Social Bookmarking or Tagging

Johann Stan, Pierre Maret

Title: Social Bookmarking or Tagging

Name: Johann Stan, Pierre Maret

Affil./Addr.: Université de Lyon, UJM-Saint-Étienne, CNRS,

Laboratoire Hubert Curien UMR 5516

F-42023 Saint- Étienne, France

E-mail: johann.stan.phd@gmail.com, pierre.maret@univ-st-etienne.fr

Social Bookmarking or Tagging

Synonyms

Social Tags, Bookmarks

Glossary

Annotation: See Tag.

Folksonomy: Whole set of tags that constitutes an unstructured collaborative knowledge classification scheme in a social tagging system.

Resource: In the context of this work, a multimedia content (e.g., text file, photos, videos, web page) available on the Internet. A resource is generally identified by an URI (Unique Resource Identifier) which enables its access using the Rest protocol.

Social Bookmarking System: Web-based systems allowing users to describe resources with tags.

Social Bookmark: Tag in the form of a link to a resource (e.g. web page) that is intentionally stored, and possibly shared, by an identified individual on a social bookmarking system, on which individuals can attach tags.

Tag: A descriptive keyword entered by a human individual with the objective to describe a resource (e.g. a photo, a web page). It is also called an annotation or user generated content.

Definition

Social Bookmarking Systems (SBS). Web-based systems allowing users to describe resources with annotations, also called tags. The fundamental unit of information in a social bookmarking system consists of three elements in a triplet, represented as (user,resource, tag) [16]. This triplet is also called a tag application (instance of a user applying a tag to a resource; this is also referred to as a tag post) [35]. The combination of elements in a tag application is unique. For example, if a user (also known as tagger) tags a paper twice with the same tag, it would only count as one tag application. Resources can mean different things for different social bookmarking systems. In the case of del.icio.us, a resource is a web site, and in the case of CiteULike, it is an academic paper.

Introduction

Social Bookmarking (or tagging) is a means for describing resources. The social bookmarks are tags attached to a resource, with the main objective to describe said resource. They describe the context or the meaning of such artifacts. Social bookmarks can be of multiple forms, depending on the semantic structure they rely on.

The manipulation of web resources involves tasks such as description, retrieval, reuse, presentation and search. All these tasks need a layer of prior knowledge, which is represented by the social bookmarks, which can be composed of different types of annotations.

Key points

Tag (or annotations) may be either structured, semi-structured or unstructured. Tags tend to be short. Hashtags are 1-word annotations. Tagging became popular on social networks.

Historical Background

The emergence of the so-called Web 2.0 (from 2004) gave rise to User Generated Contents (UGC), and therefore to web-repositories of UGCs. However, Wikipedia [2] was already launched in 2001 and it was one of the first public crowd-sourced web site. This free encyclopedia has been allowing anyone to edit the content of any article. Whereas this openness has implied many disputes on pages related to controversial subjects (e.g. facts about presidential candidates just before election, about historical events, companies, etc.), it has grown to become a major and useful reference, covering many languages. This encyclopedia has been translated to a semantic database called DBpedia [12] since 2007, enabling its user-generated content to be machine-readable, so that computer programs (and mashups) can leverage knowledge facts by formulating precise queries.

Even though wikipedia has been opened to any voluntary contributions, contributors are still few, compared to the number of readers. Participating in social bookmarking sites, like Delicious [4], have become more popular, as the contribution process was quicker, simpler and more personal. After creating a (free) account on the site, users can immediately bookmark web pages that they want to keep, because they enjoy them, they want to be able to easily find them later, and they (often) want to share them with other people. In order to make bookmarked web pages more easy to find later, users are invited to annotate them with 'tags', unconstrained words (in any language,

without even spell checking) that subjectively reflect the apparent nature, function, category and context of those web pages [19] . Web pages bookmarked (and tagged) by several people are thus described by a 'tag cloud', a displayed set of tags. The size of a tag depends on the number of people who used this tag to describe this page.

