

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 easier for end users to access resources and join in social networks. Social networks have pioneered online communities, allowing users to contribute to collective knowledge by tagging online resources. Tagging behavior increased dramatically between 2005 and 2007. This paper reports on an investigation of social tagging using data gathered from Delicious, Flickr and YouTube for the years 2005, 2006 and 2007. Preliminary findings indicate both that it is possible to profile a social network through the analysis of tagging data and that Delicious is a more representative venue for analyzing the social tagging behavior of users than either Flickr or YouTube.

Introduction

The web is undergoing a period of rapid transition from a space for presentation of syntactically formatted information to a more open and social platform that allows users to communicate knowledge and share resources. Pioneering technologies on the web make it easy for users to participate in social networks, leading to the development of online communities. In concert with Wikipedia, Google Map, RSS feeds, 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, Delicious, Flickr and YouTube are three of the best known and most popular. These networks established their prominent position either by hosting large numbers of online objects, by building a large user community, or by generating heavy web traffic. In 2007, Flickr hosted more than 2 billion images (Auchard, 2007); in 2006, Delicious 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 popularity of 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 Delicious, Flickr and YouTube for the years 2005, 2006 and 2007. It describes 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 each of these three social networks. It concludes with a discussion of the findings which indicates that it is possible to profile a social network through analysis of tagging data and that Delicious 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 Delicious, Flickr and YouTube and to store data about tags and tagging behavior in RDF triples based on the Upper Tag Ontology (UTO) (Ding, Toma, Kang, Fried & Yan, 2008; Ding, Kang, Toma, Fried & Yan, 2008).

The UTO tag crawler was developed using the Smart and Simple Webcrawler framework developed by Torunski (2008), which provides the functionalities of maximum interactions, maximum depth, filter interface, and pluggable 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. It includes two different parsers: The first parses a page and identifies links that should be visited or filtered out, while the second parses the HTML code to retrieve information about tags.

In Delicious, the crawler began with the 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 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. The crawler then repeated this process for TagB and all subsequent tags until all tags in the tag cloud had been visited.

For Flickr, the crawler accessed the tag cloud at <http://flickr.com/photos/tags> and visited each tag in the cloud. For each tag page (e.g., <http://www.flickr.com/photos/tags/party/>), information about related tags was collected. Each individual image on the tag page was visited and information about the image, the tags and the tagger was extracted. The crawling process continued with <http://www.flickr.com/photos/tags/party/?page=2>. To avoid duplicate visits, only links of 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. For each video page, it collected data about tags and taggers and then visited all links pointing to other video pages. In order to avoid visiting the same page more than once, query parts of links were ignored (e.g., <http://www.youtube.com/watch?v=X2IExa2A198> and http://www.youtube.com/watch?v=X2IExa2A198&watch_response lead to the same video). Figure 1 provides an overview of the UTO crawler. Detailed information about the UTO ontology and the UTO crawler is available in Ding, Toma et al. (2008) and Ding, Kang et al. (2008).

The UTO crawler was used to retrieve tagging data from Delicious, 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 was then converted to RDF triples based on the UTO ontology (Ding, Kang et al., 2008). In total, the crawler retrieved approximately 21 million RDF triples for Delicious, 2.3 million RDF triples for Flickr, and 2.2 million RDF triples for YouTube.

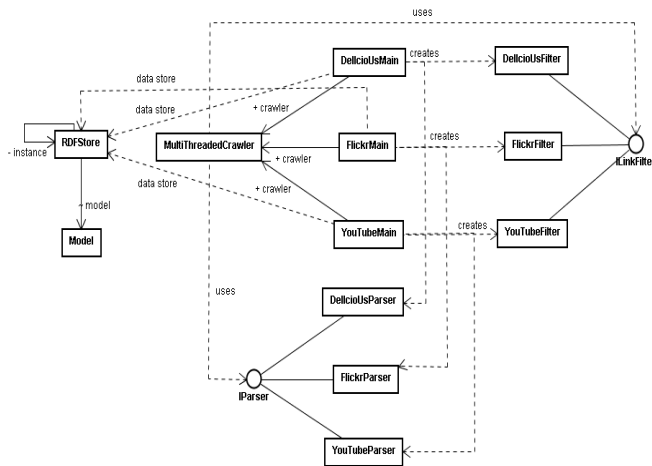


Figure 1. Overview of the Upper Tag Ontology (UTO) crawler

Table 1 presents an overview of the data collected from each social network. The total dataset contains around 1 million bookmarks, 2.8 million taggers and 9.3 million tags from Delicious; around 300,000 photographs, 150,000 taggers and 1.4 million tags from Flickr; and around 500,000 videos, 200,000 taggers and 1.35 million tags from YouTube. The average number of tags per object ranges from a low of 2.74 in YouTube to a high of 9.31 in Delicious. The average number of tags a tagger assigns ranges from a low of 3.33 in Delicious to a high of 8.79 in Flickr. The average number of objects a tagger tags ranges from 0.36 in Delicious to a high of 2.84 in YouTube. The seeming disparity reflected in the low average for objects tagged by taggers in Delicious is accounted for by the fact that, while users are required to provide a title when uploading bookmarks to Delicious, they are not required to include tags in the tag field. Thus there may be many bookmarks in Delicious that have titles but no tags. Combined data from the three social networks totals 1.8 million objects, 3.1 million taggers and 12.1 million tags, of which 648,368 tags are unique.

