Chapter 13
Folksonomy
The Collaborative Knowledge Organization System

Katrin Weller
Heinrich Heine University of Düsseldorf, Germany

Isabella Peters
Heinrich Heine University of Düsseldorf, Germany

Wolfgang G. Stock
Heinrich Heine University of Düsseldorf, Germany

ABSTRACT

This chapter discusses folksonomies as a novel way of indexing documents and locating information based on user generated keywords. Folksonomies are considered from the point of view of knowledge organization and representation in the context of user collaboration within the Web 2.0 environments. Folksonomies provide multiple benefits which make them a useful indexing method in various contexts; however, they also have a number of shortcomings that may hamper precise or exhaustive document retrieval. The position maintained is that folksonomies are a valuable addition to the traditional spectrum of knowledge organization methods since they facilitate user input, stimulate active language use and timeliness, create opportunities for processing large data sets, and allow new ways of social navigation within document collections. Applications of folksonomies as well as recommendations for effective information indexing and retrieval are discussed.

INTRODUCTION

A key problem facing today's information society is how to find and retrieve information precisely and effectively. Substantial research efforts concentrate on the challenges of information structuring and storing, particularly within different sub-disciplines of computer science and information science. In this context, information retrieval studies focus on methods and algorithms to enable precise and comprehensive searching of document collections (Frakes & Baeza-Yates, 1992; Stock, 2007a). In addition, techniques of knowledge representation have been established (Cleveland & Cleveland, 2001; Lancaster, 2003; Stock & Stock, 2008). Most prominent are approaches of document indexing:

DOI: 10.4018/978-1-60566-368-5.ch013

Figure 1. An exemplary tag cloud. Tag clouds display the most popular tags within a folksonomy based system. The bigger the font size, the more documents have been indexed with a tag.

i.e., assigning content-descriptive keywords to documents. This enhances retrieval techniques and aids users in deciding on a document's relevance. Different knowledge organization systems (KOS) are developed to support sophisticated document indexing. Common examples of KOS include classification systems (taxonomies), thesauri, and controlled keywords (nomenclatures).

Recently, a well-known problem of indexing documents with content-descriptive metadata has been addressed from a new, user centered perspective. Within the so-called "Web 2.0" (O'Reilly, 2005), web users have begun publishing their own content on a large scale and started using social software to store and share documents, such as photos, videos or bookmarks (Gordon-Murnane, 2006; Hammond, Hannay, Lund, & Scott, 2005). And they have also begun to index these documents with their own keywords to make them retrievable. In this context, the assigned keywords are called *tags*. The indexing process is called (*social*) *tagging*, the totality of tags used within one platform is called *folksonomy*. A *tag cloud* is a popular method for displaying most frequently applied tags of a folksonomy visually (Figure 1).

Thus, a folksonomy is an indexing method open for users to apply freely chosen index terms. Peter Merholz (2004) entitles this method "metadata for the masses"; the writer James Surowiecki (2004) refers to it as one example of "the wisdom of crowds." The term "folksonomy", as a combination of "folk" and "taxonomy", was introduced in 2004 by Thomas Vander Wal and cited in a blog post by Gene Smith (2004). Smith uses the term "classification" for paraphrasing folksonomies. This term arouses a misleading and faulty connotation. The same holds for the term "taxonomy." Folksonomies are not classifications or taxonomies, since they work neither with notations nor with semantic relations. They are, however, a new type of knowledge organization system, with its own advantages and disadvantages.

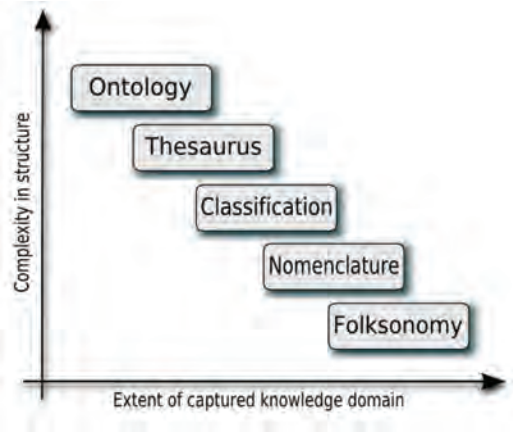
BACKGROUND

Knowledge Organization Systems

Knowledge representation methods are applied to provide a better basis for information retrieval tools. This may basically be done in two ways: by abstracting the topics of a document and by indexing a document, i.e., assigning content-descriptive keywords or placing it into a concept scheme (Cleveland & Cleveland, 2001; Lancaster, 2003). For indexing documents with content-descriptive keywords, different types of knowledge organization systems (KOS) have been developed. The most important methods – classifications, thesauri and nomenclatures – comprise a controlled vocabulary, which is used for indexing. The vocabulary of classifications and thesauri usually has the form of a structured concept hierarchy, which may be enriched with further semantic relations, e.g., relations of equivalence and concept associations (Peters & Weller, 2008; Weller & Peters, 2007).

Recently, two new developments have entered the spectrum of KOS: folksonomies and ontologies (Weller, 2007). They complement traditional techniques in different ways. Folksonomies include novel social dimensions of user involvement; ontologies extend the possibilities of formal vocabulary structuring (e.g., Alexiev et al., 2005; Davies, Fensel, & van Harmelen, 2003; Staab & Studer, 2004). Both have revived discussions about metadata on the web (Madhavan et al., 2006; Safari, 2004) and have increased the

Figure 2. Classification of KOS according to complexity and broadness of the domain



awareness of knowledge representation issues in scientific areas and even within the common web-user community.