As any URL-located resource can be bookmarked on social bookmarking sites, these descriptions can apply on various types of entities represented by those resources. For example, tags given to a page that presents a car, are most probably associated to the car, than to the page/site itself. Now that web pages exist for almost anything on earth (e.g. people, objects, places, events, etc.), social bookmarking is a promising paradigm for gathering crowd-sourced descriptions and classifications of virtual and real entities.

More specifc repositories also exist to represent and describe real word entities, and discover their involvement with people's activities. Concerning music, Musicbrainz [1] can identify the name and interpreter of a song from a sampled audio (e.g. recorded with a microphone), and tags given by people to songs and artists are gathered on web sites like Last.fm [3], which also maintains a history of the last songs that users listened to. Image sharing web sites like Flickr [5] can be considered as social bookmarking applied to photographs, as it is possible to tag one's own and other people's photographs, including the time and geographical location where the picture was taken.

Additionally, real-world places are described, reviewed by people and geographically located on various web sites (and their mobile applications) such as Yelp and Qype [6], [8].

Rattenbury et al. [33] have proven that names of places and events can also emerge by analyzing the frequency and temporal distribution of tags associated to geolocated pictures. Most web sites cited above expose public feeds that one can subscribe for being aware of last updates, and/or APIs that allow computer programs to query

information, given specifc criteria (e.g. information about a place, a topic, at a given time range).

Thousands of other APIs are referenced on sites like Programmable Web. Also note that tags are not directly available on all the web sites cited above, but keywords can be identified from the user-generated content they feature. It is also possible that pages from those sites are tagged on Delicious.

Types of annotations

Annotations may be either structured, semi-structured or unstructured:

1. *Structured Annotations.* In this case, the terms employed in the annotation are regulated by a common domain vocabulary that must be used by the members of the system. These types of annotations are currently not used in the majority of social platforms because a domain vocabulary containing the necessary terms for the annotations is needed. Although such an approach has many advantages (e.g. absence of synonyms, absence of differences in pronunciation), this is not the natural way to describe resources in web 2.0 platforms, as the domain is not well-defined and, therefore, it is very difficult to build such vocabularies and to establish a consensus for each term used. At the same time, the use of semantic annotations would be cumbersome for people, as it is time-consuming and requires additional cognitive effort to select concepts from existing domain ontologies. In addition, semantic annotations work well in systems where the domain is well-defined (e.g. a system for sharing knowledge about human genes [38]), but in social bookmarking systems this is not the case, as the shared content is generally very heterogeneous, as people can discuss without limits (i.e. covers multiple domains with no regularities and relations).

2. *Semi-Structured Annotations.* In contrast, semi-structured annotations, such as social tags, are widely used or photo tagging and bookmarking (e.g. the annotation of a web page). These annotations are generally freely selected keywords without a vocabulary in the background. However, we consider them to be semi-structured, as they represent an intermediate approach between semantic annotations (i.e. annotations that are based on concepts from domain ontologies) and free-text annotations. Besides, such collections of tags converge to a structured data organization, called a folksonomy [21]. This consists of a set of users, a set of free-form keywords (called tags), a set of resources, and connections between them. As folksonomies are large-scale bodies of lightweight annotations provided by humans, they are becoming more and more interesting for research communities which focus on extracting machine-processable semantic structures from them. These underlying data clouds of collaborative tagging systems enable Internet users to annotate or search for resources using custom labels instead of being restricted by pre-defined navigational or conceptual hierarchies (e.g. ontologies).
3. *Unstructured Annotations.* Finally, a more recent form of annotations is represented by free text annotations, also called social awareness streams, composed of status updates or microposts[30]. This can be found in the majority of social networks and microblogging systems and primarily consists of free texts in the form of short messages describing a resource, a finding, an impression, a feeling, a recent activity, mood or future plan. The limitations of this practice from the viewpoint of information retrieval and knowledge management are similar to that of social tagging, as users have complete freedom in the formulation of these messages. It is important to mention that in social awareness streams, the produced content often contains the described resource itself, in the form of an integrated hyperlink. A common practice is either to express an opinion about the resource (e.g. web page) or to provide

