Comparing Tweets and Tags for URLs

Morgan Harvey, Mark Carman, and David Elsweiler

[1] Dept. Computer Science 8 (AI), Univerisity of Erlangen-Nuremberg, Germany.
[2] Faculty of IT, Monash University, Melbourne, Australia
[3] Institute for Information and Media, Language and Culture, University of Regensburg, Germany
morgan.harvey@i8.informatik.uni-erlangen.de, mark.carman@monash.edu, david@elsweiler.co.uk

Abstract. The free-form tags available from social bookmarking sites such as Delicious[4] have been shown to be useful for a number of purposes and could serve as a cheap source of metadata about URLs on the web. Unfortunately recent years have seen a reduction in the popularity of such sites, however at the same time microblogging sites such as Twitter have exploded in popularity. On these sites users submit short messages (or "tweets") about what they are currently reading, thinking and doing and often post URLs.

In this work we look into the similarity between top tags drawn from Delicious and high-frequency terms from tweets to ascertain whether Twitter data could serve as a useful replacement for Delicious. We investigate how these terms compare with web page content, whether or not top Twitter terms converge and determine if the terms are mostly descriptive (and therefore useful) or if they are mostly expressing sentiment or emotion. We discover that provided a large number of tweets are available referring to a chosen URL then the top terms drawn from these tweets are similar to Delicious tags and could therefore be used for similar purposes.

1   Introduction

The past decade has been a time of significant evolution for the Web as it has moved from being a collection of predominantly static documents to a medium for collaboration and sharing among millions of users. In many ways this so-called "Web 2.0" movement brings the Web much closer to Tim-Berners Lee's original vision [3]. These new more social aspects of the Web include examples such as social bookmarking and microblogging. In these social systems users are expected to contribute information and content, sharing interesting events, items and web sites and their opinions about these with other users.

In social bookmarking users share URLs of interest to which they can assign free-form textual keywords or "tags", without having to adhere to a pre-defined vocabulary. This new paradigm allows users of a system to define their own

---

[4] http://delicious.com/

personal set of categories in order to organise and publicly annotate a diverse range of resources in a manner which is meaningful to them [7]. While most people tend to tag for their own benefit, the categorisations they choose can be of use to the community as a whole [6]. It has been shown that after a relatively small number of users have tagged a resource, a nascent consensus forms that remains unaffected by the addition of further tags. Over time, tag use stabilises and the community forms an unspoken group consensus of how things should be categorised, creating a shared and agreed upon vocabulary [8]. This agreed-upon and stable vocabulary can be obtained by calculating the most frequently used terms for a given URL.

Microblogging is a second form of socially contributed data that has become increasingly popular in recent years. This new form, epitomised by the seminal Twitter service, allows users to post and read short text messages - known as "tweets" - of up to 140 characters in length. In these tweets users post about what they are currently reading, thinking and doing and often post URLs [13, 15] to web sites of interest to them. As of August 2011 Twitter's users were posting over 200 million tweets per day and in 2010 over 15% of all adult web users in the US are expected to make use of the service [5]. The kinds of information posted and the sheer volume of data suggest that Twitter may be an abundant and up-to-date source of information about web sites and web pages. Research has shown that information obtained from social tagging data can be utilised to increase the performance of web search by improving term smoothing [2, 10] and can also be used to build search profiles of users in order to personalise results [9]. It is therefore useful to investigate whether or not the massive amounts of data contributed to Twitter can also be used for such purposes.

In this work we compare the tags assigned to bookmarks in Delicious with the terms used on Twitter to describe the same URLs in order to determine if tweets may serve as a replacement for social bookmarks as a source of cheap, user-generated metadata. We also compare the tweets and tags with the actual content of these sites to determine if users are simply copying the content verbatim or are making their own views and assessments of the content known. We investigate which parts of the web content these tags and tweets are being drawn from (title, metadata description, anchor text, etc). In doing so we discover interesting differences in terms of URL coverage on Twitter and Delicious, indicating different modalities of use for the services and also find interesting differences in the descriptions for URLs retrieved from these two sources. We also investigate whether or not the top terms from tweets about a single URL tend to converge in the same way that tags on Delicious have been shown to. Finally we investigate how many of the top tags and top terms from Twitter are emotional and how many are purely descriptive.