We may classify different KOS according to the complexity of their formal structure (mainly defined by the number of specified semantic relations in use for structuring the vocabulary) and the extent of the captured domain (Figure 2). Both aspects are inversely proportional: the more complex the structure, the smaller the captured domain will have to be, due to feasibility reasons. Folksonomy is a completely unstructured method of document indexing. While in most other cases trained indexers or other experts are responsible for indexing documents, folksonomies allow the producers or the users of certain content to take over this task. There is no authority which controls the terminology in use. This also means that folksonomies are in no way limited to a certain domain of interest. They can be easily applied to all given contexts, as long as a community of interest exists.

Web 2.0 and User Collaboration

The term *Web 2.0* was coined during a discussion by Tim O'Reilly and Dale Dougherty from O'Reilly Media (O'Reilly, 2005). The phrase has since been widely assumed, yet definitions still vary slightly. Generally, it describes a new era of the World Wide Web, in which the users are in the spotlight and can to easily contribute to the creation of new web-content. The borders between "consumers" and "producers" of content are blurring; we may talk of a new type of web user: the "prosumer" as envisioned by Toffler (1980).

Furthermore, the focus is on many-to-many relationships. The interrelation of groups of users (namely, *communities*) is emphasized. The collaboration of large communities enables the creation of content in new formats and of enormous scale. Thus, besides social networking and personal interconnections, the interlinking of topics and discussions plays a decisive role. Various new communication channels create a "matrix of dialogues" (Maness, 2006) across different types of content and different data formats (e.g., blogs, wikis, podcasts, multimedia content, discussions, forums, personal profiles).

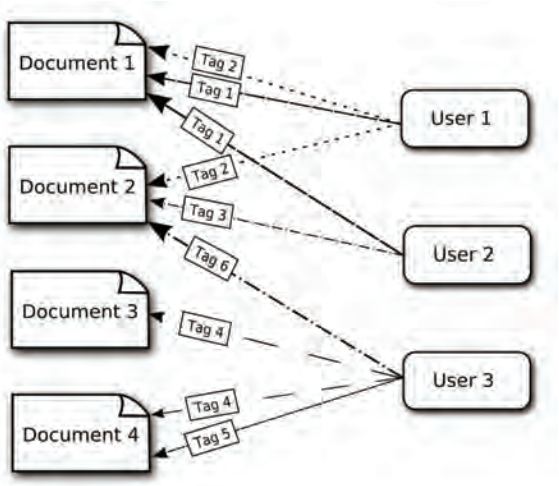
With this enormous growth of user-created content, new ways of navigating through it are needed. It was within social software platforms, that folksonomies have been introduced as an easy way to let users organize their data and make it accessible and retrievable: Everyone was allowed to tag documents with freely chosen keywords. With the success of the photo-sharing platform Flickr[1], the video-community YouTube[2], the social bookmarking tool Del.icio.us[3], and the blog search engine Technorati[4] the principles of searching documents by assigned tags became widely known.

FOLKSONOMIES AND SOCIAL TAGGING APPLICATIONS

Characteristics of Folksonomies

In folksonomies, we are confronted with three different elements (Marlow, Naaman, Boyd, & Davis, 2006): the documents to be described, the tags which are used for description, and the

Figure 3. Interrelation of tags, users, and documents (adapted fromPeters & Stock, 2007).



users who are indexing the content. These three elements enable different dimensions of interconnections that can be used for browsing and navigation (Figure 3).

Users as well as documents are interconnected with each other in a social network environment, in which the paths run along the tags. On the one hand, documents are linked "thematically" with each other when they have been indexed with the same tags. On the other hand, documents are related via users, so-called *shared users*. Finally, users are linked with each other when they use the same tags for indexing or when they index the same documents. Users are thematically related when they index with the same tags; they are coupled with shared documents when they index the same documents. Thus, tags can help to identify communities of interest.

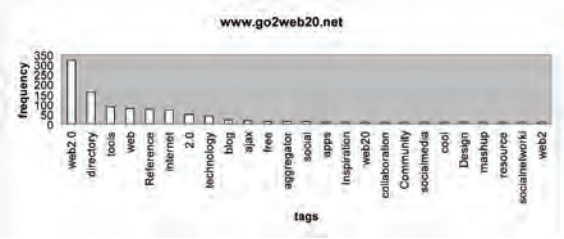
The extent of commonality may be illustrated quantitatively with similarity rates such as Cosinus, Jaccard-Sneath or Dice (Stock & Stock, 2008, p. 373), while communities of similar users can be detected by cluster analysis. All these interrelations can be used to browse a document collection. One may find interesting documents not only via tag searches, but also by following links to documents, which related users have tagged or by using the tags that they also use. This can be described as one type of *social navigation*. The most popular tags of a folksonomy (via tag clouds) are another way for entering document collections and browsing for content (Sinclair & Cardew-Hall, 2008).

According to Vander Wal (2005), one can distinguish between two types of folksonomies: *broad folksonomies*, where one document can be tagged by several users, so that tags can be assigned to each document more than once (e.g., Del.icio.us); and *narrow folksonomies*, where each tag is recorded for a document only once (e.g., Flickr, Technorati, YouTube) (Vander Wal, 2005; see also Peters & Stock, 2007). Usually the document's author provides the tags, although occasionally other users are also allowed to add tags.

