

Profiling Social Networks: A Social Tagging Perspective

Ying Ding, Elin K Jacob, James Caverlee^I,

Michael Fried^{II}, Zhixiong Zhang^{III}

School of Library and Information Science, Indiana University
E 10th Street, Bloomington, 47405, USA
{dingying, ejacob}@indiana.edu

^IDepartment of Computer Science, Texas A&M University,
3112 TAMU, College Station, TX 77843, USA
caverlee@cs.tamu.edu

^{II}STI, University of Innsbruck, Austria
michael.fried@sti2.at

^{III}Library of Chinese Academy of Science,
33 Beisihuan Xilu, Zhongguancun, Beijing, 10080, China
zhangzx@mail.las.ac.cn

Abstract

The web is rapidly becoming both more open and more social through the provision of technologies that make it easy for end users to access resources and join in social networks. Social networks have pioneered online communities, allowing users to contribute to collective intelligence through the tagging of online resources. In response, online tagging activities increased significantly between 2005 and 2007. This paper reports on an investigation of social tagging using data crawled from del.icio.us, Flickr and YouTube. The findings indicate that it is possible to profile a social network through analysis of tagging data and that del.icio.us is a more representative venue for analyzing social tagging data and the tagging behaviors of users.

Introduction

The web is experiencing a period of rapid transition from a space for presentation of syntactically formatted information to an open and social platform that allows users to communicate knowledge and share resources. The web is pioneering technologies that make it easy for users to participate in social networks, leading to the development of online communities. In concern with RSS feeds, Wikipedia, Google Map, mashup services and the interlinked social semantics that are the hallmarks of the current web revolution, social networks are actively contributing to the evolution of collective intelligence.

Among the many social networks available on the web, del.icio.us, Flickr and YouTube stand out as three of the best known and most popular. These networks established their prominent position either by hosting large amounts of online objects, building a large user community, or generating the heavy web traffic: in 2007, Flickr hosted more than 2 billion images (Auchard, 2007); in 2006, del.icio.us had more than one million users (Robinson, 2006); and a scrape of YouTube in August 2006 indicated a total of 1.73 billion viewings of videos since YouTube's inception in 2005 (Gomes, 2006). The large amount of data generated by social networks has triggered a boom in social network research (e.g., Kipp and Campbell, 2006; Mika, 2007; Li, Guo, & Zhao, 2008; Lin, Chi, Zhu, Sundaram, & Tseng, 2008; Singla & Richardson, 2008). But the question remains as to whether analysis of user tagging behaviors can be used to reveal particular and distinctive characteristics of social networks.

This paper investigates social tagging behavior in del.icio.us, Flickr and YouTube from the perspective of social tagging. It begins by describing the crawler used to harvest tagging data from each of these three social networks. It then provides an analysis of the most popular tags and the evolution of tag use in the individual social networks. Finally, it summarizes the

findings and concludes that it is possible to profile a social network through analysis of tagging data and that del.icio.us is a more representative venue for future analysis of social tagging data and the tagging behaviors of users.

UTO Tag Crawler

To integrate tagging data from different social networks, a tag crawler was developed to crawl del.icio.us, Flickr and YouTube and store the retrieved tagging data in RDF triples based on the Upper Tag Ontology (UTO) (Ding, Toma, Kang, Fried & Yan, 2008).

The UTO tag crawler was developed using the Smart and Simple Webcrawler framework developed by Torunski (2008), which provided for the functionalities of maximum interactions, maximum depth, filter interface, and plugable http connection libraries. The UTO crawler was designed as a multi-thread crawler to avoid timeouts and to make efficient use of available internet bandwidth. There are two different parsers: One parses a page and searches for links that should be visited or filtered out; the other parses HTML code to retrieve information about tags.

In del.icio.us, the crawler began with the del.icio.us tag cloud at <http://delicious.com/tag> and visited every tag in the cloud. For TagA in the tag cloud, the crawler visited <http://delicious.com/tag/tagA> and parsed the HTML code to grab information about bookmarks, taggers and related tags. For each bookmark, the crawler then went to <http://delicious.com/url/idOfURL> and crawled the history of the bookmark, focusing on which users had tagged this bookmark on which date(s). After gathering data about all of the bookmarks on the first page for TagA, the crawler visited the second and subsequent pages for TagA, performing the same tasks, until it reached the arbitrarily set threshold of 99 pages for TagA. According to our test, most of the tags only have less than 10 pages of bookmarks (assuming 100 bookmarks per page). We checked the “blog” tag which is the most frequently used tag of the whole tag set by Delicious. It says that totally there are 5465905 bookmarks for tag “blog”. But it only allows us to go to page 20 based on 100 bookmarks per page (see Figure 1). So the threshold of 99 pages based on 50 bookmarks per page for each tag in our crawler setting can ensure that the coverage of the tag set should be more than 90%. The crawler then repeated this process for TagB and all subsequent tags until all tags in the tag cloud had been visited.

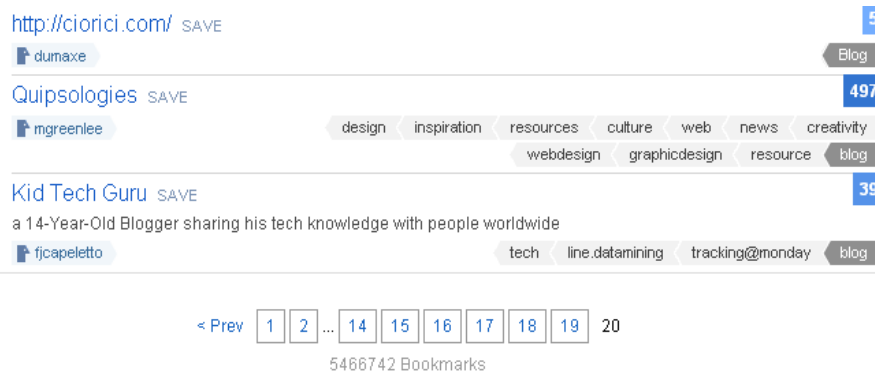


Figure 1: the available bookmarks for tag “blog”
(you can only go to page 20 based on 100 bookmarks per page)

For Flickr, the crawler started at the tag cloud at <http://flickr.com/photos/tags> and visited tag in the cloud. In each tag page (for example: <http://www.flickr.com/photos/tags/party/>), information about related tags was collected. Each photo on the tag page (20 links per page) was visited (for example: <http://www.flickr.com/photos/25612622@N08/3063428352/>) and information about the photograph, tags and tagger (one per photo) was extracted. The crawling process continued with <http://www.flickr.com/photos/tags/party/?page=2>. To avoid duplicate visits, only links to the form <http://www.flickr.com/photos/taggerID/photoID/> were accepted.

