Harnessing Folksonomies to Produce a Social Classification of Resources

Arkaitz Zubiaga, Queens College and Graduate Center, City University of New York
Víctor Fresno, NLP&IR Group @ UNED
Raquel Martínez, NLP&IR Group @ UNED
Alberto P. García-Plaza, NLP&IR Group @ UNED

Abstract—In our daily lives, organizing resources like books or web pages into a set of categories to ease future access is a common task. The usual largeness of these collections requires a vast endeavor and an outrageous expense to organize manually. As an approach to effectively produce an automated classification of resources, we consider the immense amounts of annotations provided by users on social tagging systems in the form of bookmarks. In this paper, we deal with the utilization of these user-provided tags to perform a social classification of resources. For this purpose, we have created three large-scale social tagging datasets including tagging data for different types of resources, web pages and books. Those resources are accompanied by categorization data from sound expert-driven taxonomies. We analyze the characteristics of the three social tagging systems, and perform an analysis on the usefulness of social tags to perform a social classification of resources that resembles the classification by experts as much as possible. We analyze 6 different representations using tags, and compare to other data sources by using 3 different settings of SVM classifiers. Finally, we explore combinations of different data sources with tags using classifier committees to best classify the resources.

Index Terms—social-tagging, social annotations, folksonomy, metadata, classification

1 Introduction

Organizing resources into predefined categories is a natural idea in our daily lives. Categorization effectively reduces the amount of resources one has to search. For instance, librarians organize books into subjects, and web directories such as the Open Directory Project organize web pages into categories. The process of manually categorizing resources becomes expensive as the collection of resources grows. For instance, the Library of Congress reported that the average cost of cataloging each bibliographic record by professionals was $94.58 in 2002[1], making more than $27.5 million for the 291,749 records they cataloged that year. Given the expensiveness of this task, switching to automated classifiers seems to be a good alternative to facilitate the task and keep catalogs updated by reducing manual effort.

Until now, most of the automated classifiers rely on the content of the resources, especially regarding web page classification tasks [23]. Nonetheless, the lack of representative data within many resources makes the classification task more complicated. In some cases, it may not be feasible to obtain enough data for certain kinds of resources such as books, where the full text is not available. Without sufficient data, representing the content becomes more challenging.

As a way to solve these issues, social tagging systems provide an easier and cheaper way to obtain metadata related to resources. Social tagging systems are a means to save, organize, and search resources, by annotating them with tags that the user provides. Systems like Delicious[2], LibraryThing[3] and GoodReads[4] collect user annotations in the form of tags on their respective collections of resources. These user-generated tags give rise to meaningful data describing the content of the resources [11], [29], [31]. User annotations can be useful to find out the aboutness of resources, and to help infer the categorization.

By providing tags, users are creating their own categorization system for a given resource. Given that a large number of users are providing their own annotations on each resource, our objective is focused on finding out an approach to amalgamate their contributions in such a way that resembles the categorization by professionals. In this context, our challenge lies in making the most of them in order to enhance resource categorization tasks.

In this work, we explore the social annotations provided by end users on social tagging systems as to performing a social classification of resources. This work focuses on the use of Support Vector Machines as a state-of-the-art classification algorithm. We create three large-scale social tagging datasets including different kinds of resources, web pages and books. We analyze the characteristics of these datasets in order to understand how users tag, and how the nature of a social tagging system can affect the use of social tags to automatically classify resources. We propose 6 different representations of resources using social tags, and analyze the performance of classifying resources using social tags by

1. http://www.loc.gov/loc/lcib/0302/collections.html

2. http://delicious.com
3. http://www.librarything.com
4. http://www.goodreads.com

Digital Object Indentifier 10.1109/TKDE.2012.115

comparing 3 different settings of classifiers. We find that the success of a tag-based representation greatly depends on the settings of the social tagging system, especially on whether or not the system suggests tags to the user when annotating a resource.

This paper is organized as follows. Next, we provide background on social tagging systems in Section 2. Then, we summarize the related work in Section 3. We describe and analyze the datasets used in this work in Section 4. We describe the different representations we use in Section 5, and present the utilized classification algorithms in Section 6. We show the experiments, and analyze the results in Section 7. We conclude in Section 8.

2 Social Tagging

Tagging is an open way to assign tags to resources (e.g., web pages, movies or books), enabling future retrieval in an easier way, by using tags as metadata related to resources. In addition, when a tagging system is social, tags by all the users are publicly accessible, and profitable for the community of users. The collection of tags defined by them creates a tag-based organization, so-called folksonomy. A folksonomy is also known as a community-based taxonomy, where the classification scheme is non-hierarchical, as opposed to a classical taxonomy-based categorization scheme. For instance, LibraryThing, GoodReads and Delicious are social tagging systems, where each resource can be tagged by all the users who consider it interesting.

Social tagging systems allow users to annotate resources that others have previously annotated. This enables the aggregation of annotations provided by many users on the same resource. Table 1 shows an example of tags defined by different users for the photo sharing service *Flickr* (note that a user might not add any tags, leaving the field empty, as user 5 did in the example). From tags defined by all the users (hereafter, Full Tagging Activity or FTA), a weighted list of top tags can be generated, as shown in Table 2.

User annotations: Flickr.com	
User 1:	photo, photography, images, pictures
User 2:	photo, web2.0, social, tools, blog
User 3:	cloud, pictures, sharing
User 4:	flickr, photos
User 5:	

TABLE 1
Example of annotations for the URL Flickr.com on the social bookmarking site Delicious.

In these social tagging systems, there is a set of users (U), who are posting bookmarks (B) for resources (R) annotated by tags (T). Each user $u_i \in U$ can post a bookmark $b_{ij} \in B$ of a resource $r_j \in R$ with a set of tags $T_{ij} = \{t_1, ..., t_p\}$, with a variable number p of tags. After k users posted r_j, it is described with a weighted set of tags $T_j = \{w_1 t_1, ..., w_n t_n\}$, where $w_1, ..., w_n \leq k$ represent the number of assignments of a specific tag. Accordingly, each bookmark is a triple of a user, a resource, and a set of tags: $b_{ij} : u_i \times r_j \times T_{ij}$. Thus, each user saves bookmarks of different resources, and a resource has bookmarks made by different users. The result of aggregating tags within bookmarks by a user is known as the personomy of the user: $T_i = \{w_{i1} t_{i1}, ..., w_{im} t_{im}\}$, where m is the number of different tags utilized by user u_i.

Top tags: Flickr.com (79,681 users)	
photos	22,712
flickr	19,046
photography	15,968
photo	15,225
sharing	10,648
images	9,637
web2.0	9,528
community	4,571
social	3,798
pictures	3,115

TABLE 2
Example of top tags for the URL Flickr.com on the social bookmarking site Delicious: the number associated to each tag represents the number of users annotating it.

3 Related Work

Annotations provided by users on social tagging systems have been widely deployed by researchers as metadata related to resources [9] for tasks such as (i) Information Retrieval [11], (ii) Recommender Systems [30], (iii) discovery of emergent semantics [18], and (iv) enhanced browsing and navigation through annotated resources [36], among others.

It has also attracted researchers to exploit annotations to enhance classification tasks. In an early work in the field, Noll and Meinel [20] presented a study of the characteristics of social annotations provided by end users, in order to determine their usefulness for web page classification. In this work, the authors weight the tags by normalizing the number of users annotating them. The least popular tag is given a value of 0, whereas the most popular is given a value of 1. They did not pay attention at whether or not this representation approach was appropriate to carry out the task. The authors matched user-supplied tags of a page against its categorization by the expert editors of the Open Directory Project (ODP), even though they did not perform actual classification experiments. They observed that in the power law curve formed by the popularity of social tags, not only popular tags, but also the tags in the tail provide helpful data for information retrieval and classification tasks in general. In a previous work, the same authors [19] suggested that tags provide additional information about a web page, which is not directly contained within its content. Also,

Noll and Meinel [21] studied three types of metadata about web documents: social annotations (tags), anchor texts of incoming hyperlinks, and search queries to access them. They concluded that tags are better suited for classification purposes than anchor texts or search keywords.