The basis for tag clouds are platform-specific or resource-specific tag distributions which could be also represented by graphs (see Figures 4 & 5). Vander Wal (2005), Shirky (2005) and others state that in broad folksonomies the *distribution of tags* given to a document follows a Lotka-like power law (Egghe & Rousseau, 1990, p. 293; Egghe, 2005). If this assumption is true, we see a curve with only few tags at the top of the distribution, and a "long tail" of numerous tags on the lower ranks on the right-hand side of the curve (see Figure 4). Investigations of document-specific tag distributions demonstrate that another prototypic tag distribution may appear as well. This inverse logistic distribution (Stock, 2006) shows a lot of relevant tags on the curve's left-hand side (the "long trunk") and the known "long tail" (see Figure 5). As both distributions share the characteristic "long tail," they are difficult to differentiate. Therefore, they are both often called "power law."

The evolution of tag distributions with "long tail" characteristic is commonly explained by the known "rich gets richer" or "success breeds success" phenomena. Popular tags will be used more often because of their better visibility whereas

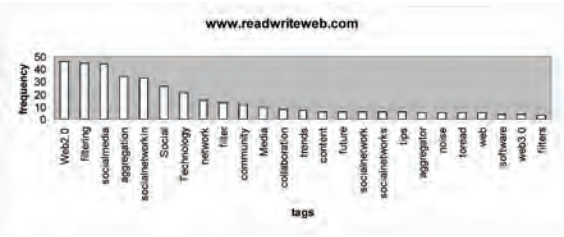
Figure 4. Power law distribution of tags (based on tagging data from Del.icio.us retrieved May 15, 2008, http://www.go2web20.net).



unpopular tags may be used seldom and will form the "long tail" (Cattuto, Loreto, & Pietronero, 2007; Halpin, Robu, & Shepherd, 2007). Several studies showed that the characteristic shape of the distributions of document-specific tags (not the absolute number of tags) will remain stable at a certain point in time (Kipp & Campbell, 2006; Golder & Huberman, 1006; Halpin, Robu, & Shepherd, 2007; Maass, Kowatsch, & Münster, 2007).

To enhance precision in folksonomy-based information retrieval systems, we may profit from this knowledge and work with *power tags* as an additional search feature (assuming that the most frequent tags are the most relevant ones and that less frequent tags can be neglected). In case of a power-law distribution we may consider only the first *n* tags (e.g., the first three tags) and in case of an inverse-logistic distribution we have to regard all tags of the long trunk up to the turning point of the curve (Peters & Stock, 2007). In most cases the "long trunk" will be shorter than the "long tail." Searches using the option "power tags only" will enhance precision of search results due to the reciprocal relationship of recall and precision. As power tags prune the tag distribution at a certain threshold, the amount of searchable and retrievable resources will decrease and recall will diminish along with it.

Figure 5. Inverse-logistic distribution of tags (based on tagging data from Del.icio.us retrieved May 15, 2008, from http://www.readwriteweb.com).



Applications of Folksonomies

By now, folksonomies are an essential part of many social software and Web 2.0-based applications. Users can tag various types of data, including scientific articles, references, bookmarks, pictures, videos, audio files, blog posts, discussions, events or even other users. The emergent content-descriptive tags can be used as an additional access-point to data collections besides traditional folder structures. They are particularly needed to improve retrievability of non-textual documents, such as videos and photos. Based on an idea by Luis von Ahn (2006), Google uses a game-like application to incite users to tag pictures on the Web. With the Google Image Labeler[5] metadata for large collections of images on the Web are collected to improve Google's image search.

Besides the various Web 2.0 applications, it is also possible to work with folksonomies in other contexts, e.g., in intranets of companies (Fichter, 2006), for indexing corporate blogs, podcasts and vodcasts (Peters, 2006), for corporate bookmarking services (Millen et al., 2006) and message boards (Murison, 2005).

Commercial online information suppliers have started to work with folksonomies as well (e.g., Engineering Village[6] by Reed Elsevier and WISO[7] by GENIOS). Folksonomies are suggested for broader use within professional databases (Stock, 2007b) as well as libraries (Kroski, 2005; Spiteri, 2006). Single libraries have begun to implement social tagging applications for their catalogue

Table 1. Benefits and problems with folksonomies

Benefits	Problems
Folksonomies • represent an authentic use of language • allow multiple interpretations • recognize neologisms • are cheap methods of indexing • are the only way to index mass information on the Web • give the quality "control" to the masses • allow searching and – perhaps even better – browsing • can help to identify communities • are sources for collaborative recommender systems • are sources for the development of ontologies, thesauri or classification systems • make people sensitive to information indexing issues	Folksonomies • have no vocabulary control and do not recognize synonyms and homonyms • do not make use of semantic relations between tags • mix up different basic levels • merge different languages • do not distinguish formal from content-descriptive tags • include spam-tags, user-specific tags, and other misleading keywords

(e.g., the University of Pennsylvania with its system PennTags[8]).

Trant (2006) analyzes folksonomies as a user centered access point to art museums collections. One can furthermore envision social tagging as an addition to classificatory folder approaches, e.g., for online shopping portals.

Some software developers have also integrated the tagging principle into their products. With Windows Vista, Microsoft enables tagging of several data formats, such as pictures, videos and Office files. A similar approach can be found with Apple's iPhoto. Yet, as long as tagging is performed by single users within their personal workspace, the social component is lacking and we cannot speak of folksonomies in a strict sense, but of *personomies* (Hotho et al., 2006).

