especially for communities of heavy information practice, such as digital libraries and Web searching. Metadata utilizes its labelling conventions to provide sufficient description of documents’ content to help users distinguish between relevant and irrelevant information (Dawson, 2004).

Subject descriptors or index terms are forms of metadata. They are manually assigned to information resources based on their content to facilitate the information retrieval process. The order and the organization of index terms and subject descriptors affect how users retrieve information. For example, terms that loosely describe a documents’ content will result in high recall at the expense of precision, while terms that are too specific to describe content will produce high precision but low recall. To ensure an appropriate balance between recall and precision, an indexing term process is used to allow more control over the creation of index terms. Controlled vocabulary is an example of an indexing process which ensures that index terms are standardized and related resources are collected for an easy access (Macgregor & McCulloch, 2006). In addition to its use in library science, controlled vocabulary guides users as they gather information resources on the internet using traditional search engines like Google, Yahoo, or Microsoft MSN.

Although offering many benefits for custom made search and browsing, traditional controlled vocabulary indexing is not flexible enough to keep pace with the rapid growth of the online digital environment. Nicholson et al. (2001) noted that controlled vocabulary indexing is sometimes too limited to describe the content of a Web document. Duval et al. (2002) added that one of the major obstacles of adopting controlled vocabulary to the Web environment is its demand for a highly structured indexing process that imposes scaling and cost difficulties. Furthermore, new social internet tools such as Wikis and blogs are producing enormous amounts of distributed information that makes it virtually impossible to have enough experts to create effective indexing for all the Web documents (Quintarelli, 2005). Research to find better aggregation and organization tools of indexing Web resources indicates that traditional problems such as efficiency, scalability, and cost can be eliminated with the use of social Web applications that would produce automatic and aggregate metadata (Shirky, 2005).
Social Tagging (Folksonomy)

Folksonomy or social tagging is recommended by (Quintarelli, 2005) as an alternative to the traditional forms of indexing. Social tagging offers a flexible, cumulative, and intuitive approach that is a better fit for the Web environment. Unlike traditional indexing, where experts are needed to catalogue and create metadata to describe search terms for efficient retrieval, social tagging can be performed by anyone by freely attaching keywords or tags to describe the content of a Webpage. With such features, social tagging has the capacity to assign metadata to the ever-expanding Web contents (Golder & Huberman, 2005).

Social tagging, or collaborative tagging, describes a practice whereby users assign keywords or tags to information resources. Social tagging applications use keywords or “tags” to organize information within users’ information space in a sharable way, thus facilitating the searching and browsing of information targets by other users. This gives users the chance to add their information resources to popular tags that have been created by other users and that can be explored by any internet user. Tags are typically single terms and are often pictured as a tag cloud where the most used tags are shown in bigger font and least used are shown as smaller font. Multiple tags can be assigned to a single source by removing punctuation and by using symbols to combine terms (Macgregor & McCulloch, 2006).

Examples of Social Tagging Applications

Tag Cloud

A tag cloud is simply a list of all the tags that belong to a certain user group where popular tags are highlighted typographically (Panke & Gaiser, 2009). A tag cloud is similar to the query mechanism in a traditional search. After a user clicks on a tag within the cloud, the tagging system takes the tag and adds it to the system algorithm as a regular query. The system matches the tags with related tags in its collections and displays a list of the results in a new page with another tag cloud with tags related to the previous search (Mesnage & Carman, 2009). Typically, users are given a choice to filter out the tag cloud to include or delete selected tags. Some systems enable users to control the properties of tag clouds through a slide bar on the side of the page to manage the size, the coarseness, and the visibility of tags within the tag cloud.
Tags, Maps, Photo, Tag Cloud

One of the most interesting trends in tag visualization is TagMaps. TagMaps is a Flash-based web application that shows key characteristics of location-based data in an easy navigable way. A project deployed by Yahoo and Flickr, TagMaps combines tag cloud, photos, and Geotags in one place available for the user. World Explorer is a beta application using the TagMap concept, where users can explore a place on the map by clicking on a highlighted tag from the tag cloud drawn on the location to view photos taken by different users for that place. For example (shown in figure 2) a user may click on the “Alamo Square” tag to get a list of the places in Alamo Square. By clicking on a particular place, a user would see photos taken of it (Yahoo, 2009).
Social Tagging at Work