For YouTube, the crawler started from the main page at <http://youtube.com> and visited every available video page. On one video page, it collected tagging data and visited the links pointing to other video pages. YouTube does not provide related tag data. In order to avoid visiting the same page more than once, the query parts of links were ignored (i.e. <http://www.youtube.com/watch?v=X2IExa2A198> and http://www.youtube.com/watch?v=X2IExa2A198&watch_response lead to the same video). Figure 2 provides an overview of the UTO crawler. Detailed information about UTO ontology and the UTO crawler is available in Ding et al. (2008, 2008a).

The UTO crawler was used to retrieve tagging data from del.icio.us, Flickr and YouTube in September 2007. For all three social networks, tagging data harvested by the crawler included object, tagger, tag, date, comment and vote. The tagging data has been converted to RDF triples based on UTO ontology (Ding et al., 2008a). In total, the crawler retrieved around 21 million RDF triples for del.icio.us, 2.3 million RDF triples for Flickr, and 2.2 million RDF triples for YouTube.

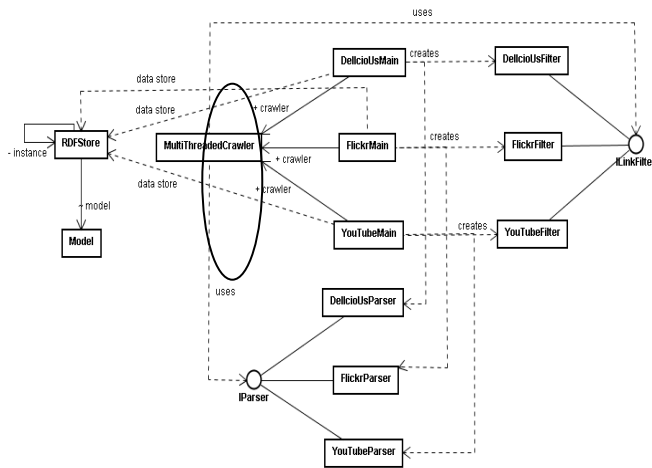


Figure 2. Overview of UTO Crawler

Table 1 shows the details of the different datasets. In total, our dataset contains around 1 million bookmarks, 2.8 million taggers and 9.3 million tags from del.icio.us, 0.3 million photos, 0.2 million taggers and 1.4 million tags from Flickr and 0.5 million videos, 0.2 million taggers and 1.4 million tags from YouTube. The average number of tags per object ranges from 2.74 (YouTube) to 9.31 (del.icio.us). The average number of tags the normal tagger uses ranges from 3.33 (del.icio.us) to 8.79 (Flickr). The average number of objects the normal tagger tags range from 0.36 (del.icio.us) to 2.84 (YouTube). Since when users upload bookmarks to del.icio.us, tag is not the required field to fill in (but the title of the URL is a required field). So there might be many bookmarks with titles but without tags.

Table1. Crawled Tag Data

Social Network	Objects	Taggers	Tags	Tag/Object	Tag/Tagger	Object/Tagger
del.icio.us	996,748	2,787,860	9,282,058	9.31	3.33	0.36
Flickr	295,837	153,778	1,351,201	4.57	8.79	1.92
YouTube	527,924	185,975	1,443,924	2.74	7.76	2.84

Social Tagging Analysis

In order to generate a portrait of each of these social networks, tagging data sets were analyzed according to different time frames (2005, 2006, 2007).

Power law distribution

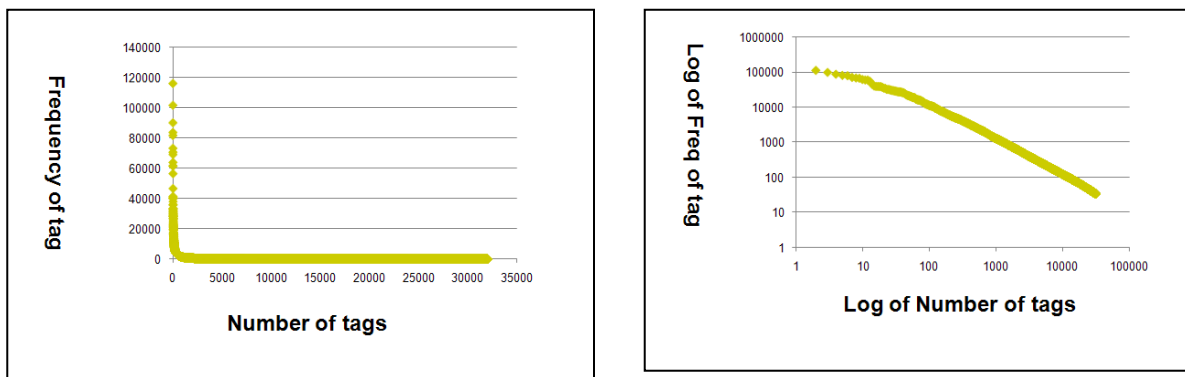


Figure 3. Distribution of Tags

We merged the tagging data from these three social networks and form one big tagging dataset. Based on this big dataset, we conduct the tag frequency analysis. Figure 3 demonstrates that tag frequency distribution is a power law distribution. Less than 0.2% of tags have frequency more than 120,000 times, while more than 55% of tags were assigned one to ten times in these three social networks. Table 2 shows the details of this distribution. It confirms the Zipf's Law which shows a few occur very often (around 97 tags having more than 120,000 frequency) while many others occur rarely (357,028 tags having less than 10 frequency). The top 1300 tags (see Appendix) already cover more than 50% of the whole corpus of the tags (around 12 million tags, see Table 1). “design” is the most frequently occurring tag, and by itself accounts for nearly 1% of all tag occurrences (101786 out of slightly over 12 million). The second-place tag “blog” counts for 0.7% of tags (90242 occurrences). So some linguistic analysis of these 1300 core tag set can reveal certain features of the social vocabulary of the social tagging. This will be certainly good topics for our future works.