its short summary for the community. Since internet took over usenet as the main computer-based means of communication, it has gone through several stages: read-only web, with large pieces of information close to magazine article size; read-write web, or web 2.0, with forums mimicking usenet, exchanging pieces of information up to half-a-page in size; blogging, close to the web page model but with a shift in authorship towards the general public; and micro-blogging, based on very short messages (e.g. 140 characters on Twitter). This shift from large, authoritative information to very short and amateur information is contemporary with the mobility evolution, with the more user-friendly web-enabled devices (e.g. the iPhone) emphasizing a particular factor: the context in which information is written. This has blurred the distinction between information and messaging, as all information on Twitter is in fact a message to followers, and all messages may be shared, thus creating information. Events and documentation on the contrary are becoming more distinct: in the traditional newspaper information model, documentation is delivered with events in a single article; in the Twitter-driven model, events are tweets, and the user is meant to seek information in more reliable and static sources, such as Wikipedia. An example of such as shift is the growing use of Twitter in the scientific community, contrasting strongly with the process of peer-reviewed publication.

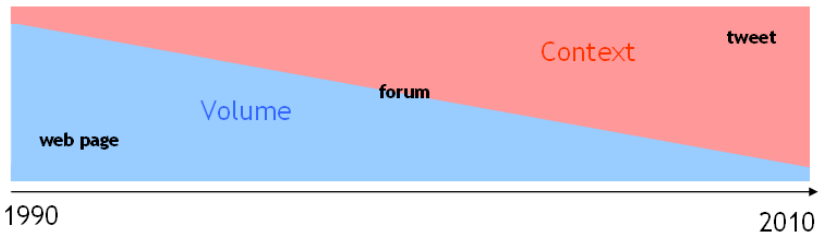

Fig. 1. Evolution of content production on the Internet: from structured documents to microposts

The decreasing size of annotations

An interesting issue about free-form posts in social platforms is their short size, which emerged as a simple, convenient way to communicate about activities or share findings. The size limitation of such posts, defined by the majority of such platforms is mostly due to the fact that users can in this way follow hundreds of friends in real-time, without an important time investment. Also, this light-weight form of communication enables users to easily broadcast opinions, activities and status [23] [31].

Common practices emerged to reduce the lengh of messages and to help users to rapidly identify the messages relevant for them. Thus, hashtags are used to identify posts relevant to a specific event or a specific topic. Also, common ways are used to synthesize information: include the source web page or reduce the amount of stop words in order to gain place for the informative terms (keywords and named entities). These practices largely depend also on the targeted user community, which can vary from a small family to the world at large.

The same applies to the composition of such posts, where common practices emerged as new means to better identify posts relevant to a specific event, also called "hashtags", or common ways to synthesize an information, such as including the source web page or reducing the amount of stop words in order to gain place for the informative terms (keywords and named entities). These practices largely depend also on the targeted user community, which can vary from a small family to the world at large. Microposts are often called "social signals" [28], and users of such systems "social sensors" [34], as they can be useful to detect important events in a given location, such as an earthquake.

Knowledge Management in Social Bookmarking Systems

A category of annotations in Social Platforms are semi-structured, also called social tags. Social bookmarking systems have become extremely popular in recent years. Their underlying data structures, known as folksonomies [27], consists of user - tag - resource triples.

Folksonomies contain peoples' structural knowledge about documents. A person's structural knowledge has been defined as the knowledge of how concepts in a domain are interrelated from the individual's point of view. According to [27], an important aspect of a folksonomy is that it is comprised of terms in a flat namespace: that is, there is no hierarchy, and no directly specified parent-child or sibling relationships between these terms. There are, however, automatically generated "related" tags, which cluster tags based on common URLs. This is unlike formal taxonomies and classification schemes where there are multiple kind of explicit relationships between terms. These relationships include functions like broader, narrower, as well as related terms. These folksonomies are simply the set of terms that a group of users tagged content with, they are not a predetermined set of classification terms or labels.

