Contextualising Tags in Collaborative Tagging Systems

Ching-man Au Yeung, Nicholas Gibbins, Nigel Shadbolt
Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science
University of Southampton
Southampton, SO17 1BJ, United Kingdom
{cmay06r,nmg,nrs}@ecs.soton.ac.uk

ABSTRACT

Collaborative tagging systems are now popular tools for organising and sharing information on the Web. While collaborative tagging offers many advantages over the use of controlled vocabularies, they also suffer from problems such as the existence of polysemous tags. We investigate how the different contexts in which individual tags are used can be revealed automatically without consulting any external resources. We consider several different network representations of tags and documents, and apply a graph clustering algorithm on these networks to obtain groups of tags or documents corresponding to the different meanings of an ambiguous tag. Our experiments show that networks which explicitly take the social context into account are more likely to give a better picture of the semantics of a tag.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Clustering; H.3.4 [**Information Storage and Retrieval**]: Systems and Software; H.3.5 [**Information Storage and Retrieval**]: Online Information Services

General Terms

Algorithms, Experimentation, Human Factors

Keywords

collaborative tagging, folksonomy, semantics, context

1. INTRODUCTION

Collaborative tagging systems [13] such as Delicious[1] and Bibsonomy[2] have emerged in recent years to become popular tools for organising and sharing resources on the Web. These systems allow Web users to use freely-chosen keywords (tags) to describe Web documents. The collaborative nature of these systems results in some continuously evolving categorisation schemes now commonly known as folksonomies [13].

While collaborative tagging offers many advantages over the use of controlled vocabularies [1], they also suffer from several limitations at the same time due to the unrestricted nature of tagging [13]. For example, polysemy is highly prevalent in Delicious [32]. The fact that many tags are ambiguous has limited the effectiveness of collaborative tagging systems in document description and retrieval. For example, when a user wants to retrieve documents about San Francisco from Delicious by using the tag `sf`, documents on various topics of science fictions – which is also abbreviated to *sf* – are also returned. While large scale clustering of tags and documents in folksonomies for discovering semantically related tags has been done in quite a number of studies (e.g. [6, 7, 24]), how different meanings of individual tags can be discovered remains largely overlooked.

In order to understand the semantics of the tags, we need to know the contexts in which they are used. While it is possible to consult some dictionaries or thesauri such as WordNet [20] for the multiple meanings of words, they do not always match with what the tags are actually used for within the system. Hence, we target for an automatic way of contextualising tags in this paper by performing clustering on network structures induced from a folksonomy. Specifically, we consider several different network representations of tags and documents, including keyword-based and user-based approaches. We use a graph clustering algorithm on the resultant networks and study whether the different contexts in which a tag is used can be discovered in different cases. Due to the lack of a 'gold standard' for evaluation, we analyse the results of our experiments both quantitatively and qualitatively, with the help of a set of manually classified Delicious data set and WordNet [20].

We first provide an overview of collaborative tagging and folksonomies in the next section, followed by a description of different network representation of subsets of folksonomies and the method of clustering in Section 3 and Section 4 respectively. We then present our experiments and discuss their implications in Section 5. We mention some related work in Section 6. Finally we give conclusions and mention future research directions in Section 7.

2. COLLABORATIVE TAGGING

Tagging originates from the idea of using keywords to describe and categorise documents. Collaborative tagging systems emerged in recent years have taken this idea further

[1]Delicious: http://delicious.com/
[2]BibSonomy: http://www.bibsonomy.org/

HT'09, June 29–July 1, 2009, Torino, Italy.

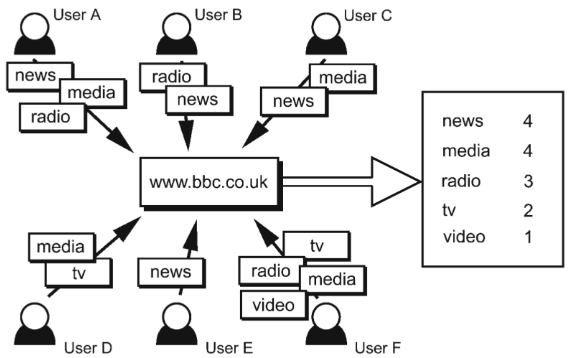


Figure 1: A collaborative tagging system aggregates tags contributed by different users to form an overall description of the document.

by allowing general users, i.e. the consumer of the information, to assign freely-chosen tags to Web documents. For example, one can post a bookmark of the homepage of BBC (http://www.bbc.co.uk/) to Delicious, and assign to it tags such as `tv`, `media` and `sports`. As the tags of different users are aggregated, the tags form a kind of signature of the document (see Figure 1), which acts as an overall description of the document from the perspective of the users, and can also be used to facilitate retrieval.

Collaborative tagging is generally considered to have a number of advantages over traditional methods of organising information [1, 26] as evidently shown by its popularity among general Web users and its application on a wide range of Web resources. In particular, the flexibility and freedom offered by these systems to Web users are what make them distinguishable from traditional systems which involve predefined taxonomies or categories. In addition, these systems are quick to adapt to changes in the vocabulary among users. However, collaborative tagging also suffers from certain limitations due largely to its unrestricted nature. Since vocabulary is uncontrolled, there is no way to make sure that a tag corresponds to a single well-defined concept. For instance, there are quite a lot of polysemous tags such as `port` and `bridge` in Delicious.

Folksonomies are products of collaborative tagging systems. The term is a combination of the terms *folk* and *taxonomy*, which emphasises on the collaborative and social nature of folksonomy as a categorisation scheme. A folksonomy is generally considered to consist of at least three types of elements [18, 19, 35], namely users, tags and Web documents. ***Users*** assign tags to Web documents in collaborative tagging systems, and are also referred to as actors as in social network analysis. ***Tags*** are keywords chosen by users to describe and categorise Web resources. Depending on the design of the systems, tags can be a single word, a phrase or a combination of symbols and alphabets. Tags are referred to as concepts in some work which focuses on extracting lightweight ontologies from folksonomies. Finally, ***documents*** refer to the objects tagged by the users in collaborative tagging systems.

In this paper, we focus on the interrelations between the three elements in a folksonomy, additional information such as the time at which a tag is assigned is less relevant. In addition, our primary source of data Delicious does not allow any subsumption relations between tags to be defined. Hence, we adopt a basic model of folksonomy which involves only the three basic elements.