Two of the most famous social tagging Web applications are Del.icio.us, pronounced Delicious, and Flickr. Delicious is “a social bookmarking service that allows users to tag, save, manage and share web pages from a centralized source. With emphasis on the power of the community, Delicious greatly improves how people discover, remember and share on the Internet,” (Del.icio.us, 2009, para. 1). Delicious allows users to choose their own expressive vocabulary as keywords or tags to bookmark Web pages and to have them available from any computer with internet access for their subsequent easy access. This ability makes Delicious unique as a bookmarking manager. With a free user’s name and a password, users can click on a bookmarking tab located in the browser tool bar that opens a form with a prompt to assign a unique label for the bookmark (Mathes, 2004). Delicious system generates a chronological order of the bookmarks assigned by users and collects them in organized categories to facilitate later access by their users. The system then uses these tags to locate similar tags throughout the Delicious system. For example, looking at the page http://del.icio.us/tag/soccer will show all bookmarks that are tagged with “soccer” by any user. Hence users can find more related information about soccer by browsing these related bookmarks. Furthermore, Delicious has important features that list the most popular tags on the right panel of the front page of the Delicious Website to invite users to explore other users’ popular favourites. Additionally, related tags identified programmatically within the system are listed on the right side of users’ tag pages (Mathes, 2004).
The second famous social tagging Web application is Flickr (www.Flickr.com), a photo sharing site that allows users to upload their own photos with unique tags for easy access and sharing. Users have the options to upload photos from their own computer or from a Web page through Flickr system. The service is free, but users have the option to pay fees for additional features, such as a bigger storage space for photos. A major difference between Delicious and Flickr is that Flickr’s users tag photos they created themselves, while Delicious users tag Web pages mainly created and posted by others (Mathes, 2004).

**Reason for Performing Social Tagging**

Hammond et al. (2005) described the reasons users perform tagging as ranging from selfish, where users tag their own contents for their own easy retrieval, to altruistic, where users tag other users’ content for easy retrieval by anyone using that system. Hammond et al explained their view through the following axis diagram:

![Figure (3) Hammond et al.’s (2005) Content Creator (p. 1)](image)

The vertical axis shows the benefits of tagging, and the horizontal axis shows the content being tagged. Social tagging tools are shown in plain text, while other traditional tools are shown in italics. Wikipedia is shown in parentheses because Hammond et al consider it as a regular taxonomy categorized by a specific community of users, but not by any one user. Therefore, it does not fall fully in the category of social tagging. As indicated in the diagram, a Flickr’s tag is created by a user primarily to benefit the user him/herself, or to share it with friends, family or other users. In Technorati, another social tagging site dedicated for blogs, a user tags his/her blog to help others find it.
Recent studies have focused on social presence as a main factor affecting the users’ tagging behavior. Lee (2006) indicated that perceived social presence has a positive impact on tagging behavior. She noted that users tend to increase their tag when perceiving that others are seeing them. Ames & Naaman (2007) found that users tagged more frequently to communicate extra context to their friends and families. The effect of social presence on tagging was supported by a study conducted by Ling et al. (2005). In their study of how people tag in MovieLens, it was shown that users increased the quality of their tags when reminded of the value of their contribution and when they were given specific tagging goals.

Nov et al. (2008) studied the motivations associated with tagging behaviour on Flickr and how it affects the level of users’ tagging activity. Basing their work on a qualitative study conducted by Ames & Naaman (2007), they confirmed that social presence is a prime motivator for users to add tags to content. According to their results, they noted that people tend to increase the number of tags if they: (1) perceive social presence of others; (2) they are self or publicly motivated; (3) and if they have a bigger collection of photos in their profile. They showed no significant correlation between family and friends as motivational candidates and the tagging level, suggesting that users are not mainly tagging content for friends and family. Accordingly, the study recommended that managers of tagging systems should feature public exposition to help attract more people to use their systems.

Some tagging motivations come from a group contribution prospective. When users get the feeling that their contribution is important to groups, they are more likely to contribute more to the system. Conversely, if they feel that their contribution is less effective, they are less likely to contribute (Karau & Williams, 1993). Beenen et al (2004) conducted a study on the “under contribution” of some users on the movie recommender system MovieLens. It was found that users, when reminded that their tag contribution will help others to choose the right movie, perceived a sense of uniqueness to their tags and hence added more movie tags to help the group.

Lee (2006) stated that although tags (posts) are largely directed towards personal organizational reasons, social presence plays an important role in getting users to tag resources. Lee’s analysis of Del.icio.us users’ behaviour showed that users, when perceiving social presence of others, tend to describe their bookmarks with tags intended to facilitate other users’ access to the bookmarks. Lee also noted that the social presence of others pushes people to tag more frequently and to reveal more information about their tags to other groups.

Becker (2007) indicated that tagging gives some users the power of naming. This identifies specific users as the initiators of new tags from the very beginning of the game: “If I sell size 8.5″ blue-striped widgets, and I’m the first to use the tag “widget”, I claim a powerful findability label for my very specific merchandise” (p. 1).
Ames & Naaman (2007) studied users’ incentives to tag in Flickr and ZoneTag—a camera phone capture tool that help users post images on Flickr. They studied users’ motivations from two dimensions: ‘social’, whether users are tagging items for individual use or for public use; and function, measuring the intended use of tags. The results of their study showed that participants were motivated by: (1) organization for general public, either to participate in public pools and help others search, or to self promote their photo collections; (2) self organization, to organize their photos for future retrieval; or (3) social communication, using tagging as a tool to add more contexts for their friends, family, and the public.