Folksonomies claim to have many advantages over controlled vocabularies or formal taxonomies. Tagging has lower costs because there is no complicated, hierarchically organized vocabulary to learn and adapt to its own one. Users simply create and apply tags. According to Wu et al., "Folksonomies are inherently open-ended and therefore respond quickly to changes and innovations in the way users categorize content" [7]. Collaborative tagging is regarded as democratic metadata generation where metadata is generated by both the creators and consumers of the content. Folksonomies can be divided into broad folksonomies, which allow different users to assign the same tag to

the same resource, and narrow folksonomies, in which the same tag can be assigned to a resource only once.

The question of why folksonomies are successful has been the subject of several studies in the literature. An important argument for this is the fact that the feedback loop is tight [27], i.e. once the user assigns a tag to an item, the cluster of items with identical or similar tags can be immediately retrieved. This can help the user decide whether to keep the tag or change it to a similar or different one. The scope of such a cluster can be expanded by showing all items from all users in the system which are tagged with the same tag. By viewing the result set, the user can decide how to better adapt the tag to the group norm or to have better visibility in the community for the tagged resource. The issue of how to influence the group norm was also studied by Udell. This tight feedback loop leads to a form of asymmetrical communication between users through metadata. The users of a system are negotiating the meaning of the terms in the folksonomy, whether purposefully or not, through their individual choices of tags to describe documents for themselves.

A folksonomy eases collaboration. Groups of users do not have to agree on a hierarchy of tags or detailed taxonomy, they only need to agree, in a general sense, on the "meaning" of a tag enough to label similar material with terms for there to be cooperation and shared value. Although this may require a change in vocabulary for some users, it is never forced, and as Udell discussed, the tight feedback loop provides incentives for this cooperation.

The main problems of social tagging systems include ambiguity, lack of synonyms and discrepancies in granularity [20]. An ambiguous word, e.g. apple, may refer to the fruit or the computer company, and this in practice can make the user retrieve undesired results for a given query. Synonyms like lorry and truck, or the lack of consistency among users in choosing tags for similar resources, e.g., *nyc* and *new york*

city, makes it impossible for the user to retrieve all the desired resources unless he/she knows all the possible variants of the tags that may have been used. Different levels of granularity in the tags may also be a problem: documents tagged 'java' may be too specific for some users, but documents tagged 'programming' may be too general for others.

Several attempts have been made to uncover the structure of this kind of data organization. Basic formal models of folksonomies include that of Mika [29] and Hotho et al. [22]. Mika proposes a model based on *tripartite hypergraphs*, while Hotho et al. on *triadic context* (term used in formal concept analysis). We present in the following the formal model of Mika, one of the most cited models in the literature for the representation of these structures.

As said before, a folksonomy is an association of users, annotations and resources. The corresponding three disjoint set of vertices are considered by Mika in the formal model: the set of actors (users) -A-, the set of concepts (tags) -C- and the set of resources -O- (e.g. photos, videos or web resources, like bookmarks, websites etc). Since in a social tagging system, users tag objects with concepts, ternary relations are created between the user, the concept and the object.

This resulting tripartite hypergraph can be transformed into several bipartite graphs, each having a very specific meaning, like AC - the graph that associates actors and concepts, CO - the graph that associates concepts and objects and AO, the graph that associates actors and resources.

Abel [9] investigates the benefits of additional semantics in folksonomy systems. Additional context can be provided to the tagging activity with an extension of the tripartite model, i.e. an association of the user, the tag and the tagged resource, that describes more precisely the particular tagging activity. For example, time stamp helps to categorize tags in a temporal manner, the mood the user had when tagging the

resource would allow to qualify opinions expressed in a tag. Other information, like background knowledge about the user, would allow to have information about the reliability of the tagger. The GroupMe! folksonomy system is proposed, which is a new kind of resource sharing system for multimedia web resources. A first extension of previous models is the introduction of the term group, which is a finite set of related resources. The folksonomy model in GroupMe! can be thus formalized in the following manner (We note with F the folksonomy model): $F = (U, T, IR, G, Y)$, where U, T, R, G are finite sets that contain instances of users, tags, resources and groups. $IR = R \cup G$ is the union set of resources and the set of groups.

