A Comparison of Social Bookmarking with Traditional Search

Beate Krause[12], Andreas Hotho[1], and Gerd Stumme[12]

[1] Knowledge & Data Engineering Group, University of Kassel
Wilhelmshöher Allee 73, D-34121 Kassel, Germany
[2] Research Center L3S, Appelstr. 9a, D-30167 Hannover, Germany

Abstract. Social bookmarking systems allow users to store links to internet resources on a web page. As social bookmarking systems are growing in popularity, search algorithms have been developed that transfer the idea of link-based rankings in the Web to a social bookmarking system's data structure. These rankings differ from traditional search engine rankings in that they incorporate the rating of users.
In this study, we compare search in social bookmarking systems with traditional Web search. In the first part, we compare the user activity and behaviour in both kinds of systems, as well as the overlap of the underlying sets of URLs. In the second part, we compare graph-based and vector space rankings for social bookmarking systems with commercial search engine rankings.
Our experiments are performed on data of the social bookmarking system Del.icio.us and on rankings and log data from Google, MSN, and AOL. We will show that part of the difference between the systems is due to different behaviour (e. g., the concatenation of multi-word lexems to single terms in Del.icio.us), and that real-world events may trigger similar behaviour in both kinds of systems. We will also show that a graph-based ranking approach on folksonomies yields results that are closer to the rankings of the commercial search engines than vector space retrieval, and that the correlation is high in particular for the domains that are well covered by the social bookmarking system.

Key words: social search, folksonomies, search engines, ranking

1 Introduction

Collaborative tagging systems such as Del.icio.us[3], BibSonomy[4], or Flickr[5] have become popular among internet users in the last years. Taggers actively index and describe Web resources by adding keywords to interesting content and storing them in a so-called *folksonomy* on a shared platform. Over the last years, a significant number of resources has been collected, offering a personalized, community driven way to search and explore the Web.

As these systems are growing, the currently implemented navigation by browsing tag clouds with subsequent lists of bookmarks that are represented in chronological order may not be the best arrangement for concise information retrieval.

[3] `http://del.icio.us/` [4] `http://www.bibsonomy.org/` [5] `http://flickr.com/`

Therefore, a first ranking approach based on the graph structure of the underlying system was proposed in [8].

In this paper, we will compare search in social bookmarking systems with traditional Web search. After a brief presentation of related work (Section 2) and of the used datasets (Google, MSN, AOL and Del.icio.us; Section 3) we will concentrate on an analysis of tagging and traditional search behaviour considering tagging and search interest: Are query terms and tags used similarily (Section 4.1)? Is tagging and search behaviour correlated over time (Section 4.2)? How strong is the overlap of the content in social bookmarking systems and search engines (Section 4.3)?

In Section 5, we turn to the comparison of the different ranking paradigms. We compare graph-based and vectors space rankings for social bookmarking systems with the rankings of commercial search engines.

2 Related Work

Search engine rankings and folksonomies have been analyzed separately in several studies. Different aspects of search were classified by [4]. In [6], temporal correlation based on the Pearson correlation coefficient is used to find similar queries. [1] calculated cross-correlation and dynamic time warping to visualize rises and falls of different terms in blogs, search engine click data and news. In [13], time series data from query logs of the MSN search engine is analyzed. A comparison of traditional search engine rankings using correlation coefficients was carried out by [3].

The vision of folksonomy-based systems and a first analysis of Del.icio.us is presented in [7]. Several studies consider social annotations as a means of improving web search. [10] conducted a user study to compare the content of social networks with search engines. [2, 14] propose to use data from social bookmarking systems to enhance Web search: [2] introduces two algorithms to incorporate social bookmarking information into Web rankings. [14] considers popularity, temporal and sentiment aspects. In [8], two of the authors presented a ranking algorithm for folksonomies, the FolkRank. It adopts the idea of PageRank [11] to the structure of folksonomies. To the best of our knowledge no work examines differences and similarity of user interactions with folksonomy and search engine systems, coverage and rankings as done in this work.

