1-1-2011

Folksonomy-based User-centric Information Organization Systems

Devendra Potnis
University of Tennessee, Knoxville, dpotnis@utk.edu

Folksonomy-based User-centric Information Organization Systems

Devendra Dilip Potnis
School of Information Sciences
University of Tennessee Suite 442 Communications Bldg
1345 Circle Park Knoxville
TN - 37996. USA

ABSTRACT: *After the rise of Web 2.0 technologies, Folksonomy emerged as one of the most popular user-centric information organization techniques in the newly reforming digital world. In the context of the rising popularity of folksonomies, this paper conceptually investigates – how can folksonomies be integrated with other ways of organizing information for building systems that could serve information users better in the future? A set of experimentation ideas of integrating folksonomy with controlled vocabularies, bibliometric maps, thesaurus-based systems, ontologies, and semantic web could inform and catalyze the design and development of folksonomy-based user-centric information organization systems in the future.*

Keywords: Web, Information Users, Folksonomy, User-centric information organization techniques

Received: 11 July 2010, Revised 19 September 2010, Accepted 29 September 2010

1. Introduction

In the 1990's, the early phase of Web 1.0 presented websites with a combination of individual posting and corporate catalogs published in the form of simple documents (W3C, 2001). Widespread usage of information and communication technologies (ICTs) led to abundant production and consumption of information in the digital format. With a number of websites doubling every 18 months (Moore's law), the explosion of digital information introduced users to fundamentally distinct ways of producing, organizing, and sharing information. By the end of the last millennium, users started realizing the power of information dissemination through websites, which encouraged Internet users to change their usage pattern and the ratio of information producer to information consumer also started shrinking. Feature like "uploading" boosted users' contribution to the digital world (Friedman, 2005). Web 2.0[1] transformed websites from an abstract space of information to digitally miscellaneous virtual world (Weinberger, 2007). Web 2.0 provides a conducive environment to users for transforming them to co-developers, promoting their collective intelligence (O'Reilly, 2005). One of the basic lessons from Web 2.0 is that users are actively willing to contribute data as well as metadata (Mika, 2007). In the Web 2.0 era, users became "active" producers of information on the Web rather than "passive" consumers of pre-existing information on the Web. Information users can now easily move anywhere on the axis of information with information consumer and information producer as its endpoints (See Figure 1).

Figure 1. Role of Users on Information Axis

Eventually, users began to experience the need to retrieve an astronomical amount of information on the Web in an effective and efficient manner. Google's page rank algorithm emerged as one of the finest full-text indexing techniques to serve the purpose (Voss, 2007). However, it could not address the need of organizing information and knowledge on the Web (Voss, 2007). This need of managing and organizing user-created information on the Web served as the foundation for inventing a new semantics called "folksonomies"[2] (Kompatsiaris and Hobson, 2008). Folksonomies (folks-taxonomies), i.e., taxonomies proposed by folks - the term "folks" refers to untrained and unskilled organizers of information - emerged as one of the ways to organize digital information. Prior to Web 2.0, information users used to rely entirely on experts, who used to decide upon filtering information that could be published as a piece of information or knowledge, approved by so-called information authorities. But, in Web 2.0 when someone can publish their content at the "press of a button" on the Internet, the relation between information and information gatekeepers have changed drastically (**Weinberger, 2007**). Clearly, information users are at the center of this folksonomy movement.

1.1 Folksonomy: A User-centric Information Organization Technique

Folksonomy is a bottom-up social classification system, essentially very different than traditional classification systems (Vander Wal, 2004). It is a user-centric indexing technique of digital documents, images, and other types of digital creations and artifacts (Matusiak, 2006). Folksonomies allow users to annotate their resources using keywords. This type of manual indexing is referred to as "tagging". These user-defined descriptors (Abbas and Graham, 2006) of original data, often known as freely-formed keywords, could be potentially formed by anybody anytime. Such descriptors are also known as social tagging (Matusiak, 2006). Folksonomies heavily rely upon tagging. Tagging demonstrates the "principle of poly-perceptions" (Hjorland, 2002), where users can reflect their perceptions about a particular item such as images, videos, bookmarks, text, document, link, etc., by means of multiple tags. Since, anyone, anywhere, anytime can contribute to folksonomies they are referred to as the "lowest-common-denominator" of classification system (Shirky, 2005). Folksonomies can be interpreted as loud thinking by users, where well established editors, publishers, and recognized authors seem to be losing gravity in terms of information management (**Weinberger, 2007**). Usually, collaborations among users lead to building folksonomies; hence they are also labeled as "assemble of concepts" (Trant and Wyman, 2006). Mathematicians define folksonomy as a Tuple (F) := (U, T, R, Y), where U, T, and R are finite sets, whose elements are called users, tags, and resources, respectively, and Y is a ternary relation among them, i.e., $Y \subseteq U * T * R$. A post is a triple (u, Tur, r) with $u \in U$, $r \in R$, and $Tur := \{t \in T | (u, t, r) \in Y\}$. Users (U) are assigned identification numbers, tags (T) are assumed as arbitrary strings, resources (R) are assumed to be a dependent variable on overall network, and finally a dependent variable Y denotes a ternary relationship among the rest of the variables in F (Cattuto et. al., 2008; Hotho et. al., 2006).

1.1.1 What Makes Folksonomies So Popular?

Along with lack of learning curve for using folksonomies, many economic, psychological, financial, and social incentives encourage users to practice folksonomies. The reliance on natural languages makes users feel very comfortable while tagging. Users can express themselves through their routine vocabulary of words. In addition, folksonomies promote a lot of individuality as well. The technique offers users the unique liberty to share, filter, organize, and even interpret meaning of information. For example, Dabble.com allows users to tag and rank preexisting songs and create their own playlists. On 43things.com, users can share their emotions, desires, and day-today activities, by revealing the next 3 things they want to do in their lives. For those next 3 things, they get cheered up as well as precautions by others in the community. Users with similar interests get clustered together on the website, and they form a group for enjoying those activities together. For example, people interested in Yoga are tagged together and this gives an opportunity for a Yoga instructor to join the pre-existing group under Yoga tag on the website. Thus tagging on 43things.com could make available an employment opportunity for that Yoga instructor.