Aliakbary et al. [3] integrated social annotations as an approach to extending web directories. They relied on the number of users using each tag as a weight. Upon that, they created a model for each category, and computed the cosine similarity for new web pages to generate predictions. They observed that the annotations provided a multi-faceted summary of the web pages, and that they represent the aboutness of web pages better than the content itself. Also, they conclude that the more users annotate a URL, the better it is classified. Xia et al. [33] studied the usefulness of social tags as a complementary source for improving the classification of academic conferences into topics. Using tagging data gathered from WikiCFP.com, and weighting the tags according to the number of users using them, they compare the classification of conferences by using only the content of the call for papers, and by integrating tagging data along with it.

With regard to the classification of resources other than web pages, Lu et al. [17] present a comparison of tagged books and their Library of Congress subject headings. Actually, no classification experiments are performed, but a statistical analysis of the tagging data shows encouraging results. By means of a shallow analysis of the distribution of tags across the subject headings, they conclude that user-generated tags seem to provide an opportunity for libraries to enhance access to resources. Using a graph-based approach, Yin et al. [34] present a method to classify products from Amazon into their corresponding categories using social tags. They conclude that social tags can enhance web products classification by representing them in a meaningful feature space, interconnecting them to indicate relationship, and bridging heterogeneous products so that category information can be propagated from one domain to another.

In a preliminary work [37], we studied different representations of social tags as to web page classification using a native multiclass classifier. We found that not only the top tags on each web page, but also tags in the tail are helpful to improve the classification performance. In a later work where social tags were exploited for the benefit of web page classification, Godoy and Amandi [7] also showed the usefulness of social tags for web page classification, which outperformed classifiers based on full-text of documents. They also compare tag-based resource representations relying on all the tags and the top tags for each resource. Their results corroborate our findings that taking into account all the tags yields better performance. Going further, they concluded that stemming the tags reduces the performance of such classification, even though some operations such as removal of symbols, compound words and reduction of morphological variants have a discrete positive impact on the task.

To the best of our knowledge, the study of classifying more than one type of resource, and studying in depth the appropriateness of different representations further considering the structure and distributions of folksonomies, are still unexplored. In this work, we further study the usefulness of social tags by (i) exploring social tagging datasets including different kinds of resources, (ii) analyzing the characteristics of large-scale social tagging datasets, and (iii) comparing 3 different settings of classifiers.

4 Datasets

This section introduces, and analyzes the three social tagging datasets we created. First, we summarize the main characteristics of the social tagging datasets we chose. Second, we describe in detail the process of generation of the datasets. Third and last, we analyze the generated datasets, and present a set of statistics and findings that help better understand both the problem and the results of the classification experiments.

4.1 Characteristics of the Selected Social Tagging Systems

For this work, we selected 3 social tagging systems: Delicious, LibraryThing, and GoodReads. These systems fulfill the following 3 requirements: (i) they have a large community of users involved, (ii) they provide full access to the triple involved in each bookmark, i.e., the user annotating it, the resource being annotated, and the tags, and (iii) the annotated resources are classified on consolidated taxonomies by experts. These 3 conditions allow us to further analyze user annotations, and their distribution across resources, users, and bookmarks.

Even though all these tagging systems have the same end of enabling users to bookmark and annotate the resources of their interest, there are several features that make each of them different from the rest. The design of the interface, constraints on the inputted tags, and other features could influence annotations given by users.

Delicious is a social tagging system that allows users to save and tag their favorite web pages, in order to ease future navigation and retrieval. Users can save any web page, so that the range of covered topics can become as wide as the Web is. However, it is known that the site is biased to some computer and design related topics. Tagging web pages is one of the main features of the site, and that is the first thing the system asks for when a user saves a URL as a bookmark. The system suggests tags used earlier for that URL when some users had used it before. Thus, new annotators can easily select tags used by earlier users without typing them. This could encourage users to reuse tags by others, reducing the number of new tags assigned to a resource.

LibraryThing and **GoodReads** are social tagging systems where users save and annotate books. Users tend

to annotate the books they own, they have read, or they are planning to read. We believe that users contributing to this kind of sites are more knowledgeable of the resources than those contributing to social bookmarking systems. Moreover, there are also writers and libraries contributing as users, who have a deep background on the field. This could yield annotations providing more detailed knowledge. The main difference among these two systems is that LibraryThing does not suggest tags when saving a book, whereas GoodReads lets the user select from tags within their personomy, i.e., tags they previously assigned to other books. The latter makes it easier to reuse tags of the user, without re-typing them. This could encourage users to keep a smaller tag vocabulary, where they barely use new tags they did not use previously. Moreover, LibraryThing brings the user to a new page when saving a book, where they can attach tags to it; GoodReads, though, requires the user to click again on the saved book to open the form to add tags. Aiello et al. [2] presented an study on a similar social tagging system for books, Anobii.

Despite the aforementioned differences, all of them have some characteristics in common: users save resources as bookmarks, a bookmark can be annotated with a variable number of tags ranging from zero to unlimited, and the vocabulary of the tags is open and unrestricted. Table 3 summarizes the main features of the three social tagging sites we study in this paper.

4.2 Generation Process of Datasets

For the generation of three large-scale datasets, we first looked for suitable resources to perform the task, and gathered several metadata afterward: (i) classification data on consolidated taxonomies, (ii) tagging data, and (iii) other data related to the resources.

4.2.1 Getting Popular Resources

As a starting point, we focused on getting a set of popular annotated resources from each site. This provided an initial list of popular resources that represented a good seed to start the gathering process from. Those resources were also more likely to be categorized by experts rather than resources in the tail with fewer annotations. We could also have started the process by looking for popular tags or active users, but starting from resources sounds reasonable when those are the goal to be classified. Next, we will focus on the process of gathering the data in such a way that those resources are well represented insofar as involved users and their annotations are taken into account. Apart from representing those resources, we were also interested in gathering additional data, in order to represent involved users and tags to a great extent.

First of all we queried the three social tagging sites for popular resources. We consider a resource to be popular if at least 100 users have bookmarked it[5]. In the case of Delicious, we found a set of 87,096 unique URLs fulfilling this requirement. As regards to LibraryThing and GoodReads, we found an intersection of 65,929 popular books. Since the latter two rely on the same kind of resource, we created parallel datasets for them, i.e., including the same books.

4.2.2 Looking for Classification Data

In the next step, we looked for classification labels assigned by experts for both kinds of resources. For the URLs gathered from Delicious, we used the Open Directory Project[6] (ODP) as a classification scheme. ODP is an open web directory, constructed and maintained by a community of volunteer editors, and it includes categorization data on a hierarchical structure for more than 4 million URLs. A matching between popular URLs on Delicious and those in the ODP returned a set of 12,616 URLs with a category assigned. The ODP is made up by 17 categories. For the set of books, we fetched their classification for the Library of Congress Classification (LCC) system. We found that 24,861 books had an LCC category assigned. The LCC comprises 20 categories. For both taxonomies, we rely on the top level of the hierarchy. We kept the structure of the taxonomies as they were, but made a little change for LCC: we merged E (*History of America*) and F (*History of the United States and British, Dutch, French, and Latin America*) categories into a single one, as it is not clear that they are disjoint categories.

4.2.3 Gathering Tagging Data

Finally, we queried (a) Delicious for gathering all the personomies involved in the set of categorized URLs, and (b) LibraryThing and GoodReads for gathering all the personomies involved in the set of categorized books. By personomy, we consider the whole list of bookmarks posted by a user, including an identifier of the resources and the tags attached by them. All three sites present no restrictions on the bookmarks shown in personomies, so that they return all available public bookmarks for the queried users.

Within the gathered data, we focus on the following information for each bookmark: an identifier of the **User (U)**, an identifier of the **Resource (R)**, and a set of **Tags (T)**. That is, the triple of $U \times R \times T$ involved in a bookmark. We consider all the tags attached to each bookmark, except for GoodReads. In this case, a tag is automatically attached to each bookmark depending on the reading state of the book: `read`, `currently-reading` or `to-read`. As this is not part of the tagging process, but just an automated step that does not provide useful information for classification, we removed all their appearances in our dataset.