3 Experimental Setup

3.1 Basic notions

Tags in Folksonomies. The central data structure of a social bookmarking system like Del.icio.us is called *folksonomy*. It can be seen as a lightweight classification structure which is built from *tag* annotations (i. e., freely chosen keywords) added by different users to their resources. A folksonomy consists thus of a set of users, a set of tags, and a set of resources, together with a ternary relation between them.

Query Terms in Search Engines. For comparing tagging and search behaviour, we need similar structures on both sides. In the search engine log data,

Table 1. Overview of datasets

Dataset Name	Date	Terms/Tags	Nb. of different URLs
Del.icio.us 2005	until July 2005	456,697	3,158,435
Del.icio.us May only	May. 06	375,041	1,612,405
Del.icio.us complete	until Oct. 06	2,741,198	17,796,405
MSN click data	May 06	2,040,207	14,923,285
MSN crawl	Oct. 06	29,777	19,215,855
AOL click data	March - May 06	1,483,186	19,215,858
Google crawl	Jan. 07	34,220	2,783,734

we will therefore split up each query into the terms that constitute it. *Query terms* are thus all substrings of a query that are separated by blanks.

Items. We will use the term *item* to subsume tags and query terms.

3.2 Data collection

We consider a MSN log data set and data of the social bookmarking system Del.icio.us to compare the search behaviour with tagging. To compare folksonomy rankings to search engine rankings, we use crawls of commercial search engines (MSN, Google). A log dataset from AOL [12] is further used to find out about overlaps between different systems. Table 1 presents an overview of the datasets' dates, numbers of queries and numbers of different URLs.

Social bookmarking data. In summer 2005 and November 2006 we crawled Del.icio.us to obtain a comprehensive social bookmarking set with tag assignments from the beginning of the system to October 2006. Based on the time stamps of the tag assignments, we are able to produce snapshots. In this paper, we use a snapshot of May 2006 for Section 4 and the entire dataset to compute rankings in Section 5. The first 40,000 tags of the Summer 2005 dataset served as queries in our search engine crawls.

Click data. We obtained a click data set from Microsoft for the period of May 2006. To make it comparable to tags, we decomposed a query into single query terms, removed stop words and normalized them. Sessions which contained more than ten queries with the same query terms in a row were not included into our calculations. A second click data set was obtained from AOL.

Search engine data. Two crawls from MSN and Google are used. While we retrieved 1000 URLs for each query in the MSN dataset, we have 100 URLs for each query in Google.

All query terms and all tags were turned to lowercase.

4 Tagging and Searching

Both search engines and bookmarking systems allow users to interact with the Web. In both systems, the fundamental resources are URLs. In search engines, a user's information need is encoded in a query being composed of one or more terms. In social bookmarking systems, the users themselves assign in a proactive fashion the tags – which later will be used in searches — to the resources.

Table 2. Statistics of Del.icio.us and MSN in May, 2006

	MSN	**Del.icio.us**	**MSN - Del.**
items	31,535,050	9,076,899	—
distinct items	2,040,207	375,041	96,988
average	15.46	24.20	—
frequent items	115,966	39,281	18,541
frequent items containing "_"	90	1,840	1
frequent items containing "-"	1,643	1,603	145
frequent items cont. "www.", ".com", ".net" or ".org"	17,695	136	30

In this section, we will compare the search and tagging behaviour. Search behaviour is described by the query terms submitted to a search engine. We use the number of occurrences of a term in the queries of a certain period of time as an indicator for the users' interests. The interests of taggers are described by the tags they assigned to resources during a certain period. We start our exploration with a comparison of the overlap of the set of all query terms in the MSN log data with the set of all tags in Del.icio.us in May 2006. This comparison is followed by an analysis of the correlation of search and tagging behaviour in both systems over time. Query log files were not available for bookmarking systems, hence we study the tagging (and not the search) behaviour only.