Benefits and Problems with Folksonomies

Metadata produced by broad communities are cost-efficient and can easily be applied to large data collections. The pros and cons of folksonomies are well discussed (e.g., Kroski, 2005; Peters & Stock, 2007; Smith, 2008). Discussion on the quality of folksonomies often focuses on a comparison with other KOS. While traditional techniques are based on elaborated knowledge representations techniques, controlled vocabularies, and expert skills in indexing, folksonomies rely on the principles of "collective intelligence" (Weiss, 2005) and wisdom of the crowds. This leads to the following key aspects of a critical reflection on folksonomies, which will be discussed in more detail (Weller, 2007): (a) the confrontation of user's language versus vocabulary control; (b) the social and personal objectives in tagging behavior; and (c) the contrast between retrieval and exploration. Table 1 summarizes the main benefits and problems with folksonomies.

The main property of a folksonomy is that it authentically captures the language-use of its user community and reflects the prosumers' conceptual model of information (Quintarelli, 2005). People are free to use whichever tags they want and do not depend on a predefined set of terms. This freedom in the choice of tags however means that folksonomies are entirely uncontrolled vocabularies, which leads to the well known "vocabulary problem" (Furnas et al., 1987; Furnas et al., 2006; Golder & Huberman, 2006; Mathes, 2004): Different people use different words to describe the same object. Synonyms, trans-language synonyms, spelling variants and abbreviations are not bound together. Thus, someone searching for "United States of America" will not find documents tagged with "US", "USA", "United States" or "America." Homonyms and polysems are not distinguished, thus for example searching for "trunk" in Flickr's folksonomy will retrieve photos of trees as well as suitcases and probably also elephants. As different languages are used within most folksonomies, additional trans-language homonyms may occur. Misspellings and encoding limitations are serious problems for folksonomies (Guy & Tonkin, 2006). All these peculiarities have to be kept in mind by the user when searching in folksonomy based systems. Alternatively, additional techniques of

vocabulary control may be applied to avoid some of these problems.

But still the flexibility in the choice of tags is probably also the greatest advantage of folksonomies: It enables timeliness and multiple perspectives. A controlled vocabulary is always bound to a certain point in time and to a certain point of view. Folksonomy users can create tags quickly in response to new developments and changes in terminologies (Kroski, 2005).

Some tags may be neologisms. Mathes (2004) discusses the words "sometaithurts" (for "so meta it hurts") and "flicktion" on Flickr. "Although small, there is a quick formation of new terms to describe what is going on, and others adopting that term and the activity is describes" (Mathes, 2004). Such an unanticipated and unexpected use of tags reflects a "communication and ad-hoc group formation facilitated through metadata" (Mathes, 2004).

Collaborative tagging of documents leads to "multiple interpretations", different and sometimes disparate opinions and "multicultural views" of the same piece of information (Peterson, 2006). Folksonomies "include everyone's vocabulary and reflect everyone's needs without cultural, social, or political bias" (Kroski, 2005), even niche interests can be represented. "Shared intersubjectivities" enable the users "to benefit, not just from their own discoveries, but from those of others" (Campbell, 2006, p. 10). Tags can be used as basis for recommender systems (Szomszor et al., 2007). Yet, the intentions of tags in social tagging systems are not always social. Users who tag documents do not necessarily do this with the objective of helping a community in finding relevant documents. Many users simply use tags to organize their own private documents. Vander Wal (2008) describes social tagging as being "collective" work rather than "collaborative." Thus, many tags in use are personal rather than social (Guy & Tonkin, 2006).

Some tags do not describe the document, but give a judgment ("stupid"). User-specific tags describe or evaluate a document only from the user's very own perspective so that some tags "are virtually meaningless to anybody except their creators" (Pluzhenskaia, 2006, p. 23). Some other tags can be called "performative": Often a planned or done activity is tagged, for example "toread" on Del.icio.us (Kipp, 2006a). Additionally, there are syncategorematic tags – terms which can only be understood in the specific context. A good example of this type of tag is the term "me" on Flickr, which describes a photo of the document's author. Some keywords are even mere spam-tags.

Overall, research has analyzed the nature of tags as well as the different functions of tags (see Al-Khalifa & Davis, 2007; Golder & Huberman, 2006; Kipp, 2006b). The strength of folksonomies lies in "serendipity" (Mathes, 2004), in discovering information via different paths, and in easy to handle search mechanisms (Quintarelli, 2005). "The long tail paradigm is also about *discovery of information*, not just about finding it", Quintarelli (2005) adds. Folksonomies provide different entry points to document collections; as described above, users may browse along relations between tags, users and documents. Searching with tags is much easier for non-information professionals than searching with elaborated retrieval tools such as, for example, the International Patent Classification system. On the other hand, professionally generated metadata are usually segmented into different fields, such as the document type and the notations of classification systems. Here indexing distinguishes formal aspects from content-descriptive information ("aboutness"). In folksonomies a strict boundary between different metadata is lacking. There are tags that identify what a document is about. At the same time, one can find tags referring to formal descriptions at the same level: i.e., tags identifying the owner of the document or tags referring to file format (Golder & Huberman, 2006, p. 203). Within a library catalog, this can cause problems, as one could not, for example, clearly distinguish between books written by William Shakespeare and books about him.

SOLUTIONS AND RECOMMENDATIONS

There are basically three different approaches aiming to solve the present problems of folksonomies. All approaches complement each other. First, one can focus on the actors and try to educate users to improve "tag literacy" (Guy & Tonkin, 2006). The second approach comprises combinations of social tagging with other knowledge organization systems (Weller, 2007). And finally we may generally consider tags as elements of natural language and treat them by means of automatic methods of natural language processing (NLP) for better retrieval results (Peters & Stock, 2007; Stock, 2007a, chapter 13-18).