5. It was shown that the tag set of a resource tends to converge when 100 users contribute to it [8]. Thereby we consider it as a threshold for a resource to be popular.

6. http://www.dmoz.org

	Delicious	LibraryThing	GoodReads
Resources	web documents	books	books
Tag suggestions	based on earlier bookmarks on the resource	no	based on user's personomy
Users	general	readers, writers & libraries	readers, writers & libraries
Vocabulary	open	open	open
Tag insertion	space-separated	comma-separated	one by one text-box
When saving a resource	prompts user to add tags	prompts user to add tags at second step	user needs to click again to add tags

TABLE 3
Characteristics of the studied social tagging systems.

4.2.4 Gathering Additional Data

Besides the aforementioned tagging data, we gathered some more data about the categorized resources. These enable to compare other data sources to tagging data as to the resource classification task.

On one hand, we got the following data for the categorized URLs:

- **Self-content:** it is the content of the web page itself, i.e., the HTML code fetched from the original URL.
- **User reviews:** a review can be defined as a free text commenting on the content of a web page. It can be an objective description of the content of the web page, or it can be the user's opinion on the web page, providing a subjective bias. We gathered user reviews for web pages from Delicious and StumbleUpon[7].

On the other hand, with regard to the categorized books, there is no easy way to get the content of the book. We did not have access to the content of the books, since most of them are not freely available. Thus, we got the following metadata associated to the books:

- **Synopses:** a synopsis is a brief summary of the content of a book, which is usually printed on the back cover. We fetched synopses from the book retailer Barnes&Noble[8].
- **Editorial reviews:** summaries written by the publisher or other professionals are considered as editorial reviews. We gathered them from Amazon[9].
- **User reviews:** we also collected reviews written by users on LibraryThing, GoodReads and Amazon. With reviews, users comment on the books providing their summaries and thoughts.

As we do not have access to self-content of the books, we will consider both synopses and editorial reviews as a summary of their contents.

4.3 Statistics and Analysis of the Datasets

To understand the nature and characteristics of each dataset, and to analyze how the settings of each social tagging system affect the folksonomies, we present and study statistics of the datasets.

Note that attaching tags to a bookmark is optional, so depending on the social tagging site, a number of bookmarks may remain without tags. Table 4 presents the number of users, bookmarks and resources we gathered for each of the datasets, as well as the percent with attached annotations. In this work, as we rely on tagging data, we only consider annotated data, ruling out bookmarks without tags. Thus, from now on, all the results and statistics presented are based on annotated bookmarks. From these statistics, it stands out that most users (above 87%) provide tags for bookmarks on Delicious, whereas there are fewer users who tend to assign tags to resources on LibraryThing and GoodReads (roughly 38% and 17%, respectively). This shows the importance of Delicious' encouragement to adding tags, and GoodReads' disencouragement to this end, requiring the user to click twice on the book in order to add tags. The latter makes the tagging process cumbersome, and yields a large number of untagged bookmarks. LibraryThing is halfway between those two, which automatically conveys the user to the tagging form, but at a skippable second step after saving the book.

Of the resources collected for the datasets, not all of them have categorization data provided by experts. Table 5 shows the statistics on the number of categorized and uncategorized resources, according to the categorization data we gathered from expert-driven taxonomies. It can be seen that the subset categorized by experts is small as compared to the whole set. This enables to analyze a larger folksonomy as a whole for finding out tagging patterns on each site, as well as encourages the study so we are able to classify many more resources using social tags.

A first glance at the vocabulary employed in each folksonomy can be given by looking at the top tags on each site. The top 10 of tags set by users for each of the datasets is listed in Table 6. On one hand, top tags on Delicious include tags like `design`, `software` and `blog`, showing its computer and design related bias. On the other hand, top tags on LibraryThing and GoodReads share some similarities, where tags related

7. http://www.stumbleupon.com
8. http://www.barnesandnoble.com
9. http://www.amazon.com

Delicious			
	Annotations	Total	Ratio
Users	1,618,635	1,855,792	87.22%
Bookmarks	273,478,137	300,571,231	91.00%
Resources	92,432,071	102,828,761	89.89%
Distinct Tags		11,541,977	-
LibraryThing			
	Annotations	Total	Ratio
Users	153,606	400,336	38.37%
Bookmarks	22,343,427	44,612,784	50.08%
Resources	3,776,320	5,002,790	75.48%
Distinct Tags		2,140,734	-
GoodReads			
	Annotations	Total	Ratio
Users	110,344	649,689	16.98%
Bookmarks	9,323,539	47,302,861	19.71%
Resources	1,101,067	1,890,443	58.24%
Distinct Tags		179,429	-

TABLE 4
Statistics on availability of tags in users, bookmarks, and resources for the three datasets.

Resources			
	Categ.	Uncateg.	Ratio
Delicious (ODP)	12,616	92,419,455	0.014%
LibraryThing (LCC)	24,861	3,751,459	0.636%
GoodReads (LCC)	24,861	1,076,206	2.310%

TABLE 5
Ratio of categorized and uncategorized resources. The ratio value represents the percent of categorized bookmarks as compared to the uncategorized ones.

to literary genres stand out.

Delicious	**LibraryThing**	**GoodReads**
design	fiction	fiction
blog	non-fiction	fantasy
tools	fantasy	non-fiction
software	history	own
webdesign	mystery	young-adult
web	science fiction	classics
reference	read	mystery
programming	biography	romance
music	poetry	wishlist
web2.0	novel	nonfiction

TABLE 6
Top 10 most popular tags on the datasets.

Regarding the distribution of tags across all the resources, users and bookmarks in the datasets, there is a clear difference of behavior among the three collections. Figure 1 shows the usage percents of tags, ordered by their usage rank (note the logarithmic scale). The 3 lines represent the usage of tags by users, on resources, or on bookmarks. The X axis refers to the percent of the tag rank, whereas the Y axis represents the percent of appearances in resources, users and bookmarks. For instance, if the tag ranked first had been used on the half of the resources, the value for the top ranked tag on resources would be 50%. Thus, these graphs enable to analyze how popular are the tags in the top as compared to the tags in the tail on each site. On the other hand, Figure 2 shows the average usage of tags in a given rank for resources for each dataset. That is, we give a value of 1 to the tag used the most on a resource, hence ranked first for that resource. The second tag is given the value according to the fraction of users utilizing it as compared to the first one. And so on for tags ranked third, fourth,... on resources. Finally, we compute the average of tags ranked on each position, which is shown in the graph. It helps infer the popularity gap between top tags on resources and tags ranked lower. Looking at those two figures together, it stands out that GoodReads has the highest usage of tags in the tail, while Delicious presents the highest usage of tags in the top. Delicious is the site with highest diversity of tags, where a few tags become really popular (both in the whole collection and on resources), and many tags are seldom-used. We believe that the reasons for these differences on tag distributions are:

- Since Delicious suggests tags that have been utilized by previous users to a resource, it is obvious that those tags on the top are likely to happen more frequently, whereas others may barely be used.
- LibraryThing and GoodReads do not suggest tags used by earlier users and, therefore, tags other than those in the top tend to be used more frequently than on Delicious.
- GoodReads suggests tags from previous bookmarks of the same user, instead of tags that others assigned to the resource being tagged. Thus, this encourages reusing tags in their personomy, making it remain with a smaller number of tags (see Table 7). In addition, users tend to assign fewer tags to a bookmark on average, leveraged by the one-by-one tag insertion method of site's interface.

# of tags	**Delicious**	**LibraryThing**	**GoodReads**
Per resource	33.35	14.53	13.33
Per user	632.714	357.15	131.03
Per bookmark	3.75	2.46	1.55

TABLE 7
Average counts of different tags.