The section ends with an analysis of the coverage of URLs considering again the bookmarking system Del.icio.us, and the search engines Google, MSN and AOL. As we do not have access to the indexes of the search engines, we approximate their content by the results of the most prominent queries.

4.1 Query Term and Tag Usage Analysis

By comparing the distribution of tags and query terms we will get some first insights into the usage of both systems. The overlap of the set of query terms with the set of tags is an indicator of the similarity of the usage of both systems. We use the Del.icio.us data from May 2006 to represent social bookmarking systems and the MSN 2006 click data to represent search engines.

Table 2 shows statistics about the usage of query terms in MSN and tags in Del.icio.us. The first row reflects the total number of queried terms, and the total number of used tags in Del.icio.us. The following row shows the number of distinct items in all systems. As can be seen, both the total number of terms and the number of distinct terms is significantly larger in MSN compared to the total number of tags and the number of distinct tags in Del.icio.us. Interestingly, the average frequency of an item is quite similar in all systems (see third row). These numbers indicate that Del.icio.us users focus on fewer topics than search engine users, but that each topic is, in average, equally often addressed.

Figure 1 shows the distribution of items in both systems on a log-log scale. The x-axis denotes the count of items in the data set, the y-axis describes the number of tags that correspond to the term/tag occurrence number. We observe a power law in both distributions.

Power law means in particular that the vast majority of terms appears once or very few times only, while few terms are used frequently. This effect also explains

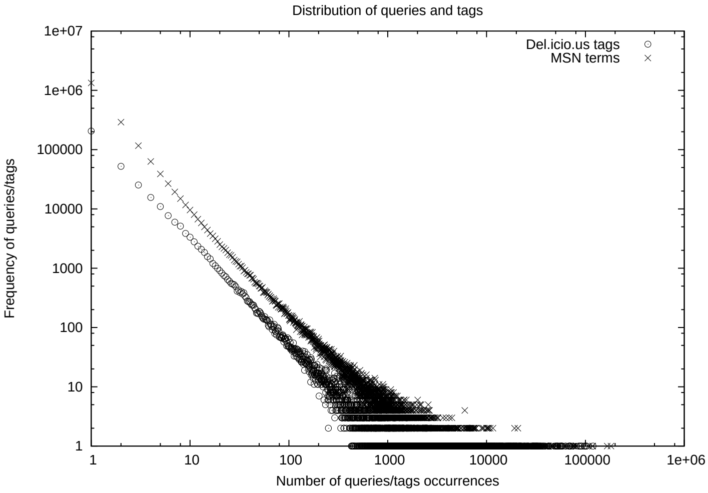

Fig. 1. Item distribution

the relatively small overlap of the MSN query terms with the Del.icio.us terms, which is given in the 2nd row/3rd column of Table 2. In order to analyse the overlap for the more central terms, we restricted both sets to query terms/tags that showed up in the respective system at least ten times.[6] The resulting frequencies are given in the first line of the second part of Table2. It shows that the sizes of the reduced MSN and Del.icio.us datasets become more equal, and that the relative overlap increases.

When browsing both reduced data sets, we observed that the non-overlapping parts result very much from the different usages of both systems. In social bookmarking systems, for instance, people frequently encode multi-word lexems by connecting the words with either underscores, hyphens, dots, or no symbol at all. (For instance, all of the terms 'artificial_intelligence', 'artificial-intelligence', 'artificial.intelligence' and 'artificialintelligence' show up at least ten times in Del.icio.us). This behaviour is reflected by the second and third last rows in Table 2. Underscores are basically used for such multi-word lexemes only, whereas hyphens occur also in expressions like 'e-learning' or 't-shirt'. Only in the latter form they show up in the MSN data.

A large part of the query terms in MSN that are not Del.icio.us tags are URLs or part of URLs, see the last row of Table 2. This indicates that users of social bookmarking systems prefer tags that are closer to natural language, and thus easier to remember, while users of search engines (have to) anticipate the syntactic appearance of what they are looking for.