Table 2: Tag Frequency Distribution

Tag Freq Range	No. of tags	Cumulative %
1-10	357028	55.07%
11-20	217746	88.65%
21-30	27404	92.88%
31-40	11524	94.65%
41-50	6656	95.68%
51-60	4454	96.37%
61-70	3387	96.89%
71-80	2461	97.27%
81-90	2066	97.59%
91-100	1597	97.83%
101-200	1348	98.04%

201-300	6193	99.00%
301-400	2151	99.33%
401-500	1044	99.49%
501-1000	645	99.59%
1001-120,000	1301	99.79%
>120,000	97	100.00%

del.icio.us

Table 3 shows the 20 most frequently assigned tags in del.icio.us for the years 2005, 2006, and 2007. These tag sets appear to be relatively stable across the three years. The tags *xml*, *science*, *search*, *games*, *technology*, and *security* appear among the top 20 tags for 2005 but are dropped from the lists of top 20 tags for 2006 and 2007; and the tags *imported*, *research*, and *internet* are dropped from the list of top 20 tags for 2007. The tags *development*, *howto*, *tutorial* and *Web2.0* appear in the lists for both 2006 and 2007, and *webdesign*, *free* and *opensource* are introduced in 2007, pointing to the emergence of new trends in user interests. Overall, 85% of the top 20 tags are stable across 2006 and 2007, indicating that a shared social vocabulary may be emerging in del.icio.us.

A profile of del.icio.us can be generated through analysis of the lists of popular tags. The tags *web*, *programming*, *video*, *music* and *news* indicate key interests of del.icio.us users when tagging bookmarks to store or share. As indicated by the tags introduced in 2006 and 2007, users evince greater interest in opensource or free resources as well as tutorials for learning programming languages or skills. While the tags *music*, *art* and *news* indicate a level of general interest that spans all three years, actual tagging evidence for 2006 and 2007 strongly supports the popular assumption that del.icio.us is a social network for individuals interested in the web and programming skills (see Table 3).

Table 3: Top 20 Tags in del.icio.us in 2005, 2006 and 2007

Rank	2005	2006	2007
1	blog	blog	blog
2	programming	programming	design
3	software	software	software
4	music	design	programming
5	design	reference	reference
6	web	music	tools
7	reference	web	Web2.0
8	java	tools	web
9	art	art	video
10	tools	java	music
11	linux	video	art
12	news	Web2.0	linux
13	xml	linux	webdesign
14	science	news	howto
15	search	tutorial	free
16	games	howto	tutorial
17	research	imported	news
18	technology	development	development
19	security	research	opensource
20	video	internet	java

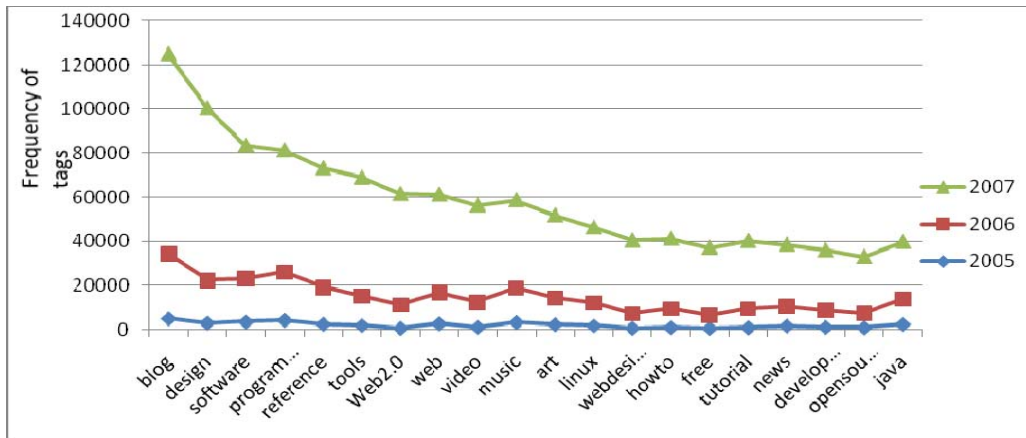


Figure 4. Evolution of Top 20 Tags in del.icio.us for the Years 2005, 2006 and 2007.

Table 4: Top 20 tags in Delicious 2007 and their frequency in 2005, 2006 and 2007

Top 20 Tags in Delicious 2007	2005	2006	2007	2006/2005	2007/2005	2007/2006
blog	5046	29485	90474	6	18	3.1
design	3045	19273	78115	6	26	4.1
software	3558	19533	60405	5	17	3.1
programming	4295	21789	55237	5	13	2.5
reference	2541	16643	53971	7	21	3.2
tools	1943	13340	53772	7	28	4.0
Web2.0	658	10620	50279	16	76	4.7
web	2743	14115	44406	5	16	3.1
video	1114	11383	43847	10	39	3.9
music	3325	15523	39859	5	12	2.6
art	2344	12043	37518	5	16	3.1
linux	1799	10434	34241	6	19	3.3
webdesign	688	6542	33224	10	48	5.1
howto	962	8588	31701	9	33	3.7
free	643	5793	30750	9	48	5.3
tutorial	895	8683	30648	10	34	3.5
news	1712	8854	28086	5	16	3.2
development	1107	7588	27322	7	25	3.6
opensource	872	6468	25735	7	30	4.0
java	2449	11606	25732	5	11	2.2

Note: 2006/2005 means the figures in column 2006 divided by the figures in column 2005. Actually, it shows the increase of the frequency about the tag from 2005 to 2006. The same are for column 2007/2005 and 2007/2006.