Wu et al. [36] identify the key challenges in collaborative tagging systems. The three identified challenges are the following: (i) the identification of communities, i.e. groups of users with similar interests, (ii) preventing information overload by filtering out high quality documents and users (e.g. experts in a domain) and (iii) how to create scalable, navigable structures from folksonomies. Folksonomies are criticized to have flaws that formal classification systems are designed to eliminate, including polysemy, words having multiple related meanings, and synonymy, multiple words having the same or similar meanings.

Information Retrieval from Folksonomies: Social Information Retrieval

In the previous section we have seen the general definition and structure of folksonomies, the data organization in social tagging systems. In this section we go further and review existing techniques of information retrieval in folksonomies.

The biggest challenge in folksonomies is information retrieval, i.e. the question of how to efficiently rank items (e.g. tags, resources, users) for a given user query. In traditional Internet applications the search and navigation process serves two vital

functions: retrieval and discovery. Retrieval incorporates the notion of navigating to a particular resource or a resource containing particular content. Discovery incorporates the notion of finding resources or content interesting but therefore unknown to the user. The success of collaborative tagging is due in part to its ability to facilitate both these functions within a single user-centric environment. Reclaiming previously annotated resources is both simple and intuitive, as most collaborative tagging applications often present the user's tag in the interface. Selecting a tag displays all resources annotated by the user with that tag. Users searching for particular resources they have yet to annotate may select a relevant tag and browse resources annotated by other users. However, the discovery process can be much more complex. A user may browse the folksonomy, navigating through tags, resources, or even other users. Furthermore, the user may select one of the results of a query (i.e. tag, resource, or user) as the next query itself. This ability to navigate through the folksonomy is one reason for the popularity of collaborative tagging.

In order to provide efficient retrieval mechanisms, a formal model of folksonomies is required. There are several models in the literature, e.g. that of Mika [29] and Hotho et al. [22]. Mika proposes a model based on *tripartite hypergraphs*, while Hotho et al. on *triadic context* (term used in formal concept analysis).

Hotho et al. adapt the well-known PageRank algorithm in order to apply it on folksonomies, called *FolkRank*. The impossibility of applying *PageRank* has its origins in the fact that a folksonomy is different from the web graph (undirected triadic hyperedges instead of directed binary edges). By modifying the weights for a given tag, FolkRank can compute a ranked list of relevant tags.

The original formulation of PageRank [13] reflects the idea that a page is important if there are many pages linking to it, and if those pages are important themselves (recursive aspect of importance). The distribution of weights can thus be described as

the fixed point of a weight passing scheme on the web graph. This idea was extended in a similar fashion to bipartite subgraphs of the web in HITS [25] and to n-ary directed graphs [37]). The same underlying principle is employed for the ranking scheme in folksonomies. The basic notion is that a resource which is tagged with important tags by important users becomes important itself. The same holds, symmetrically, for tags and users. Such a ranking schema can help the emergence of a common vocabulary in collaborative tagging systems, by recommending to the user tags that have a bigger visibility in the community and that is also semantically close to the user-defined tag.

Abel et al. [10] perform an in-depth analysis of ranking algorithms specially designed for folksonomies: FolkRank, SocialSimRank [11], and SocialPageRank and adapts them to the GroupMe! social bookmarking system, where an additional dimension is added to folksonomies, i.e. groups of resources.

Gemmel et al. [18] propose a method to personalize a user's experience within a folksonomy using unsupervised clustering of social tags as intermediaries between a query and and a set of items. Terms in the query are weighted based upon their affinities to particular clusters to help disambiguate queries.

Bao et al. [11] propose different algorithms, such as SocialSimRank and SocialPageRank to optimize web search using social annotations. The undelying hypothesis of the proposed algorithms are the following: (i) social annotations about web pages are good summarizations of the given web page and can be used for efficient computation of similarity between a search query and a web page and (ii) the amount of annotations assigned to a web page is a good indication of its popularity.