Improvement of tag literacy would require a broader understanding of indexing principles within the folksonomy community. For training the user in selecting "good" tags, systems that suggest tags to the users (based on co-occurrences, lexical similarity or semantic relations) may be useful (MacLaurin, 2005; Xu et al., 2006). Yet, providing tag suggestions has influences on the classical "wisdom of the crowds" approach; the social component of folksonomies may get lost and the "success breeds success"-effect (Egghe & Rousseau, 1995) may adulterate the tag distributions.

Combinations of folksonomies and other KOS are very promising: e.g., approaches of using clustering mechanisms to apply some structure for search result presentations or methods of automatic query expansion or query refinement (Grahl et al., 2007; Gruber, 2007; Kolbitsch, 2007). Some research is done in the field of emergent semantics: i.e., gradually growing semantic structures from folksonomies to more complex KOS (see Zacharias & Braun, 2007) for instance by identifying existing semantic interrelations between concepts (Angeletou et al., 2007; Peters & Weller, 2008). Related approaches of editing and improving unstructured folksonomies with basic vocabulary control are discussed as *tag gardening* (Governor, 2006; Weller & Peters, 2008).

The development and updating of structured KOS can profit from folksonomies (Aurnhammer et al., 2006; Christiaens, 2006; Gendarmi & Lanubile, 2006; Macgregor & McCulloch, 2006; Mika, 2005; Spyns et al., 2006; Zhang et al., 2006), because the tags, their frequency and their distribution are sources for new controlled terms, for modifications of terms and perhaps for deleting concepts in the sense of a "bottom-up categorization" (Vander Wal, 2005). In this way tags guarantee a fast response to changes and innovations in the knowledge domain. Generally, folksonomies should not be regarded as competitors for classical KOS but rather as a complement.

According to Peters (2006) it is not advisable to work exclusively with folksonomies in professional environments (e.g., intranets, commercial online services), but to mix them with other indexing methods. Here, a layer model (Krause, 1996) for the combined use of folksonomies, thesauri, classification systems, etc. will work well.

To revise applied tags for effective information indexing and retrieval, it is useful to treat them by means of NLP. After language identification and parsing of tags, typical NLP-tasks, including error detection, word form conflation, identification of named entities, phrase recognition, and decompounding, can be executed. In this way, the variety and ambiguity of tags can be reduced considerably.

FUTURE TRENDS

The topics of information retrieval and relevance ranking within folksonomies have not yet been discussed exhaustively. First approaches try to implement a PageRank-like relevance algorithm, the *FolkRank*, for the ranking of tagged documents (Hotho et al., 2006): "The basic notion is that a resource which is tagged with important tags by important users becomes important itself" (Hotho et al., 2006, 417). A patent application by Yahoo!

for its photo-sharing service Flickr proposes an "interestingness" ranking which takes into account, for instance, the user's behavior in clicking and tagging or the number of assigned tags (Butterfield et al., 2006). All in all, three sets of applicable ranking factors can be determined: (1) tags, (2) collaboration, and (3) prosumers (Peters & Stock, 2007).

Besides research efforts on improving the quality of folksonomies, some work is also done to use them as a basis for new applications (or as a source for data mining). For example, different methods for identifying communities of interest with the help of folksonomies are considered (Diederich & Iofciu, 2006; Wu et al., 2006), and analyses on how people tag documents on the web might lead to a better understanding of how humans organize and process information (Lodwick, 2005).

CONCLUSION

Folksonomies present a valuable addition to the spectrum of knowledge representation methods. They appear in the context of user collaboration in Web 2.0 environment and provide easy and comprehensive access to large data collections. With web users taking control over document indexing, folksonomies offer an inexpensive way of processing large data sets. User centered approaches to tagging have multiple benefits, as they can actively capture the authentic language of the user, are flexible and allow new ways of social navigation within document collections. Yet some problems derive from the unstructured nature of tags which may be solved by improving the users' tag literacy, by (automatic) query refinements, or by processing tags through natural language processing.

In the future, the advantages and shortcomings of folksonomies will be considered more closely as advanced approaches to the use of social tagging applications are emerging. Folksonomies and traditional knowledge representation methods are not to be viewed as rivalling systems; additionally, new options for combinations of different techniques will be designed. This will also be particularly beneficial in specialized contexts, since the number of professional database providers, libraries, and museums that have adapted folksonomies continues to grow.

REFERENCES

Al-Khalifa, H. S., & Davis, H. C. (2007). Towards better understanding of folksonomic patterns. In S. Harper et al. (Eds.), *Proceedings of the 18th Conference on Hypertext and Hypermedia* (pp. 163-166). New York: ACM Press.

Alexiev, V., Breu, M., de Bruijn, J., Fensel, D., Lara, R., & Lausen, H. (Eds.). (2005). *Information integration with ontologies: Experiences from an industrial showcase.* Chichester, England: Wiley & Sons.

Angeletou, S., Sabou, M., Specia, L., & Motta, E. (2007). Bridging the gap between folksonomies and the Semantic Web: An experience report. In B. Hoser & A. Hotho (Eds.), *Bridging the gap between Semantic Web and Web 2.0* (pp. 30-43). Innsbruck, Austria: International Workshop at the 4th European Semantic Web Conference (SemNet 2007).