Using a collection of tags on books and music, Zollers (2007) studied users’ tagging motivations in Amazon and Last.fm. Through snowball sampling and random sampling, the study deduced three motivations for users to tag a resource: (1) opinion expressing; (2) performance, using tags as a show-off technique to others; and (3) activism, to broadcast a group view on a certain topic to a certain audience. With regard to activism as a motivational factor, the study found 480 users tagged 1054 Amazon products with the defective design tag.

Intending to identify the most important motivations for users to tag video material, Velsen & Melenhorst (2009) classified users’ motivations for tagging into three paths: indexing, socializing, and communicating:

- Motivations related to indexing: indicate that users tag an item to re-find this item in the future, make it easier for others to find that item, add more information for the item to be discovered by others, or to help users find information related to that item.

- Motivations related to socializing: suggest that users tag an item to recommend it to others, to find friends or peers, and to connect a personal profile with other’s matching profiles for topic recommendations.

- Motivations related to communication: indicate that users may tag to express personal opinion, or to contact similar taggers to share opinions about an object.

Based on their findings, Velsen & Melenhorst (2009) noted that motivations related to indexing outweighed motivations related to socializing or communication.

In their study of users’ media tagging, Cunningham & Nichols (2008) indicated that individuals use tags in media items for the following reasons: mental, to pass the time or to enjoy something; visual, to have a look at a media item from a friends’ referral; audio, to listen to a media item; learning, to learn a new technique for a certain task; social, to follow a friend’s preference or to share something interesting with friends; MSM, to get updated on Mainstream Media; temporal, to follow up on a future event; and others, to decide if this item will be important for future purposes or not.
Marlow et al. (2006) stated that some users begin tagging for personal purposes and later appreciate its social aspect, thereby encouraging participation in the social tagging process. They indicate that tagging motivations vary with differing levels of user interaction with the tagging system. They explain that motivation for tagging can be categorized in two different directions: organizational and social. Tagging intended for the purpose of organization is used as a tool for structured filing, wherein users utilize their personal structure standards or other’s tagging standards to organize their links. Alternatively, some users consider tagging as a communicative tool to express their opinion and issue and their views to others in the tagging community.

The following are motivations identified by Marlow et al. (2006) directing tagging behaviour: (1) contribution and sharing, to add more tags to a cluster of tags for the purpose of sharing and contributing resources (e.g., tagging vacation photos for a partner or a friend); (2) attracting attention, to have others look at their tags to make their tag popular either for personal reasons or for spamming reasons; (3) play and competition, to tag others’ resources for competition as in ESP Game, or to change how the tags in system look via playing with tag clouds; (4) self presentation, to leave personal marks on selected resources; (5) opinion expressing, to express view and opinion to others; and (6) future retrieval, to mark items for personal retrieval (e.g. tagging a group of pages on Del.icio.us in preparation to write a book). The author noted that these types of tags could be used to incite an activity or work as a reminder of others. Additionally, they are sometimes extremely helpful in providing metadata about objects that have no other tags associated with them.

Panke & Gaiser (2009) conducted two quantitative and qualitative surveys to study the impact of social tagging on personal information management. They found that the use of tagging as a personal management tool was much more important to participants than using it as an information sharing tool. They identified four types of taggers: (1) ego taggers, which include those users who consider themselves as the information elite and who seek the publicity of being taggers; (2) archivers, users who use tagging mainly to organize their daily web activities; (3) broadcasters, which includes users who tag items to share them with the public; and (4) team players, those users who use tags to exchange information with personal networks. The authors also found that taggers were highly sensitive to privacy issues and utilized tags to protect their privacy. In their interview with tagging experts, the authors noticed that even though tagging systems had many additional features, few of these were used. Examples of these features were building a contact network, using bundles to organize tags, or renaming keywords. This study concluded that participants used tags as a mnemonic device under the rule of “out of sight, out of mind”.

Zhan et al. (2009) indicated that different information foragers have different knowledge backgrounds and hence different mental activation processes for the same web content. They
considered social tagging as a “socio-cognitive artefact”, where tags and web content intersect to influence the activation of the user’s mental state and the action of tagging. The concept of socio-cognitive artefact also entails users tagging items intentionally as a supplement to make sense of information when they revisit it and to reduce the cost of operation. Zhang et al indicated that the generation of users’ tags is determined by user’s interests, salient activation, and the expected reuse of tag.