Vocabulary Construction and Emergence of Semantics

In this section we present different approaches for extracting and constructing a hierarchical structure of tags in collaborative tagging systems. Several papers proposed

different approaches to construct conceptual hierarchies from tags collected from social Web sites. Mika [29] uses a graph-based approach to construct a network of related tags, projected from either a user-tag or object-tag association graphs. Although there is no evaluation of the induced broader/narrower relations, the work provides a good suggestion to infer them by using betweenness centrality and set theory. Other works apply clustering techniques to keywords expressed in tags, and use their co-occurrence statistics to produce conceptual hierarchies [14] [39].

Brooks et al. [14] argue that hierarchical structures which seems to match that created by humans can in fact be inferred from existing tags and articles in collaborative tagging systems. This may imply that folksonomies and traditional structured representations are not so opposed after all, rather, tags are a first step in helping an author or reader to annotate her information. Automated techniques can then be applied to better categorize specific articles and relate them more effectively to other articles. The method used is agglomerative clustering and consists of the following steps: the comparison of each tag cluster to every other tag cluster, using the pairwise cosine similarity metric. Each article in cluster one is compared to each article in cluster two and the average of all measurements is computed. The two closest-similarity clusters from the list of tag clusters is removed and replaced with a new abstract tag cluster, which contains all of the articles in each original cluster. This cluster is annotated with an abstract tag, which is the conjunction of the tags for each cluster.

This procedure is followed until there is a single global cluster that contains all of the articles. By recording the order in which clusters are grouped into progressively more abstract clusters, a tree that shows the similarity of tags can be constructed. Plangprasopchok et al. [32] proposes a different approach for constructing folksonomies from user-specified relations on Flickr by statistically aggregating tags from different collections. This approach uses the shallow hierarchies created through the collection-

set relations on Flickr. Authors argue that partial hierarchies are a good source information for generating folksonomies and propose a simple statistical approach to resolve hierarchical relation conflicts in the aggregation process.

Another approach for the extraction of hierarchical semantics from social annotations is proposed by Zhou et al. [39]. A probabilistic unsupervised method is proposed, called Deterministic Annealing. This method performs a top-down approach on the flat tag space, beginning with the root node containing all annotations and splitting it to obtain clusters with narrower semantics.

[17] performs an analysis on a large-scale snapshot of the popular social bookmarking system Delicious. To provide a semantic grounding of the folksonomy-based measures, tags of delicious are mapped to synsets of WordNet [26] and use the semantic relations of WordNet to infer corresponding semantic relations in the folksonomy. In WordNet, the similarity is measured by using both the taxonomic path length and a similarity measure by Jiang and Conrath [24] that has been validated through user studies and applications [15]. The use of taxonomic path lengths, in particular, allows to inspect the edge composition of paths leading from one tag to the corresponding related tags, and such a characterization proves to be especially insightful. Co-occurence is a measure that extracts from the folksonomy a graph for tags, where edges are weighted with the number of times they co-occur (i.e. tags on the same resource).

The results can be taken as indicators that the choice of an appropriate relatedness measure is able to yield valuable input for learning semantic term relationships from folksonomies, i.e. (i) synonym discovery, (ii) concept hierarchy extraction and (iii) the discovery of multi-word lexemes. The cosine similarity is clearly the measure to choose when one would like to discover synonyms. Cosine similarity delivers not only spelling variants but also terms that belong to the same WordNet synset. Both FolkRank and co-occurrence relatedness yields more general tags. This could be a

proof that these measures provide valuable input for algorithms to extract taxonomic relationships between tags.

Key applications

The main application domain is the area of social networking systems. Howeer, this techniques related to bookmarkig has been already used in information retrieval and classification systems. More generally, it is used in entreprise knowledge management.

Future directions

The social bookmarking activity became very popular on the web from the second part of the 2000 decade with the rise of Web 2.0 and social networks. On the user side, the free-text bookmaking strategy may not change in the next years, and freedom of users will even increase. The main novelties will certainly arise from the backoffice processing of the annotations. One can expect higher precision and an increasing dynamicity in the management of vocabularies, especially thanks to large scale tag collection process, higher processing power, learning algorithmsn and the better consideration of the semantic.