Thus folksonomy is a gift by Web 2.0 which keeps on giving to information users. A facility of forming a variety of tagging offers users a sense of creativity, and hence, folksonomy is often perceived as a revolution at an intersection of arts and science of categorization (Sterling, 2005). For example, Geotagging is a new way of tagging where users combine the longitude and latitude of their location as an identifier or tag. A user can form "74-19" as his/her geotag, indicating a location at the intersection of 74 degrees of longitude and 19 degrees of latitude. User-defined description (i.e. tag) of the snap could potentially reveal a mental heuristic held by the user for that snap. For example, if a user comes across a beautiful snap of Himalayas which he/she was searching for a long time, then he/she would attempt to associate that snap with his/her ways of thinking or memories. The association of tags with emotions help to keep "found things found" (Morville, 2005). Moreover, the website on which the snap was found could also be bookmarked by the user for future visits. Therefore, folksonomy enjoys the reputation of being one of the most popular digital information organization techniques in the world.

This newly booming technique of digital information organization makes it essential to study if folksonomies could be incorporated with established techniques of information organization to devise information organization systems offering better services to users in the future. The paper presents a conceptual roadmap of designing and developing folksonomy based user-centric information organization systems for the future. With the help of various applications of folksonomies, the next section illustrates the significance of designing and developing information organization systems based on folksonomies. The third section elaborates a combination of ideas and existing implementations for merging folksonomies with mainstream information organization techniques such as controlled vocabularies, bibliometric maps, thesaurus-based systems, ontologies, and the semantic web. The fourth section stresses the importance of knowing users well while designing and developing user-centric information organization systems. 3

2. Applications of Folksonomies, Highlighting the Significance

Folksonomy-based information organization systems would definitely be a commercial success in the areas where folksonomies have already established themselves. Folksonomies find its applications in areas including but not limited to business, marketing, cutting-edge research domains, organizational management and museum management. In addition, consensus-building and intranets are two other potential applications of folksonomies, which are still to a large extent in the theoretical stage. This section covers both established and conceptual applications of folksonomies to highlight the significance of designing and developing folksonomy-based user-centric information organization systems in the future.

2.1 Established Applications of Folksonomies

In business, there are many lucrative financial incentives associated with tagging. For example, a website like rawsugar.com significantly increases the profit levels of enterprises by applying innovative ways of sales and marketing through tagging (Damianos, 2006; Millne et. al., 2006). Social book-marking websites (e.g. del.icio.us and technocrati.com) are studied by sales and marketing agencies to find out current trends in markets. Nowadays, special discovery engines trace tagging in terms of bookmarks made by potential consumers for gauging the pulse of markets. Enterprises also receive advantages like shared corporate learning and information management through tagging (John and Seligmann, 2006). Tags generated by consumers on specific product lines are stored in databases. Those tags can be then used by database designers and developers for facilitating complex querying in the field of information retrieval (Udell, 2004).

Research shows that the corpora of knowledge, in cutting-edge knowledge domains, can be very well explained by their own users and contributors, using their own set of jargons rather than traditional controlled vocabularies (Morville, 2005). Interestingly, tagging has already taken a significant lead over traditional information organization techniques in such disciplines of cutting-edge knowledge. For example, genetics is an interdisciplinary research area. It has a lot of acronyms, and the field is extensively engaged into cross-boundary information sharing. The subject-indexing for such a cutting-edge knowledge domain can be developed better, by contributors and users from Genetics itself, rather than traditional catalogers. Information users from Genetics community have a great incentive to start tagging, reflecting their own vocabulary as a part of their communication in the community.

By setting up light-weight folksonomies, the corporate sector and individual users can overcome the problems of "information acquisition bottleneck" (Hotho et. al., 2006). For example, on a social networking website like LinkedIn.com, people are tagged according to their skills, level of expertise, number of years of experience, geographical location, and other virtues (Farrell and Lau, 2006). This also helps job-seekers to locate future employers who satisfy required criteria. Tagging on such websites promotes creation of "information pools" with different network ties, helping users to reach and convey their ideas and information content to heterogeneous audiences (Reagans and McEvily, 2003). Thus, tagging adds value to social networking efforts through the Internet.

Some recent trends indicate that employees of huge organizations are encouraged to develop "work-related" folksonomies, also referred as "emergent enterprise taxonomy," for their own workgroups in the organizations (Millne et. al., 2006). Researchers believe that such folksonomies have the potential to facilitate "workplace democracy," by managing distribution of work in workgroups. However, folksonomies in organizations can facilitate workflow, only if employees tag consistently without indulging into any personal interests or biases. Malicious intentions of employees towards organizations may harm the noble goals of developing folksonomies in organizations. Also, highly mechanistic organizations with structured hierarchies of governance are the toughest to be penetrated for developing work-democracies with the help of employee-designed folksonomies.

As part of museum management, visitors are encouraged by museum authorities or museum curators to tag resources (mainly artifacts), owned by museums (Trant and Wyman, 2006). Visitors are given incentives or motivation to tag museum artifacts, for increasing user engagement as well as to understand user perspectives towards artifacts managed by museums (Trant, 2006). Visitors who are willing to invest their time and resources help to make art collections more accessible for future visitors. It is also believed that tags created by users at museums could bridge the semantic gap between professional curators and popular language of museum visitors (Trant and Wyman, 2006). All of the above mentioned established applications of folksonomies suggest lucrative proposition for folksonomy-based information organization systems.