Table 1. Data collected from Delicious, Flickr and YouTube for the years 2005-2007.

Social Network	Objects	Taggers	Tags	Tag/Object	Tag/Tagger	Objects/Tagger
Delicious	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
Total	1,820,509	3,127,613	12,077,183	5.54	6.63	1.71

Note: Cells in the column labeled *Tags* represent the total number of tags assigned by taggers (e.g., when TagA is assigned by Tagger X and by Tagger Y, it is counted as two tags).

Power law distribution

Tagging data from Delicious, Flickr and YouTube was merged to form a single, comprehensive dataset. Using this combined dataset, an analysis of tag frequency was conducted. Figure 2 demonstrates that the distribution of tag frequency follows a power law distribution that conforms to Zipf's Law. Table 2 shows the details of this distribution: Only 1363 out of 648,368 unique tags (or approximately 0.2% of all tags assigned between 2005 and 2007) were assigned more than 1,00 times each, while 357,028 (or approximately 55% of all tags) were assigned only once.

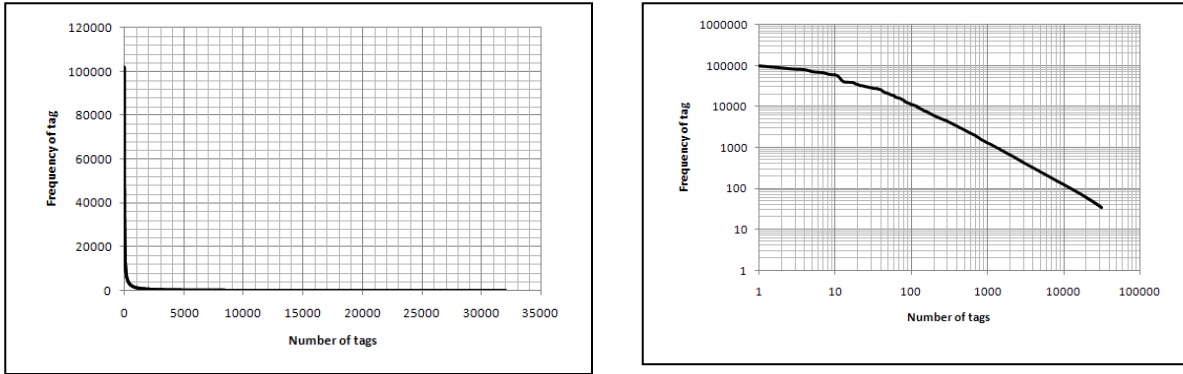


Figure 2. Distribution of tag frequency

In the combined dataset, the most frequently occurring tag is *design*, which accounts for 101,786 or nearly 1% of all tag occurrences. The second most frequently occurring tag is *blog* and accounts for 90,242 or 0.7% of the total tags assigned between 2005 and 2007. The 1,363 most frequently occurring tags account for a total of 6,210,163 tagging instances; these 1363 tags comprise a core tagging vocabulary that represents more than 50% of the entire corpus of 12,077,183 tagging instances. (See Appendix for a list of the 1363 tags that make up the combined core tagging vocabulary of Delicious, Flickr and YouTube). It is hoped that linguistic analysis of this core set of tags will be able to reveal features of the evolving vocabulary of tags in each social tagging network.

Table 2: Tag frequency distribution

Tag Frequency Range	No. of unique tags	Cumulative %
1	357028	55.07%
2-10	217746	88.65%
11-20	27404	92.88%
21-30	11524	94.65%
31-40	6656	95.68%
41-50	4454	96.37%
51-60	3387	96.89%
61-70	2461	97.27%
71-80	2066	97.59%
81-90	1597	97.83%
91-100	1348	98.04%
101-200	6193	99.00%
201-300	2151	99.33%
301-400	1044	99.49%
401-500	645	99.59%
501-1,000	1301	99.79%
1,001-120,000	1363	100.00%

Note: Cells in the column labeled *No. of unique tags* represent the total of unique tags (e.g., when TagA is assigned by Tagger X and by Tagger Y, it is counted as one tag).

Social Tagging Analysis

In order to generate individual portraits of tag use and the composition of tag vocabularies in Delicious, Flickr and YouTube, the data from each social network were analyzed independently using three time frames (2005, 2006, 2007).

Delicious

Table 3 shows the 20 most frequently assigned tags in Delicious 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 Delicious.

A profile of Delicious users can be generated through analysis of the lists of popular tags. The dominance of tags such as *blog*, *web*, *programming*, and *design* indicate key interests of Delicious users who are tagging bookmarks to store or share. While the tags *music*, *video*, *art* and *news* indicate a level of general interest that spans all three years, actual tagging evidence strongly supports the popular assumption that Delicious is a social network for individuals interested in the web and programming skills. Furthermore, the tags introduced in 2006 and 2007 indicate a growing interest in free or open source resources as well as tutorial and how-to resources that support learning programming languages or applications and developing new computer skills.