The top five tags of Del.icio.us and the top five terms of MSN in May 2006 can be seen in Table 3 with their frequencies. One can see that Del.icio.us has a strong bias towards IT related terms. Eleven of the 20 top tags are computer terms (such as web, programming, ajax or linux). The top terms of MSN are more difficult to interpret. "yahoo" and "google" may be used when people have the MSN search

[6] The restriction to a minimum of 5 or 20 occurrences provided similar results.

Table 3. Top items in May 2006

Tags Del	Frequency	Query terms MSN	Frequency
design	119,580	yahoo	181,137
blog	102,728	google	166,110
software	100,873	free	118,628
web	97,495	county	118,002
reference	92078	myspace	107,316

interface as a starting point in their internet explorer, or when they leave Microsoft related programs such as hotmail, and want to use another search engine. “county” is often part of a composed query such as “Ashtabula county school employees credit union” or “county state bank”. We lack a good explanation for the high frequency of this term. This might result from the way Microsoft extracted the sample (which is unknown to us).

4.2 Correlation of Search and Tagging Behaviour over Time

Up to now, we have considered both data collections as static. Next we analyze if and how search and tagging behaviour are correlated over time. Again we use the MSN query data and the Del.icio.us data of May 2006. Each data set has been separated into 24-hour bins, one for each day of May 2006. As the unit of analysis, we selected those tags from Del.icio.us that also appeared as a query term in the MSN click data. In order to reduce sparse time series, we excluded time series which had fewer than five daily query or tagging events. In total, 1003 items remained.

For each item i, we define two time series. The Del.icio.us time series is given by $X_i^d = (x_{i,1}^d, ..., x_{i,31}^d)$, where $x_{i,t}^d$ is the number of assignments of tag i to some bookmark during day $t \in \{1, \ldots, 31\}$. For MSN, we define $X_{i,t}^m = (x_{i,1}^m, ..., x_{i,31}^m)$, where $x_{i,t}^m$ is the number of times this term was part of a query on day t according to the MSN data.

To reduce seasonal effects, we normalized the data. We chose an additive model for removal of seasonal variation, i. e., we estimated the seasonal effect for a particular weekday by finding the average of each weekday observation minus the corresponding weekly average and substracted this seasonal component from the original data [5]. The model underlies the assumption that no substantial (i. e., long-term) trend exists which otherwise would lead to increasing or decreasing averages over time. As our time period is short, we assume that long term trends do not influence averages. We also smoothed the data using simple average sine smoothing [9] with a smoothing window of three days to reduce random variation. Other smoothing techniques delivered similar results.

In order to find out about the similarity of the two time series of an item i, we used the correlation coefficient between the two random variables $x_{i,t}^d$ and $x_{i,t}^m$ which is defined as $r = \frac{\sum_t (X_{i,t}^d - \mu(X_i^d))(X_{i,t}^m - \mu(X_i^m))}{\sigma(X_i^d)\sigma(X_i^m)}$ where $\mu(X_i^d)$ and $\mu(X_i^m)$ are the expected values and $\sigma(X_i^d)$ and are $\sigma(X_i^m)$ the standard deviations.

We applied the t-test for testing significance using the conventional probability criterion of .05. For 307 out of 1003 items, we observed a significant correlation. We take this as indication that tagging and searching behaviour are indeed triggered by similar motivations.