2.2 Folksonomy-based Information Organization Systems for Consensus-building

Folksonomies revolve around the power of communities and individual uploading, which in turn depend upon users' level of expertise and experience of tagging. Individuals who are avid surfers and frequent contributors to folksonomies are likely to introduce a lot more tags to the existing folksonomies than beginners or intermittent users (Mathes, 2006). The most used tags are expected to be more visible to all the users; this is also supported by the principle of the least efforts where all users are expected to pick up the most visible tags that are readily available on folksonomy. Naturally, the higher degree of visibility of most used tags has a higher possibility of being reinforced in the existing folksonomies. This continuous reinforcement of most visible and most used tags eventually create converged tags, also interpreted as "consensus-building" among users. Consensus-building is predicted to work only in closely-knit communities of users, where collaboration for tagging takes hardly any efforts (Bates, 2006). User communities dedicated for a particular cause would attain a matured closure on tagging in a lesser duration compared to a community having users with competing interests.

In contrast, the research exploring the application of folksonomies for consensus-building reports that as the popularity of a folksonomy increases it starts experiencing a flattening of tags i.e. the most used – most visible tags start losing their popularity. As the popularity of a folksonomy increases, more and more untrained users are attracted to that popular folksonomy carrying few most visible tags. The deep penetration of masses of user populations do not pick up the most visible – the most used tags in that folksonomy instead start "littering" the folksonomy with their own tags that could potentially introduce a lot of folksonomic flaws (Guy and Tonkin, 2006). Contradictory research evidences about the latent application of folksonomies for consensus-building among users keep commercial usage of information organization systems for consensus-building in the nascent stage.

2.3 Folksonomy-based Information Organization Systems for Intranets

The ability of folksonomies to offer immediate feedback to users empowers users for either strengthening existing folksonomies by reusing the most visible tags or diminishing the popularity of tags by introducing new tags. As soon as users assign tags to their items, clusters of items tagged by the same keywords are formed. The types of items or information content clustered around tags provide an opportunity for users to verify whether they meant the same. If users do not find clustered items under their tags as relevant, they can immediately rectify their tags, and place their items under appropriate baskets of tags.

Immediate feedback mechanism gives rise to virtual communities of users who are often guided by their passion, common interests and common conflicts (Burnett, 2001). Through an ongoing process of immediate feedback, gradually though subtly, these virtual communities could end up developing a complex set of norms, and influencing users' information behavior, attitudes, interests and language used for tagging. All non-confidential items available online and matching with users' tags can be organized into the same baskets, and most importantly, those items can be managed by many users at the same time. Users can gain access to items that are not created or uploaded by them. Users can also actively contribute to general resource discovery and information organization (MacGregor 2006). This application of folksonomies matches with the concept of *intranet* which is known as a collection of private networks pertinent to groups engaged for specific purposes. Collaborative tagging allows users to participate in exciting and highly interactive purpose-specific activities. People collectively enrich such shared data items in small worlds of their user communities formed under collaborative tagging, but: *how* small is their small world? From practitioners' point of view, can user communities (formed due to folksonomies) be converted to intranets?

Folksonomies when depicted as networks, show a very small average path length clustering coefficient (Cancho and Sol´e 2001, R´eka and Barab´asi 2002, Shen and Wu 2005). This mathematical evidence could be used to infer that at least as of now, users approach folksonomies with a very narrow-minded approach which leads them to form a number of small worlds within the Web. Another way to analyze above questions is entropy-based Link Analysis on Mining Informative Structure (LAMIS) algorithmic approach which assesses power taggers (i.e. taggers who tag the most or taggers whose tags are the most visible

or taggers whose maximum information content are clustered under the majority of tags), if any, who determine the behavior of user communities practicing folksonomy (Kao et al. 2002). LAMIS is an automatic informative structure extracting system based on the weighted link analysis of webpages, which reveals distinct degrees of influence held by power taggers in any different pockets of folksonomies. However, the nascent stage of research on algorithmic assessment of folksonomies makes it difficult to predict accurately the possible application of folksonomies for intranets.

According to circumstantial behavioral evidences of folksonomic communities, they do not share anything in common other than their own personal interests (MacGregor 2006). Once their immediate needs are answered, users are gratified in their small worlds (Chatman 1991). It is possible that users will not share data beyond their interests and expectations through social tagging. Even in collaborative environments, the focus on self-interest is found to lead the opportunistic behavior (Kumar et al. 1998). In the context of collaborative tagging, if users don't find any incentives other than their original interests for sharing data, their behavior could lead them to form a community promoting very narrow personal interests. There exists uncertainty about the possible application of folksonomies for intranets; however, such developing issues need not hinder the integration of folksonomy with other information organization techniques depicted in the next section.

3. Integration of Folksonomy with Other Information Organization Techniques

Various issues related to inconsistent user behavior lead to a growing *concern* over the unreliability of folksonomies in terms of findability and usability; hence, there is a growing *demand* for folksonomies to become more formal with proficient tags. Different kinds of hybrid designs of metadata, where folksonomies could be combined with other techniques of information organization are expected to support findability in newly designed information organization systems (Rosenfeld 2005). The following sub-sections discuss conceptual ideas that could be implemented in folksonomy-based information organization systems.

3.1 Folksonomies + Controlled Vocabularies in Hierarchical Representation

Incorporation of hierarchies and tagging leads to *tag bundles*. Applying the general concepts of hierarchies to tagging create *hierarchical folksonomies* also known as *pseudo hierarchical folksonomies* (Guy and Tonkin 2006). Collective annotation of content (tags) presented in hierarchical relationships offer a high-quality source of evidence for folksonomies (Plangprasopchok and Lerman 2008). Figure 2 depicts an illustration of integration of folksonomies with hierarchies. Each box could potentially represent a set of controlled vocabularies labeled as Flat World, Internet, Uploading and Folksonomies.