The paper is structured as follows: we first briefly discuss related work including other investigations of social media. We then proceed to describe the data collected for our experiments and some techniques applied in order to clean it. We then present the main analyses of this work; namely the application of similarity metrics to determine the overlap between tweets, tags and web page

content. Next, using SentiWordNet 3 [1], we investigate how many of the top terms from tweets and the tags are descriptive and how many are emotive. This provides further insight into how useful these lexical terms may be for web page categorisation. Finally we conclude the paper with a discussion of the important findings and how these may influence future use of social data.

2   Related Work

The work presented in this paper draws from the techniques presented in past publications which have compared lexical content from various sources in order to determine their similarity and usefulness. For example work by Carman et al. [4] investigated the similarity of search queries to tag data from Delicious and found that while there is some similarity, the two sets of data did not derive from the same underlying distribution over lexical terms. They also showed that queries were more similar to page content than tags and that the tags from Delicious could be a useful source of extra information for smoothing language models, thus being potentially useful for improving search systems. We re-use some of their techniques in order to make our comparisons in this work, however we go beyond simply analysis of lexical overlap and investigate the actual terms used and how their probabilities change as more data is added.

More recently Huang et al. [11] investigated the use of hashtags (terms prefixed with a hash symbol, thought to be similar to tags) in Twitter over time, contrasting them with the use of tags on Delicious. They found differences in the use of tags caused mostly by the design and intended usage of the systems. Twitter is intended to allow people to express their views, communicate and share information in the short-term. Delicious on the other hand provides a means for people to collect URLs of interest to them over much longer periods of time. Tags are used as a means to facilitate recall of one's own bookmarks but also as a way to search, filter and browse bookmarks from the whole community. As a result, hashtags on Twitter were found to be frequently used to link posts to discussions on popular topics and their use was therefore found to generally be short-lived. On the other hand, popular Delicious tags have a much longer lifespan being used over very long periods of time to describe related URLs.

Heymann et al. [10] considered whether tags from Delicious could be used to improve the performance of web search. It was found that tags are almost always highly descriptive of the web pages they are used to annotate and are not usually simply terms drawn verbatim from the content. This suggests that tags could be used to partially overcome the issues of vocabulary mismatch in search. However the authors also conclude that the coverage of web sites on Delicious is not particularly high, thus holding back its use for improving web search. This is one of our motivations for performing this research as tweets may cover URLs for which there is no data available on Delicious. In similar work Bao et al. [2] use tags to improve search performance and in doing so are able to implement a system which outperforms a BM25 baseline.

In this work we build upon the existing literature to gain a more complete understanding of how tweets and tags are related and how much they represent the content of the web pages they relate to. We investigate the top terms from whole tweets, rather than using simply the short-lived and conversational hashtags, and look into exactly which parts of a web page's HTML content is replicated in tags and tweets. In doing so we attempt to answer the question of whether or not nascent consensus on Twitter is a valid replacement for Delicious tags and can therefore be used in the same way; namely as a cheap source of meta-data for web pages.

3   Data Collection and Cleaning

We are interested in comparing lexical similarity between three different sources of data; tweets, Delicious tags (from bookmarks) and web page content. In order to build a dataset to make comparisons it was necessary to first collect a list of URLs posted to either Delicious or Twitter and then attempt to locate the same URLs in the other service. A first attempt at this made use of the TREC microblog track data set[5] of tweets collected between the 23rd of January and 8th of February 2011 and extracted all tweets containing URLs. For each of these URLs we searched Delicious and downloaded the top 10 tags and also attempted to download the web page content. In doing so we discovered that out of the 15,777 URLs obtained from the TREC collection only 623 were bookmarked by Delicious users; a coverage ratio of just 3.9%. Since the overlap for this data is so small we decided to investigate whether the opposite relationship was true: i.e. do URLs posted on Delicious also have poor coverage on Twitter?