Table 3: Top 20 tags in Delicious for the years 2005, 2006 and 2007

Rank	2005	2006	2007
1	blog/blogs	blog/blogs	blog/blogs
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

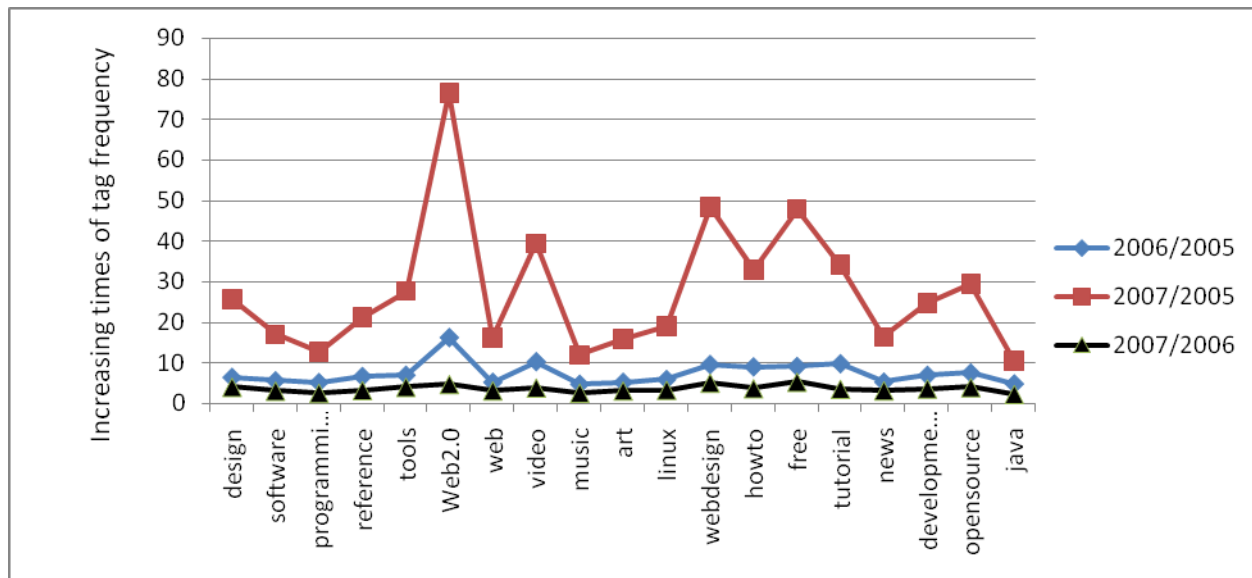


Figure 3. Evolution of the top 20 tags in Delicious for the period 2005-2007.

The line with diamonds represents the increase in tag frequency from 2005 to 2006 (tag frequency for 2006/tag frequency for 2005). The line with squares represents the increase in tag frequency from 2005 to 2007 (tag frequency for 2007/tag frequency for 2005). The line with triangles represents the increase in tag frequency from 2006 to 2007 (tag frequency for 2007/tag frequency for 2006).

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

Top 20 tags in Delicious for 2007	2005	2006	2007	2006/2005	2007/2005	2007/2006
blog/blogs	6731 (1)	29485 (1)	90474 (1)	4	13	3.1
design	3045 (5)	19273 (4)	78115 (2)	6	26	4.1
software	3558 (3)	19533 (3)	60405 (3)	5	17	3.1
programming	4295 (2)	21789 (2)	55237 (4)	5	13	2.5
reference	2541 (7)	16643 (5)	53971 (5)	7	21	3.2
tools	1943 (10)	13340 (8)	53772 (6)	7	28	4.0
Web2.0	658 (-)	10620 (12)	50270 (7)	16	76	4.7
web	2743 (6)	14115 (7)	44406 (8)	5	16	3.1
video	1114 (20)	11383 (11)	43847 (9)	10	39	3.9
music	3325 (4)	15523 (6)	39859 (10)	5	12	2.6
art	2344 (9)	12043 (9)	37518 (11)	5	16	3.1
linux	1799 (11)	10434 (13)	34241 (12)	6	19	3.3
webdesign	688 (-)	6542 (-)	33224 (13)	10	48	5.1
howto	962 (-)	8588 (16)	31701 (14)	9	33	3.7
free	643 (-)	5793 (-)	30750 (15)	9	48	5.3
tutorial	895 (-)	8683 (15)	30648 (16)	10	34	3.5
news	1712 (12)	8854 (14)	28086 (17)	5	16	3.2
development	1107 (-)	7588 (18)	27322 (18)	7	25	3.6
opensource	872 (-)	6468 (-)	25735 (19)	7	30	4.0
java	2449 (8)	11606 (10)	25732 (20)	5	11	2.2

Note: Numbers in parentheses in the columns labeled 2005, 2006 and 2007 reflect the ranking of a term for that particular year. The column labeled 2006/2005 indicates that the value in each cell is the result of dividing the value for 2006 by the value for 2005. The result indicates the increase in raw numbers of frequency of tag assignment from 2005 to 2006. This also applies to the columns labeled 2007/2005 and 2007/2006.