Definition 1. *A folksonomy $\mathcal{F}$ is a tuple $\mathcal{F} = (U, T, D, A)$, where U is a set of users, T is a set of tags, D is a set of Web documents, and $A \subseteq U \times T \times D$ is a set of annotations.*

3. FOLKSONOMY AS NETWORKS

A folksonomy is primarily a set of associations between three basic elements, namely users, tags and documents. Hence a folksonomy can always be represented in the form of a graph or a network, with vertices representing the elements and edges representing their associations. As folksonomies involves three different types of nodes, the underlying networks are usually represented in the form of tripartite hypergraphs [10, 16, 19, 24]: $\mathcal{H} = \langle V, E\rangle$ where $V = (U \cup T \cup D)$ and $E = \{(u, t, d) \mid (u, t, d) \in A\}$.

Depending on the types of element in a folksonomy we would like to focus on, different types of sub-networks can be generated from the tripartite hypergraph of a folksonomy. For example, a network of tags with weights of edges determined by co-occurrence is considered in quite a number of works (e.g. [6, 30]). On the other hand, Mika [19] considers the bipartite graphs of user-tag associations and document-tag associations, which are further folded into a one-mode network of tags as a lightweight ontology.

In this paper, as we are focusing on individual tags, we will always be working on a subset of the folksonomy that is associated with a particular tag. In this section we describe several different kinds of networks on which clustering algorithms can be applied such that we are likely to discover the different contexts in which a tag is used. We first introduce the following notations.

Given a tag t, we denote by U_t the set of users who have used the tag t on one or more documents: $U_t = \{u \mid \exists d \in D, (u, t, d) \in A\}$; by D_t the set of documents which have been assigned the tag t: $D_t = \{d \mid \exists u \in U, (u, t, d) \in A\}$; and by T_t the set of tags which have been used together with t on some documents by the same users: $T_t = \{t' \mid \exists(u, d) \in U \times D, (u, t, d) \in A \wedge (u, t', d) \in A \wedge t \neq t'\}$. In addition, we employ the cosine similarity measure between two vectors in the following discussion.

$$csim(\boldsymbol{v}_1, \boldsymbol{v}_2) = \frac{\boldsymbol{v}_1 \cdot \boldsymbol{v}_2}{||\boldsymbol{v}_1|| \times ||\boldsymbol{v}_2||} \tag{1}$$

where $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbf{R}^n$.

For all the networks described in the following sections, we always assume that we are focusing on a particular tag t, and we refer to the sets U_t, D_t and T_t for the construction of the networks.

3.1 Tag-based Document Networks

Tagging can be considered as an act of indexing the documents on the Web. A weighted term (tag) vector $\boldsymbol{v}_d$, which is commonly used in document clustering and information retrieval (e.g [11, 31, 34]), can be constructed to represent a document d, with each element of the vector corresponding to the number of times a tag has been assigned to it.

$$\boldsymbol{v}_d = (v_{d,1}, v_{d,2}, ..., v_{d,|T_t|}) \tag{2}$$

where $v_{d,i} = |\{u \mid (u, t_i, d) \in A\}|$.

A similarity matrix $\mathbf{A} = \{a_{ij}\}$ can be constructed to represent the pairwise similarity of each document by using the cosine similarity measure:

$$a_{ij} = csim(\boldsymbol{v}_{d_i}, \boldsymbol{v}_{d_j}). \quad (3)$$

.

In this way, a network with $|D_t|$ vertices representing each of the documents and edges weighted by the similarity between these documents can be constructed. It can be hypothesised that, if the tag t is used by the users to refer to different concepts in different contexts, we should find vertices representing documents which correspond to the same context to be highly connected with each other, resulting in different clusters of vertices. To obtain a label for each of the clusters, one can extract the tags which are most frequently used among the documents within the clusters.

It should be noted that the tag t which we are looking at is not included in the term vectors given the definition of T_t. This is actually desirable because t has been assigned to every document (and probably by many users) in the set D_t, therefore the inclusion of the tag in the vectors will probably result in all the documents being very similar to each other.

In addition, when constructing the term vectors we only consider the frequencies of the tags, while some weighting schemes such as the TF-IDF (term frequency-inverse document frequency) scheme [28] can also be used (as in [7, 8] for example). Our reason is that we are trying to group documents which are about similar topics (e.g. San Francisco), instead of trying to identify keywords which are most important to a document. It is found that tags are more likely to be broad terms rather than specific terms [25], suggesting that tags are more likely to be used to categorise a document. Hence, by considering the frequencies of tags in term vectors, we should be able to group documents into different categories, which correspond to the different contexts in which the tag t is used.

3.2 User-based Document Networks

The second type of network is based on the consideration that documents in a tagging system can also be characterised by the users who have assigned tags to them, and that documents tagged by similar users can be considered as similar to each other. Formally, this can be represented by a similarity matrix $\mathbf{B} = \{b_{ij}\}$:

$$b_{ij} = |\{u \mid (u, t, d_i) \in A \wedge (u, t, d_j) \in A\}|. \quad (4)$$

In this way, a network with $|D_t|$ vertices representing each of the documents and edges weighted by the number of users who have assigned the tag to both of them can be constructed.

While this type of network that characterises documents simply by the users who have assigned a particular tag to them has not been considered in the literature, it does provide valuable insight into how the tag is used among the users by putting them into the social context. In a preliminary studies [2] on this type of network, we reveal that most users are consistent in using a certain tag, meaning that they are unlikely to use the same tag to refer to different concepts. Hence, documents which are linked to each other in this network are likely to be about similar topic and constitute the same context in which the tag is used.

If clustering algorithms applied to this type of network reveal any clusters of documents, it is very likely that they will correspond to different contexts in which the tag is used. To obtain a label for each cluster, we can extract the tags that are most frequently used among the documents within the clusters.

3.3 Tag Co-occurrence Networks

Besides document networks, we can also look directly at the set T_t of tags which are used together with the tag t. The most common way of constructing a network of tags which reflects the relations between them is to consider their co-occurrence [6]. Intuitively two tags are more related to each other if they are used together more frequently on the same documents and/or by the same users. Mathematically, a matrix $\mathbf{C} = \{c_{ij}\}$ representing the strength of relations between tags can be constructed by counting the number of times two tags are used together using one of the following two methods:

$$\begin{aligned} c_{ij} &= |\{d \mid \exists u_a, u_b, (u_a, t_i, d) \in A \wedge (u_b, t_j, d) \in A\}| \quad (5) \\ c'_{ij} &= |\{(u, d) \mid (u, t_i, d) \in A \wedge (u, t_j, d) \in A\}| \quad (6) \end{aligned}$$

with the exception that $c_{ij} = 0$ and $c'_{ij} = 0$ if $i = j$.