We next collected a second data set by instead crawling Delicious for URLs first by downloading the latest URLs posted to Delicious. This approach ensures a random sample of the sites bookmarked by Delicious users. Then for each of these URLs we downloaded the top 10 Delicious tags (as this is the most the API will allow) and also queried Twitter for any tweets made about them (up to a maximum of 100). We collected 9,462 Delicious URLs over a period of two weeks in early September 2011, of which 7,748 had tags available (a total of 59,874 unique tags). Of the 9,462 URLs collected, 4,013 were found on Twitter resulting in 62,299 tweets being downloaded; a much better coverage ratio of 42.4%. In total there were 3,240 URLs for which we were able to retrieve both tags and tweets and this is the dataset used in the following analysis. While this dataset may seem quite small, it is a similar size to those used in previous work [4] and is still more than large enough to provide statistically significant results.

This massive difference in availability between URLs on the two services is interesting. In manually analysing the tweeted URLs that don't appear in Delicious it became clear that the vast majority were specific web pages, for example individual news stories or blog posts. On the other hand, a large proportion of URLs submitted to Delicious are the root domains of web sites and not specific

---

[5] TREC microblog track web site: https://sites.google.com/site/microblogtrack/

articles or posts. This ties in with the conclusions of Huang et al. [11] that tweets are more conversational and temporal in nature. In their work they show that many single articles will peak in popularity very quickly and then very quickly fall into obscurity, whereas the main index pages are likely to have more general relevancy. Since Delicious is a more permanent store of URLs it is more likely to cover a large range of popular root domains, even if they are no longer frequently discussed or bookmarked. Twitter, on the other hand, will tend to have mostly URLs which are currently popular or frequently discussed, particularly web memes such as YouTube videos and stories currently in the news.

Due to the conversational and messy nature of tweets it is necessary to perform some cleaning of the data before proceeding with analysis. To do this we remove a standard list of English stop words as well as some additional Twitter-specific stop words. These were obtained by identifying terms in the Twitter data with extremely low IDF values, i.e. terms which appear in the vast majority of tweets and are therefore poor descriptors. Examples include terms such as "lmao", "haha" and "rofl". We perform standard data cleaning by removing any instances of URLs and punctuation from the data and convert all characters to lowercase. Finally we remove all references to other users as these will not be related to the URL about which the tweet is made. These can be easily identified in Twitter as they are pre-fixed with an @ symbol.

4   How Similar Are Tweets and Tags?

In keeping with previous analysis of lexical similarity [4] we first calculate the overlap coefficient between sets of terms. The overlap coefficient is a metric that describes how much of the smaller of the vocabularies is included in the larger and is not sensitive to the relative sizes of the two vocabularies [14]. Given two sets of vocabulary words V_{tag} and V_{tweet} it is defined as follows:

$$Overlap(V_{tag}, V_{tweet}) = \frac{|V_{tag} \cap V_{tweet}|}{min(|V_{tag}|, |V_{tweet}|)}$$

Due to the sparsity and small size of the first dataset we will only describe the overlap between the tags and top tweet terms for the second dataset. As discussed before user-contributed content tends to quickly stabilise in terms of its term distribution as more users describe it and as such we are interested in comparing the terms which are stable. To do this we select the top-k terms by frequency within the tweets for a given URL to compare with the top tags from Delicious. This method also mitigates issues of randomness in the terms, ensuring that only those reused by a large number of users (which therefore can be expected to be good descriptive terms) are compared. Also the relative difference in the sizes of the term vocabularies is often quite large as we have a maximum of 10 Delicious tags but potentially very many Twitter terms. If we choose only the top-k tweet terms then we can ensure that the vocabularies are of a similar or even the same size.
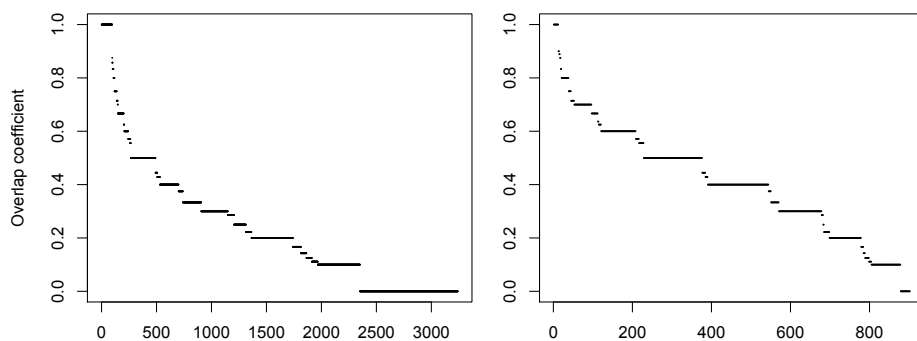
Fig. 1. Vocabulary overlap (overlap coefficient) between Delicious tags and Twitter tweets. Data points are ordered from most to least overlap.