Figure 3 shows the evolution of dominant topical tags used in the Delicious social network for the period 2005-2007. The tag *Web2.0* shows the highest peak in both 2006 and 2007: The raw frequency with which *Web2.0* was

used to tag bookmarks increased 16 times in 2006 and 76 times in 2007 when compared with its raw tagging frequency in 2005. The tags showing the most dramatic increase in raw tagging frequency from 2006 to 2007 were *webdesign*, *free* and *Web2.0*, indicating growing interest in these topics on the part of Delicious taggers. The three tags with the least impressive increase in raw tagging frequency from 2006 to 2007 were *java*, *programming*, and *music*. While this might seem to indicate waning interest in these topics, only the ranking for *java*, which dropped from eighth most popular tag in 2005 to twentieth most popular in 2007 (Table 4), appears to support this conclusion. The tag *programming* drops from second position in 2005 and 2006 to fourth position in 2007; however, this is not a drop in popularity significant enough to justify any conclusions about waning interest on the part of Delicious taggers. The tag *music* does demonstrate a more dramatic drop in popularity -- from fourth position in 2005, to sixth position in 2006, and to tenth position in 2007 -- but the fact that Last.fm became one of the more popular social networks for sharing music during this period may help to explain why tagging with *music* decreased from 2005 through September 2007.

Flickr

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

Flickr taggers frequently assign informal tags to photographs (e.g., *me*), indicating that users may be tagging photographs for purposes of storing and retrieving them for their own use rather than with any intent to share them with others. When tagging photographs, 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*). Nonetheless, time (i.e., year, season or month), locations and colors are the major features of images tagged by users. It could be useful to analyze the tagging culture of Flickr in greater detail given that annotating images is an important area for image retrieval.¹

Table 5. Top 20 tags in Flickr for the years 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

¹ An interesting example of ongoing research on social annotation of images and videos is GWAP, the "games with a purpose" project at Carnegie Mellon available at <http://www.gwap.com/gwap/>

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 4 and Table 6 show the temporal history of tag popularity in Flickr for the period 2005-2007. In 2005 and 2006, tagging was not particularly popular in the Flickr community with total tags of 3,598 in 2005 and 23,066 in 2006. However, as tagging became more popular on the web, tagging behavior changed dramatically in Flickr. There were 1,324,537 tags assigned by Flickr taggers through September 2007, approximately 50 times more tags than were assigned for all of 2006. Raw tagging frequency for *canon*, the second most popular tag in 2007, increased 203.5 times over its total use in 2006; but *fall*, the thirteenth most popular tag in 2007, showed the greatest jump, increasing 672.5 times over its raw frequency of assignment in 2006.

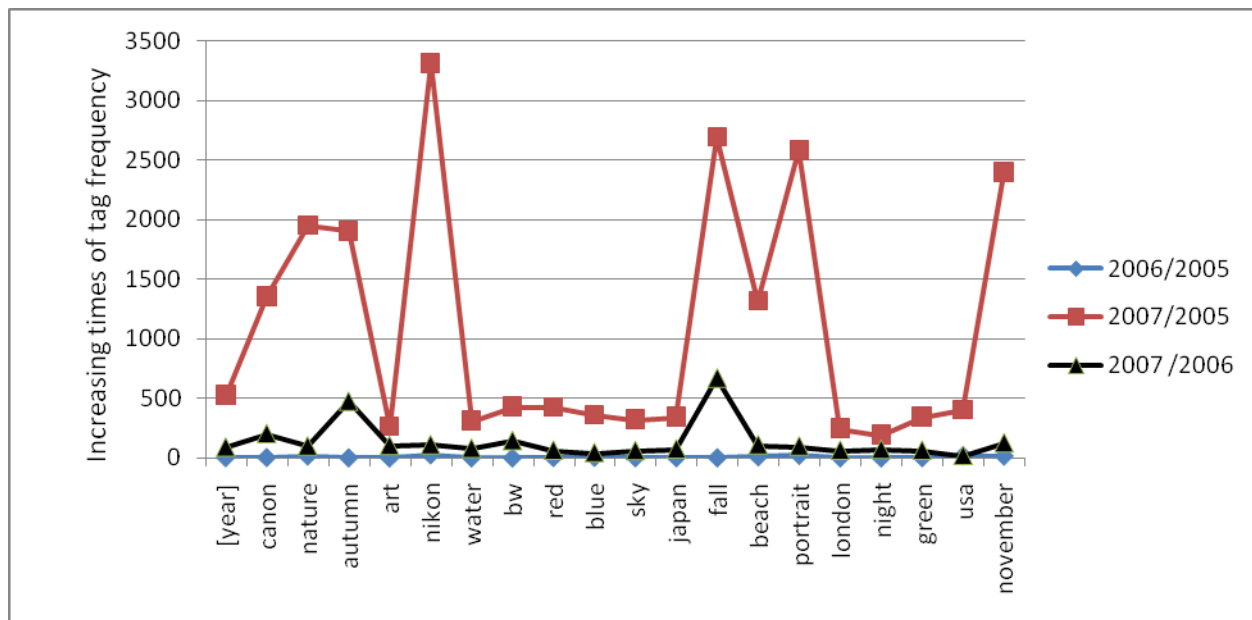


Figure 4. Evolution of the top 20 tags in Flickr for the period 2005-2007.

The line with diamonds represents the increase in tag frequency from 2005 to 2006 (tag frequency for 2006/tag frequency for 2005). The line with squares represents the increase in tag frequency from 2005 to 2007 (tag frequency for 2007/tag frequency for 2005). The line with triangles represents the increase in tag frequency from 2006 to 2007 (tag frequency for 2007/tag frequency for 2006).