**Advantages of Social Tagging**

Before getting into social tagging benefits, it is worth referring to some of the unique features offered by social tagging applications such as those adopted by Delicious and Flickr. Hammond et al. (2005) pointed out that traditional link managers used to be located in the user’s browser. However, social tagging managers are server side based that allows links to be publicly tagged by users to facilitate efficient retrieval. The following are examples of the new features that exist in tagging managers: users can create personal user accounts and group accounts; users are offered tools like prompts forms for entering links, titles and comments about a Web page that help them remember the content when they re-access it; users can browse bookmarks to facilitate entry; users can search by tag, user, or by doing a combination of both; users can add social tagging plug-ins to their browser to facilitate the bookmarking process.

Hammond et al. (2005) argued that social tagging is more beneficial than traditional search engines, noting that

> This ability to sort out the wheat from the chaff is an important win over a web-based search engine. Search engines, at this point, tend to index and search a global space – not my local space. My space comprises the documents I am interested in and the documents of other users that I want to follow. (p. 1)

Quintarelli (2005) indicated that although the concepts of bookmarks and keywords are not new to the Web, social tagging applications, like the one adopted by Delicious and Flickr, offer some relevant new features that make bookmarks more appealing to users. For instance, the feedback is immediate. As users assign a tag to an item, they see clusters of items with the same tags. Further, users are given incentive to change the tag or add another so they do not have to worry about making the right choice when assigning tags. Quintarelli continued by indicating that social tagging adds a sense of serendipity. As users start with their own tag, they end up clicking on similar tags that might lead them to get unexpected but related information about a certain topic. Quintarelli concluded by referring to advanced social tagging features like ‘tag cloud’ that enhance the browsing and searching process by displaying popular tags based on the frequency of users’ visits to these tags.
In Sinha’s blog (2005) titled "A cognitive analysis of tagging", she indicates that social tagging has an advantage over traditional indexing by not adding to users’ cognitive load. It uses existing cognitive process by allowing users to see how others have tagged similar items. According to Sinha (2005) people experience a “state of fear” to lose a Web page forever if they make the wrong decision to bookmark a Web page with the right keyword. By doing that, users may fall into the so called “post-activation analysis paralysis”. Sinha noted the potential benefits of social tagging in preventing such a problem by suggesting that Social tagging applications can offer immediate tagging feedback to give users suggestions as tagging options to better describe Web content.

Social Tagging Limitations

Hammond et al. (2005) pointed out that social tagging tools may cause users to give up some of their privacy by publishing their bookmarks and interests. However, these tools offer more utilities at the price of giving personal data. Hammond et al. noted that social tagging applications may expose users to serious issues like spam attack. However, these attacks do not seem to deter people from using these services, mainly because the risk of spy ware and spamming, like in any other computer application, can be managed by vigilant legitimate use.

Mathes (2004) pointed out some problems with the use of uncontrolled vocabulary in social tagging or folksonomies. Due to the lack of systematic guidelines and explicit rules in tagging documents, terms in social tagging can be ambiguous as different users assign diverse terms to describe the same content. Taking into account that each user may have a different interpretation of the same content, bookmarks can be labelled with potentially misleading tags. For example, the “filtering” tag in Delicious includes Web pages whose theme is “filtering” but referring to totally different domains as shown in the following web pages:

- Last.FM - Your personal music network - Personalized online radio station
- InfoWorld: Collaborative knowledge gardening
- Wired 12.10: The Long Tail
- Introduction to Bayesian Filtering

Conclusion and Future Research

It is worth concluding that social tagging websites like Delicious, YouTube, and Flickr are attracting many internet users. Unlike traditional websites, social tagging allows internet users (authors) to build and organize internet contents according to their own preferences, then share them easily with others. One of the main advantages of social tagging is its unique way to link web resources through the semantic association of keywords that users can click on to
know more about related topics. Some may argue that giving users the freedom to click on related words may distract them from getting to their search goal. While this could be true, many users still surf these websites even if it costs them some distraction from their main search goal. But why would users keep surfing social tagging websites knowing that they may drift from their search target? Perhaps this is a different kind of distraction, one that is enjoyable, adds to users’ knowledge, or that leads users closer to what is relevant to them, although it was not expressed in the original query. It is also possible that users like to stumble into unplanned but related topics because the related tags trigger their memory or their curiosity. This raises intriguing questions including: why do users like surfing these websites? When do users’ consider drifting created by social tagging useful? When is it a distraction? Is it possible to measure distraction in social tagging in a way that differentiates between useful exploration and disruptive intrusion? How can we use the advantages of social tagging websites without getting users into the distraction zone? Is it possible that features of social tagging (e.g. semantic association) resonate with how the mind functions in interpreting things and hence can be used to enhance the internet learning process?
References