Aurnhammer, M., Hanappe, P., & Steels, L. (2006). Augmenting navigation for collaborative tagging with emergent semantics. *Lecture Notes in Computer Science*, *4273*, 58–71. doi:10.1007/11926078_5

Butterfield, D. S., Costello, E., Fake, C., Henderson-Begg, C. J., & Mourachow, S. (2006). *Interestingness ranking of media objects* (Patent Application No. US 2006/0242139 A1). Washington, DC: U.S. Patent and Trademark Office.

Campbell, D. G. (2006). A phenomenological framework for the relationship between the Semantic Web and user-centered tagging systems. In *Proceedings of the 17th Workshop of the American Society for Information Science and Technology, Special Interest Group in Classification Research*, Austin, TX. Retrieved June 23, 2008, from http://dlist.sir.arizona.edu/1838/

Cattuto, C., Loreto, V., & Pietronero, L. (2007). Semiotic dynamics and collaborative tagging. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(5), 1461–1464. doi:10.1073/pnas.0610487104

Christiaens, S. (2006). Metadata mechanisms: From ontology to folksonomy… and back. *Lecture Notes in Computer Science*, *4277*, 199–207. doi:10.1007/11915034_43

Cleveland, D. B., & Cleveland, A. (2001). *Introduction to indexing and abstracting*. Englewood, CO: Greenwood Press.

Davies, J., Fensel, D., & van Harmelen, F. (Eds.). (2003). *Towards the Semantic Web: Ontology-driven knowledge management*. Chichester, England: Wiley & Sons.

Diederich, J., & Iofciu, T. (2006). Finding communities of practice from user profiles based on folksonomies. In *Proceedings of the 1st International Workshop on Building Technology Enhanced Learning Solutions for Communities of Practice (TEL-CoPs'06)*. Retrieved June 23, 2008, from http://palette.cti.gr/workshops/telcops06.htm

Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Amsterdam: Elsevier Academic Press.

Egghe, L., & Rousseau, R. (1990). *Introductions to informetrics.* Amsterdam: Elsevier.

Egghe, L., & Rousseau, R. (1995). Generalized success-breeds-success principle leading to time-dependent informetric distributions. *Journal of the American Society for Information Science and Technology*, *46*(6), 426–445. doi:10.1002/(SICI)1097-4571(199507)46:6<426::AID-ASI3>3.0.CO;2-B

Fichter, D. (2006). Intranet applications for tagging and folksonomies. *Online*, *30*(3), 43–45.

Frakes, W. B., & Baeza-Yates, R. (Eds.). (1992). *Information retrieval: Data structure and algorithms.* Upper Saddle River, NJ: Prentice Hall.

Furnas, G. W., Fake, C., von Ahn, L., Schachter, J., Golder, S., Fox, K., et al. (2006). Why do tagging systems work? *CHI'06 Extended Abstracts on Human Factors in Computing Systems* (pp. 36-39). New York: ACM.

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communication: An analysis and a solution. *Communications of the ACM*, *30*, 964–971. doi:10.1145/32206.32212

Gendarmi, D., & Lanubile, F. (2006). Community-driven ontology evolution based on folksonomies. *Lecture Notes in Computer Science*, *4277*, 181–188. doi:10.1007/11915034_41

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, *32*(2), 198–208. doi:10.1177/0165551506062337

Gordon-Murnane, L. (2006). Social bookmarking, folksonomies, and Web 2.0 tools. *Searcher - The Magazine for Database Professionals*, *14*(6), 26-38.

Governor, J. (2006, October 1). *On the emergence of professional tag gardeners*. Retrieved June 23, 2008, from http://www.redmonk.com/jgovernor/2006/01/10/on-the-emergence-of-professional-tag-gardeners/

Grahl, M., Hotho, A., & Stumme, G. (2007). Conceptual clustering of social bookmarking sites. In *Proceedings of I-Know'07*, Graz, Austria (pp. 356-364).

Gruber, T. (2007). Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Web Semantics and Information Systems, 3*(1), 1–11.

Guy, M., & Tonkin, E. (2006). Folksonomies: Tidying up tags? *D-Lib Magazine, 12*(1). Retrieved June 23, 2008, from http://www.dlib.org/dlib/january06/guy/01guy.html

Halpin, H., Robu, V., & Shepherd, H. (2007). The complex dynamics of collaborative tagging. In *Proceedings of the 16th Conference on World Wide Web* (pp. 211-220). New York: ACM.

Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005). Social bookmarking tools: A general review. Part I. *D-Lib Magazine, 11*(4). Retrieved June 23, 2008, from http://www.dlib.org/dlib/april05/hammond/04hammond.html

Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006). Information retrieval in folksonomies: Search and ranking. *Lecture Notes in Computer Science, 4011*, 411–426. doi:10.1007/11762256_31

Kipp, M., & Campbell, D. (2006). Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. In *Proceedings of the 17th Annual Meeting of the American Society for Information Science and Technology,* Austin, TX.

Kipp, M. E. I. (2006a). @toread and cool: Tagging for time, task and emotion. In *17th ASIS&T SIG/CR Classification Research Workshop*: *Abstracts of Posters*, Austin, TX (pp. 16-17).

Kipp, M. E. I. (2006b). Exploring the context of user, creator and intermediate tagging. In *Proceedings of the IA Summit 2006*, Canada. Retrieved June 23, 2008, from http://www.iasummit.org/2006/files/109_Presentation_Desc.pdf

Kolbitsch, J. (2007). WordFlickr: A solution to the vocabulary problem in social tagging systems. In *Proceedings of I-MEDIA '07 and I-SEMANTICS'07*, Graz, Austrai (pp. 77-84).

Krause, J. (1996). Informationserschließung und -bereitstellung zwischen Deregulation, Kommerzialisierung und weltweiter Vernetzung. *IZ-Arbeitsbericht, 6*.