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

Top 20 tags in Flickr for 2007	2005	2006	2007	2006/2005	2007/2005	2007 /2006
[year]	21 (1)	124 (3)	11112 (1)	6	529	89.6
canon	3 (-)	20 (-)	4070 (2)	7	1357	203.5
nature	2 (-)	39 (16)	3899 (3)	20	1950	100.0
autumn	2 (-)	8 (-)	3804 (4)	4	1902	475.5
art	13 (9)	33 (-)	3416 (5)	3	263	103.5
nikon	1 (-)	30 (-)	3312 (6)	30	3312	110.4
water	10	39 (15)	3126 (7)	4	313	80.2
bw	7 (-)	21 (-)	3028 (8)	3	433	144.2
red	7 (-)	47 (12)	2988 (9)	7	427	63.6
blue	8 (-)	66 (8)	2888 (10)	8	361	43.8

sky	9 (-)	48 (10)	2878 (11)	5	320	60.0
japan	8 (-)	37 (-)	2738 (12)	5	342	74.0
fall	1 (-)	4 (-)	2690 (13)	4	2690	672.5
beach	2 (-)	24 (-)	2636 (14)	12	1318	109.8
portrait	1 (-)	26 (-)	2581 (15)	26	2581	99.3
london	10 (17)	39 (18)	2503 (16)	4	250	64.2
night	13 (11)	35 (-)	2489 (17)	3	191	71.1
green	7 (-)	38 (19)	2417 (18)	5	345	63.6
usa	6 (-)	126 (1)	2406 (19)	21	401	19.1
november	1 (-)	19 (-)	2394 (20)	19	2394	126.0

Note: Numbers in parentheses in the columns labeled 2005, 2006 and 2007 reflect the ranking of a term for that particular year. The column labeled 2006/2005 indicates that the value in each cell is the result of dividing the value for 2006 by the value for 2005. The result indicates the increase in raw numbers of frequency of tag assignment from 2005 to 2006. This also applies to the columns labeled 2007/2005 and 2007/2006.

Interestingly, an analysis of tagged photographs indicates that there are two major communities of Flickr taggers: One community contains non-professional photographers who appear to use Flickr as a platform for sharing photographs with friends and family and tag images so that they can be retrieved by others; the second community consists of professional photographers who do not tag often but who frequently provide comments on photographs taken by other professionals.

YouTube

Table 7 shows the 20 most popular tags in YouTube for the years 2005, 2006 and 2007. The topics that are most frequently tagged in this social network are music, videos, humor, sex and girls, apparently 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	dancing	dance	live
9	live	rock	dj
10	AMV	cat	2007
11	fun	Halloween	dance
12	guitar	love	hot
13	hot	girl	comedy
14	girl	movie	rock
15	japan	dj	love
16	anime	in	and
17	Halloween	sexy	sex
18	cat	and	in
19	halo	fight	new
20	of	you	cat

Tagging activity in YouTube increased dramatically between 2005 and 2007. The total number of tags assigned in YouTube increased from 4,735 in 2005, to 366,147 in 2006, to 1,073,042 in 2007: Tag use was 78.7 times greater

in 2006 and 236.7 times greater in 2007 than it was in 2005. Compared with 2005, the tag *[year]* had the greatest increase in use in 2007, followed by *new* and *sex/sexy*, while *dance* showed the least increase between 2005 and 2007. The tag set in YouTube appears to be more stable than that of 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 5 and Table 8).