Next, we analyze the number of tags that are used more, equal or less frequently in an item (i.e., resources, users or bookmarks) than in another (see Figure 3). By definition, a tag cannot appear in a smaller number of bookmarks than users or resources. Looking at the rest of data, it stands out that tags tend to appear in more bookmarks than users ($b > u$) and more resources than users ($r > u$) for GoodReads, due to the same feature that allows users to select among tags in their personomy. However, LibraryThing and Delicious have many tags

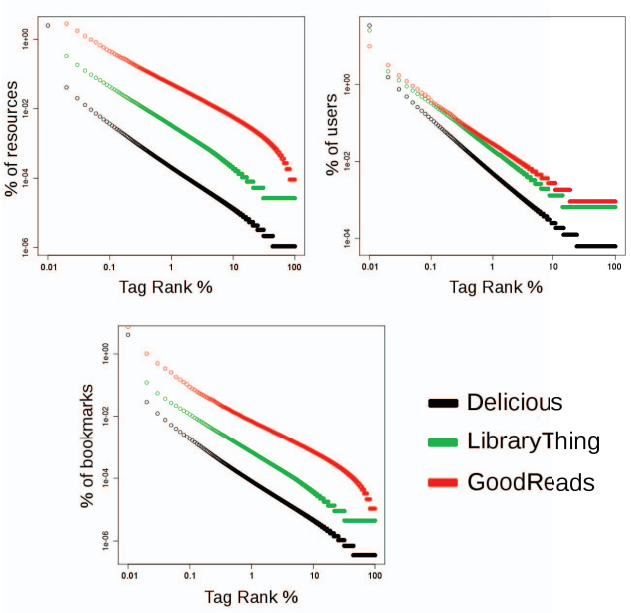

Fig. 1. Tag usage percentages in the collection. These 3 graphs represent, on a logarithmic scale for both *x* and *y* axes, the percent of annotations to resources, users, and bookmarks per tag rank.

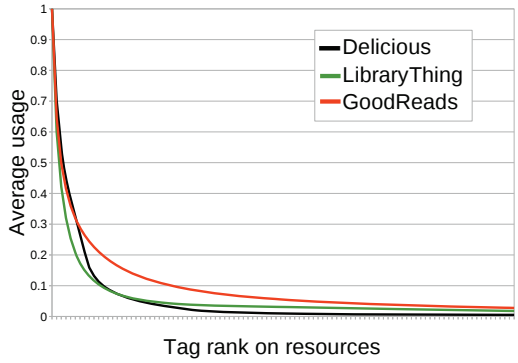

Fig. 2. Tag usage percentages on resources. Each tag rank represents the average usage of tags appearing in that position on resources as compared to the top ranked tag.

present in the same number of bookmarks and users ($b = u$), and resources and users ($r = u$), even though the difference is more marked for LibraryThing. This reflects the large number of tags that users utilize just once on these sites. All three sites have two features in common: there are a few exceptions of tags utilized by more users than the number of resources it appears in ($r < u$), and almost all the tags are present in the same number of bookmarks and resources ($b = r$). The latter, combined with the lower ($b = u$) values, means there is a large number of users spreading personal tags across resources that only have a bookmark with that tag, especially on GoodReads, but also for the other two sites.

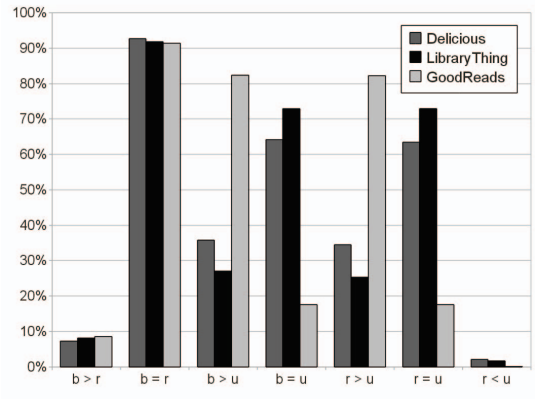

Fig. 3. Tag distribution across resources (r), users (u) and bookmarks (b). Each bar represents the percent of tags that match the condition on X axis.

Finally, we analyze the extent to which a bookmark introduces new tags into a resource that were not present in earlier bookmarks. Figure 4 shows these statistics for Delicious and LibraryThing. The same graph for GoodReads is not shown because neither the timestamp nor the ordering of the bookmarks is available in our dataset. The graph shows, on average, the ratio of new tags, not present in earlier bookmarks of a resource, assigned in bookmarks that rank from first to 100th bookmark, i.e., if $\texttt{tag}_1$ and $\texttt{tag}_2$ were utilized in the first bookmark of a resource, and $\texttt{tag}_2$ and $\texttt{tag}_3$ in the second bookmark for the same resource, the ratio of novelty for the second bookmark is of 50%. It stands out the marked inferiority of tag novelty on Delicious as against to LibraryThing. This is, again, due to the tag suggestion policy of Delicious, what brings about a higher likelihood of reusing previously existing tags.

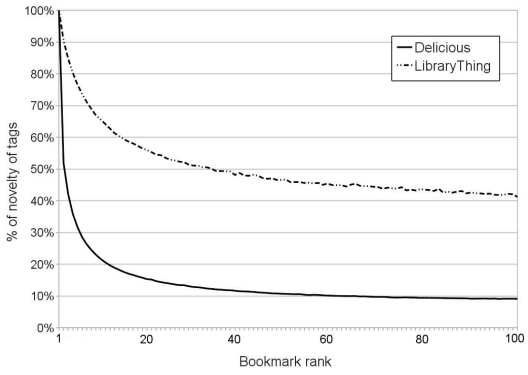

Fig. 4. Novelty ratio of tags per rank of bookmark.

5 Representation of Resources

We use different data sources to represent resources to classify. We focus especially on the representation using social tags, but also analyze and compare to other data sources: content, and user reviews.

5.1 Representing Resources Using Tags

The tagging activity of a community of users creates an aggregated list of tags on each resource. A resource annotated by p users will have a list of n different tags, where each tag could have been utilized by at least 1 user and p users at most. The number of users who utilized a certain tag–w_t–defines a value that allows to infer an ordered list of tags for a resource. Given that in this work we rely on the Vector Space Model to represent resources, this aggregation of annotations performed by different users could be represented in several ways, especially when it comes to assigning weights to tags.

On one hand, different data can be considered or ignored for the weighting of tags. For instance, the number of users utilizing a tag–w_t–can be considered for the weighting, or it can be ignored by just considering the absence or presence of a tag. The total number of users annotating the resource–p–can also be considered for the weighting, or it can be ignored not to depend on the popularity of the resource. On the other hand, when representing a resource, one could only consider the values within the resources for the weightings, or additional data from the rest of the collection could also be used to this end.

Next, we present 6 representations using social tags, organized by type of weighting. First, we present 3 local weightings, which only consider data from the resource itself, and then we present 3 global weightings, which take into account information from the whole collection.

5.1.1 Local Tag Weighting

We present 3 representations that rely on the annotations on the resource being annotated, and do not consider information from other resources in the collection.

- **Fraction-based Tag Weighting:** the weight is computed according to the fraction of users who utilize a tag, w_t/p, i.e., the number of users utilizing a tag on a resource, divided by the total number of users who annotated the resource. Taking into account both the number of users who bookmarked a resource r and the weight of each tag w_t, it is possible to define the fraction of users assigning each tag. A tag would have been utilized by the totality of the users when its weight matches the user count of a resource, getting a value of 1 as the fraction. According to this, each tag is set a value between 0 and 1. This representation approach is similar to that by Noll and Meinel [20] for their analysis of the similarity between social tags and the classification by experts. However, they ignore the least popular tags, what may give rise to the removal of useful tags from the representation.
- **Binary Tag Weighting:** in a binary way, the presence of a tag represents a value of 1, and its absence a value of 0. The only feature considered for this representation is the occurrence or non-occurrence of a tag in the annotations of a resource. This approach thereby ignores the weights of tags, and assigns a binary value to each feature in the vector.
- **Frequency-based Tag Weighting (TF):** it considers the number of users assigning the tag (w_t) as a weight. The weight for each of the tags of a resource ($w_1, ..., w_n$) is considered as it is in this approach. Now, by definition, the weights of the tags are fully respected, although the amount of users bookmarking a resource is ignored. Note that different orders of magnitude are mixed up now, since the count of bookmarking users range within very different values. For instance, Ramage et al. [24] used this approach in their work for clustering web pages, but they assumed it without comparing it to other representations.