Cross references

Folksonomies, 110

Tag Clouds, 126

Sentiment Quantification on User-Generated Content, 110170

Semantic Social Networks Analysis, 381

References

1. (2001) Musicbrainz - the open music encyclopedia. URL http://musicbrainz.org
2. (2001) Wikipedia knowledge base. URL http://wikipedia.org
3. (2002) Last.fm - music website. URL http://last.fm
4. (2003) Delicious social bookmarking system. URL http://delicious.com
5. (2004) Flickr - image and video hosting website. URL http://flickr.com
6. (2004) Yelp - local directory service with social networking and user reviews. URL http://yelp.com
7. (2006) Harvesting social knowledge from folksonomies, vol Novel systems and models
8. (2006) Qype - local directory service with social networking and user reviews. URL http://qype.com
9. Abel F (2008) The benefit of additional semantics in folksonomy systems. In: Proceedings of the 2nd PhD workshop on Information and knowledge management, ACM, New York, NY, USA, PIKM '08, pp 49–56, DOI 10.1145/1458550.1458560, URL http://doi.acm.org/10.1145/1458550.1458560
10. Abel F, Henze N, Krause D (2008) Ranking in folksonomy systems: can context help? In: CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management, ACM, New York, NY, USA, pp 1429–1430
11. Bao S, Xue G, Wu X, Yu Y, Fei B, Su Z (2007) Optimizing web search using social annotations. In: Proceedings of the 16th international conference on World Wide Web, ACM, New York, NY, USA, WWW '07, pp 501–510
12. Bizer C, Lehmann J, Kobilarov G, Auer S, Becker C, Cyganiak R, Hellmann S (2009) Dbpedia - a crystallization point for the web of data. J Web Sem 7(3):154–165
13. Brin S, Page L (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks and ISDN Systems 30(1-7):107–117, DOI 10.1016/S0169-7552(98)00110-X
14. Brooks CH, Montanez N (2006) Improved annotation of the blogosphere via autotagging and hierarchical clustering. In: Proceedings of the 15th international conference on World Wide Web (WWW2006), Edinburgh, Scotland, pp 625–632

15. Budanitsky A, Hirst G (2006) Evaluating wordnet-based measures of lexical semantic relatedness. Comput Linguist 32(1):13–47, DOI 10.1162/coli.2006.32.1.13, URL http://dx.doi.org/10.1162/coli.2006.32.1.13
16. Cattuto C, Loreto V, Pietronero L (2006) Semiotic dynamics in online social communities. In: In The European Physical Journal, Springer-Verlag, pp 33–37
17. Cattuto C, Benz D, Hotho A, Stumme G (2008) Semantic grounding of tag relatedness in social bookmarking systems. In: Proceedings of the 7th International Conference on The Semantic Web, Springer-Verlag, Berlin, Heidelberg, ISWC '08, pp 615–631, DOI 10.1007/978-3-540-88564-1_39, URL http://dx.doi.org/10.1007/978-3-540-88564-1_39
18. Gemmell J, Shepitsen A, Mobasher M, Burke R (2008) Personalization in folksonomies based on tag clustering. In: Proceedings of the 6th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems
19. Golder SA, Huberman BA (2006) Usage patterns of collaborative tagging systems. J Inf Sci 32(2):198–208, DOI 10.1177/0165551506062337
20. Golder SA, Huberman BA (2006) Usage patterns of collaborative tagging systems. J Inf Sci 32(2):198–208, DOI 10.1177/0165551506062337, URL http://dx.doi.org/10.1177/0165551506062337
21. Gruber T (2007) Ontology of folksonomy: A mash-up of apples and oranges. International Journal on Semantic Web & Information Systems 3(2):1–11
22. Hotho A, Jschke R, Schmitz C, Stumme G (2006) Information retrieval in folksonomies: Search and ranking. In: Proceedings of the 3rd European Semantic Web Conference, Springer, Budva, Montenegro, LNCS, vol 4011, pp 411–426
23. Java A, Song X, Finin T, Tseng B (2007) Why We Twitter: Understanding Microblogging Usage and Communities. Procedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007
24. Jiang J, Conrath D (1997) Semantic similarity based on corpus statistics and lexical taxonomy. In: Proc. of the Int'l. Conf. on Research in Computational Linguistics, pp 19–33
25. Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. J ACM 46(5):604–632, DOI 10.1145/324133.324140, URL http://doi.acm.org/10.1145/324133.324140
26. Markines B, Cattuto C, Menczer F, Benz D, Hotho A, Stumme G (2009) Evaluating similarity measures for emergent semantics of social tagging. In: Proceedings of the 18th international