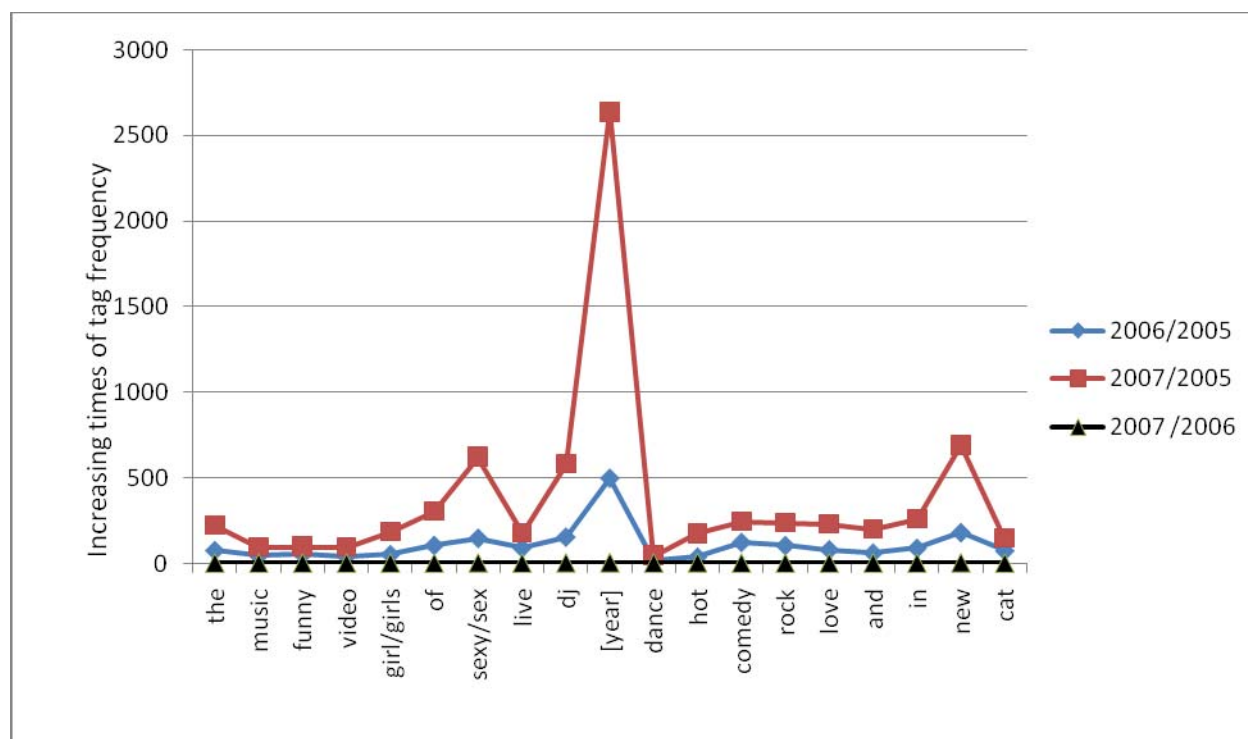


Figure 5. Evolution of the top 20 tags in YouTube for the period 2005-2007.

The line with diamonds represents the increase in tag frequency from 2005 to 2006 (tag frequency for 2006/tag frequency for 2005). The line with squares represents the increase in tag frequency from 2005 to 2007 (tag frequency for 2007/tag frequency for 2005). The line with triangles represents the increase in tag frequency from 2006 to 2007 (tag frequency for 2007/tag frequency for 2006).

Table 8. Top 20 tags in YouTube for 2007 and their frequency of assignment in 2005, 2006 and 2007.

Top 20 Tags in Youtube for 2007	2005	2006	2007	2006/2005	2007/2005	2007 /2006
the	42 (4)	3240 (1)	9371 (1)	77	223	2.9
music	67 (1)	3080 (3)	6452 (2)	46	96	2.1
funny	58 (2)	3091 (2)	5784 (3)	53	100	1.9
video	53 (3)	2234 (3)	5065 (4)	42	96	2.3
girl/girls	25 (14)	1334 (13)	4647 (5)	53	186	3.5
of	13 (20)	1390 (6)	3955 (6)	107	304	2.8
sexy/sex	9 (-/-)	1338 (17/-)	5601 (7/17)	149	622	4.2
live	17 (9)	1563 (5)	3028 (8)	92	178	1.9
dj	5 (-)	777 (15)	2920 (9)	155	584	3.8
[year]	1 (-)	498 (-)	2641 (10)	498	2641	5.3
dance	56 (5)	1061 (8)	2526 (11)	19	45	2.4
hot	14 (13)	552 (-)	2467 (12)	39	176	4.5
comedy	10 (-)	1245 (7)	2461 (13)	125	246	2.0
rock	10 (-)	1059 (9)	2380 (14)	106	238	2.2
love	10 (-)	817 (12)	2294 (15)	82	229	2.8
and	11 (-)	689 (18)	2190 (16)	63	199	3.2

in	8 (-)	723 (16)	2095 (18)	90	262	2.9
new	3 (-)	544 (-)	2079 (19)	181	693	3.8
cat	13 (18)	977 (10)	1906 (20)	75	147	2.0

Note: Numbers in parentheses in the columns labeled 2005, 2006 and 2007 reflect the ranking of a term for that particular year. The column labeled 2006/2005 indicates that the value in each cell is the result of dividing the value for 2006 by the value for 2005. The result indicates the increase in raw numbers of frequency of tag assignment from 2005 to 2006. This also applies to the columns labeled 2007/2005 and 2007/2006.

Summary and Conclusion

When comparing these three social networks, Delicious demonstrates the tightest connection to the use of tags as extended information about resources. In Delicious, every user can tag an object with the tag(s) of his own choice; and an object can be tagged many times and by many different users, thereby indicating that it “belongs” (or is highly relevant) to the Delicious community as a whole. Delicious exemplifies community tagging where anyone can tag (or bookmark) any online resource (Marlow et al., 2006). Other similar social networks include CiteULike and Connotea, where tagged resources are bibliographical records, and LibraryThing, where tagged resources are books.

Social networks such as Delicious, CiteULike, and LibraryThing are very different from Flickr, where a resource (photograph) is generally tagged only by the individual who uploads it. The major activity of other members of the Flickr community is to “comment on” or “vote for” resources by indicating that a particular photograph is a favourite image. Flickr also provides users with the ability to allow friends to tag photos they have uploaded; but this functionality limits tagging behavior and thus the development of a sense of community in that it prohibits open tagging by Flickr users at large. Because tagging a resource in Flickr is not generally open to everyone, Flickr cannot be considered a true community-based tagging system; rather, it is better thought of as a self-tagging system for users and their close friends. YouTube operates in a manner very similar to that of Flickr, allowing an individual to tag the resources (videos) he has uploaded while limiting the participation of other Flickr users to voting for resources by assigning “stars”.

These differences in tagging rights have created differences not only in the role tags play in each system but also in the nature of the tags that are assigned (Marlow et al., 2006). Based on analyses of the top 20 tags in each of the three social networks, it is apparent that tags in Delicious are more content-oriented in that they are generally related to the topics of the resources bookmarked. The tags used in Flickr are more annotation-oriented in that they are generally related to the physical features of the photographs themselves, such as colors, lighting and location. While tags in Delicious are likely to reflect the intellectual content of resources and those in Flickr generally represent the physical features of photographs, tags in YouTube tend to focus on the medium or genre of resources (e.g., *music*, *video*, *comedy*, *movie*, *tv*) and on affective judgments (e.g., *funny*, *sexy*, *hot*, *love*, *new*).