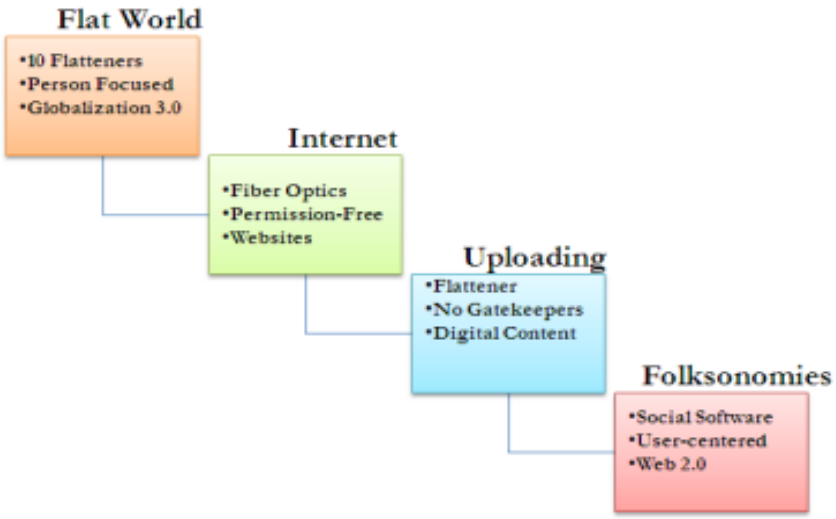

Figure 2. Combination (of Folksonomies + Hierarchies) for Today's World of Information Organization

Controlled vocabulary titled Flat World has three subject-indexed terms namely, 10 flatteners, Person Focused and Globalization 3.0. Similarly, a controlled vocabulary titled Uploading has three subject-indexed terms such as Flatteners, No Gatekeepers and Digital Content. Now let's visit the possible process of deriving the above-mentioned *pseudo-hierarchical folksonomy*

(Plangprasopchok and Lerman 2008). Initially, users will be asked to come up with a set of tags for today's world of information organization. Let's assume that users come with 12 overall tags such as: 10 flatteners, Person Focused, Globalization 3.0, Fiber Optics, Permission-Free, Websites, Flatteners, No Gatekeepers, Digital Content, Social Software, User-centric and Web 2.0. The second phase involves categorization of tags developed in the first round. Let's assume that users categorize above mentioned 12 freely-formed keywords (i.e. tags) into four categories such as Flat World, Internet, Uploading and Folksonomies. Thus we derive a more user-centric system of information organization with all user-defined descriptors categorized into buckets (or well-thought categories), integrating folksonomies with hierarchies, the best from both the worlds. The following table (see Table 1) synthesizes strengths and weaknesses of folksonomony-based controlled vocabularies.

Strengths	Weaknesses
• Limits the number of terms • Maintains authority control over names, spellings, etc. • Aids in information retrieval • Reduced ambiguity of terms in different languages so as to be more consistent • Redirects user to synonyms and relevant terms to aid retrieval • Provides guidelines for organization	• Terms can appear artificial at times • Gap between controlled and natural language can be very wide – people are more used to searching and more used to the natural language focus of Internet search engines, which may impede their ability or wish to use controlled vocabulary • If a term or subject is tangential to an area, a different term and placement within a scheme may impede retrieval • May not be sufficiently exhaustive • Become outdated quickly, especially as knowledge grows • User has to be familiar with scheme in order to use it effectively

Table 1. Strengths and Weaknesses of Folksonomy-based Controlled Vocabularies

3.2 Folksonomies + Hierarchies (Tagging Digital Resources and NOT Just Items)

Another way of implementing *integration of folksonomies and controlled vocabularies in the form of hierarchies* includes tagging of digital *resources* and not just items. This is an outcome of users' *improvised learning* (Boudreau and Robey 2005). For example, when del.icio.us started offering folksonomic services of bookmarking, it never intended to use its website for this reinvented purpose. Eventually, many users from del.icio.us began tagging URLs from one website to other websites. This is an innovative idea of organizing information where users bring themselves to the messy crossroads of information to get splashed with lots of other ideas and knowledge created by others (Weinberger 2007).

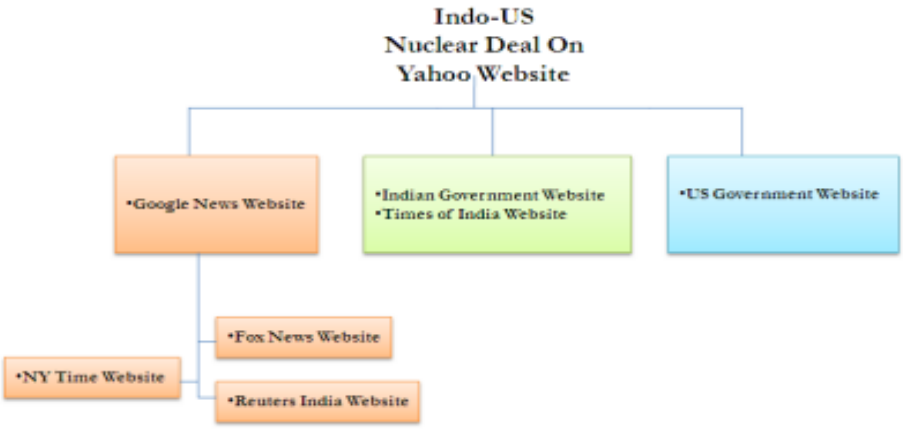

Figure 3. URLs[3] from one web-resource TAGGED with URLs from other web-resources

Figure 3 above illustrates this idea. Let us assume that a user wants to know more about the Indo-US Nuclear Deal and relevant repercussions around the world. So, a user visits a Yahoo website and finds a news item that discusses the Indo-US Nuclear Deal. While reading the news item, the user comes across various news items published on other websites that are, in turn, published by a number of news agencies and both the governments (India and the US). A smart user tags that specific webpage on Yahoo (which talks about Indo-US Nuclear Deal) with all other webpages that cover the Indo-US Nuclear deal.

This kind of tagging is an innovative way of integrating digital resources and not just digital items. Since webpages published on news agencies' websites and governments' websites contain meticulously framed information content, all those web-resources resemble to the concept of controlled vocabularies. Moreover, tagging of webpages (from various sources) by users represents *user-defined folksonomies*. The following table (see Table 2) synthesizes strengths and weaknesses of folksonomony-based hierarchies.