conference on World wide web, ACM, New York, NY, USA, WWW '09, pp 641–650, DOI 10.1145/1526709.1526796, URL http://doi.acm.org/10.1145/1526709.1526796

27. Mathes A (2004) Folksonomies – cooperative classification and communication through shared metadata. Computer Mediated Communication - LIS590CMC URL http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html-accessed25/02/2013
28. Mendes PN, Passant A, Kapanipathi P, Sheth AP (2010) Linked open social signals. In: Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01, IEEE Computer Society, Washington, DC, USA, WI-IAT '10, pp 224–231, DOI 10.1109/WI-IAT.2010.314, URL http://dx.doi.org/10.1109/WI-IAT.2010.314
29. Mika P (2007) Ontologies are us: A unified model of social networks and semantics. Web Semant 5(1):5–15, DOI 10.1016/j.websem.2006.11.002, URL http://dx.doi.org/10.1016/j.websem.2006.11.002
30. Naaman M, Boase J, Lai CH (2010) Is it really about me?: message content in social awareness streams. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work, ACM, New York, NY, USA, CSCW '10, pp 189–192, DOI 10.1145/1718918.1718953, URL http://doi.acm.org/10.1145/1718918.1718953
31. Naaman M, Boase J, Lai CH (2010) Is it really about me?: message content in social awareness streams. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work, ACM, New York, NY, USA, CSCW '10, pp 189–192, DOI http://doi.acm.org/10.1145/1718918.1718953, URL http://doi.acm.org/10.1145/1718918.1718953
32. Plangprasopchok A, Lerman K (2009) Constructing folksonomies from user-specified relations on flickr. In: Quemada J, Len G, Maarek YS, Nejdl W (eds) WWW, ACM, pp 781–790
33. Rattenbury T, Good N, Naaman M (2007) Towards automatic extraction of event and place semantics from flickr tags. In: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, New York, NY, USA, SIGIR '07, pp 103–110, DOI 10.1145/1277741.1277762
34. Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web, ACM, New York, NY, USA, WWW '10, pp 851–860, DOI http://doi.acm.org/10.1145/1772690.1772777, URL http://doi.acm.org/10.1145/1772690.1772777

35. Sen S, Lam SK, Rashid AM, Cosley D, Frankowski D, Osterhouse J, Harper FM, Riedl J (2006) tagging, communities, vocabulary, evolution. In: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, ACM, New York, NY, USA, CSCW '06, pp 181–190, DOI 10.1145/1180875.1180904

36. Wu H, Zubair M, Maly K (2006) Harvesting social knowledge from folksonomies. In: Proceedings of the seventeenth conference on Hypertext and hypermedia, ACM Press New York, NY, USA, pp 111–114

37. Xi W, Zhang B, Chen Z, Lu Y, Yan S, Ma WY, Fox EA (2004) Link fusion: a unified link analysis framework for multi-type interrelated data objects. In: WWW '04: Proceedings of the 13th international conference on World Wide Web, ACM Press, New York, NY, USA, pp 319–327

38. Yeh I, Karp PD, Noy NF, Altman RB (2003) Knowledge acquisition, consistency checking and concurrency control for gene ontology (go). Bioinformatics 19(2):241–248

39. Zhou M, Bao S, Wu X, Yu Y (2007) An unsupervised model for exploring hierarchical semantics from social annotations. In: Aberer K, Choi KS, Noy NF, Allemang D, Lee KI, Nixon LJB, Golbeck J, Mika P, Maynard D, Mizoguchi R, Schreiber G, Cudr-Mauroux P (eds) ISWC/ASWC, Springer, Lecture Notes in Computer Science, vol 4825, pp 680–693