Kroski, E. (2005). *The hive mind: Folksonomies and user-based tagging*. Retrieved June 23, 2008, from http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging

Lancaster, F. W. (2003). *Indexing and abstracting in theory and practice* (3rd ed.). Urbana, IL: University of Illinois Urbana.

Lodwick, J. (2005). *Tagwebs, Flickr, and the human brain*. Retrieved January 10, 2008, from http://www.blumpy.org/tagwebs/

Maass, W., Kowatsch, T., & Münster, T. (2007). Vocabulary patterns in free-for-all collaborative indexing systems. In *Proceedings of International Workshop on Emergent Semantics and Ontology Evolution*, Busan, Korea (pp. 45-57).

Macgregor, G., & McCulloch, E. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, *55*(5), 291–300. doi:10.1108/00242530610667558

MacLaurin, M. B. (2005). *Selection-based item tagging* (Patent Application No. US 2007/0028171 A1). Washington, DC: U.S. Patent and Trademark Office.

Madhavan, J., Halevy, A., Cohen, S., Dong, X., Jeffery, S. R., Ko, D., & Yu, C. (2006). Structured data meets the Web: A few observations. *A Quarterly Bulletin of the Computer Society of the IEEE Technical Committee on Data Engineering*, *29*(4), 19–26.

Maness, J. M. (2006). Library 2.0 theory: Web 2.0 and its implications for libraries. *Webology*, *3*(2). Retrieved January 10, 2008, from http://www.webology.ir/2006/v3n2/a25.html

Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). HT06, tagging paper, taxonomy, Flickr, academic article, to read. In *Proceedings of the 17th Conference on Hypertext and Hypermedia* (pp. 31-40). New York: ACM.

Mathes, A. (2004). *Folksonomies – cooperative classification and communication through shared metadata*. Urbana, IL: University of Illinois Urbana-. Retrieved June 23, 2008, from http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

Merholz, P. (2004, October 19). *Metadata for the masses*. Retrieved January 10, 2008, from http://www.adaptivepath.com/publications/essays/archives/000361.php

Mika, P. (2005). Ontologies are us: A unified model of social networks and semantics. *Lecture Notes in Computer Science*, *3729*, 522–536. doi:10.1007/11574620_38

Millen, D. R., Feinberg, J., & Kerr, B. (2006). DOGEAR: Social bookmarking in the enterprise. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 111-120). New York: ACM.

Murison, J. (2005). Messageboard topic tagging: User tagging of collectively owned community content. In *Proceedings of the 2005 Conference on Designing for User eXperience* (Article no. 5). New York: American Institute of Graphic Art.

O'Reilly, T. (2005). *What is Web 2.0: Design patterns and business models for the next generation of software*. Retrieved January 10, 2008, from http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

Peters, I. (2006). Against folksonomies: Indexing blogs and podcasts for corporate knowledge management. In H. Jezzard (Ed.), *Proceedings of Preparing for Information 2.0, Online Information 2006* (pp. 93-97). London: Learned Information Europe.

Peters, I., & Stock, W. G. (2007). Folksonomy and information retrieval. In *Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology, 45*, 1510-1542.

Peters, I., & Weller, K. (2008). Paradigmatic and syntagmatic relations in ontologies and folksonomies. *Information . Wissenschaft und Praxis*, *59*(2), 100–107.

Peterson, E. (2006). Beneath the metadata: Some philosophical problems with folksonomies. *D-Lib Magazine*, *12*(11). doi:10.1045/november2006-peterson

Pluzhenskaia, M. (2006). Folksonomies or fauxsonomies: How social is social bookmarking? In *17th ASIS&T SIG/CR Classification Research Workshop: Abstracts of Posters*, Austin, TX (pp. 23-24).

Quintarelli, E. (2005). *Folksonomies: Power to the people*. Paper presented at the ISKO Italy UniMIB Meeting, Milan.

Safari, M. (2004): Metadata and the Web. *Webology, 1*(2). Retrieved January 10, 2008, from http://www.webology.ir/2004/v1n2/a7.html

Shirky, C. (2005). *Ontology is overrated: Categories, links, and tags.* Retrieved January 10, 2008, from www.shirky.com/writings/ontology_overrated.html

Sinclair, J., & Cardew-Hall, M. (2008). The folksonomy tag cloud: When is it useful? *Journal of Information Science*, *34*(1), 15–29. doi:10.1177/0165551506078083

Smith, G. (2004, August 3). *Folksonomy: Social classification*. Retrieved January 10, 2008, from http://atomiq.org/archives/2004/08/folksonomy_social_classification.html

Smith, G. (2008). *Tagging: People-powered metadata for the social Web*. Berkeley, CA: New Riders.

Spiteri, L. F. (2006). The use of folksonomies in public library catalogues. *The Serials Librarian*, *51*(2), 75–89. doi:10.1300/J123v51n02_06

Spyns, P., de Moor, A., Vandenbussche, J., & Meersman, R. (2006). From folksonomies to ontologies: How the twain meet. *Lecture Notes in Computer Science*, *4275*, 738–755. doi:10.1007/11914853_45

Staab, S., & Studer, R. (Eds.). (2004). *Handbook on ontologies*. Berlin, Germany: Springer.

Stock, W. G. (2006). On relevance distributions. *Journal of the American Society for Information Science and Technology*, *57*(8), 1126–1129. doi:10.1002/asi.20359

Stock, W. G. (2007a). *Information retrieval: Informationen suchen und finden*. München, Germany: Oldenbourg.