5.1.2 Global Tag Weighting

TF-IDF (Term Frequency-Inverse Document Frequency) is an inverse weighting function, which has been widely applied to text collections, and has proven to be beneficial for a large number of tasks. TF-IDF is a term weighting function that serves as a statistical measure that defines the importance of a word to a document in a collection [26]. When computing the TF-IDF value for the term i within the document j as a part of a document collection D, it comprises two underlying measures: (1) the term frequency (TF), i.e., the number of appearances of the term i within the document j, and (2) the inverse document frequency (IDF), i.e., the logarithm of the number of documents in the whole set (D) divided by the number of documents in which the term i occurs, which refers to the general importance of the term i in the collection (see Equation 1). The product of these two measures defines the TF-IDF weight of term i in the document j (see Equation 2).

$$idf_i = \log \frac{|D|}{|\{d : t_i \in d\}|} \tag{1}$$

$$tf\text{-}idf_{ij} = tf_{ij} \times idf_i \tag{2}$$

Integrating the IDF factor allows to rate lower or higher such a term depending on its distribution across the collection. This weighting function yields a higher value when the term i occurs in a few documents, considering that it is of utmost representativity to those documents. On the other hand, the value will be lower when the term i occurs in many documents of the collection, considering that it rather spreads across the collection instead of focusing in a few documents. In the latter case, the value becomes null when the term i occurs in all the documents.

Unlike classical collections of web pages or library catalogs made up by texts, social tagging systems comprise more dimensions to explore into. Besides the distribution of tags across documents or annotated resources, different users set those tags within different bookmarks. These two characteristics are new on social tagging with respect to classical text document collections.

Next, we propose 3 tag weighting approaches, taking the TF-IDF approach to the social tagging scenario, and adapting it to rely on resources, users and bookmarks. These three dimensions suggest the definition of that many tag weighting functions considering inverse resource frequency (IRF), inverse user frequency (IUF), and inverse bookmark frequency (IBF) values, respectively. These three approaches follow the same function for the tag i within the resource j (see Equation 3).

$$TF\text{-}IxF_{ij} = tf_{ij} \cdot ixf \tag{3}$$

where tf_{ij} is the number of occurrences of the tag i in the resource j, and ixf is the inverse frequency function considered in each case, irf, iuf or ibf, thus x being r, u, or b.

Based on the TF-IDF weighting scheme, and thus using information from other resources in the collection, we propose the following 3 representations:

- **TF-IRF Tag Weighting:** This is the application of the TF-IDF approach to a social tagging system with annotated resources, considering that resources are analogous to documents in this case. Tags that are widely spread across resources are penalized with low weights and, vice versa, tags within fewer resources are considered relevant with a higher weight. Thus, the function outputs the logarithm of the total number of resources divided by the number of resources in which the tag is present (see Equation 4).

$$irf_i = \log \frac{|R|}{|\{r : t_i \in R\}|} \tag{4}$$

  It has previously been used in a few works in the social tagging literature, even though they usually referred to this approach as TF-IDF. Angelova et al. [4] rely on this measure to infer similarity of tags by creating a tag graph, weighting the TF-IDF value of each user to a tag. Shepitsen et al. [27] and Liang et al. [16] use this measure to represent resources in a recommendation system where resources are recommended to users. The latter concluded that although both TF-IDF and TF have identical trends, the former provides superior results in their recommendation task. Likewise, Ramage et al. [24] compared TF-IDF and TF for clustering web pages, and showed a superiority for the former. However, they did not pay attention at the effect of tag distributions on these weightings, and they showed the usefulness of TF-IDF just for a specific case. Li et al. [15] create tag vectors using TF-IDF to compute the similarity between two documents annotated on Delicious. They assumed this weighting measure, and they did not pay attention at whether or not it was appropriate.
- **TF-IUF Tag Weighting:** As a new dimension present in social tagging systems, the number of users using each of the tags could also be significant to know whether a tag is representative within a collection of resources. Thus, we consider that a tag used by many users is not as representative as a tag that fewer users are utilizing (see Equation 5).

$$iuf_i = \log \frac{|U|}{|\{u : t_i \in U\}|} \tag{5}$$

  This function was inferred from a previous application to a collaborative filtering system [5]. With the aim of recommending resources to users, Diederich and Iofciu [6] and Liang et al. [16] rely on the IUF for discovering similarities among users. The latter use both IUF and IRF to represent users and resources, respectively, but no comparison is performed among their characteristics. Abbasi et al. [1] use TF-IUF along with TF-IRF over Flickr tags and user groups for finding landmark photos. They concluded that their approach was effective to find landmark photos on Flickr, but they did not study whether or not relying on those weighting measures was appropriate.
- **TF-IBF Tag Weighting:** This is a similar inverse weighting function relying on the third dimension in which tags are distributed: bookmarks. This function considers that a tag that has been used in many bookmarks is not as relevant to represent a resource as others that have been assigned to fewer bookmarks (see Equation 6).

$$ibf_i = \log \frac{|B|}{|\{b : t_i \in B\}|} \tag{6}$$

  To the best of our knowledge, this tag weighting scheme has never been used so far.

Even though all three frequencies can somehow be related, there are substantial differences among them. A tag used by many users can spread across many resources, or it can just congregate in a few resources. Likewise, this might affect the number of bookmarks.

5.2 Representing Resources Using Other Data Sources

As data sources to compare performance to that by social tags, we rely on two types of data: content and user reviews. Regarding the collection of web pages, we strip HTML tags from their source in order to get the textual content. We also utilize user reviews about the web pages fetched from social media sites. In the case of books, since we do not have access to the full content of books, we consider synopses and editorial reviews as a summary of their content. On the other hand, we utilize user-generated reviews gathered from social media sites (see Section 4.2.4).

In order to get a vectorial representation of resources from content and user reviews, we rely on the bag-of-words model [10]. To produce the bag-of-words corresponding to a resource, we follow the same process both for content and for user reviews. In the case of user

reviews, we first merge them into a single text. After that, we clean up both content and user reviews. In order to clean up those texts, we first stripped HTML tags to get plain texts. Afterward, we remove stopwords contained in texts, and stem the remaining words using the Porter algorithm [22]. Then, we weight the words according to values given by the TF-IDF function.

6 Classification Algorithms

As a state-of-the-art classification algorithm, we rely on Support Vector Machines (SVM) [13]. This algorithm looks for a hyperplance that separates the classes in a vector space model; this hyperplane should maximize the distance between it and the nearest resources, which is called the margin. Basically, an SVM looks for the optimal hyperplane that minimizes the outcome of Equation 7.

$$\min \left[\frac{1}{2}||\mathbf{w}||^2 + C \sum_{i=1}^{l} \xi_i^d \right] \tag{7}$$

Subject to: $y_i(\mathbf{w} \cdot \mathbf{x_i} + b) \geq 1 - \xi_i,\ \xi_i \geq 0$

where *C* is the penalty parameter, ξ_i is an stack variable for the i^{th} resource, l is the number of labeled resources in the training set, and *d* is the sigma parameter that defines the non-linear mapping from the input space to some high-dimensional feature space.

Upon this, several settings can be used in a SVM [12], [14]. Even though SVM only solves binary classification problem by default, different approaches have been proposed to work with multiclass problems. In this work, we use the 3 most popular settings for supervised multiclass SVMs: (i) a native multiclass approach, (ii) one-against-all binary SVMs, (iii) and one-against-one binary SVMs. We set them up to work with a linear kernel and the default parameters.