While Equation (5) only requires two tags to be assigned to the same document for the situation to be considered a co-occurrence, Equation (6) defines co-occurrence between two tags as a situation in which they have to be used on the same document by the same user. This distinction is not explicitly considered and discussed in previous studies. For example, the former is used by Belgeman et al. [6] in tag clustering, and the latter is used by Cattuto et al. [9] when studying various tag similarity measures.

In this paper we consider both methods and want to find out whether the associations established by the users are significant in understanding the semantics of a tag. Equation (6) will produce tag relations with smaller weights because obviously $c'_{ij} \leq c_{ij}$ (tags can be assigned to the same document but not necessarily by the same user). However, it can be hypothesised that Equation (6) will produce tag relations of higher significance because the viewpoints of the users are explicitly taken into account. In addition, (6) should be less vulnerable to spamming in collaborative tagging systems, as tags assigned by spammers are much less likely to be associated with tags assigned by other users (more on spamming in tagging systems can be found in [15]).

It should be noted that there is also one more type of tag co-occurrence, which is the situation in which two tags have been used by the same user, regardless of whether they are used on the same document or on different documents. However, we believe that this is of small value in this paper as tags used on different documents are much less likely to be semantically related, as we have found in a separate study that user interests can be very diverse [4].

3.4 Tag Context Similarity Networks

The last type of network we consider in this paper is based on the distributional measure of tag relatedness proposed in [9]. In order to use this measure, we have to define a tag co-occurrence vector $\boldsymbol{v}_{t_i}$ for each tag $t_i \in T_t$:

$$\boldsymbol{v}_{t_i} = (v_{t_i,1}, v_{t_i,2}, ..., v_{t_i,|T_t|}) \quad (7)$$

where $v_{t_i,j} = c_{ij}$ or $v_{t_i,j} = c'_{ij}$ depending on which of the aforementioned method is used. A matrix $\mathbf{D} = \{d_{ij}\}$ representing a network of tags can then be constructed by calculating the similarity between two tags with the cosine similarity measure:

$$d_{ij} = csim(\boldsymbol{v}_{t_i}, \boldsymbol{v}_{t_j}) \quad (8)$$

The tag co-occurrence vector reflects the context in which a tag is used because it encodes the co-occurrence frequencies of other tags which are used with this tag. Hence, the cosine similarity used in Equation (8) is actually performing a comparison of the contexts in which two tags are used (cf. [29]). This is different from the tag co-occurrence network in which tags are considered to be related or similar simply when they are used together.

The networks we consider in this paper are constructed based on the *fully-connected* approach [17], which means that any pair of vertices with a positive similarity between them will be connected by an edge. In fact there are other ways to construct these networks. In particular, we can choose to discard certain edges if their weights are too small. For example, we can adopt the ϵ-neighbourhood approach, which removes edges with weights lower than ϵ. Or we can adopt the k-nearest neighbour approach, in which each vertex in the graph connects to at most k neighbours which are most similar to it. However, as it is not clear at present which approach is the most suitable for our tasks and we would like to take as much information as we have into consideration, the fully-connected approach is used. We will consider comparing the different approaches in our future work.

4. TAG CONTEXTUALISATION BY GRAPH CLUSTERING

By tag contextualisation we mean the process of finding out the different contexts in which a tag is used. The result of such process will be one or more sets of tags which when presented with the tag in question point to different concepts it represents.

Given the above networks, graph clustering algorithms are expected to return a set of clusters, with each of them hopefully corresponds to one context in which a tag is used. There are actually many different kinds of clustering algorithms available, such as the k-means and spectral graph clustering. Our early experiments show that simple k-means method does not produce satisfactory results. Besides, it requires the value of k to be specified beforehand, which is not possible in our case.

In recent years, community discovery algorithms for networks have attracted attentions of researchers from different disciplines [22, 27]. The aim is to identify groups of vertices in a network which are highly connected with those in the same group but loosely connected with those in other groups. This notion is highly relevant to our task because documents or tags related to the same meaning of a tag should be highly connected in the above networks. In addition, *modularity* [23] has been proposed as a measure of the quality of a division of a network.

In this paper, we choose the fast greedy algorithm for optimising modularity generalised to handle weighted networks [21] as a basis of our tag contextualisation process. The algorithm is chosen because of its efficiency and good performance in a wide range of problems. While different clustering algorithms would produce different results, we believe our experiments provide useful qualitative insights into the differences between the aforementioned network representations of folksonomies. The process involves the following three steps.

1. Firstly, we construct either a network of documents or tags based on one of the methods mentioned in the previous section. This is represented as an adjacency matrix.
2. Secondly, we apply the clustering algorithm to the network and obtain a set of clusters of nodes.
3. Thirdly, we extract labels for each of the clusters. For document networks, we extract the N most popular tags among the documents in a cluster. For tag networks, we extract the top N tags which are most frequently used with the tag in question. These tags constitute the different contexts the clusters correspond to.

As an illustrating example, consider the tag `wine`. We observe that the tag has been used by users in Delicious in two different contexts: (1) as a kind of alcoholic drinks and (2) as the name of a software application. The result of the clustering process on a document network may contain two clusters, with one corresponding to the first context and another corresponding to the second. The top five tags extracted as labels for the two clusters may be something like these:

1. {food, shopping, drink, vino, cooking}
2. {linux, ubuntu, emulation, windows, software}

Although this method of extracting sets of tags as labels for the clusters does not produce exactly the different meanings of a tag, the most frequently used tags actually constitute a coherent context from which the exact meaning of the tag can be easily deduced. This form of representation also facilitates further utilisation of the information in other applications in which comparisons between sets of tags are required.