Strengths	Weaknesses
• Shows the relationship of a term to others within a scheme • Encompasses of broad and narrow terms • Conforms to a strict scheme • May be easier for machines to retrieve results, because of that strict control	• Limits where the information can be categorized, because it must fit within a hierarchy • Only shows relationships among a hierarchy, but not relationship between relevant terms outside of that hierarchy • Doesn't always portray real-world relationship

Table 2. Strengths and Weaknesses of Folksonomy-based Hierarchies

3.3 Folksonomies + Ontologies

Ontologies are defined as "…explicit specification of the conceptualization of a domain…(Gruber 1993)." They are often developed to represent the conceptual view of the existing knowledge domains. Folksonomies and ontologies are both lightweight and easy to access. Social network (folksonomies) and semantic networks (ontologies) are in fact flip sides of the same coin (Mika 2007). In Figure 4, the process of developing folksonomies derived from ontologies is a *feedback based* knowledge organization technique (Voss 2007). *Conceptual analysis in the figure* represents a process of conceptualizing chosen sub-field of knowledge. Usually, ontologies are products of conceptual analysis done by skilled ontology engineers. The outcome of conceptual analysis is highly dependent on: *Why* is ontology engineer asked to develop ontology for a particular body of knowledge? *Who* are the target users for final product in the form of ontology? And *what* sub-areas of chosen domain of knowledge are perceived relevant by users of the final ontology?

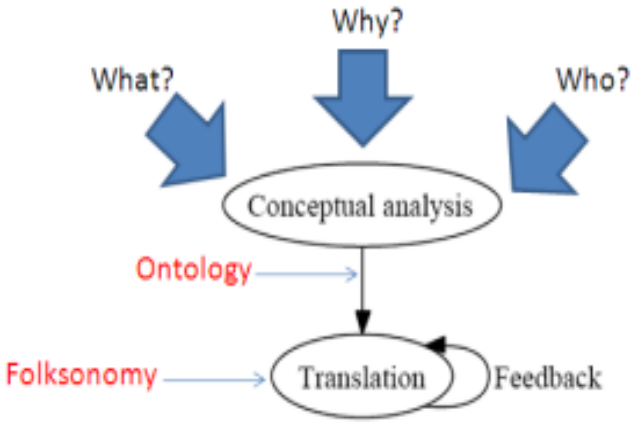

Figure 4. Subject Indexing: Integration of Ontologies with Folksonomies
Source: Adapted from Voss 2007

Once the ontology is ready to use, users can choose tags for developed ontology. *Translation* represents a process of selecting *tags*. There is a continuous feedback mechanism expected to take place among users who choose free keywords for pre-existing ontologies. Feedback reduces the drawbacks of uncontrolled indexing; this gives rise to a better quality indexing developed by untrained users (Voss 2007). The following table (see Table 3) synthesizes strengths and weaknesses of folksonomony-based ontologies.

3.4 Folksonomies + Bibliometric Maps

Bibliometric methods, including bibliometric maps, have been extensively used in the past for creating efficient user-centric knowledge organization systems (Broughton et al. 2005). Bibliometric maps can be combined with folksonomies to cater user needs in better ways. Usually, users access online bibliographic metadata developed using traditional techniques of information

organization and maintained at websites of libraries. In addition to subject indexing done by skilled catalogers, users can add another level of indexing in the form of tags. Users can create tags for carefully designed bibliographic metadata by skilled catalogers. This type of bibliographic metadata *uploaded* by libraries on the Web can be shared, organized and managed by users through tagging (BibSonomy 2008, Connotea 2008, CiteULike 2008, LibraryThing 2008). Moreover, a group of users can form clusters of various user-defined tags, which leads to managing information associated with catalogs, collaboratively (UPenn 2005). This technique in which third ordered metadata (tags) can be built for carefully developed second order metadata (bibliographic metadata) is known as *social reference management* or *collaborative cataloging* (Voss 2007).

Strengths	Weaknesses
• Allows the user to see the bigger picture • Also reliant on feedback so it eliminates some of the dangers of uncontrolled vocabulary • Terms express concepts within a domain and the relationships between concepts • Provides shared vocabulary • Explains a domain of knowledge • Allows multiple perspectives • Because of the different languages used (ontology languages, that is), it is an important facet in realizing the Semantic web – especially expressing the same meanings represented by different terms in different concepts	• Uses specific languages that need to be acquired • Many different languages instead of a single encoding language

Figure 5. A Tag Cloud on Amazon.com: An Example of Collaborative Cataloging
Source: Amazon.com

For example, Amazon, an online store selling books and other entertainment items, applies the same principle of *collaborative cataloging* in identifying most popular categories of merchandise, books and other items. The website allows its users to create a tag cloud (See Figure 5), where typically a cloud consists of just titles or labels for most popular categories of merchandise including best seller categories for books. No restrictions to avoid synonyms, homographs, lexical anomalies are imposed on users. This technique is useful for business as well as for future customers to trace an ongoing trend in online business. Users have the facility to surf and choose online articles through tagging (Voss 2006). The following table (see Table 4) synthesizes strengths and weaknesses of folksonomony-based bibliometric maps.

3.5 Folksonomies + Thesaurus-based System

Under this integration technique, folksonomies are embedded in the pre-existing thesaurus developed either by experts or by untrained users. One of the best examples of this technique is Wikipedia, the free encyclopedia. Wikipedia emerged as a unique categorization system where anybody could enter any relevant information in it. It hardly allows its users to be neutral but to communicate and negotiate. It has also been acclaimed as the first application of collaborative tagging with a thesaurus (Voss 2005).