Figure 1 shows the overlap between tags and tweet terms with the top 20 tweet terms being selected for comparison. The figure on the left shows the overlap for all URLs in the dataset, where the mean and median overlap is only 0.231 and 0.2 respectively. This is much lower than that reported by Carman et al. [4], however this is perhaps not so surprising since in this dataset there is no guarantee that the Delicious tags have stabilised and that there is enough density in the tweet data to form good top terms. The relatively large number of instances where the overlap is 0 (890 of 3240) and also where it is 1 (96) may be because there are only a very small number of tags or tweets to compare with. For instances where the overlap is 1 we find that the median number of tags is only 1 (mean 2.73) compared with 10 (mean 7.93) over the whole dataset.

To deal with this problem we queried the Delicious API for the number of bookmarks made for each URL. Then we ran the comparison again, but this time only comparing term distributions where the number of Delicious bookmarks was greater than or equal to 50, since it is a fair assumption that the top tags will have converged by this point. This gave us a set of 3,369 Delicious URLs. We do a similar thing for the tweet terms by only choosing URLs for which we have 20 or more tweets, giving a much denser and richer set of data to draw terms from. For these denser URLs we have a mean of 60.1 tweets and a median of 56, we later refer to these URLs as having "high-density" tweet terms.

The results from this analysis of much denser data is shown on the right of Figure 1. It is clear from this plot that the overlap in this case is quite significantly higher and is closer (but still not nearly as close) to the overlap between tags and queries reported by Carman et al. They report that "well over half" of the URLs in their sample have an overlap of 0.5 of more, whereas for our data only around a quarter of URLs have an overlap of 0.5 or more. For this dataset the mean and median of the overlap is 0.409 and 0.4 respectively. This suggests that if enough tweets are available, the top terms will have reasonably high overlap with top tags (again, assuming enough bookmarks are present for

socialmedia google marketing social book libros strategy zmot top ebook
googles free google marketing truth moment zmot book zero ebook 55
mail app wx google email e-mail chrome gmail offline
app mail via google store chrome gmail googlemail web offline
twitter tools aggregator news social facebook web2.0 rss socialmedia newspaper
news twitter stories rss paper facebook read todays daily top
auction shop auctions search sell buy popular ebay shopping online
auction buyme new forsale art ebay us date quot end

Table 1. Examples of top Delicious tags (top) and top tweet terms (bottom) for high-density URLs.

the distribution to have stabilised). Even with the much denser data, the overlap is still not nearly as high as between queries and tags. However, Carman et al. also calculate similarity between top 20 terms, which is a fairer comparison with the analysis we have performed, and find that it is less than for all terms. Table 1 shows some examples of top Delicious tags and top tweet terms for some high-density URLs. These examples illustrate that the tweets terms are quite similar to the tags and also that they are generally good descriptive terms for the URL.

4.1   Do Twitter Terms Stabilise?

An important question arises from this analysis. Do we know if Twitter terms stabilise over time and how many are required for them to become stable? Since it is known that Delicious tags stabilise over time (i.e. as more users contribute tags to a URL) we wanted to find out if the same behaviour occurred with top terms from Twitter tweets. We therefore investigated whether or not the distribution of top tweet terms converges as more tweets are added. To do this we need to be able to calculate the term distribution before and after new tweets have been added and then compare these distributions. Halpin et al. [8] attempt to perform a similar analysis on tags to show convergence, however - based on their description it appears that - they compare the difference between subsequent distributions in time (i.e. before and after new tags have been added). This introduces a significant bias since the previous set of tags must be a subset of the next set and therefore it is hardly surprising that the distribution over terms converges so quickly. To get around this bias we instead keep odd and even tweet additions separate and calculate the difference between these two sets of term distributions after a new tweet has been added to one of them in order. This means that we are not at any point comparing sets of terms with their own subsets and therefore any convergence of term use is likely to be genuine.