Figure 4 and Table 4 show the evolution of tags (topics) in the del.icio.us social network. The tag *Web2.0* shows the highest peak in both 2006 and 2007. The frequency with which *Web2.0* was used to tag bookmarks increased 16 times in 2006 and 76 times in 2007 when compared with its tagging frequency in 2005. The three tags showing the most dramatic increase in tagging frequency from 2006 to 2007 are *webdesign*, *free* and *Web2.0*, indicating growing interest in these topics on the part of del.icio.us taggers. The three tags with the least impressive increase in tagging frequency from 2006 to 2007 are *java*, *programming*, and *music*. This would seem to indicate waning interest in these topics on the part of del.icio.us taggers. However, given that Last.fm has recently become one of the most popular social networks for sharing music, this may explain why music bookmarks are decreasing in del.icio.us.

Flickr

Table 5 shows the 20 most frequently assigned tags in Flickr in the years 2005, 2006, and 2007. In sharp contrast to the tagging culture in del.icio.us, Flickr taggers like to use year, locations, colors, and seasons to tag photos. Favorite locations in Flickr include Hong Kong (2005), Germany (2005), USA (2006 and 2007), London (2005-2007), California (2006), and Japan (2007). Favourite colors are orange (2005), blue (2006 and 2007), red (2006 and 2007), green (2006 and 2007), and black-and-white (*bw* in 2007). Most frequently used tags for seasons are *autumn* and *fall* (in 2007). In addition, users also favor tagging photographs with time of day (or lighting conditions), especially when they are night views. With the exception of the tagging categories of colors and locations, tag sets differ widely across these three years.

Flickr taggers also use informal tags to tag photos (e.g., the years 2005, 2006 and 2007 and tags such as *me*), indicating that users tag many objects for purposes of storing and retrieving photographs for themselves, rather than for purposes of sharing. When tagging images, users tend to emphasize the eye-catching features of an image such as color, subject (e.g., *sky*, *water*, *beach* and specific locations), and light conditions (e.g., *night* and *nightview*), but time (i.e., year, season or month), locations and colors are the major features of images tagged by users. It would be useful to analyze the tagging culture of Flickr in greater detail given that annotating images is an important area for image retrieval. An interesting example of ongoing research on social annotation of images and videos is GWAP,¹ the "games with a purpose" project at Carnegie Mellon.

Table 5. Top 20 Tags in Flickr in 2005, 2006 and 2007

Rank	2005	2006	2007
1	2005	usa	2007
2	d70	california	canon
3	tsimshatsui	2006	nature
4	hongkong	cameraphone	autumn
5	nightview	celltagged	art
6	germany	zonetag	nikon
7	newkie	sanfrancisco	water
8	ragbrai	blue	bw
9	art	light	red
10	wonder	sky	blue
11	night	urban	sky
12	buttersweet	red	japan
13	15fav	sea	fall
14	central	me	beach
15	light	water	portrait
16	marco	nature	london
17	london	marco	night
18	apargioides	london	green
19	orange	green	usa
20	ads1	music	november

Figure 5 and Table 6 show the temporal features of tags in Flickr. In 2005 and 2006, tagging was not popular in the Flickr community; but this changed dramatically in 2007 as tagging became more popular on the web in general. Nearly 100 times more Flickr users tagged photographs in 2007 than did in 2006. Tagging frequency for *canon*, the second most popular tag in 2007, shows the greatest jump in use, increasing 203.5 times over its tagging frequency in 2006 (see Figure 5). Analysis of photographs tagged by Flickr users also indicates that there are two major communities: One community contains non-professional photographers who use Flickr as a platform for sharing photographs with friends and family and tag images so that they can be found by others; the other community consists of professional photographers who do not tag often but who do provide comments on photographs taken by other professionals.

¹ <http://www.gwap.com/gwap/>

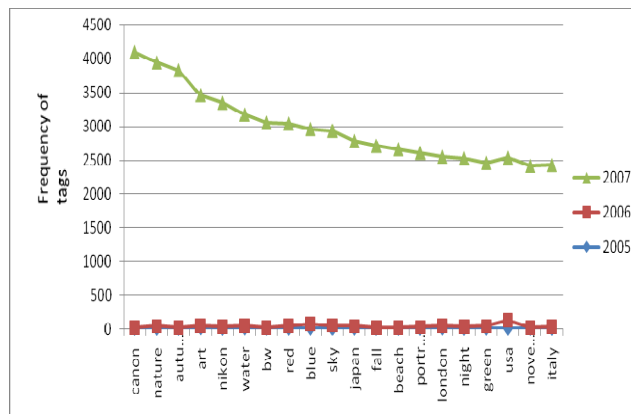


Figure 5. Evolution of Top 20 Tags in Flickr for the Years 2005, 2006 and 2007.

Table 6. Top 20 tags in Flickr 2007 and their frequency in 2005, 2006 and 2007

Top 20 Tags in Flickr 2007	2005	2006	2007	2006/2005	2007/2005	2007 /2006
canon	5	20	4070	4	814	203.5
nature	5	39	3899	8	780	100.0
autumn	5	18	3804	4	761	211.3
art	13	33	3416	3	263	103.5
nikon	5	30	3312	6	662	110.4
water	10	39	3126	4	313	80.2
bw	7	21	3028	3	433	144.2
red	7	47	2988	7	427	63.6
blue	8	66	2888	8	361	43.8
sky	9	48	2878	5	320	60.0
japan	8	37	2738	5	342	74.0
fall	5	18	2690	4	538	149.4
beach	5	24	2636	5	527	109.8
portrait	5	26	2581	5	516	99.3
london	10	39	2503	4	250	64.2
night	5	35	2489	7	498	71.1
green	7	38	2417	5	345	63.6
usa	6	126	2406	21	401	19.1
november	5	18	2394	4	479	133.0
italy	5	33	2391	7	478	72.5

Note: 2006/2005 means the figures in column 2006 divided by the figures in column 2005. Actually, it shows the increase of the frequency about the tag from 2005 to 2006. The same are for column 2007/2005 and 2007/2006.