Strengths	Weaknesses
• User-generated and user-friendly • Managed by users • Communities may develop their own tags in keeping with their vocabulary • Enables companies, users, etc. to keep track of trends • Evaluates and compares trends and user behavior • Highly collaborative	• Terms may be uncontrolled and redundant • Unknowledgeable users may generate incorrect terms • Not as much control over entries • May be difficult to achieve precision in retrieval • Most-used tags appear bigger, which may be descriptive of their popularity rather than their usefulness

Table 4. Strengths and Weaknesses of Folksonomy-based Bibliometric Maps

Varieties of categories in the form of topics for discussions on Wikipedia serve as tags for users to enter information. Most of the users tend to enter information for already created topics for discussion rather than creating new categories or re-categorizing existing structure. The continuously-entered new information serves as tags thereby offering an entry point to other types of digital resources *uploaded* on the Internet. Even less number of users are found to change or introduce new categories of topics under which discussions are published. This type of user behavior conforms to power – law distribution offered by Zipf's law (R´eka and Barab´asi 2002, Voss 2006). If a user enters any information under a topic, it means that the user agrees with the label or tag of that topic. The above set of user behavior proves tagging in Wikipedia to be *in between controlled vocabularies and folksonomies.* Wikipedia offers a unique facility to users to be consumers as well as producers of the same information content. Further, research proves that "…domain- specific thesauri can be enriched and created with Wikipedia's category and link structure… (Voss 2007)." The following table (see Table 5) synthesizes strengths and weaknesses of combination of folksonomony-based and thesaurus-based systems.

Strengths	Weaknesses
• Use pre-existing terms so there is a measure of control • Controlled vocabulary used to improve indexing, tagging, retrieval • Existing relationships between concepts made explicit • Terms are placed in contexts so users can see relationships • Constructed by specialists for control • Scope notes that help define meanings	• Narrower than controlled vocabularies • May be difficult for un-initiated users • Not user-centered or generated • Strictly hierarchical, which may not always express all relationships

Table 5. Strengths and Weaknesses of Folksonomies + Thesaurus-based Systems

3.6 Folksonomies + Semantic Web

It is predicted that user experience of tomorrow's information organization systems rests upon today's semantic web technologies (Morville 2005). Currently, the evolution of the Web is believed to be at Web 2.0 mark where accelerated user-collaborations have enormously increased the opportunities for creating and sharing more metadata by users (Mika 2007). For example, folksonomies offer a platform to users for sharing their digital data as well as digital metadata. Practicing folksonomies is just the first step towards receiving data as well as metadata from users in digital media.

The idea of Semantic Web was proposed at the end of the last millennium to enable computers to share any information among them without any human intervention. Semantic web is "…a mesh of information linked up in such a way as to be easily processable by machines, on the global scale" (Semantic Web Introduction 2008). This requires a common machine language understood by all the machines (computers) in this world. Data aspect of Semantic Web consists of structure, content, format; and metadata consists of representation languages for metadata interchange (Gangolly 2007). XML schema is used to implement structure and content of Semantic Web. Ontologies are precisely developed to build the common platform with common set of conceptualizations for all machines to communicate.

Semantic Web extends new technological opportunities in combining data and various services from various sources either automatically (machine learning) or by users. The Semantic Web is a solution proposed for executing information organization through machine-level learning. Logic-based languages such as Prolog (logic programming) hailed from artificial intelligence, a sub-domain of computer science, form the core of the Semantic Web (Mika 2007). The key reason for representing information in a logical form is because it efficiently helps to locate and eliminate imprecise information (Sergot et al. 1986). The original vision for the Semantic Web believed that users will enrich the pre-existing as well as future content on the web by learning core technologies of the Semantic Web. Currently, through folksonomies users share their metadata developed in natural language. These recent developments in collaborative annotations through Web 2.0 seem promising enough to gearing up users for this type of hybrid information organization technique.

Though it is a great challenge to acquire the skill, it is also a requirement for all web-users to grasp and then master the Semantic Web languages such as resource description framework (RDF, a metadata language) and web ontology languages (OWL, a data language). OWL is an ontology representation language representing data aspect of Semantic Web. RDF, Ontology Exchange Language (XOL), Ontology Markup Language (OML), RiboWeb, Simple HTML Ontology Extensions are some of the metadata languages developed for Semantic Web (Gangolly 2007). They are primarily used for building structured metadata. From the lessons and observations from Web 2.0, we can infer that users are willing to provide structured metadata nowadays. However, in the future, highly user-friendly interfaces for Semantic Web programming languages could overcome the requirement of grasping and then mastering RDF and OWL by average users (Mika 2007). It is expected that social adoption of the Semantic Web might take-off soon due to exponentially rising social networking under Web 2.0 (Kompatsiaris and Hobson 2008).

Spam detection is another incentive for integrating folksonomies with semantic web technologies. Folksonomies could help to efficiently classify emails in mailboxes using advanced data classification techniques. For instance, Capocci and Caldarelli's research (2007) found that "…Establishing a relationship between clustering and semantics may suggest tools and algorithms for technological tasks such as automatic categorization of resources, recommendation and spam detection techniques."

Structured metadata like ontologies created with techniques like RSS, OAI-PMH and RDF act as supplementary methods to user-centric indexing mechanisms like folksonomies (Voss 2007). The application of Concept-Actor-Instance model of Ontologies on del.icio.us website reveals that ontologies can not only emerge from folksonomies but can also be enriched by folksonomies (Mika 2007). Especially, slowly evolving linguistic ontologies (e.g. WordNet) have a greater potential to receive nourishment from folksonomies (Fellbaum 1998). The following table (see Table 6) synthesizes strengths and weaknesses of folksonomony-based semantic web techniques.

Strengths	Weaknesses
• Users are creators and consumers • Information can be shared and linked without human intervention • Enrich content on the web • Spam detection enhanced • Improve usability and function of search engines • Link data from different sources around the world – more holistic • Uses natural language	• Uses different web languages, which need to be learned and mastered for encoding • Practical and economic feasibility • Humans may mislead machine-readable documents intentionally or unintentionally • Censorship and privacy – because of the public availability of information • Governments and organizations may be able to retrieval more and more intimate information of citizens and clients • More time-consuming to create and publish content because you need two formats: 1) human-viewable; 2) machine viewable

Table 6. Strengths and Weaknesses of Folksonomy-based Semantic web

Research studies hypothesize that without developing folksonomies, community-based ontologies can be extracted from the web (Mika 2007). In other words, lightweight ontologies can replace folksonomies from the web. Using basic co-occurrence analysis method of text mining, concept-instance ontology can be built, where a webpage is tagged if a desired concept

resides on that particular webpage. There are two other types of ontologies proposed: namely, Actor-Concept Ontology and Community-based Ontologies which could extract the social network of people and associate them with concepts (or topics) contributed by them on the Web (Mika 2007).