5. EXPERIMENTS

5.1 Data Preparation

We conduct experiments on data obtained from Delicious, which is by far the most popular collaborative tagging system on the Web. Evaluation of tag contextualisation is challenging due to a lack of a 'gold standard' or a ground truth. Similar studies in word sense discrimination usually resolve to a small set of manually-examined samples, or the use of pseudowords – artificially ambiguous words created by combining two different words together [29]. The use of pseudowords is less suitable in our case as the user groups of two different tags may be very different that results may not be useful in general. In addition, the use of an established dictionary such as WordNet [20] may not be as helpful as one would expect. This is because it is very possible that not all meanings of a tag defined in the dictionary are used by users in Delicious, and there may be new meanings of the tag which do not necessarily appear in the dictionary (see Section 5.5). Hence, in this paper we rely on a small set of data crawled from Delicious and perform both quantitative and qualitative analyses as described below.

Tag	# Users	# Docs	# Tags
architecture	7,963	689	1,891
bridge	1,392	819	1,294
language	7,577	621	1,689
opera	3,851	668	1,138
sf	2,578	1,009	2,085
soap	6,130	1,051	1,263
sun	3,928	697	1,566
tube	2,882	769	1,798
wine	5,437	790	1,029
xp	4,474	529	1,231

Table 1: Statistics of the collected data sets.

We first identify 20 ambiguous tags on Delicious, i.e. tags used to represent two or more concepts by the users. We complement the list with 30 tags randomly selected from a set of 100 popular tags crawled from the front page of Delicious. For each of the tags, we collect a set of users who have used the tag, a set of documents to which the tag has been assigned, and a set of other tags which have been used together with the tag. Although the interface of Delicious does not allow access to the full dataset with respect to a particular tag, we still have over 500 documents for each of the tags, which are quite sufficient for our analysis.

We ask ten users who have basic understanding of collaborative tagging systems to classify bookmarks randomly collected from Delicious by examining the intended meaning of a specific tag. For example, for the tag `sf`, a participant would put documents about San Francisco in one group, and those about science fictions in another. Each user examines the data of two tags, each containing 50 randomly selected documents. Hence, every dataset are examined by two participants. We obtain the final outcomes by combining the classifications given by two participants. From these 50 tags, we choose 10 tags of which the classifications of the two participants most agreed on, i.e. having two or more common contexts and two or less different contexts. These tags represent a good range of topics from different domains. Table 1 summarises the statistics of the datasets of the final 10 tags. We use sets of tags extracted from different groups of classified documents as their labels. Table 2 gives the result of this manual classification process.

It should be noted that while the manual classification process does not necessarily return all the contexts in which the selected tags are used, they do provide a reasonably good common ground for the comparison of the different networks described in the previous section. In the following section, we describe the performance measures used in our quantitative analysis.

5.2 Performance Measures

We denote the set of contexts discovered automatically by the clustering algorithms by $\mathbf{S}_t^A = \{s_{t,i}^A\}$, and the set of manually discovered contexts by $\mathbf{S}_t^M = \{s_{t,i}^M\}$. In addition, we define a match function which, given the set $\mathbf{S}_t^A$ of automatically discovered contexts and a particular manually discovered context $s_{t,i}^M$, returns the number of automatically discovered contexts which match the manually discovered one (two automatically discovered contexts may correspond to the same manually discovered context):

$$\text{match}(\mathbf{S}_t^A, s_{t,i}^M) = \sum_{s \in \mathbf{S}_t^A} \delta(s, s_{t,i}^M) \qquad (9)$$

where $\delta(s_i, s_j) = 1$ if s_i refers to the same context as s_j, $\delta(s_i, s_j) = 0$ otherwise.

We introduce two performance measures here which will be used to study the differences between the aforementioned networks. Note that we do not measure the precision of the contextualisation process. This is because we do not really have a clear idea of what is an incorrect outcome. Given the limited data in the manual classification process, the clustering process is very likely to discover contexts that have not been identified in the former. Hence, we believe it would be more useful to study the following two measures, namely recall and redundancy, as well as to qualitatively look into the results to see if unexpected contexts are meaningful ones.

$$\text{Recall} = \frac{|\{s_{t,i}^M | \text{match}(\mathbf{S}_t^A, s_{t,i}^M) > 0\}|}{|\mathbf{S}_t^M|} \qquad (10)$$

$$\text{Redundancy} = \frac{\sum_{s_{t,i}^M \in \mathbf{S}_t^M} F(\mathbf{S}_t^A, s_{t,i}^M)}{|S_t^A| - 1} \qquad (11)$$

where

$$F(\mathbf{S}_t^A, s_{t,i}^M) = \begin{cases} \text{match}(\mathbf{S}_t^A, s_{t,i}^M) - 1 & \text{if match}(\mathbf{S}_t^A, s_{t,i}^M) > 1 \\ 0 & \text{otherwise} \end{cases} \qquad (12)$$

As Equation (10) suggests, ***recall*** measures the fraction of contexts discovered by the automatic process with respect to the contexts which are manually discovered. High recall means that the automatic process is able to discover more contexts in which a tag is used. ***Redundancy***, on the other hand, measures how many clusters returned by the clustering algorithms actually correspond to the same context. Higher redundancy means that extra effort is needed to combine similar contexts. A good result should achieve high recall (returning all the contexts discovered by the manual process), and low redundancy (all contexts returned are unique).

5.3 Quantitative Analysis

We apply the fast greedy algorithm for optimising modularity to each of the different types of network for the ten tags. We manually calculate the recall and redundancy measures by examining the tags extracted from the clusters. In most cases, the tags alone clearly reveal what the contexts are, with large overlap with the tags extracted in the manual process. There are also cases in which the tags constitute contexts which are not discovered in the manual process. The results are shown in Table 3 and Figure 2.

Figure 2 shows that the user-based document networks (UD) and tag co-occurrence networks (TC and TC') produce the largest number of clusters. By comparing the tag-based document networks (TD) and UD, we find out that edge weights in TD are usually higher than in UD. This is because documents sharing a similar set of tags may not be tagged by a similar set of user. In particular, there can be no edges between two documents if there are no users who have assigned the tag t to both of them, even though they are about a similar topic. Hence, the clustering algorithm breaks down UD into more and smaller clusters. For TC and TC', the relatively large number of clusters can be explained by the fact that many documents share only a few popular tags, and the other tags thus are less connected with each other, forming small groups of tags in the networks.