The role of tags in Delicious is to represent bookmarked resources not only for future retrieval but also for sharing them with the larger community. Tags play a major role in Delicious: Without the tags assigned by users of the social network, there would be no means either to share bookmarks or to identify and retrieve resources, which are the main functions of Delicious. In contrast, tagging does not play a major role in Flickr. Because the decisions as to whether or not to tag a photograph and who may tag it are left to the individual uploading a photograph, tagging in Flickr is more of a secondary activity or side effect. Furthermore, photographs on Flickr can be searched for and retrieved by their titles and are ranked by comments or votes rather than by the number of tags assigned. This is also the case with YouTube in that videos are most frequently shared based on comments and votes rather than assigned tags. Indeed, it appears that many YouTube users may not understand the purpose of tagging: Instead of adding specific tags, users often enter descriptions of their videos in the tagging field, which accounts for the occurrence of helping words such as articles, prepositions and conjunctions (e.g., *the*, *of*, *in*, *and*).

among the more popular tags in YouTube. Table 9 summarizes the characteristics of social networks that were identified in the analysis of Delicious, Flickr and YouTube.

Social tagging behaviors are also related to the community of users in each social network. Delicious gathers a community interested in IT-related topics. These individuals are interested in the content of bookmarked resources and tagging provides a way for them to summarize this content. In such a situation, tagging becomes the key function of the system and plays a major role in sharing and retrieving bookmarks. Users of the Flickr are more interested in commenting on and sharing their photographs with family and friends. Thus, rather than comprising a single, cohesive community, users in Flickr appear to be divided into two primary communities: Professional photographers who upload photographs for comment and feedback from other professionals and non-professional users for whom Flickr provides a place to store personal photographs and share them with close friends. Alternatively, the community of YouTube can be viewed as a snapshot of the entire Web community. YouTube is populated by individuals from all over the world who are of different ages and have many different interests. They come to YouTube with many different purposes and expectations, and many of them do not tag their videos because the role of tagging is overshadowed by rating and commenting.

After analyzing social tagging behavior in Delicious, Flickr and YouTube, it is apparent that tagging activities have increased tremendously from 2005 to 2007. More and more individuals are using online social networks to tag resources for purposes of storage, access, and retrieval, both for themselves and for the purpose 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 emerging or waning topical interests among users.

Table 9. Summary of social tagging in Delicious, Flickr and YouTube

Features	Delicious	Flickr	YouTube
Community	People interested in sharing bookmarks about the web, programming, etc.	Professional and non-professional photographers interested in sharing photographs	People interested in sharing videos on any subject.
Main Topics	IT-related resources, video, music and news	Features of photographs, including colors, locations, years, seasons	Genres (music, humor) and affective responses (sexy, hot, new)
Tagging Behavior	A key activity with many users participating in tagging activities	Few users participate in tagging	Few users participate in tagging
Dynamic	Yes	No	No
Emerging Topics (Trends)	Web design, tutorial, web2.0	None apparent	None apparent
Declining Topics	Internet, research, xml, security, java	None apparent	None apparent
Stability	Set of most popular tags changes 10%-20% each year on average	Tag set changes but some tag categories (color, location, year) are relatively stable	Tag set changes but tag categories (genre, affective reaction) are relatively stable
Tagging Vocabulary	Relatively stable	Stable categories of tags	Stable categories of tags
Tagging Focus	Content of the bookmarked resources	Feature(s) of photographs	Genres, affective responses to videos
Primary Tagging Purpose	Accessing and sharing	Storing and retrieving	Describing and rating

While tag sets in Delicious appeared to become more stable across the time frame of this study, it was also apparent that collective tagging vocabularies could benefit from both syntactic and semantic normalization of

tags: For example, in YouTube in 2007 there were 2,796 uses of the tag *girl* and 1,851 uses of the tag *girls*. Normalization of singular and plural forms as well as acronyms and full names would increase the effectiveness of tags for retrieval purposes, as would standardization of the syntactical formation of tags (e.g., tag phrases with or without a space between individual terms). Perhaps as important is the introduction of user education regarding the choice of tags and their potential utility in social networks (Ackerman, James & Getz, 2007).

This study demonstrates that it is possible to profile a social network by analyzing data about tags and tagging behaviors in social networks. Thus, analysis confirms the popular assumption that the Delicious community is largely comprised of individuals interested in IT-oriented topics such as design and programming. In contrast, the Flickr community appears to contain two primary groups of users: professional photographers interested in feedback and non-professional photographers interested in sharing photographs with family and friends. In contrast to Delicious and Flickr, the YouTube community is very broad and can be best viewed as a self-selected subset of the general social web community. Tagging is a major activity in Delicious but not in Flickr and YouTube. Tagging in Delicious 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 in YouTube is undertaken primarily for identifying the genre of a video and for indicating the tagger's affective reaction to it. Taggers want to represent the content of a resource in Delicious, but they tend to focus on the specific features of an image in Flickr and the genre of a video in YouTube.