YouTube

Table 7 shows the 20 most popular tags in YouTube for the years 2005, 2006 and 2007. The main topics in this social network are music, videos, humor, sex and girls, reflecting the broad interests of the general web community.

Table 7. Top 20 Tags in YouTube for the Years 2005, 2006 and 2007

Rank	2005	2006	2007
1	music	the	the
2	funny	funny	music
3	video	music	funny
4	the	video	video
5	dance	live	girl
6	crazy	of	of
7	commercial	comedy	sexy
8	live	dance	live
9	dancing	rock	dj
10	guitar	cat	2007
11	fun	halloween	dance
12	AMV	love	hot

13	girl	girl	comedy
14	japan	movie	rock
15	hot	dj	love
16	anime	in	and
17	Halloween	sexy	sex
18	halo	and	in
19	of	fight	new
20	cat	you	cat

Tagging behavior increased significantly in YouTube between 2006 and 2007, with tag use 78.7 times greater in 2006 and 236.7 times greater in 2007 than it was in 2005. The tag *dj* had the greatest increase in use in 2007, followed by *sex/sex*, while *dance* showed the least increase between 2005 and 2007. The tag set in YouTube appears to be more stable than it was in Flickr for the same time period, seemingly indicating that areas of user interest have remained fairly steady for the social web community as a whole (see Figure 6 and Table 8).

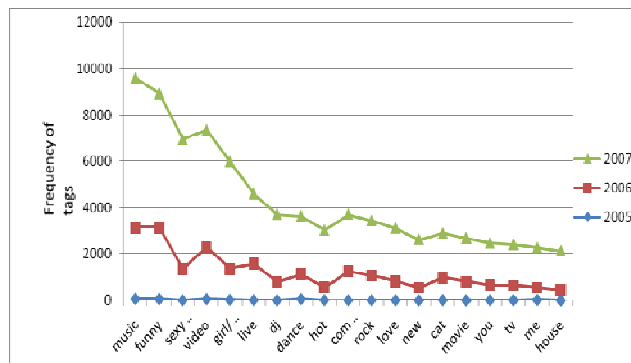


Figure 6. Evolution of Top 20 Tags in YouTube for the years 2005, 2006 and 2007.

Table 8. Top 20 tags in Flickr 2007 and their frequency in 2005, 2006 and 2007

Youtube tag2007	2005	2006	2007	2006/2005	2007/2005	2007 /2006
music	67	3080	6452	46	96	2.1
funny	58	3091	5784	53	100	1.9
sexy/sex	9	1338	5601	149	622	4.2
video	53	2234	5065	42	96	2.3
girl/girls	25	1334	4647	53	186	3.5
live	17	1563	3028	92	178	1.9
dj	4	777	2920	194	730	3.8
dance	56	1061	2526	19	45	2.4
hot	14	552	2467	39	176	4.5
comedy	10	1245	2461	125	246	2.0
rock	10	1059	2380	106	238	2.2
love	10	817	2294	82	229	2.8
new	4	544	2079	136	520	3.8
cat	13	977	1906	75	147	2.0
movie	6	795	1887	133	315	2.4
you	11	652	1805	59	164	2.8
tv	12	629	1768	52	147	2.8
me	21	520	1729	25	82	3.3
house	4	451	1666	113	417	3.7

Note: 2006/2005 means the figures in column 2006 divided by the figures in column 2005. Actually, it shows the increase of the frequency about the tag from 2005 to 2006. The same are for column 2007/2005 and 2007/2006.

Summary and Conclusion

Table 9 summarizes the characteristics of social networks identified in the analysis of del.icio.us, Flickr and YouTube. When comparing these three social networks, del.icio.us demonstrates the tightest connection to the use of tags as extended information about resources. In del.icio.us, every user can tag an object with the tag(s) of his own choice; and an object can

be tagged many times and by multiple users, thereby indicating that it “belongs” (or is more relevant) to the community as a whole. So del.icio.us is a kind of community-tagging where anyone can tag any available online resources (here in del.icio.us are bookmarks) (Marlow, et.al., 2006). Similar social networks also include CiteULike and Connotea (online resources are bibliographical records), LibraryThing (online resources are books) and so on.

This is very different from Flickr, where content is mainly tagged by the user who uploads the photograph; the major community activities of other users can just “comment” or “vote” for resources by indicating that a particular photograph is a favourite image. Flickr also provides functions to allow users to tag photos uploaded by their friends. But this limits the community-tagging only to the users and their closed friends. So Flickr is not a fully community-based tagging system rather a kind of self-tagging system for the users and their closed friends as not anyone can tag any photo in Flickr. YouTube has a system similar to that of Flickr. A user can tag the content (videos) he has uploaded and the public is able to vote for them by assigning “stars”.

So the different tagging rights have created the difference among the nature and types of resultant tags and the role of tags in the systems (Marlow, et.al., 2006). Based on the analysis of the top 20 tags in each social network, we found out that the tags in del.icio.us are more content-oriented which are related either to topics of the bookmarks. While the tags in Flickr are more annotation-oriented which are related to the features of the photos, such as colour, year and location. While the tags in YouTube are content and feature oriented which are somewhere related to the content and the feature of the videos. The role of tags in del.icio.us is to organize the bookmarks and help to retrieve and share the bookmarks. The tags play the major role in del.icio.us as del.icio.us counts on them for the users to share and find bookmarks which are the major functions of del.icio.us. While the tags in Flickr are a kind of side-effect which means that it is not necessary to tag your photo and it is up to the choice of the users. Photos can be searched via title of the photos and ranked by comments and votes. Tagging does not play a major role in Flickr. The same for YouTube, many YouTube users do not actually tag their videos. Videos are shared through comments and votes.