Stock, W. G. (2007b). Folksonomies and science communication. A mash-up of professional science databases and Web 2.0 services. *Information Services & Use*, *27*(3), 97–103.

Stock, W. G., & Stock, M. (2008). *Wissensrepräsentation: Informationen auswerten und bereitstellen*. München, Germany: Oldenbourg.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few*. New York: Anchor Books.

Szomszor, M., Cattuto, C., Alani, H., O'Hara, K., Baldassarri, A., Loreto, V., et al. (2007, June). Folksonomies, the Semantic Web, and movie recommendation. In B. Hoser & A. Hotho (Eds.), *Bridging the gap between Semantic Web and Web 2.0* (pp. 71-84). Innsbruck, Austria: International Workshop at the 4th European Semantic Web Conference (SemNet 2007).

Toffler, A. (1980). *The third wave*. New York: Morrow.

Trant, J. (2006). Exploring the potential for social tagging and folksonomy in art museums: Proof of concept. *New Review of Hypermedia and Multimedia*, *12*(1), 83–105. doi:10.1080/13614560600802940

Vander Wal, T. (2005, February 21). *Explaining and showing broad and narrow folksonomies*. Retrieved January 10, 2008, from http://www.vanderwal.net/random/entrysel.php?blog=1635

Vander Wal, T. (2008). Keeping up with social tagging. In *Workshop Good Tags – Bad Tags, Social Tagging in der Wissensorganisation*, Institut für Wissensmedien, Tübingen, Germany. Retrieved June 23, 2008, from http://www.e-teaching.org/community/taggingcast

von Ahn, L. (2006). Games with a purpose. *IEEE Computer Magazine*, 96-98.

Weiss, A. (2005). The power of collective intelligence. *netWorker, 9*(3), 16-23.

Weller, K. (2007). Folksonomies and ontologies: Two new players in indexing and knowledge representation. In H. Jezzard (Ed.), *Online Information Conference Proceedings* (pp. 108-115). London: Learned Information Europe.

Weller, K., & Peters, I. (2007). Reconsidering relationships for knowledge representation. In *Proceedings of I-KNOW '07*, Graz, Austria (pp. 493-496).

Weller, K., & Peters, I. (2008). Seeding, weeding, fertilizing – different tag gardening activities for folksonomy maintenance and enrichment. In *Triple-I Conference, Proceedings of I-Semantics*, Graz, Austria (pp. 110-117).

Wu, H., Zubair, M., & Maly, K. (2006). Harvesting social knowledge from folksonomies. In *Proceedings of the 17th Conference on Hypertext and Hypermedia* (pp. 111-114). New York: ACM.

Xu, Z., Fu, Y., Mao, J., & Su, D. (2006). Towards the Semantic Web: Collaborative tag suggestions. In *Proceedings of the 15th International WWW Conference*.

Zacharias, V., & Braun, S. (2007). SOBOLEO. Social bookmarking and lightweight ontology engineering. In *Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge (CKC), 16th International World Wide Web Conference (WWW 2007)*, Banff, Alberta, Canada.

Zhang, L., Wu, X., & Yu, Y. (2006). Emergent semantics from folksonomies: A quantitative study. *Lecture Notes in Computer Science, 4090*, 168–186. doi:10.1007/11803034_8

KEY TERMS AND DEFINITIONS

Broad and Narrow Folksonomies: Broad and narrow folksonomies differ in whether multiple assignments of identical tags are possible or not. Systems with broad folksonomies allow to assign the same tag to one document several times (thus the tag frequency can be counted), whereas narrow folksonomies record every tag only once.

Folksonomy: An indexing method open for users to apply freely chosen index terms. The term "folksonomy" was introduced in 2004 by Thomas Vander Wal as a combination of "folk" and "taxonomy."

Knowledge Representation and Indexing: In the context of information storage and retrieval techniques, knowledge representation is concerned with providing methods for organizing and representing knowledge domains and sorting documents accordingly. A traditional way to do this is by document indexing: i.e., by assigning keywords or notations (usually taken from a controlled vocabulary or classification scheme) to a document to describe its content.

Knowledge Organization Systems (KOS): Knowledge Organization Systems are (structured) representations of a knowledge domain, used for document classification and indexing. Common classical knowledge organization systems include classifications (taxonomies), thesauri, and nomenclatures. Folksonomies and ontologies are new forms of KOS.

Tag: Within a given context, a tag is a keyword assigned to a document to describe it. Tags can be used for document retrieval. Folksonomy tags can be freely chosen by the users of a folksonomy-based system.

Tag Cloud: A tag cloud displays the popularity of tags, either for tags assigned to one single document or for all tags within a complete folksonomy-based platform. The bigger and broader a tag is displayed in a tag cloud, the more often has it been used.

Tag Distribution: The frequency of tags assigned to one document (or within a platform) can be counted and visualized as a tag distribution graph. Some specific forms of tag distributions are dominant within folksonomies: for example, the emergence of a "long tail", which reacts to the rules of the power law. A "long trunk" may appear as well; the curve then follows an inverse-logistic distribution.

ENDNOTES

1 Flickr: http://www.flickr.com
2 You Tube: http://www.youtube.com
3 Del.icio.us: http://del.icio.us
4 Technorati: http://www.technorati.com
5 Google Image Labeler: http://images.google.com/imagelabeler/
6 Engineering Village: http://www.engineeringvillage.com
7 WISO: http://www.wiso-net.de
8 PennTags: http://tags.library.upenn.edu/