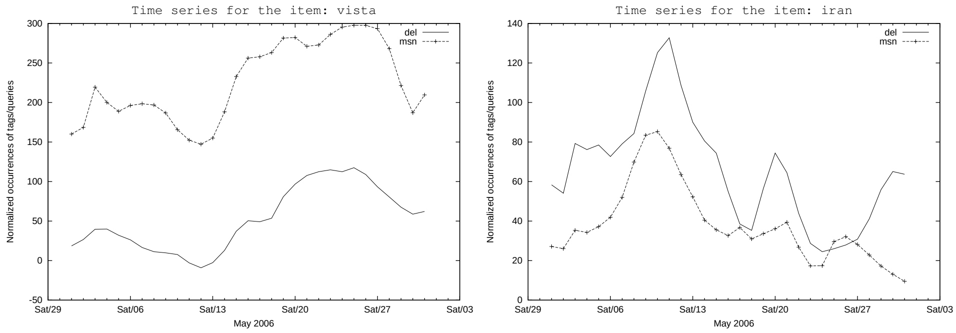

Fig. 2. Time series of highly correlated items

The highest correlation has the item 'schedule' ($r = 0.93$), followed by 'vista' ($r = 0.91$), 'driver', 'player' and 'films'. While both 'schedule' time series are almost constant, the following item 'vista' has a higher variance, since a beta 2 version of Microsoft's Vista operating system was released in May 2006 and drew the attention of searchers and taggers. The 'vista' time series are given in the left of Figure 2. Another example where the peaks in the time series were triggered from an information need after a certain event is "iran" ($r = 0.80$), which has the 19th highest correlation of all tags. The peaks show up shortly after the confirmation of the United States White House that Iran's president sent a letter to the president of the US on May 08, 2006; and are strongly correlated. A similar peak for 'iran' can be observed in Google Trends[7] showing Google's search patterns in May 2006. These examples support the hypothesis that popular events trigger both search and tagging close to the event.

4.3 Coverage of Del.icio.us with MSN, Google and AOL

In this section we shift our focus from query terms and tags to the underlying resources, i. e., the URLs. Considering today's size of the Web, both search engines (in particular the part we can crawl) and folksonomies constitute only a small fraction of the Web. An interesting question is thus if there is any significant overlap between the URLs provided by both systems.

To compare the coverage of the different data sets, we compute the overlaps between MSN crawl, Google crawl, AOL click data and the Del.icio.us dataset of October 2006. As we had no access to the indices of the search engines, we crawled all search engines with 1,776 queries to obtain comparable datasets. These queries were determined by taking the 2000 most popular tags of the Del.icio.us 2005 dataset and intersecting them with the set of all AOL items. The resulting datasets are described in more detail in Section 3.2.

[7] http://www.google.com/trends?q=Iran&geo=all&date=2006-5

Table 4. Averages of all Del.icio.us URLs (full / normalised) with the search datasets

Dataset	top 25	top 50	top 75	top 100
Google	19.91 / 24.17	37.61 / 47.83	54.00 / 71.15	69.21 / 85.23
MSN	12.86 / 20.20	22.38 / 38.62	30.93 / 56.47	39.09 / 74.14
AOL	— / 19.61	— / 35.57	— / 48.00	— / 57.48

In order to see whether Del.icio.us contains those URLs that were judged relevant by the traditional search engines, we computed a kind of "recall" for folksonomy-URLs on the other data sets as follows: First we cut each of the 1,776 rankings of each search data set after the first 25, 50, 75 and 100 URLs. For each ranking size, we computed the intersection with all Del.icio.us URLs. As the AOL log data consist of domain names only (and not of full URLs), we also pruned the URLs of the other systems in a second step to the domain names.

Table 4 shows the results. The first number in each cell is the average number of overlaps for the original URLs, the second for the pruned URLs. Google shows the highest overlap with Del.icio.us, followed by MSN and then AOL. For all systems, the overlap is rather high. This indicates that, for each query, both traditional search engines and folksonomies focus on basically the same subset of the Web. The values in Table 4 will serve as upper bounds for the comparison in Section 5.

Furthermore, the top rankings show more coverage: While in average 24.17 URLs in the top Google 25 ranking are represented in Del.icio.us, only 85.23 are represented in the top 100 URLs in average. This indicates that the top entries of search engine rankings are – in comparison with the medium ranked entries – also those which are judged more relevant by the Del.icio.us users.