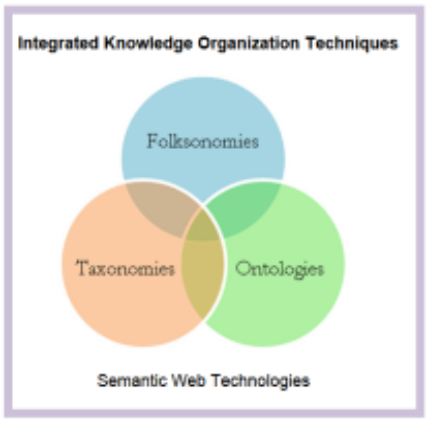

Figure 6. Future beyond Imagination[4]

Currently, we know possibilities of combinations between folksonomies and ontologies, and folksonomies and taxonomy (See Figure 6) but: Do we need to build information organization system that would inherit the best of all the three worlds - folksonomies, taxonomies and ontologies? How can we achieve it? We are promoting machine learning with an expectation of building a *web of data* where machines would communicate with each other without human intervention. Thus, the world is swiftly moving towards that deep intersection of all the three circles. After visiting various combinations of information organization techniques, the next section describes the significance and ways to understand users' information needs, and involve users into the actual process of designing and developing user-centric information organization systems.

4. Conclusion

Folksonomy, a user-centric technique of organization digital information has the potential to radically change the ways in which users make sense of their lives. Hence, future efforts to design and develop user-centric information organization systems need to consider blending folksonomy with established techniques of information organization. The strengths and weaknesses of various combinations of folksonomy-based information organization techniques reveal practical applications and limitations of the proposed user-centric information organization systems. This paper is expected to inform researchers and practitioners looking forward to advance folkosnomy-based information organization systems.

Reference

[1] Abbas, J.,Graham, J. (2006). "So, let's talk about tagging, user-defined/supplied descriptors ... A research and curricular agenda, *Journal of the American Association for Information Science*, 52(6) 487-498.
[2] Bates, M. (2006). Tag, You're it!, *Online*, 30 (1) 64.
[3] BibSonomy (2008) Tagging on BibSonomy, available at: http://bibsonomy.org
[4] Boudreau, M. C., Robey, D. (2005). Enacting integrated information technology: A human agency perspective," *Organization Science*, 16 (1) 3-18.
[5] Broughton V., Hansson J., Hjørland B., López-Huertas M. J. (2005). Knowledge Organization, *In*: Proceedings of European Curriculum Reflections on Library and Information Science Education, Copenhagen, Denmark.
[6] Burnett, G., Besant, M., Chatman, E. A. (2001). Small Worlds: Normative Behavior in Virtual Communities and Feminist Bookselling, *Journal of Association for Information Systems*, 52 (7) 536–547.
[7] Burnkrant, R. E. (1976), A motivational model of information-processing intensity, *Journal of Consumer Research,* 3, 21 30.
[8] Cancho F.R., Sol´e R. V. (2001). The small world of human language, Technical Report, Santa Fe Working Paper, 01-03-016.
[9] Capocci, A., Caldarelli, G. (2007).Folksonomies and clustering in the collaborative system CiteULike, *Journal of Physics A: Mathematical and Theoretical*, 41, *Digital Libraries*, http://www.iop.org/EJ/abstract/1751-8121/41/22/224016
[10] Cattuto C., Benz D., Hotho A., Stumme G. (2008). Semantic Analysis of Tag Similarity Measures in Collaborative Tagging Systems Digital Libraries, *In*: Proceedings of 3rd Workshop on Ontology Learning and Population, Patras, Greece.