Tag	Context	Label
architecture	physical structures	design, home, art, travel, urban
	programming	design, software, reference, development, webdev
bridge	networking	networking, network, wifi, wireless, linux
	card game	games, cardgame, poker, resources
	architecture	architecture, structure, travel, photos, blog
language	human	reference, education, learning, english, dictionary
	computer	programming, research, reference, software, miscrosoft
opera	music	music, classical, tickets, theatre, woman
	browser	browser, web, software, tools, javascript
sf	science fiction	scifi, fiction, science, literature, sci-fi
	city	sanfrancisco, san, francisco, bayarea, california
soap	cleaning agent	soapmaking, diy, recipes, making, organic
	web services	webservices, programming, xml, web, soa
sun	computer company	solaris, java, linux, programming, unix
	astronomy	science, astronomy, space, photography, solar
tube	video sharing	video, youtube, you, videos, web2.0
	electronics	diy, amplifier, audio, electronics, amp
	underground	london, travel, transport, map, uk
wine	beverage	food, drink, cooking, alcohol, shopping
	software	linux, ubuntu, software, windows, tools
xp	operating system	windows, software, computer, tools, microsoft
	programming	software, development, extremeprogramming, process, agile

Table 2: Results of the manual classification process. The names of the context are added for easier comprehension. Due to limited space, we only show the top five tags for each context.

The tag context similarity networks (CS and CS') return the fewest number of clusters. This is because they do not only incorporate co-occurrence information but also involve the comparison of the contexts of each tag in calculating their similarity (also known as second-order co-occurrence [29]. This is actually similar to the idea of latent semantic analysis [12]. In other words, tags are not connected only because they have been used together directly, but because they have been used with other similar tags (having similar contexts). This increases the number of edges and edges weights in a tag network, thus vertices are more connected with each other, resulting in a smaller number of clusters.

Recall, on the other hand, is generally high in all of the cases. In particular, both UD and TC' achieve 100% recall. In fact, the manually discovered meanings of the tags can be identified for most of the time except in the cases of a few tags such as *bridge* and *tube* which have more different meanings than the other tags. A closer look at the clusters in CS and CS', which achieves relatively lower recall, reveals that tags related to the missing meanings are included in a cluster which corresponds to a different meaning. This means that the context similarities between some less related tags are too strong such that the clustering algorithm is unable to split them into two groups.

The chart of redundancy levels has a similar shape as that of number of clusters. This is because when the clustering algorithm returns more clusters it is more likely that two or more clusters correspond to the same context, especially when the number of contexts in which a tag is used in Delicious is limited. However, we also note that high redundancy levels in some cases are also due to the fact that the contexts discovered in the manual classification process are too general, such that some more specific contexts discovered in the clustering process are mapped to the same contexts. We will discuss more about this in the next section.

Redundancy is important because when it is too high the results can not be directly used in other applications. Some post-processing steps will be needed to combine clusters which correspond to the same context. Given that we label the clusters with sets of tags extracted from the clusters, one way to combine the clusters is to compare the sets of tags and perform a merge if there is significant overlap. The second option is to filter away clusters of small size. For example, if we remove clusters of size less than 5% of the total number of nodes in the user-based document networks (UD), it achieves a redundancy level of 0.3, similar to that of TD, while maintaining a recall level of 1.0.

5.4 Qualitative Analysis

Firstly, we look at the extra contexts discovered by the clustering algorithm. The use of UD returns the largest number of 'new meanings' of the tags we examine. For example, it reveals that the tag `sf` is also used by Delicious users to refer to 'Sourceforge', an open source software repository on the Web. It also reveals that the tag `soap` is also used to refer to 'TV dramas'. These meanings are not identified by tag-based networks such as TD and TC. A closer look at the documents and tags in the corresponding clusters reveals that only a relatively small number of users are using the tags for those meanings (about 5% of users for 'Sourceforge' and less than 2% for 'TV dramas').

If we perform clustering based on tags, tags which are used in those contexts are likely to be mixed up with other tags if they co-occur in some other documents. On the other hand, by connecting documents based only on the users (as in the case of UD), it is more likely that documents which are about the same topic would be grouped together, causing also tags used in the same context to be grouped together as well.

In addition, the existence of subtopics among the clusters is another aspect which is not reflected in the quantitative performance measures. The meanings discovered in the manual classification process (Table 2) are actually rather general. For example, while `sun` is found to be used to refer to the computer company, there are some clusters which point to particularly the Java programming language devel-

Tag	TD				UD				TC				TC'				CS				CS'			
	N	Rl	Ry	E	N	Rl	Ry	E	N	Rl	Ry	E	N	Rl	Ry	E	N	Rl	Ry	E	N	Rl	Ry	E
architecture	5	1.0	0.6	0	3	1.0	0.3	0	6	1.0	0.3	1	8	1.0	0.3	1	2	1.0	0.0	0	2	1.0	0.0	0
bridge	3	0.8	0.0	0	14	1.0	0.6	0	6	0.5	0.3	0	7	1.0	0.3	0	2	0.5	0.0	0	4	0.8	0.0	0
language	3	1.0	0.3	0	6	1.0	0.7	0	7	1.0	0.7	0	8	1.0	0.8	0	3	1.0	0.3	0	2	1.0	0.0	0
opera	3	1.0	0.3	0	9	1.0	0.8	0	5	1.0	0.6	0	7	1.0	0.7	0	3	0.5	0.3	0	3	1.0	0.3	0
sf	2	1.0	0.0	0	8	1.0	0.6	1	6	1.0	0.7	0	7	1.0	0.4	0	2	1.0	0.0	0	3	1.0	0.3	0
soap	5	0.5	0.6	0	11	1.0	0.7	1	9	1.0	0.8	0	8	1.0	0.6	0	3	1.0	0.3	0	4	1.0	0.0	1
sun	3	1.0	0.3	0	10	1.0	0.8	1	9	1.0	0.7	0	8	1.0	0.6	0	4	1.0	0.5	1	3	1.0	0.3	0
tube	3	1.0	0.0	0	8	1.0	0.6	0	5	0.7	0.0	0	11	1.0	0.7	0	4	1.0	0.3	0	3	0.0	0.0	0
wine	3	1.0	0.3	0	12	1.0	0.8	0	3	1.0	0.3	0	5	1.0	0.4	0	2	1.0	0.0	0	2	1.0	0.0	0
xp	3	1.0	0.3	0	7	1.0	0.7	0	6	1.0	0.7	0	7	1.0	0.7	0	3	1.0	0.3	0	2	1.0	0.0	0

Table 3: Results of tag meaning disambiguation. The network types are TD (tag-based document network), UD (user-based document network), TC (tag co-occurrence network), TC' (tag co-occurrence network with user information), CS (tag context similarity network), and CS' (tag context similarity network with user information). N stands for number of clusters, Rl for recall, Ry for redundancy, and E is the number of extra contexts discovered.