6.1 Multiclass SVM (mSVM)

As a native multiclass approach, we use the approach by Weston and Watkins [32], which modifies the optimization function getting into account all the *k* classes at once (see Equation 8).

$$\min \left[\frac{1}{2} \sum_{m=1}^{k} ||\mathbf{w_m}||^2 + C \sum_{i=1}^{l} \sum_{m \neq y_i} \xi_i^m \right] \tag{8}$$

Subject to: $\mathbf{w_{y_i}} \cdot \mathbf{x_i} + b_{y_i} \geq \mathbf{w_m} \cdot \mathbf{x_i} + b_m + 2 - \xi_i^m, \xi_i^m \geq 0$

This native multiclass approach considers the task with a single classifier, and thus it learns a model for all the classes at the same time.

The native multiclass approach we use in our experiments has been implemented by using *svm-multiclass*[10], a multiclass SVM classifier by Joachims [13].

6.2 One-Against-All SVM (oaaSVM)

One-against-all is a method that combines binary SVM classifiers [25]. Instead of considering the multiclass problem as a single task, *one-against-all* splits it into smaller binary ones. Specifically, it constructs a binary classifier for each of the classes, where the considered class is the positive case, and the rest correspond to the negative case. For a problem with k classes, *one-against-all* defines k different classifiers. In the training phase, each of the k classifiers learns a model to separate a class from the rest $k-1$. This model creates a hyperplane to separate the class from the rest. For instance, for a task with 4 classes, the following classifiers would be created: *1 vs 2-3-4*, *2 vs 1-3-4*, *3 vs 1-2-4* and *4 vs 1-2-3*. In the process of categorizing, each classifier provides an output for each resource, which refers to the margin–i.e., distance to the hyperplane–as a reliability value. Each resource j is predicted as a negative or positive case by each classifier with a reliability value. Thus, having a reliability value for each of the classes, the classifier maximizing the output defines the final class predicted by the system (see Equation 9).

$$\hat{C}_j = \underset{i=1,\ldots,k}{\operatorname{argmax}} \{m_{ij}\} \tag{9}$$

where $\hat{C}_j$ is the class predicted by the classifier for the resource j, and m_{ij} is the margin outputted by the classifier i for the resource j.

We implemented it using a binary SVM classifier by Joachims [13] available for research purposes: *svm-light*[11].

6.3 One-Against-One SVM (oaoSVM)

One-against-one is another method that combines binary SVM classifiers. Specifically, it constructs all possible pairwise discriminating classifiers. This way, it allows to compute, between each pair of classes, which class is more likely to belong to the considered resource. Thus, it creates $n = \frac{k(k-1)}{2}$ pairwise classifiers when it comes to a problem with k classes. For instance, for a problem with 4 classes, the following classifiers would be created: *1 vs 2*, *1 vs 3*, *1 vs 4*, *2 vs 3*, *2 vs 4* and *3 vs 4*. After that, it classifies each new resource by using all the classifiers, where a vote is added for the winning class over each classifier (see Equation 10); the method will propose the class with more votes as the result. Combining the outputs of classifiers, the class winning more frequently is predicted by the final classifier (see Equation 11). The main problem of this approach is the exponential growth of the number of classifiers as the number of classes increases. In our experiments, where 20 classes are considered at most, it comes up to 190 classifiers.

$$\hat{C}_j = \underset{i=a,b}{\operatorname{argmax}} \{m_{ij}\} \rightarrow V_{\hat{C}_j j} = V_{\hat{C}_j j} + 1 \tag{10}$$

where $\hat{C}_j$ is the class maximizing the output for the

10. http://svmlight.joachims.org/svm_multiclass.html

11. http://svmlight.joachims.org

pairwise *a vs b* classifier for the resource j. A vote is added on $V_{\hat{C}_j j}$ for the predicted class.

$$\hat{C}_j = \underset{i=1,\ldots,n}{\operatorname{argmax}}\{V_{ij}\} \quad (11)$$

We implemented it using the same binary SVM classifier as above: *svm-light*.

7 Experiments

We use the 3 SVM settings and the representations introduced above for the experiments. Regarding the selection of training and test sets, we use randomly selected training sets of 6,000 web pages for Delicious (6,616 for the test set), and 15,000 books for the book datasets (9,861 for the test set). In order to make the results more independent of the specific selections of training sets, we perform 10 different random selections. We run 10 different SVMs with those pairs of training and test sets, and get the average of all the runs. Since both LibraryThing and GoodReads datasets include the same books, the content and user reviews are the same, and so are the results we show for those. The results for tags are different, as they are based on tags from each site.

7.1 Classification Using Tags and Other Data Sources

Table 8 shows the results of the resource classification using different data sources and representations. Specifically, Table 8(a) shows the results for Delicious, Table 8(b) for LibraryThing, and Table 8(c) for GoodReads.

(a) Delicious

		mSVM	oaaSVM	oaoSVM
Content		.610	.438	.419
Reviews		.646	.582	.558
Tags	**Fractions**	.464	.706	.693
	Binary	.572	.702	**.698**
	TF	**.680**	.615	.629
	TF-IRF	.639	.625	.628
	TF-IBF	.641	.628	.630
	TF-IUF	.661	**.638**	.638

(b) LibraryThing

		mSVM	oaaSVM	oaoSVM
Content		.806	.732	.725
Reviews		.827	.718	.677
Tags	**Fractions**	.714	.886	.875
	Binary	.649	.884	.878
	TF	.861	.883	.878
	TF-IRF	.895	.887	.880
	TF-IBF	**.896**	.887	.881
	TF-IUF	.893	**.890**	**.883**

(c) GoodReads

		mSVM	oaaSVM	oaoSVM
Content		.806	.732	.725
Reviews		**.827**	.718	.677
Tags	**Fractions**	.659	**.768**	.708
	Binary	.647	.738	.734
	TF	.734	.733	**.750**
	TF-IRF	.802	.706	.725
	TF-IBF	.802	.649	.725
	TF-IUF	.797	.633	.719

TABLE 8
Accuracy results for tag-based classification, and comparison to classification using other data sources.

The results show that the use of social tags almost always outperforms the other data sources. The exception is the multiclass classifier–mSVM–for GoodReads. In general, GoodReads is the system that shows the worst performance of tags as compared to the other data sources. This happens because GoodReads does not encourage users to attach tags to books. GoodReads requires users to add tags by following a 2-step process, what makes the task less accessible. Consequently, fewer users provide tags, and books tend to remain annotated with fewer tags. This makes tags from GoodReads not to be sufficient to yield an outperformance as Delicious and LibraryThing do. Tags from these two systems clearly outperform classification using content or reviews. Between these two data sources, user reviews usually outperform content, but not even reviews are enough to reach the performance of social tags.

The performance of different representations of social tags greatly depends on the classifier and dataset utilized. The use of inverse weighting functions on tag-based representations does not seem to be helpful in most cases. However, they show a great improvement in a few cases. Specifically, the use of inverse weighting functions helps tag-based representation when using mSVM on the book datasets, LibraryThing and GoodReads. The feature that affects most the performance of inverse weighting functions on Delicious is the suggestion policy of the system. Suggesting tags to users based on earlier annotations to the same resource leads users to reuse popular tags, and reduces the likelihood of adding new tags. This produces a forced folksonomy with different tag distributions from a natural behavior, what leverages the low performance of inverse weighting functions. In fact, among the 3 inverse weighting functions, IUF shows a higher performance than IRF and IBF on Delicious. This corroborates the effect of system suggestions, since those independent users who get rid of suggestions and annotate tags of their choice reflect on higher IUF values. Accordingly, higher IUF values make the difference, and give rise to better performance than IRF and IBF. On the other hand, in the case of the book datasets, LibraryThing and GoodReads, the 3 inverse weighting functions show very similar performance, and none of them can be determined to be the best.

Comparing the local weighting representations of social tags–Fractions, Binary, and TF–there is also a clear difference among classifiers. TF is clearly the best solution when an mSVM classifier is used. However, TF performs worse or similar to Fractions and Binary

approaches when combinations of binary classifiers–oaaSVM and oaoSVM–are used. This suggests that a native multiclass classifier as mSVM rather uses detailed weightings where the relevance of each tag is explicitly defined with the number of annotators. However, in the case of combinations of binary classifiers–oaaSVM and oaoSVM–, where only two classes are considered at a time, it is enough to rely on simpler weightings.