Social tagging behaviours are also related to the community of the social networks. Del.icio.us gathers a community interested in IT-related topics. These people are interested in the content of the bookmarks and tagging provides a good way for them to summarize the content of the bookmarks. Naturally, tagging becomes the key function of the system and plays a major role for sharing and retrieving. While in Flickr, the major community are professional photographers who would like to share their pieces of arts for comments and feedbacks and other users who just use the Flickr as a space to manage their own personal photos and share with their closed friends. Searching photos in Flickr is based on the title and tag of the photos. The community of Flickr are interested in commenting and sharing. While in YouTube, its community can be viewed as a snapshot of the whole community of the Web. They are people from all of the world with all kinds of different interests and from different age ranges. Many of them do not tag their videos. They come to YouTube with different purposes and expectations. The role of the tagging is shadowed by the rating and commenting. Searching videos in YouTube is mainly based on the titles of the videos. So the tagging in YouTube is fading.

After analyzing social tagging behavior in del.icio.us, Flickr and YouTube, it was readily apparent that tagging activities have increased tremendously across 2005, 2006 and 2007. More and more people are using online social networks to tag online objects for purposes of storage, access, and retrieval, both for themselves and for purposes of sharing those resources with others. Through tag analysis, it is possible to develop a portrait of the social culture of a network and, in some cases, to identify trends of increasing (and waning) topical interests among users.

Table 9. Summary of social tagging in del.icio.us, flickr and YouTube

Features	del.icio.us	flickr	YouTube
Community	People interested in sharing bookmarks about the web and programming	People interested in sharing photos (including both professional and non-professional photographers)	People interested in sharing videos on any subject.
Main Topics	Web, programming, videos, music, news	Colors, location, years, seasons	Music, videos, humor, sex and girls
Tagging Behaviour	A key activity with many users participating in tagging	Not many users tagging	Not many users tagging
Dynamic	Yes	No	No
Emerging Topics (Trends)	Web design, tutorials, web2.0	None apparent	None apparent
Declining Topics	Internet, research, xml, security	None apparent	None apparent
Stability	Set of most popular tags is changing on average 10%-20% each year	Tag set is stable	Tag set is stable
Tagging Vocabulary	Relatively stable	Stable categories of tags	Relatively stable
Tagging Focus	Content of the bookmark	Feature(s) of photographs	Subject content and feature of video
Primary Tagging Purpose	Access and sharing	Access	Sharing

While tag sets in del.icio.us appear to become more stable across the timeframe of this study, it is also apparent that collective tagging vocabularies could benefit from both syntactic and semantic normalization of tags: for example, in del.icio.us, there were 90,242 uses of *blog* as a tag and 35,791 uses of *blogs*. Normalization of singular vs. plural forms as well as acronyms vs. full names would increase the effectiveness of tags for retrieval purposes as would normalization of (or user education regarding) normative syntactical formations of tags (e.g., tag phrases with or without a space between individual terms).

This study demonstrates that it is possible to profile a social network through analysis of social tagging data. Thus, the del.icio.us community is seen to be comprised in large part of people interested in IT-oriented topics such as the web and programming. In contrast, the Flickr community appears to contain two primary groups of users: professional photographers and non-professional photographers. Finally, the YouTube community is very broad and can be best viewed as a subset of the general social web community. Tagging is a major activity in del.icio.us, but not in Flickr and YouTube. Tagging in del.icio.us is used primarily for purposes of storing, retrieving and sharing online resources across the community; tagging in Flickr emphasizes indexing objects for retrieval by the tagger and his friends and associates; and tagging activities in YouTube are undertaken primarily for sharing resources with the online community as a whole. Taggers intend to capture the content of a resource (i.e., bookmark or video) in del.icio.us and YouTube, but they focus on representing the specific features of an image in Flickr. In del.icio.us, emerging and/or waning trends in user interests can be identified by analyzing increases and decreases in tag frequency across time; but, in both Flickr and YouTube, such trends are not obvious.

Acknowledgements

The authors would like to thank the University of Innsbruck for its support of data collection and analysis. The authors are also very grateful for the technical support provided by Ioan Toma of the University of Innsbruck.

References

- Auchard, E. 2007. Flickr to map the world's latest photo hotspots, *Reuters*. Retrieved September 30, 2008.
- Ding, Y., Toma, I. Kang, S., Fried M. and Yan Z. 2008. Data Mediation and Interoperation in Social Web: Modeling, Crawling and Integrating Social Tagging Data. *WWW2008 Workshop on Social Web Search and Mining (SWSM2008)*, 17th International World Wide Web Conference, April 21-25, 2008, Beijing, China.

Ding, Y., Kang, S., Toma, I., Fried M. and Yan, Z. 2008a. Integrating Social Tagging Data: Upper Tag Ontology. In *Proceedings of the 2008 IEEE International Conference on Systems, Man and Cybernetics*, 12-15 October 2008, Singapore.

Gomes, L. 2006. Will All of Us Get Our 15 Minutes On a YouTube Video?. *The Wall Street Journal Online*, Dow Jones & Company. Retrieved September 26, 2008.

Kipp, M. E.; Campbell, D. G. 2006. Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices. In *Proceedings Annual General Meeting of the American Society for Information Science and Technology*, Austin, Texas, USA.

Li, X.; Guo, L.; Zhao, Y. 2008. Tag-based Social Interest Discovery. In *Proceedings of the World Wide Web 2008*, April 2008, Beijing, China.

Lin, Y.; Chi, Y.; Zhu, S.; Sundaram, H. and Tseng, B. 2008. FacetNet: A Framework for Analyzing Communities and Their Evolutions in Dynamic Networks. In *Proceedings of the World Wide Web 2008*, April 2008, Beijing, China.

Marlow, C.; Naaman, M.; Boyd, D. and Davis, M. (2006). HT06, tagging paper, taxonomy, flickr, academic article, to read. *Proceedings of the 17th Conference on Hypertext and Hypermedia*.

Mika, P. 2007. Ontologies are us: A unified model of social networks and semantics. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 5(1), 5-15.

Robinson, B. 2006. Del.icio.us reports 1 million users. TechCrunch. Available at <http://www.techcrunch.com/2006/09/25/del.icio.us-reports-1-million-users-post-yahoo-growth-tops-all-of-digg/> Retrieved on September 26, 2008.

Singla, P. and Richardson, M. 2008. Yes, There is a Correlation – From Social Networks to Personal Behavior on the Web. In *Proceedings of the World Wide Web 2008*, April 2008, Beijing, China.