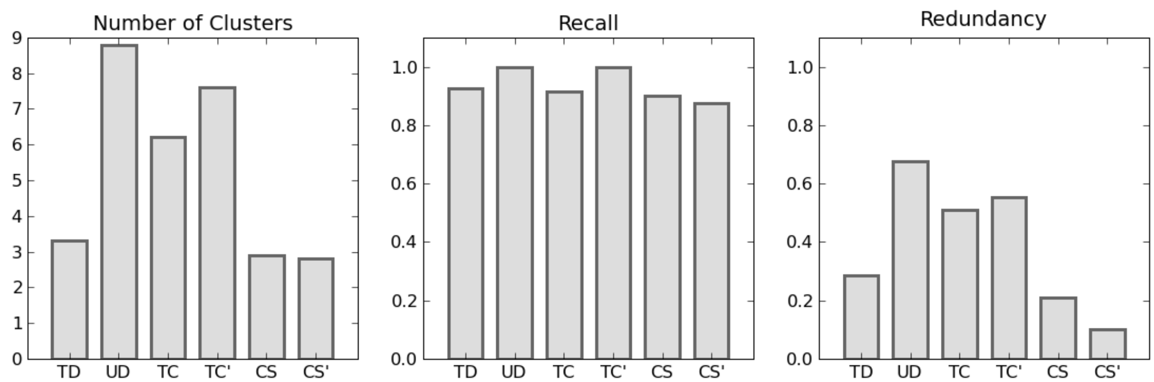


Figure 2: Average number of clusters, recall and redundancy of the tag meaning disambiguation process.

oped by the company, and some others which point to the company's Solaris operating system. In this respect, clustering of UD returns more subtopics than the other networks. For example, for the tag `sf`, there are clusters about food and restaurants in San Francisco, while others are about hiking and outdoor activities in the city. For the tag `language`, there are clusters which correspond to different languages such Chinese, English and Japanese. The context similarity networks CS and CS', which return the least number of clusters, return the least number of subtopics. This is probably because the context similarity tends to group tags into as general groups as possible.

Finally, we also notice that clustering on UD also returns some language-specific clusters. There are clusters with only Chinese documents described mainly by Chinese tags, and some others with only Japanese documents described mainly by Japanese tags. This suggests that user-based networks are also able to identify specific user communities of different languages in a collaborative tagging system.

5.5 Comparison with WordNet

As a further qualitative analysis, we compare the results of tag contextualisation with meanings returned by WordNet [20]. WordNet is an English lexicon which groups words into sets of synonyms called synsets. It also distinguishes between different senses of a polysemy word by associating the word with different synsets. By querying WordNet, it is possible to find out the different meanings of an ambiguous word.

We submit each of the 10 tags as queries to the online interface of WordNet.[3] Each query returns a set of synsets in which the tag appears. For example, submitting a query to WordNet using the tag `opera` would obtain the following: (1) a drama set to music, consists of singing with orchestral accompaniment and an orchestral overture and interludes; (2) a commercial browser; and (3) opera house, a building where musical dramas are performed.

We manually compare these synsets with the results of clustering described in the previous section. It should be noted that it is not trivial to perform the comparison as there may not be a one-to-one mapping between the contexts discovered in the clustering process and the synsets returned by WordNet. For example, WordNet returns four synsets for the tag `architecture`, three of which are related to building physical structures and refer to the structures, the discipline and the profession respectively. When performing the comparison, we consider the above three synsets are all matched by one of the contexts discovered in the clustering process.

Due to space limitation, we only briefly report important findings here. Firstly, the number of synsets returned by WordNet for a tag is usually larger than that of contexts discovered in the clustering process. WordNet also returns more fine-grained results. For example, *soap* can be used to

[3]http://wordnetweb.princeton.edu/perl/webwn

refer to money offered as a bribe, and is used as the street name of a drug. While this suggests that the contexts discovered in the clustering process may not be comprehensive enough, it is also possible that these additional meanings of the tag are never or rarely used in Delicious. If the aim of tag contextualisation is to enhance organisation and retrieval of documents in tagging systems, additional meanings of a tag would not be very useful.

A more important finding is that quite a number of contexts discovered in the clustering process cannot be found in WordNet. These include 'programming' (architecture), 'networking' (bridge), 'web services' and 'TV dramas' (soap), 'computer company' (sun), 'video sharing' (tube) and 'software' (wine). In addition, WordNet does not offer any information about abbreviations such as *sf* and *tube*. One may suggest that these meanings can be found on Wikipedia's disambiguation pages. However, Wikipedia offers mainly textual information and it is difficult to query Wikipedia for structured data (DBpedia [5] and YAGO [33], which are attempts to construct ontologies by extracting information from Wikipedia, do not contain all the disambiguation information). In addition, the disambiguation pages of Wikipedia usually contain a lot of meanings of a term, most of which are not found to be used in Delicious.

In summary, the above analysis suggests that querying external resources may not be a suitable way of obtaining the different contexts in which ambiguous tags are used. In this respect, unsupervised clustering methods are more suitable especially when we want to find out how the tags are actually used within a collaborative tagging system.

5.6 Summary

Our experiments suggest that the use of graph-based clustering algorithms to perform tag contextualisation on an individual-tag level produces promising results. We summarise our findings as follows.

- Tag-based document networks, while being one of the simplest forms of network derived from a folksonomy, do not favour the identification of meanings used by only a small number of users or a specific user group.

- Tag context similarity networks tend to capture the most general concepts represented by the tags being disambiguated. It provides the most clear-cut results among all the network types. However, it also tends to miss some contexts in some cases.

- User-based document networks facilitate the identification of many sub-topics with are actually interested by the users in the folksonomy, and it even helps to identify user communities with respect to a particular topic. This is probably due to the fact that these networks ground the relationship between documents on the social context, i.e. the group of users who are interested in them.

- Automatic clustering of folksonomy networks for tag contextualisation produces satisfactory results. Compared to the use of external resources such as dictionaries and ontologies, it is more likely to identify the different contexts in which the tags are actually used within the system.