In Delicious, changing trends in user interests can be identified and tracked by analyzing tag frequencies across time; in both Flickr and YouTube, however, such trends are not obvious, perhaps because the focus of tagging activities is not on the intellectual content of resources but on more superficial features such as color (in Flickr) or affective reactions (in YouTube). Thus, even though YouTube has been characterized as a subset of the general web population, the results of this research indicate that Delicious is a more representative venue for analyzing social tagging vocabularies and the tagging behaviors of users. This conclusion is supported by the finding that the community of users in Delicious is more cohesive than in Flickr or YouTube; by the dynamic behavior of users that supports tracking of emerging and waning interests within the Delicious community; and by the participatory focus on sharing that characterizes user tagging activity in Delicious.

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

- Ackerman, G., James, M., & Getz, C. T. (2007). The application of social bookmarking technology to the national intelligence domain. *International Journal of Intelligence and Counterintelligence*, 20, 678-698.
- Auchard, E. (2007, November 19). Flickr to map the world's latest photo hotspots. *Reuters*. Retrieved September 30, 2008, from <http://www.reuters.com/article/technologyNews/idUSHO94233920071119>
- Ding, Y., Kang, S., Toma, I., Fried M., & Yan, Z. (2008). Integrating Social Tagging Data: Upper Tag Ontology. *Proceedings of the 2008 IEEE International Conference on Systems, Man and Cybernetics, Singapore*.
- Ding, Y., Toma, I. Kang, S., Fried, M., & Yan, Z. (2008). Data Mediation and Interoperation in Social Web: Modeling, Crawling and Integrating Social Tagging Data. *Proceedings of the Workshop on Social Web Search and Mining (SWSM2008), 17th International World Wide Web Conference, Beijing, China*.

Gomes, L. (2006, August 30). Will All of Us Get Our 15 Minutes On a YouTube Video? *The Wall Street Journal*. Retrieved September 26, 2008, from <http://online.wsj.com/public/article/SB115689298168048904.html>

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, November 3-8, 2006, Austin, Texas*. [S.I.]: Richard B. Hill. Available from <http://eprints.rclis.org/archive/00008315/>

Li, X., Guo, L., & Zhao, Y. (2008). Tag-based social interest discovery. In *Proceedings of the 17th International World Wide Web Conference*, April 21-25, 2008, Beijing, China (pp. 675-684). Retrieved January 31, 2009, from <http://www2008.org/papers/pdf/p675-liA.pdf>

Lin, Y., Chi, Y., Zhu, S., Sundaram, H., & Tseng, B. (2008). FacetNet: A framework for analyzing communities and their evolutions in dynamic networks. In *Proceedings of the 17th International World Wide Web Conference*, April 21-25, 2008, Beijing, China (pp. 685-694). Retrieved January 31, 2009, from <http://www2008.org/papers/pdf/p685-linA.pdf>

Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). HT06, tagging paper, taxonomy, flickr, academic article, to read. In *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia*, August 22-25, 2006, Odense, Denmark (pp. 31-40). New York: ACM. Available from <http://portal.acm.org/citation.cfm?id=1149949>

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

Robinson, B. (2006, September 25). Del.icio.us reports 1 million users – post Yahoo! growth tops all of Digg. Message posted to <http://www.techcrunch.com/2006/09/25/del.icio.us-reports-1-million-users-post-yahoo-growth-tops-all-of-digg/>

Singla, P., & Richardson, M. (2008). Yes, there is a correlation – From social networks to personal behavior on the web. In *Proceedings of the 17th International World Wide Web Conference*, April 21-25, 2008, Beijing, China (pp. 655-664). Retrieved January 31, 2009, from <http://www2008.org/papers/pdf/p655-singla.pdf>

Torunski, L. (2008). Smart and simple Web crawler (Version 1.1) [Computer software]. Santa Clara, CA: Sun Microsystems. Available from <https://crawler.dev.java.net>

Appendix

Core tag vocabulary in social networks: Top 1363 tags in Delicious, Flickr and YouTube

	Core Tag Vocabulary
Numbers & others	1, 2, 3, 2005, 2006, 2007, -, .net, 3d
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
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
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
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
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
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
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
H	Hack, hacking, hacks, hair, Halloween, halo, happiness, happy, hardware, Haskell, hci, hdr, health, healthcare, heart, Hebrew, help, het, hibernate, high, hip, hiphop, history, holiday, home, hop, horror, hosting, hot, hotel, hotels, house, housing, how, howto, hp, html, http, human, humor, humour
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
J	j2ee, jabber, jack, james, japan, Japanese, java, javascript, jazz, jesus, jewelry, job, jobs, john, joomla, journal, journalism, journals, jsf, json, juegos
K	Kernel, keyboard, kid, kids, king, kiss, knitting, knowledge, korea, korean
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
M	Mac, macbook, macintosh, macosx, macro, Madrid, magazine, magazines, magic, mail, make, man, management, manga, manual, mÃ¡sica, 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, mspace, mysql
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
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
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
Q	quotes
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
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
T	Tags, 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
U	Ubuntu, ui, uk, uml, uni, university, unix, unread, up, upload, urban, us, usa, usability, usb, useful, usenet, utilities, utility, ux
V	Vacation, validation, Vancouver, vc, vector, vegetarian, viajes, video, videogames, videos, vim, vintage, vinyl, viral, virtual, virtualization, vista, visual, visualization, vmware, voip, vs
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

X	X, xbox, xhtml, xml, xp, xslt, xxx
Y	Yahoo, yellow, York, you, young, your, youth, youtube
Z	Zombie, zoo