Torunski, L. 2008. Smart and Simple Webcrawler: <https://crawler.dev.java.net> Retrieved on September 30, 2008.

Appendix:

Core Tag Vocabulary for Social Networks (top 1300 tags in three social networks – del.icio.us, Flickr and YouTube)

Core Tag Vocabulary	
Numbers & others	1, 2, 3, 2005, 2006, 2007, -, .net, 3d
Starting from a	a, academia, academic, accessibility, accessories, acoustic, action, actionscript, activism, ad, admin, administration, adobe, ads, adsense, adult, advertising, advice, Africa, agency, aggregator, agile, ai, air, airline, airlines, airplane, airport, ajax, algorithm, algorithms, all, alternative, amateur, amazing, amazon, America, American, Amsterdam, analysis, analytics, and, angel, angst, animal, animals, animation, anime, anonymous, anthropology, apache, api, apple, application, applications, apps, architecture, archive, archives, argentina, art, arte, article, articles, artist, artists, arts, as3, asia, asian, asp.net, ass, asterisk, astronomy, at, atheism, atom, au, audio, audiobooks, Australia, authentication, auto, automation, autumn, awards, awesome
Starting from b	Baby, backup, bad, ball, band, bands, bandslash, bank, banking, bar, Barcelona, baseball, bass, bbc, beach, beatles, beautiful, beauty, beer, berlin, best, bible, bibliography, bicycle, big, bike, bioinformatics, biology, bird, birds, birthday, bit200f06, bit200w07, bittorrent, black, blackandwhite, blog, blogger, blogging, blogs, blood, blue, Bluetooth, boat, body, boobs, book, bookmarking, bookmarks, books, boston, boy, boys, bpm, brain, branding, brasil, brazil, bridge, Britney, Brooklyn, brown, browser, browsers, Buddhism, building, bus, bush, business, buy, bw, by
Starting from c	C, c#, c++, calculator, calendar, California, camera, cameraphone, camping, Canada, canon, car, card, cards, career, cars, cartoon, cartoons, cat, cats, cd, celebrity, cell, cellphone, celltagged, censorship, change, charity, charts, chat, cheap, cheatsheet, chemistry, Chicago, chicken, child, children, chile, china, Chinese, chocolate, chords, chris, Christian, Christianity, Christmas, church, ciencia, cine, cinema, city, class, classic, classification, climate, clip, clothes, clothing, clouds, club, cluster, clustering, cms, cocoa, code, coding, coffee, collaboration, collection, college, color, colors, colour, comedy, comic, comics, commercial, communication, community, company, comparison, competition, compiler, complexity, computer, computers, computing, concert, concurrency, conference, conferences, conspiracy, consumer, content, contest, control, conversion, convert, converter, cooking, cool, copyright, corporate, country, course, courses, cover, crack, craft, crafts, crazy,