4.4 Conclusions of Section 4

The overlap of the whole set of the MSN query terms with the set of all Del.icio.us tags is only about a quarter of the size of the latter, due to a very high number of very infrequent items in both systems (Section 4.1, Table 2). Once the sets are reduced to the frequent items, the relative overlap is higher. The remaining differences are due to different usage, e. g., to the composition of multi-word lexems to single terms in Del.icio.us, and the use of (parts of) URLs as query terms in MSN.

In Section 4.2, we have seen that for a relatively high number of items the search and tagging time series were significantly correlated. We have also observed that important events trigger both search and tagging without significant time delay, and that this behaviour is correlated over time.

Considering the fact that both the available search engine data and the folksonomy data cover only a minor part of the WWW, the overlaps of the sets of URLs of the different systems (as discussed in Section 4.3) are rather high, indicating that users of social bookmarking systems are likely to tag web pages that are also ranked highly by traditional search engines. The URLs of the social bookmarking system cover over-proportionally the top results of the search engine rankings. A likely explanation is that taggers use search engines to find interesting bookmarks.

5 Comparison of Social and Traditional Rankings

In the previous section we compared the user interaction in social bookmarking systems and search engines and the coverage of URLs of folksonomies in search engines. In this section we focus on ranking algorithms. Are overlapping results different when we introduce a ranking to the folksonomy structure? Are important URLs in search engines similar to important URLs in social bookmarking systems? Is the ranking order within the overlap the same? These questions will be answered below.

For the commercial search engines, we rely on our crawls and the data they provided, as the details of their ranking algorithms are not published (beside early papers like [11]). To rank URLs in social bookmarking systems, we used two well-known ranking approaches: the traditional vector space approach with TF-IDF weighting and cosine similarity, and FolkRank [8], a link-based ranking algorithm similar to PageRank [11], which ranks users, resources or tags based on the tripartite hypergraph of the folksonomy.

5.1 Overlap of ranking results

To compare the overlap of rankings, we start with an overview of the average intersection of the top 50 URLs calculated for all of our datasets. In this case we based the analysis on the normalized URLs of the same datasets as used in Section 4.3. Table 5.1 contains the average overlap calculated over the sets of normalized URLs and the TF, TF-IDF and FolkRank rankings of the Del.icio.us data. We see that the overlap of Del.icio.us Oct. 2006 with the result sets of the three commercial search engines is low. The average overlap of the MSN and Google crawl rankings is considerably bigger (11.79) – also compared to the AOL results which are in a similar range with the Del.icio.us data. The two major search engines therefore seem to have more in common than folksonomies with search engines.

The TF and TF-IDF based rankings show a surprisingly low overlap with Google, MSN and AOL, but also with the FolkRank rankings for Del.icio.us. This indicates that – as for web search – graph-based rankings provide a view on social bookmarking systems that is fundamentally different to pure frequency-based rankings.

Although the graph-based ranking on Del.icio.us has a higher overlap with the search engine rankings than TF-IDF, it is still very low, compared to the potential values one could reach with a 'perfect' folksonomy ranking, e. g., an average overlap of 47.83 with the Google ranking as Table 4 shows. The remaining items are contained in the Del.icio.us data, but FolkRank ranked them beyond the top 50.

To investigate this overlap further, we have extended the Del.icio.us result sets to the top 100 and top 1,000, resp.

Table 6 shows the average overlap of the top 100 and the top 1,000 normalized URLs of the FolkRank computations in Del.ico.us data of Oct. 2006 to the top 50 normalized URLs in the Google crawl, MSN crawl and AOL log data. It extends thus the middle column of Table 5.1. For Google, for instance, this means that the relative average overlap is $\frac{6.65}{50} \approx 0.13$ for the top 50, $\frac{9.59}{100} \approx 0.10$ for the top 100, and only $\frac{22.7}{1000} \approx 0.02$ for the top 1000. This supports our finding of Section 4.3,