[11] Chatman, E. (1991). Life in a small world: Application of gratification theory to information-seeking behavior, *Journal of American Society for Information Science*, 42(6) 438–449.
[12] CiteULike (2008) CiteULike, available at: http://citeulike.org
[13] Connotea (2008) Connotea, available at: http://connotea.org
[14] Damianos D. C. L., Griffith J. (2006). Onomi: Social bookmarking on a corporate intranet, *In*: Proceedings of WWW 2006 Collaborative Web Tagging Workshop, Edinburgh, Scotland
[15] Fellbaum, C. (1998). WordNet: An Electronic Lexical Database, Boston, MA: MIT Press
[16] Friedman, T. (2005). World is Flat: The Globalized World in the Twenty-First Century, New Delhi, India: Penguin Books India Pvt. Ltd.
[17] Farrell S., Lau T. (2006). Fringe contacts: People, *In*: Proceedings of the WWW 2006 Collaborative Web, Edinburgh, Scotland.
[18] Gangolly, J. (2007). Lecture Notes on Markup Languages and Semantic Web, Class offered through INF 722 Lecture Series, College of Computing and Information, State University of New York at Albany, NY.
[19] Gruber, T. (1993). Towards Principles for the Design of Ontologies Used for Knowledge Sharing, *In:* Formal Ontology in Conceptual Analysis and Knowledge Representation (Guarino, N., and Poli, R., eds.), Kluwer Academic Publishers, Deventer, Netherlands.
[20] Hjorland B. (2002). Domain Analysis in Information Science: Eleven Approaches –Traditional as well as Innovative, *Journal of Documentation*, 58. 422-462
[21] Hotho A., J¨aschke R., Schmitz C., Stumme G. (2006). Information retrieval in folksonomies: Search and ranking, *The Semantic Web: Research and Applications*, Springer, Berlin, Germany, p. 411– 426
[22] Ideant. (2005). Tag Literacy, http://ideant.typepad.com/ideant/2005/04/tag_literacy.html
[23] John, A., Seligmann, D. (2006). Collaborative tagging and expertise in the enterprise, *In*: Proceedings of the WWW 2006 Collaborative Web Tagging Workshop, Edinburgh, Scotland.
[24] Julien, H., Michels, D. (2004). Intra-individual information behaviour in daily life, *Information Processing and Management,* 40, p. 547-562.
[25] Kao H.Y., Lin, S.H., Ho, J.M., Chen M.S. (2002). Entropy-Based Link Analysis for Mining Web Informative Structures, *In*: Proceedings of 11th ACM International Conference on Information and Knowledge Management, McLean, Virginia.
[26] Koehn, P., Knight, K. (2003). Empirical methods for compound splitting, *In*: Paper presented at the Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics, April 12-17, Budapest, Hungary.
[27] Kompatsiaris, Y., Hobson, P. (2008). *Semantic Multimedia and Ontologies*, London, UK: Springer–Verlag Limited.
[28] Kumar, K., Dissel, H. G. V., Bielli, P. (1998). The merchants of Prato-revisited: toward a third rationality of information systems, *MIS Quarterly*, 22 (2) 199-226.
[29] LibraryThing. (2008) Library Thing, available at: http://librarything.com
[30] Macgregor, G., MacCulloch, E. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool, *Library Review*, 55 (5) 291-300.
[31] Mathes A. (2006). Folksonomies - Cooperative Classification and Communication Through Shared Metadata, AdamMathes.com, http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html
[32] Matusiak, K. K. (2006). Features Towards user-centered indexing in digital image collections, *International Digital Library Perspectives*, 22 (4) 283-298.
[33] Merholz, P. (2004). Ethnoclassification and vernacular vocabularies, Peterme.com, www.peterme.com/archives/000387.html
[34] Mika, P. (2007). Social networks and the Semantic Web, New York, NY: Springer Publications
[35] Millen D., Medelyan O., Witten, I. H. (2006). Mining Domain-Specific Thesauri from Wikipedia: A case study, *In*: Proceedings of the Conference on Web Intelligence, Hong Kong, China.
[36] Morville, P. (2005). *Ambient Findability: What We Find Changes Who We Become*, Sebastopol, CA: O'Reilly
[37] O'Reilly T. (2005). What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software, Oreillynet.com, http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html
[38] Plangprasopchok, A., Lerman, K. (2008). Constructing Folksonomies from User-specified Relations on Flickr, Track: Social Networks and Web 2.0 / Session: Photos and Web 2.0, *In*: International World Wide Web Conference Committee (IW3C2), WWW 2009, April 20–24, 2009, Madrid, Spain.
[39] Reagans, R., McEvily, B. (2003).Network Structure and Knowledge Transfer: The Effects of Cohesion and Range, *Administrative Science Quarterly*, 48, 240-267.
[40] R´eka, A., Barab´asi, A. L. (2002). Statistical Mechanics in Complex Networks, *Review of Modern Physics*, 74, 47-97. [41] Rosenfeld, L. (2005). Folksonomies? How about Metadata Ecologies? http://louisrosenfeld.com/home/bloug_archive 000330.html

[42] Semantic Web Introduction. (2008) http://infomesh.net/2001/swintro/
[43] Sergot, M. J., Sadri, F., Kowalski, R. A., Kriwaczek, F., P. Hammond, and Cory, H. T. (1986). The British Nationality Act as a Logic Program, *Communications of the ACM*, 29, 370-386.
[44] Shen, K., Wu, L. (2005). Folksonomy as a Complex Network," Working Paper, Department of Computer Science, Fudan University, China
[45] Shirky, C. (2005). Folksonomies are a forced move: A response to Liz, *Corante*, Online Edition.
[46] Solomon, P. (2002). Discovering information in context, *Annual Review of Information Science and Technology (ARIST)*, 36, p. 229-264.
[47] Sterling, B. (2005). Order out of chaos, *Wired*, 13 (4) Online Edition.
[48] Trant, J. (2006) "Social classification and folksonomy in art museums: Early data from the steve.museum tagger prototype, *In*: Proceedings of the 17th SIG Classification Research Workshop, Austin, Texas
[49] Trant, J., Wyman, B. (2006). Investigating social tagging and folksonomy in art museums with steve.museum, *In*: Proceedings of the WWW 2006 Collaborative Web Tagging Workshop, http://www.rawsugar.com/www2006/4.pdf
[50] Udell, J. (2004). Collaborative knowledge gardening, *InfoWorld*, Online Edition
[51] UPenn. (2005). Library Tags, http://tags.library.upenn.edu
[52]Vander Wal (2004). You Down with Folksonomy?, Vanderwal.net, http://www.vanderwal.net/random/entrysel.php?blog=1529
[53] Voss, J. (2005). Measuring Wikipedia, *In*: Paper presented at the Proceedings the International Society for Scientometrics, http://eprints.rclis.org/archive/00003610
[54] Voss, J. (2006). Collaborative thesaurus tagging the Wikipedia way, *Wikimetrics Research Papers*, 1(1), http://arxiv.org abs/cs.IR/0604036
[55] Voss, J. (2007). Tagging, Folksonomy & Co –Renaissance of Manual Indexing?, *In*: 10th International Symposium for Information Science, Cologne, Germany, http://arxiv.org/abs/cs/0701072
[56] Weinberger, D. (2007), Everything is Miscellaneous: The power of the new digital disorder, New York, NY: Times Books
Wilson, T.D. (1997). Information behaviour: an interdisciplinary perspective, *Information Processing and Management*, 33(4) 551-572.
[57] W3C (2001). World Wide Web Consortium Semantic Web Activity, W3C, http://www.w3c.org
[58] Zhang, P., Benbasat, I., Carey, J., Davis, F., Galletta, D., Strong, D. (2002). Human-Computer Interaction Research in the MIS Discipline, *Communications of the AIS*, 9, 334-355.