With a few exceptions, the use of a multiclass classifier–mSVM–that considers the task as a single classifier performs the best. This shows that a classifier that has deeper knowledge of the whole task at the same time performs better than those classifiers that combine more limited knowledge of smaller binary tasks. Despite the success of mSVM, the main difference is on representations. It is better to rely on inverse weighting functions in the absence of resource-based tag suggestions, i.e., on LibraryThing and GoodReads. However, a TF-based weighting is superior when those suggestions are given by the system, i.e., on Delicious.

7.2 Getting the Most Out of All Data Sources

While the tag-based classifier is doing it well for a great part of the resources, the other classifiers may help improve the results by fixing some mispredictions. As an SVM classifier outputs a margin for each class over each resource, a ranking of predictions for all the classes can be inferred for each resource. Combining classifiers appropriately may help obtain better results. The combination of SVM classifiers may be done by using the so-called classifier committees [28]. Classifier committees rely on the predictions of several classifiers, and combine them by means of a decision function, which serves to define the weight or relevance of each classifier in the final prediction [35]. After applying the decision function on the predictions of all classifiers, a single unified prediction can be inferred.

An SVM classifier outputs a margin for each resource over each class in the taxonomy, as a confidence to belong to that class. The class maximizing the margin for each resource is then selected by the classifier. The larger is the gap between the maximum margin and the rest, the more reliable can be considered the prediction of the classifier. Thus, the combination of predictions of different SVM classifiers could be done by means of adding up their margins or reliability values for each class. Each resource will then have a new reliability value for each class, i.e., the sum of margins by different classifiers for a resource. Nonetheless, in this case, since each of the three classifiers work with different type of data, the range and scale of the margins they output differ. To solve this, we propose the normalization of the margins based on the maximum margin value outputted by each classifier, $\max(m_i)$ (see Equation 12).

$$m'_{ijc} = \frac{m_{ijc}}{\max(m_i)} \tag{12}$$

where m_{ijc} is the margin by the classifier i between the resource j and the hyperplane for the class c, and m'_{ijc} is its value after normalizing it. The class maximizing this sum of margins will be predicted by the classifier. Then, the sum of margins between the class c and the resource j using a committee with n classifiers is defined by Equation 13.

$$S_{jc} = \sum_{i=1}^{n} m'_{ijc} \tag{13}$$

If the classifiers work over k classes, then the predicted class for the resource j will be defined by Equation 14.

$$C_j^* = \underset{i=1,\ldots,k}{\operatorname{argmax}}\{S_{ji}\} \tag{14}$$

Table 9 shows an example of outputs in the form of margins of two classifiers for a resource in a taxonomy with 3 categories. Let this resource belong to the category #2. The example shows that, even though the classifier A predicted the category #1, and the classifier B says that the resource should be classified in category #3, the committees get the largest margin value for the category #2 by adding up margins from both classifiers.

	Category #1	**Category #2**	**Category #3**
Classifier A	**1.2**	1.1	0.6
Classifier B	0.5	1.0	**1.2**
Committees	1.7	**2.1**	1.8

TABLE 9
Example of committees: both classifiers mispredict the category, but committees guess it correctly.

Next, we show the results of using classifier committees on separate tables for each dataset. Note that the accuracy of the best single classifier in each case is also included, in order to enable to compare the performance of committees to it. Apart from that, we show the results by 4 different classifier committees, i.e., all 3 possible double combinations of data sources, and the combination of all 3 data sources. When combining tags with classifier committees, we use the representation that scored the best performance as a single classifier.

Table 10 shows the results of the resource classification using classifier committees. Specifically, Table 10(a) shows the results for Delicious, Table 10(b) for LibraryThing, and Table 10(c) for GoodReads.

The native multiclass classifier–mSVM–clearly outperforms the combinations of binary classifiers–oaaSVM and oaoSVM–for all 3 datasets. This outperformance is much clearer than for the use of single classifiers. This confirms that besides obtaining the best results, mSVM also provides better margin values than the others, what helps correctly guess some resources that were mispredicted by single classifiers. Going in more

(a) Delicious

	mSVM	oaaSVM	oaoSVM
Best single classifier	.680	.638	**.698**
Content + Reviews	.669	.561	.572
Content + Tags	.700	.599	.619
Reviews + Tags	.697	**.662**	.652
Cont. + Rev. + Tags	**.706**	.630	.650

(b) LibraryThing

	mSVM	oaaSVM	oaoSVM
Best single classifier	.861	**.890**	**.883**
Content + Reviews	.835	.775	.776
Content + Tags	.892	.844	.845
Reviews + Tags	**.893**	.808	.839
Cont. + Rev. + Tags	.888	.822	.844

(c) GoodReads

	mSVM	oaaSVM	oaoSVM
Best single classifier	.827	.768	.750
Content + Reviews	.835	.775	.776
Content + Tags	.831	.779	.789
Reviews + Tags	**.843**	.758	.788
Cont. + Rev. + Tags	**.843**	**.793**	**.810**

TABLE 10
Accuracy results of classifier committees. Best single classifier is included to enable comparison.

depth in the results of the classifier committees using mSVM classifiers, the performance of those committees including tags perform best. In fact, classifier committees that include tags always outperform the best single classifier. This shows the great potential of social tags, not only for working on their own, but also to be combined with other data sources using classifier committees. The use of content in classifier committees, however, does not always seem to be helpful. Content shows a slight improvement for Delicious, but the contribution is not that clear, and it is sometimes even harmful for the book datasets, LibraryThing and GoodReads. The main reason could be that we had to use synopses and editorial reviews as a summary of the content of books, because we did not have access to their textual content. Using synopses and editorial reviews may not work properly as a replacement of the content of books. Finally, reviews seem to be also useful to some extent, both for Delicious and for the book datasets, LibraryThing and GoodReads.

8 Conclusion

We have explored the usefulness of tags provided by users on social tagging systems for a social classification of resources. We have used three large-scale datasets of different types of annotated resources to compare the classification using social tags to that by experts on consolidated taxonomies. To the best of our knowledge, this is the first research work comparing tags from different systems and for different resources as to a social classification of resources. With this work, we complement our earlier work [37] by extending the study to different resources, analyzing in depth the characteristics of different tagging systems, and by exploring different settings of classifiers.

The results of our experiments show the great potential of social tags not only as a single classifier, but also to combine with other data sources. These results are best when a native multiclass classifier is used as the SVM setting. For the selection of an appropriate representation using social tags, the settings of the studied social tagging system should be taken into account. Among settings, we have shown that systems providing resource-based tag suggestions greatly alter folksonomies, and condition the success of certain representations.

As a future work, we plan to perform semantic analyses of tags to further understand the role of each tag. Likewise, this would help perform a study on dimensionality reduction, to check if fewer tags can yield similar performance. We will also explore how the popularity of a resource affects to its classification accuracy.

Acknowledgments

This work has been part-funded by the Government of Madrid under the Research Network MA2VICMR (S-2009/TIC-1542), the Spanish Ministry of Science and the Innovation project Holopedia (TIN2010-21128-C02-01).

References

[1] R. Abbasi, S. Chernov, W. Nejdl, R. Paiu, and S. Staab. Exploiting flickr tags and groups for finding landmark photos. In *ECIR '09: Proceedings of the 31th European Conference on IR Research*, pages 654–661, Berlin, Heidelberg, 2009. Springer-Verlag.

[2] L. M. Aiello, A. Barrat, C. Cattuto, G. Ruffo, and R. Schifanella. Link creation and profile alignment in the anobii social network. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing*, SOCIALCOM '10, pages 249–256, Washington, DC, USA, 2010. IEEE Computer Society.

[3] S. Aliakbary, H. Abolhassani, H. Rahmani, and B. Nobakht. Web page classification using social tags. *Computational Science and Engineering, IEEE International Conference on*, 4:588–593, 2009.