Table 5. Average overlap of top 50 normalized URLs

	Google	MSN	Del FolkRank	Del TF-IDF	Del TF
AOL	2.39	1.61	2.30	0.30	0.21
Google		11.79	6.65	1.60	1.37
MSN			3.78	1.20	1.02
Del FolkRank				1.46	1.79
Del TF-IDF					49.53

Table 6. Average overlap with top 100/1,000 normalized Del.icio.us URLs

	Google top 50	MSN top 50	AOL top 50
Del 100	9.59	5.00	1.65
Del 1000	22.72	13.43	5.16

that the similarity between the FolkRank ranking on Del.icio.us and the Google ranking on the Web is higher for the top than for the lower parts of the ranking.

5.2 Correlation of rankings

After determining the coverage of folksonomy rankings in search engines, one question remains: Are the rankings obtained by link analysis (FolkRank) and term frequencies / document frequencies (TF-IDF) correlated to the search engine rankings? Again, we use the rankings of the 1,776 common items from Section 4.3. As we do not have interval scaled data, we select the Spearman correlation coefficient $r_s = 1 - \frac{6 \sum d^2}{n(n^2-1)}$, where d denotes the difference of ranking positions of a specific URL and n the size of the overlap.[8]

In Section 5.1 we showed that the overlap of the rankings is generally low. We therefore only compared those rankings having at least 20 URLs in common. For each such item, the Spearman coefficient is computed for the overlap of the rankings. Table 7 shows the results. The AOL comparisons to Del.icio.us (using the link-based method as well as TF-IDF) do not show sufficient overlap for further consideration. The Google and MSN comparisons with the link-based FolkRank ranking in Del.icio.us yield the highest number of ranking intersections containing more than 20 URLs (Google 361, MSN 112). Both Google and MSN show a large number of positive correlations. For instance, in Google, we have 326 positive correlations, whereby 176 are significant. This confirms our findings from Section 5.1.

From the results above we derive, that if overlap exists, a large number of rankings computed with FolkRank are positively correlated with the corresponding search engine rankings. In order to find out the topics on which the correlation is high, we extracted the top ten correlations of the Del.icio.us FolkRank with Google and MSN, resp., see Table 8. We found that most items in this set are IT related. As a major part of Del.icio.us consists of IT related contents, we conclude that link-based rankings for topics that are specific and sufficiently represented in a folksonomy yield results similar to search engine rankings.

[8] In [3], enhancements to Kendall's tau and Spearman are discussed to compare rankings with different URLs. These metrics are heavily influenced if the intersection between the rankings is small. Because of this we stick to the Spearman correlation coefficient.

Table 7. Correlation values and number of significant correlations

Datasets	# overlap > 20)	Avg. correlation	Avg.of significant correlations	# correlated rankings	# significant correlated rankings
			pos/neg	pos/neg	pos/neg
Google/FolkRank	361	0.26	0.4/-0.17	326/37	176/3
Google/TF-IDF	17	0.17	0.34/0	15/2	5/0
MSN/FolkRank	112	0.25	0.42/-0.01	99/13	47/1
MSN/TF-IDF	6	-0.21	-/-	2/4	0/0
AOL/FolkRank	1	0.25	-/-	1/0	0/0
AOL/TF-IDF	1	0.38	0.38/-	1/0	1/0

Table 8. Top Correlations of Delicious FolkRank with Google (left) and MSN (right), based on top 100 of Del.icio.us.

Item	**Intersection**	**Correlation**	**Item**	**Intersection**	**Correlation**
technorati	34	0.80	validator	21	0.64
greasemonkey	34	0.73	subversion	22	0.60
validator	34	0.71	furl	23	0.59
tweaks	22	0.68	parser	27	0.58
metafilter	24	0.67	favicon	28	0.57
torrent	29	0.65	google	25	0.57
blender	22	0.62	blogosphere	21	0.56
torrents	30	0.62	jazz	26	0.56
dictionaries	21	0.62	svg	23	0.55
timeline	21	0.62	lyrics	25	0.54

5.3 Conclusions of Section 5

In Section 5.1, we have seen that a comparison of rankings is difficult due to sparse overlaps of the data sets. It turned out that the top hits of the rankings produced by FolkRank are closer to the top hits of the search engines than the top hits of the vector based methods. Furthermore we could observe that the overlap between Del.icio.us and the search engine results is larger in the top parts of the search engine rankings.