	creative, creativecommons, creativity, credit, crime, crossover, cryptography, cs, css, cultura, culture, curiosidades, custom, cute, cycling
Starting from d	Daily, dance, dancing, dark, data, database, datamining, dating, david, day, dc, de, dead, deals, death, debian, del.icio.us, delicious, demo, democracy, design, designer, desktop, deutsch, Deutschland, dev, developer, development, dhtml, dictionary, diet, dig, digital, directory, diseÃ±o, Disney, distributed, distro, diy, dj, django, dns, do, documentary, documentation, dog, dogs, dom, domain, dotnet, download, downloads, drawing, driver, drm, drugs, drunk, drupal, duesouth, dvd
Starting from e	Earth, ebay, ebook, ebooks, eclipse, ecology, ecommerce, economia, economics, economy, editing, editor, edtech, educaciÃ³n, educacion, education, effects, el, elearning, e-learning, electronic, electronics, email, embedded, employment, emulation, en, encryption, encyclopedia, energy, engine, engineering, England, English, enterprise, enterprise2.0, entertainment, entrepreneur, entrepreneurship, environment, erlang, esl, espaÃ±a, espaÃ±ol, essay, ethics, eu, europa, Europe, event, events, evolution, examples, excel, exchange, exercise, experimental, extension, extensions, eyes
Starting from f	f1, face, facebook, fall, family, fanfic, fanfiction, fantasy, faq, fashion, fat, feed, feeds, female, feminism, festival, fetish, fic, fiction, fight, file, files, filesharing, filesystem, film, films, finance, financial, fire, firefox, firefox:bookmarks, firefox:rss, firefox:toolbar, firewall, fish, fitness, flash, flex, flickr, flight, flights, florida, flower, flowers, fob, folksonomy, font, fonts, food, football, for, forms, forum, forums, foto, fotografia, fotos, framework, france, free, freedom, freelance, freeware, French, friends, from, fuck, fun, functional, funny, furniture, future
Starting from g	Gadget, gadgets, gallery, game, games, gaming, garden, gardening, gay, gear, geek, gen, gender, genealogy, generator, genetics, geo, geography, George, geotagged, german, germany, ghost, gifts, girl, girls, gis, glass, global, gmail, gnome, gnu, go, god, good, google, googlemaps, government, gps, graffiti, grammar, graph, graphic, graphicdesign, graphics, gratis, great, green, grid, gtd, gui, guide, guitar
Starting from h	Hack, hacking, hacks, hair, Halloween, halo, happiness, happy, hardware, Haskell, hci, hdr, health, healthcare, heart, Hebrew, help, het, hibernate, high, hip, hip-hop, history, holiday, home, hop, horror, hosting, hot, hotel, hotels, house, housing, how, howto, hp, html, http, human, humor, humour
Starting from i	I, ia, ibm, ical, icon, icons, ict, ide, idea, ideas, identity, ie, illustration, illustrator, im, image, images, imported, in, india, indie, info, informatica, information, innovation, inspiration, install, installation, insurance, intel, intelligence, interaction, interactive, interesting, interface, interior, international, internet, interview, investing, investment, ip, iphone, ipod, iptv, iran, Iraq, irc, Ireland, is, islam, island, Israel, it, italia, Italian, Italy, itunes
Starting from j	j2ee, jabber, jack, james, japan, Japanese, java, javascript, jazz, jesus, jewelry, job, jobs, john, joomla, journal, journalism, journals, jsf, json, juegos
Starting from k	Kernel, keyboard, kid, kids, king, kiss, knitting, knowledge, korea, korean
Starting from l	La, lake, landscape, language, languages, laptop, latex, latin, law, layout, learn, learning, leaves, lectures, legal, lego, lesbian, lessons, libraries, library, library2.0, libros, life, lifehack, lifehacker, lifehacks, lifestyle, light, lighting, lights, linguistics, link, links, linux, lisp, list, lists, literacy, literature, literature, little, live, local, logic, logo, lol, London, long, los, losangeles, love, lyrics
Starting from m	Mac, macbook, macintosh, macosx, macro, Madrid, magazine, magazines, magic, mail, make, man, management, manga, manual, mÃºsica, map, mapas, mapping, maps, market, marketing, mashup, math, mathematics, maths, mckay/Sheppard, me, media, medical, medicine, memory, men, menu, messaging, metadata, metal, mexico, Michael, microformats, Microsoft, midi, military, mind, mindmap, misc, mit, mix, mobile, model, modeling, models, modern, module, money, monitor, monitoring, motion, motiongraphics, motivation, mountain, movie, movies, Mozilla, mp3, multimedia, museum, music, Musica, musik, my, myspace, mysql
Starting from n	Naked, naruto, nasa, national, nature, navigation, nc-17, Netherlands, network, networking, networks, new, newmedia, news, newspaper, newspapers, newyork, night, Nikon, Nintendo, nlp, no, nokia, nonprofit, notes, noticias, November, nptech, nude, nutrition, nyc
Starting from o	Ocean, October, of, office, oil, old, on, one, online, ontology, open, opened, openoffice, opensource, open-source, opera, opinion, optimization, oracle, orange, organic, organization, origami, os, osx, out, outdoors, outlook, owl
Starting from p	p2p, painting, palm, paper, papers, parenting, paris, park, parody, parser, parsing, party, password, pattern, patterns, paul, pc, pda, pdf, peace, people, performance, perl, personal, personality, pet, pets, philosophy, phone, photo, photographer, photography, photos, photoshop, php, physics, piano, picture, pictures, pink, planning, plants, play, player, plugin, plugins, pocketpc, podcast, podcasting, podcasts, poetry, poker, Poland, police, policy, polish, politics, politik, pop, porn, portable, portal, portfolio, portrait, Portugal, post, power, powerpoint, pr, presentation, presentations, print, printing, privacy, process, processing, product, production, productivity, products, programming, project, projectmanagement, projects, property, prototype, proxy, psychology, public, publishing, punk, puppy, pussy, puzzle, python
Starting from q	quotes
Starting from r	r, race, racing, radio, rails, random, rap, rdf, read, reading, real, realestate, recherchÃ©, recipe, recipes, recording, records, recovery, recursos, red, reference, reflection, regex, religion, remix, remote, repair, research, resource, resources, rest, restaurant, restaurants, retro, review, reviews, rights, river, road, robot, robotics, robots, rock, roma, rome, rpg, rps, rss, ruby, rubyonrails, running, Russia, russian
Starting from s	Safari, safari_export, sam/dean, san, sanfrancisco, satellite, scary, scheme, school, science, scifi, Scotland, screen, script, scripting, scripts, sculpture, sea, search, searchengine, searchengines, seattle, secondlife, security, seguridad, self, semantic, semanticweb, semweb, seo, series, server, service, services, sewing, sex, sexy, sf, sga, share, sharepoint, sharing, shell, shoes, shop, shopping, short, show, simulation, singing, site, sky, skype, slash, sleep, slideshow, smallville, sms, snow, soa, soap, soccer, social, socialmedia, socialnetworking, socialnetworks, socialsoftware, society, sociology, software, solar, song, songs, sony, sound, source, south, space, spain, spam, Spanish, spears, speech, speed, spirituality, spn, sport, sports, spring, sql, ssh, standards, star, startup, starwars, statistics, stats, stock, stocks, storage, store, stories, story, strategy, streaming, street, streetart, studio, study, stuff, stupid, style, subversion, summer, sun, sunset, super, supernatural, support,

	sustainability, svn, Sweden, sweet, swing, Switzerland, symbian, sync, sysadmin, system
Starting from t	Tabs, tag, tagging, tags, Taiwan, teaching, tech, techno, technology, tecnologia, telephone, television, template, templates, terrorism, test, testing, texas, text, the, theme, themes, theory, thesis, time, tips, to, todo, Tokyo, tom, tool, tools, top, toread, Toronto, torrent, torrents, tour, tourism, toy, toys, trabajo, tracking, trading, traffic, trailer, train, training, translation, transport, transportation, travel, tree, trees, trends, tricks, trip, tuning, tutorial, tutorials, tutorials, tv, twitter, type, typography
Starting from u	Ubuntu, ui, uk, uml, uni, university, unix, unread, up, upload, urban, us, usa, usability, usb, useful, usenet, utilities, utility, ux
Starting from v	Vacation, validation, Vancouver, vc, vector, vegetarian, viajes, video, videogames, videos, vim, vintage, vinyl, viral, virtual, virtualization, vista, visual, visualization, vmware, voip, vs
Starting from w	w3c, wall, wallpaper, wallpapers, war, Washington, water, weather, web, web2.0, webapp, webcam, webcomic, webdesign, webdev, webdevelopment, weblog, webmaster, webservice, webservices, website, websites, webstandards, webtools, wedding, weird, white, widget, widgets, wifi, wii, wiki, Wikipedia, wikis, window, windows, wine, winter, wireless, wishlist, with, woman, women, wood, word, wordpress, words, work, workflow, world, wow, writing, wysiwyg
Starting from x	X, xbox, xhtml, xml, xp, xslt, xxx
Starting from y	Yahoo, yellow, York, you, young, your, youth, youtube
Starting from z	Zombie, zoo