The technique of contextualising tags in a folksonomy has many applications on the Web. For example, the identified contexts as well as the corresponding relevant tags can be used directly to classify documents in a folksonomy. When a user searches for documents with the tag `sf`, the system can use the sets of tags which correspond to the different contexts of the tag to partition the result into two or more groups of documents of different topics, thus facilitating the user in locating documents most relevant to his needs. We have already applied the technique to provide classification to Web search results and obtained satisfactory results [3].

6. RELATED WORKS

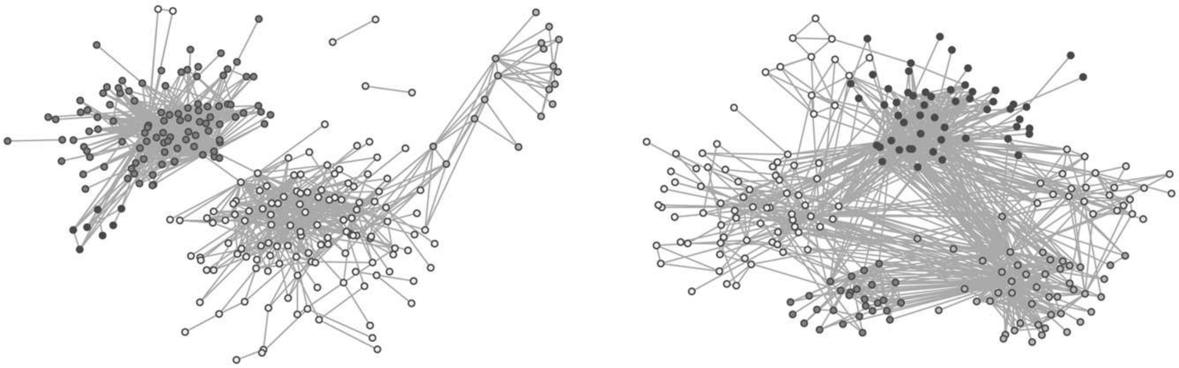
To our best understanding, there have been no studies in the literature that directly address the problem of tag ambiguity. Early work on folksonomy analysis focuses on clustering of tags in order to discover semantically related tags. Begelman et al. [6] employ clustering techniques to find groups of tags which are related to each other. Brooks and Montanex [7] also study hierarchical clustering of tags in Technorati, a collaborative tagging system for tagging blogs.[4] However, the fact that tags can possess multiple meanings in different contexts is not explicitly considered.

Cattuto et al. [8] perform spectral clustering on bookmarks, with the similarity between two bookmarks determined by a weighted comparison of the tags associated with them. They discover that different clusters of bookmarks corresponding to different contexts in which a tag is used can be obtained. However, the paper focuses only on bookmarks associated with two tags (`design` and `politics`) and does not provide a thorough analysis of the method. Mika [19] reveals that the associations between tags are best captured when the social context in which these tags are used are considered. While the studies mentioned above focus on discovering significant associations between tags as a means of revealing the semantics of tags, the differences between tag associations in different contexts are not considered. Most proposed method are unable to tell the differences between the tags associated with, for example, the tag `sf` when it is used in different contexts such as 'San Francisco' and 'science fiction'.

Nevertheless, tag contextualisation can be observed as a by-product in some studies. For example, Wu et al. [35] use latent semantic analysis to study the co-occurrence relations between different tags. Tags are modelled as vectors in a multidimensional space, and ambiguous tags are found to attain high scores in multiple predefined dimensions. In a similar work, Zhou et al. [37] use deterministic annealing to derive hierarchical structures of tags in del.icio.us and Flickr. The authors report that tags with multiple meanings are found to appear in different branches of the resulting hierarchy. While these studies provide valuable insight into how implicit semantics of tags can be extracted from a folksonomy, they do not provide a targeted solution to the problem of tag ambiguity.

Although our work addresses word ambiguity, it is different from studies conducted under the category of word sense disambiguation [14]. While word sense disambiguation aims at identifying and labelling the exact meaning of a word by examining the context in which it is situated, we aim at discovering the different contexts in which an ambiguous tag is

[4] Technorati: http://www.technorati.com/



(a) User-based Document Network (UD) (b) Tag Co-occurrence Network (TC')

Figure 3: Examples of clustering results. (a) shows the user-based document network of sf after clustering. The clusters correspond to either the context of 'San Francisco' or 'science fiction'. In particular, the upper right cluster seems to be related to art studios in San Francisco, while the large cluster towards the bottom is related to food and restaurants in the city. (b) shows the tag co-occurrence network (with user information) of the same tag. The clusters are again found to correspond to either one of the two contexts.

used. We are therefore focusing on the discrimination part of the disambiguation task [29]. Our work is in principle more similar to studies on document clustering techniques which aim at identifying groups of documents corresponding to different meanings of a word [31, 36]. The main difference between our proposed method and these techniques is that we make use of tags assigned by users and the associations between elements in a folksonomy instead of keywords appeared in the documents to perform clustering, and therefore our method is likely to produce more meaningful (clusters.

7. CONCLUSION

We study the problem of tag contextualisation in this paper. We consider several different kinds of network representation of a subset of a folksonomy for the purpose. While the intentions of the users in using a particular tag to describe some documents are already implicitly considered in a tag-document relation, networks which incorporate explicitly the factor of users are found to give more useful results. We also show that automatic clustering techniques are more suitable for the task than the use of external resources.

This work reveals that the social context in which the tags are used should be taken into consideration when attempting to understand the semantics of the tags. Although our discussion focuses on collaborative tagging systems, the findings are actually applicable to the more general area of social applications that enjoy much popularity nowadays. Tags or other forms of text contributed by users can be better understood and categorised if the dynamic user behaviour and interactions are taken into consideration.

In terms of future work, there are two main directions. Firstly, we plan to further investigate how we can refine the results of clustering to eliminate redundancy, such as by measuring similarity between sets of tags, by threshold filtering, or by aggregation of two or more types of networks so as to combine their advantages. We also plan to use other graph clustering algorithms for the task to see if there are more suitable algorithms for the types of networks we have considered. Secondly, we plan to apply this technique of contextualisation to other applications on the Web. For example, online forums and question-answering sites involve users exchanging messages on different topics. We plan to investigate how this can be used to classify messages by identifying the different meanings of the keywords appearing in the messages.

8. ACKNOWLEDGMENTS

Ching-man Au Yeung is supported by the R C Lee Centenary Scholarship (2006-2009) generously offered by the Drs Richard Charles and Esther Yewpick Lee Charitable Foundation, Hong Kong. The authors would like to thank the anonymous reviewers for their invaluable comments and suggestions.