In order to calculate the difference between the term distributions we use the KL divergence which measures the relative entropy between 2 distributions. To obtain two distributions to compare we calculate counts of the top 100 terms
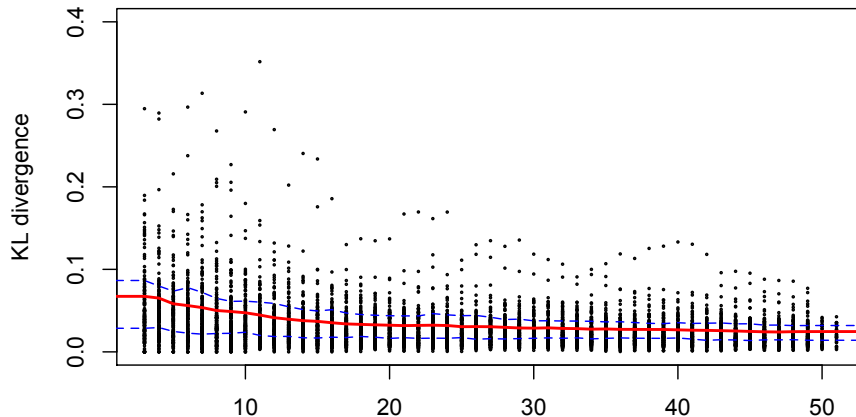
Fig. 2. Vocabulary overlap as number of tweets increases.

and then from these calculate a multinomial distribution using the empirical maximum likelihood estimate. The KL divergence between two probability distributions P and Q over x components is defined as:

$$D_{KL} = \sum_x P(x) \, log \left(\frac{P(x)}{Q(x)} \right)$$

Laplace smoothing is applied to terms to ensure that KL divergence will be finite. Figure 2 shows the KL divergence between the top terms in the odd and even sets of tweets as the number of tweets in each increases. These comparisons were made only for URLs for which we had 100 tweets to ensure that the total number of data points is the same over all comparison points. The thick red line shows the mean KL divergence and the dashed blue lines show the upper (75%) and lower quartiles (25%). We can see that the speed of convergence is not nearly as rapid as reported for tags by Halpin et al. [8], however given the bias in their choice of comparison this is perhaps not surprising. Nevertheless the top terms drawn from the tweets do appear to converge as more data is added, especially over the first 10 or so additions. By the 25th addition of new information the maximum KL divergence compared to the previous distribution is only 0.121, whereas for the first comparison the maximum is 0.703. At the 25th addition only 5% of term distributions have a KL of more than 0.05, on the other hand for the first comparison 55% of distributions meet this criteria.

5   How Well Do They Describe Content?

Having downloaded the content of the URLs we can compare the Delicious tags and the top tweet terms with various HTML fields to determine where the terms

Fig. 3. Overlap of tags and tweet terms and 12 main HTML fields.

are coming from and if they are indeed just being copied verbatim from the web site itself. Figure 3 shows the overlap of tags and top tweet terms with 12 main HTML fields. These are: page title, metadata description, metadata keywords, anchor text, bold text, strong text, underlined text, italic text, main header, secondary header, tertiary header, image alternate text. We compared the content with the following:

All Tags tags for all URLs with at least 1 tag
Min10 Tags tags for URLs with 10 tags
All Tweets top 10 tweet terms for all tweeted URLs
Min 20 Tweets top 10 tweet terms for tweeted URLs with 20 or more tweets