In Section 5.2 we observed that the folksonomy rankings are stronger correlated to the Google rankings than to MSN and to AOL, whereby the graph-based FolkRank is closer to the Google rankings than TF and TF-IDF. Again, we assume that taggers preferably use search engines (and most of all Google) to find information. A qualitative analysis showed that the correlations were higher for specific IT topics, where Del.ico.us has a good coverage.

6 Discussion and Outlook

In this paper, we conducted an exploratory study to compare social bookmarking systems with search engines. We concentrated on information retrieval aspects by analyzing search and tagging behaviour as well as ranking structures. We were able to discover both similar and diverging behaviour in both kinds of systems, as summarized in Sections 4.4 and 5.3. An open question is whether, with more data available, the correlation and overlap analyses could be set on a broader basis. A key question to be answered first though is what is to be considered a success? Is it desirable that social search tries to approximate traditional web search? Is Google the measure of all things? Computing overlap and comparing correlations helped us finding out about the similarities between systems. However, we have no information which approach offers more relevant results from a user's perspective. A user study in which users create a benchmark ranking and performance measures might be of benefit. Further investigation also has to include a deeper analysis of where URLs show up earlier and the characteristics of both system's URLs not being part of the overlap.

Acknowledgment. This work has partially been supported by the Microsoft Live Labs Award "Accelerating Search in Academic Research".

References

1. E. Adar, D. Weld, B. Bershad, and S. Gribble. Why we search: Visualizing and predicting user behavior. In *Proc. WWW'07*, pages 161–170, Banff, Canada, 2007.
2. S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. Optimizing web search using social annotations. In *Proc. WWW'07*, pages 501–510, Banff, Canada, 2007.
3. J. Bar-Ilan, M. Mat-Hassan, and M. Levene. Methods for comparing rankings of search engine results. *Comput. Networks*, 50(10):1448–1463, 2006.
4. A. Z. Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002.
5. C. Chatfield. *The analysis of time series: an introduction*. CRC Press, Florida, US, 6th edition, 2004.
6. S. Chien and N. Immorlica. Semantic similarity between search engine queries using temporal correlation. In *Proc. WWW'05*, pages 2–11, New York, NY, USA, 2005. ACM Press.
7. S. Golder and B. A. Huberman. The structure of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, 2006.
8. A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In *Proc. ESWC'06*, volume 4011 of *LNAI*, pages 411–426, Heidelberg, 2006. Springer.
9. Ivorix. Sine-weighted moving average, 2007. `http://www.ivorix.com/en/products/tech/smooth/smooth.html`.
10. A. M. Krishna P. Gummadi and P. Druschel. Exploiting social networks for internet search. In *Proc. HotNets-V*, pages 79–84, 2006.
11. L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the web. In *Proc. WWW'98*, pages 161–172, Brisbane, Australia, 1998.
12. G. Pass, A. Chowdhury, and C. Torgeson. A picture of search. In *Proc. 1st Intl. Conf. on Scalable Information Systems*. ACM Press New York, NY, USA, 2006.
13. M. Vlachos, C. Meek, Z. Vagena, and D. Gunopulos. Identifying similarities, periodicities and bursts for online search queries. In *Proc. SIGMOD '04*, pages 131–142, New York, NY, USA, 2004. ACM Press.
14. Y. Yanbe, A. Jatowt, S. Nakamura, and K. Tanaka. Can social bookmarking enhance search in the web? In *Proc. JCDL'07*, pages 107–116, New York, NY, USA, 2007. ACM Press.