9. REFERENCES

[1] M. Adam. Folksonomies - cooperative classification and communication through shared metadata. *http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html*, 2004.

[2] C. M. Au Yeung, N. Gibbins, and N. Shadbolt. Understanding the semantics of ambiguous tags in folksonomies. In *The International Workshop on Emergent Semantics and Ontology Evolution (ESOE2007) at ISWC/ASWC 2007*, November 2007.

[3] C. M. Au Yeung, N. Gibbins, and N. Shadbolt. A k-nearest-neighbour method for classifying web search results with data in folksonomies. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 70–76, 2008.

[4] C. M. Au Yeung, N. Gibbins, and N. Shadbolt. A study of user profile generation from folksonomies. In *Proceedings of the Workshop on Social Web and Knowledge Management at WWW2008*, 2008.

[5] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. Dbpedia: A nucleus for a web of open data. In *Proceedings of the 6th International Semantic Web Conference (ISWC)*, volume 4825 of *LNCS*, pages 722–735. Springer, 2008.

[6] G. Begelman, P. Keller, and F. Smadja. Automated tag clustering: Improving search and exploration in the tag space. In *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland*, 2006.

[7] C. H. Brooks and N. Montanez. Improved annotation of the blogosphere via autotagging and hierarchical clustering. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 625–632, New York, NY, USA, 2006. ACM Press.

[8] C. Cattuto, A. Baldassarri, V. D. P. Servedio, and V. Loreto. Emergent community structure in social tagging systems. In *Proceedings of the European Conference on Complex Systems*, Dresden, Germany, October 2007.

[9] C. Cattuto, D. Benz, A. Hotho, and G. Stumme. Semantic analysis of tag similarity measures in collaborative tagging systems. In *Proceedings of the 3rd Workshop on Ontology Learning and Population (OLP3)*, Patras, Greece, July 2008.

[10] C. Catutto, C. Schmitz, A. Baldassarri, V. D. P. Servedio, V. Loreto, , A. Hotho, M. Grahl, and G. Stumme. Network properties of folksonomies. *AI Communications Journal, Special Issue on "Network Analysis in Natural Sciences and Engineering"*, 2007.

[11] D. R. Cutting, D. R. Karger, J. O. Pedersen, and J. W. Tukey. Scatter/gather: a cluster-based approach to browsing large document collections. In *Proceedings of the 15th Annual ACM SIGIR Conference*, pages 318–329, New York, NY, USA, 1992. ACM.

[12] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407, 1990.

[13] S. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, April 2006.

[14] N. Ide and J. Véronis. Introduction to the special issue on word sense disambiguation: the state of the art. *Computational Linguistics*, 24(1):1–40, 1998.

[15] G. Koutrika, F. A. Effendi, Z. Gyöngyi, P. Heymann, and H. Garcia-Molina. Combating spam in tagging systems. In *AIRWeb '07: Proceedings of the 3rd international workshop on Adversarial information retrieval on the web*, pages 57–64, New York, NY, USA, 2007. ACM.

[16] R. Lambiotte and M. Ausloos. Collaborative tagging as a tripartite network. In *Proceedings of the International Conference on Computational Science*, page 1114-1117. Springer-Verlag, 2006.

[17] U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[18] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Ht06, tagging paper, taxonomy, flickr, academic article, to read. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 31–40, New York, NY, USA, 2006.

[19] P. Mika. Ontologies are us: A unified model of social networks and semantics. *Web Semantics*, 5(1):5–15, 2007.

[20] G. A. Miller. Wordnet: a lexical database for english. *Commun. ACM*, 38(11):39–41, 1995.

[21] M. E. J. Newman. Analysis of weighted networks. *Physical Review E*, 70:056131, 2004.

[22] M. E. J. Newman. Mathematics of networks. In *The New Palgrave Encyclopedia of Economics, 2nd edition*. Palgrave Macmillan, Basingstoke, 2007.

[23] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69:026113, 2004.

[24] S. Niwa, T. Doi, and S. Honiden. Folksonomy tag organization method based on the tripartite graph analysis. In *IJCAI Workshop on Semantic Web for Collaborative Knowledge Acquisition*, January 2007.

[25] M. G. Noll and C. Meinel. Exploring social annotations for web document classification. In *SAC '08: Proceedings of the 2008 ACM symposium on Applied computing*, pages 2315–2320, New York, NY, USA, 2008. ACM.

[26] E. Quintarelli. Folksonomies: power to the people. ISKO Italy-UniMIB meeting, June 2005.

[27] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks. *Proc. of the National Academy of Sciences of the United States of America*, 101:2658, 2004.

[28] G. Salton and C. Buckley. Term weighting approaches in automatic text retrieval. Technical report, Cornell University, Ithaca, NY, USA, 1987.

[29] H. Schütze. Automatic word sense discrimination. *Comput. Linguist.*, 24(1):97–123, 1998.

[30] K. Shen and L. Wu. Folksonomy as a complex network. *The Computing Research Repository*, September 2005.

[31] D. W. J. Stefanowski. Carrot^2 and language properties in web search results clustering. In *Proceedings of the International Atlantic Web Intelligence Conference*, 2003.

[32] F. Suchanek, M. Vojnovic, and D. Gunawardena. Social tagging: Meaning and suggestions. In *Proceedings of ACM 17th Conference on Information and Knowledge Management*. ACM, 2008.

[33] F. M. Suchanek, G. Kasneci, and G. Weikum. Yago - a core of semantic knowledge. In *Proceedings of the 16th International World Wide Web Conference*, 2007.

[34] C. J. van Rijsbergen. *Information Retrieval*. Butterworth, 2nd edition, 1979.

[35] X. Wu, L. Zhang, and Y. Yu. Exploring social annotations for the semantic web. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 417–426, New York, NY, USA, 2006. ACM Press.

[36] O. Zamir and O. Etzioni. Grouper: a dynamic clustering interface to web search results. In *WWW '99: Proceeding of the eighth international conference on World Wide Web*, pages 1361–1374, New York, NY, USA, 1999. Elsevier North-Holland, Inc.

[37] M. Zhou, S. Bao, X. Wu, and Y. Yu. An unsupervised model for exploring hierarchical semantics from social annotations. In *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007*, volume 4825 of *LNCS*, pages 680–693. Springer, 2007.