[4] R. Angelova, M. Lipczak, E. Milios, and P. Pralat. Characterizing a social bookmarking and tagging network. In *Proc. of the ECAI 2008 Workshop on Mining Social Data*, pages 21–25. IOS, 2008.

[5] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 43–52. Morgan Kaufmann, 1998.

[6] J. Diederich and T. Iofciu. Finding communities of practice from user profiles based on folksonomies. In *In Proceedings of the 1st International Workshop on Building Technology Enhanced Learning solutions for communities of Practice*, 2006.

[7] D. Godoy and A. Amandi. Exploiting the social capital of folksonomies for web page classification. In *Software Services for E-World*, volume 341 of *IFIP Advances in Information and Communication Technology*, pages 151–160. Springer, 2010.

[8] S. Golder and B. A. Huberman. The structure of collaborative tagging systems. *Journal of Information Science, 32(2)*, pages 198–208, 2006.

[9] M. Gupta, R. Li, Z. Yin, and J. Han. Survey on social tagging techniques. *SIGKDD Explorations*, 12(1):58–72, 2010.

[10] Z. Harris. Distributional structure. In *Papers in Structural and Transformational Linguistics*, pages 775–794. D. Reidel Publishing Company, Dordrecht, Holland, 1970.

[11] P. Heymann, G. Koutrika, and H. Garcia-Molina. Can social bookmarking improve web search? In *WSDM '08: Proceedings of the international conference on Web search and web data mining*, pages 195–206, New York, NY, USA, 2008. ACM.

[12] C.-W. Hsu and C.-J. Lin. A comparison of methods for multi-class support vector machines. *IEEE Trans. on Neural Netw.*, 13(2), 2002.
[13] T. Joachims. Text categorization with support vector machines: learning with many relevant features. In C. Nédellec and C. Rouveirol, editors, *Proceedings of ECML-98, 10th European Conference on Machine Learning*, number 1398, pages 137–142, Chemnitz, DE, 1998. Springer Verlag, Heidelberg, DE.
[14] U. Kreßel. Pairwise classification and support vector machines. In *Advances in kernel methods*, pages 255–268. MIT Press, 1999.
[15] X. Li, L. Guo, and Y. E. Zhao. Tag-based social interest discovery. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 675–684, New York, NY, USA, 2008. ACM.
[16] H. Liang, Y. Xu, Y. Li, R. Nayak, and X. Tao. Connecting users and items with weighted tags for personalized item recommendations. In *HT '10: Proceedings of the 21st ACM conference on Hypertext and hypermedia*, pages 51–60, New York, NY, USA, 2010. ACM.
[17] C. Lu, J.-r. Park, and X. Hu. User tags versus expert-assigned subject terms: A comparison of librarything tags and library of congress subject headings. *Journal of Information Science*, 36(6):763–779, 2010.
[18] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and S. Gerd. Evaluating similarity measures for emergent semantics of social tagging. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 641–650, New York, NY, USA, 2009. ACM.
[19] M. G. Noll and C. Meinel. Authors vs. readers: A comparative study of document metadata and content in the www. In *2007 ACM symposium on Document engineering*, pages 177–186, Winnipeg, Manitoba, Canada, 2007. ACM.
[20] M. G. Noll and C. Meinel. Exploring social annotations for web document classification. In *2008 ACM symposium on Applied computing*, pages 2315–2320, Fortaleza, Ceara, Brazil, 2008. ACM.
[21] M. G. Noll and C. Meinel. The metadata triumvirate: Social annotations, anchor texts and search queries. In *WI-IAT '08*, volume 1, pages 640–647, 2008.
[22] M. F. Porter. An Algorithm for Suffix Stripping. *Program*, 14(3):130–137, 1980.
[23] X. Qi and B. D. Davison. Web page classification: Features and algorithms. *ACM Comput. Surv.*, 41(2):1–31, 2009.
[24] D. Ramage, P. Heymann, C. D. Manning, and H. Garcia-Molina. Clustering the tagged web. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pages 54–63, Barcelona, Spain, 2009. ACM.
[25] R. Rifkin and A. Klautau. In defense of one-vs-all classification. *Journal of Machine Learning Research*, 5:101–141, December 2004.
[26] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24:513–523, August 1988.
[27] A. Shepitsen, J. Gemmell, B. Mobasher, and R. Burke. Personalized recommendation in social tagging systems using hierarchical clustering. In *RecSys '08: 2008 ACM conference on Recommender Systems*, pages 259–266, New York, NY, USA, 2008. ACM.
[28] B.-Y. Sun, D.-S. Huang, L. Guo, and Z.-Q. Zhao. Support vector machine committee for classification. In *Advances in Neural Networks - ISNN 2004*, pages 648–653, 2004.
[29] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos. A unified framework for providing recommendations in social tagging systems based on ternary semantic analysis. *IEEE Transactions on Knowledge and Data Engineering*, 22:179–192, 2010.
[30] P. Vatturi, W. Geyer, C. Dugan, M. Muller, and B. Brownholtz. Tag-based filtering for personalized bookmark recommendations. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 1395–1396. ACM, 2008.
[31] M. Vojnovic, J. Cruise, D. Gunawardena, and P. Marbach. Ranking and suggesting popular items. *IEEE Transactions on Knowledge and Data Engineering*, 21:1133–1146, 2009.
[32] J. Weston and C. Watkins. Multi-class support vector machines. In *Proceedings of the 1999 European Symposium on Artificial Neural Networks*, 1999.
[33] J. Xia, K. Wen, R. Li, and X. Gu. Optimizing academic conference classification using social tags. *Computational Science and Engineering, IEEE International Conference on*, 0:289–294, 2010.
[34] Z. Yin, R. Li, Q. Mei, and J. Han. Exploring social tagging graph for web object classification. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '09, pages 957–966, New York, NY, USA, 2009. ACM.
[35] Z. Zheng. Naive bayesian classifier committees. *Machine Learning: ECML-98*, pages 196–207, 1998.
[36] A. Zubiaga. Enhancing Navigation on Wikipedia with Social Tags. In *Wikimania 2009: 4th Annual Conference of the Wikimedia Community*, August 2009.
[37] A. Zubiaga, R. Martínez, and V. Fresno. Getting the most out of social annotations for web page classification. In *Proceedings of the 9th ACM symposium on Document engineering*, DocEng '09, pages 74–83, New York, NY, USA, 2009. ACM.

Arkaitz Zubiaga Arkaitz Zubiaga is a post-doctoral researcher at the Queens College and Graduate Center of the City University of New York. Arkaitz earned his Ph.D. in Computer Science in July 2011 under the supervision of Prof. Raquel Martínez and Prof. Víctor Fresno at the NLP&IR Group of the UNED University in Madrid, Spain. His main research interests lie in the fields of social media mining, machine learning, and data mining.

Víctor Fresno Dr. Víctor Fresno is Assistant Professor at UNED. He received the Bs. D. of theoretical physics from Universidad Autnoma de Madrid in 1999, the M.Sc. in telecommunication engineering from Universidad Politcnica de Madrid in 2004, and the Ph.D. in Computer Science from Universidad Rey Juan Carlos in 2006. His research interests include web page characterization for classification/clustering and Information Retrieval, Fuzzy Logic and NLP tools for Text Mining and social media mining.

Raquel Martínez Dr. Raquel Martínez Unanue is Associate Professor at the Department of Lenguajes y Sistemas Informáticos (LSI) at UNED, in Madrid, Spain. She is member of the NLP&IR Research Group and her research lines include Multilingual Document Clustering, Named Entities Disambiguation and Text mining in social media.

Alberto Pérez García-Plaza Alberto Pérez García-Plaza earned a B.S. and M.S. degree in Computer Engineering from Universidad Rey Juan Carlos, Madrid (Spain) in 2003 and 2006 respectively. He works as Teaching Assistant at the Department of Computer Systems and Languages at UNED in Madrid, Spain, where he is also Ph.D. candidate and member of the NLP&IR Group. His main research interests are web page clustering, document representation, fuzzy logic and social media mining.