Some of the HTML fields have very little overlap in general with both tags and tweet terms, for example italic text, bold text and tertiary headers. Looking at the title overlap we see that for all URLs with tweets (dark red) the overlap with the title is very high, in fact it is the highest overlap over all of the comparisons. However once top terms have been derived from a large number of tweets this overlap decreases and becomes the same as for tags (no significant difference, p-value=0.326[6]). This suggests that in many cases tweets referring to a URL simply copy the title of the page verbatim and do not attempt to explain the posting or describe the content of the page. When we only have a small number of tweets these tweets overwhelm the rest and cause the top terms to be very

---

[6] Independent 2-group Mann-Whitney U test.

similar to the title. However if a large number of tweets for the same URL are conflated together then it is less likely that the tweets containing copied titles will overwhelm the other terms, allowing the most frequently agreed-upon descriptive terms to dominate. Notice also that the overlap between tweet terms (for all URLs) and the main header text is also quite high and in fact the values look to be in very similar ratios as for the title. This is likely because a large number of web sites will repeat the title text in the main header. By analysing individual tweets, as opposed to top terms over all tweets for a single URL, we find that nearly a third (28.4%) of tweets have an overlap of greater than 0.9 with the title. If we wish to single-out "useful" tweets it may make sense to choose those which have only a small overlap with the page title.

Looking at the tags we see that they tend to have a high overlap with the metadata keywords, hinting that taggers are likely to choose words that describe the web page, rather than just copying the title. This is also true for the high-density tweet terms, hinting that the converged terms also describe the content quite well without simply copying the title. There is no significant difference in the overlaps between content keywords and tags and content keywords and high-density tweet terms (p-value=0.6532^6). However there is a highly significant difference between the overlaps for tweet terms for all URLs and for high-density tweet terms (p-value$\ll 0.01^6$). This relationship is also true with the metadata descriptions, however it is much less pronounced. This analysis hints that both tags and "high-density" tweet terms may be useful as a replacement for web page metadata for sites which do not provide it.

6   Are The Terms Descriptive or Emotional?

As noted in previous research both tweets [12] and tags [7] often contain a large amount of emotional content, where people express their opinions about something rather than describing it. Clearly if terms derived from tagging and tweets are to be used as good classifiers of web site content they should be descriptive, rather than opinionated. While emotive terms may be useful for determining the quality of web pages they are not ideal for use as keywords. In order to get a sense of how many of the top tweet terms and Delicious tags are emotive we looked each term up in SentiWordNet 3 [1]. SentiWordNet assigns a score to each term indicating how positive or negative it is, where a positive number indicates a positive sentiment and vice-versa. Terms which are mostly descriptive and therefore non sentiment-bearing are assigned a score of 0. For example the term "excellent" has a score of 0.75, "evil" is assigned a score of -0.875 and "algorithm" is not sentiment-bearing and is therefore scored 0.

For each Delicious tag and each top tweet term we found the associated SentiWordNet score, using the mean over all synsets for eat word. Overall 18% of top tags and 22% of top tweet terms were not available in the SentiWordNet and could not be assigned a score. These terms are therefore ignored for this analysis. Table 2 shows the main results of this analysis. For both data types (tags and tweet terms) the vast majority of terms are non sentiment-bearing and

	Top Delicious tags	Top tweet terms
Total found	5,356	4,547
# positive (%)	788 (14.7)	816 (17.9)
# negative (%)	196 (3.7)	291 (6.4)
# descriptive (%)	4,372 (81.6)	3,440 (75.7)
mean score	0.031	0.044
mean score (!=0)	0.168	0.179
mean score (positive)	0.285	0.352
mean score (negative)	-0.302	-0.307

Table 2. Results of sentiment analysis on tags and tweet terms

are therefore likely to be useful as resource descriptors. Overall the tweet terms are a little more likely to be sentiment-bearing than tags which is perhaps not surprising given the conversational nature of the medium. However the number of the tweet terms which are purely descriptive is still very high, suggesting that they would be useful as metadata terms for the URLs they relate to. The mean scores for positive sentiment words show that that tweets are significantly more likely to score higher than tags (p-value$\ll 0.01^6$), whereas for negative sentiment words there is no difference (p-value$=0.827^6$). This is perhaps because people frequently post items that they particularly like on Twitter as recommendations to friends and to generally express their views. In such cases they are likely to use very positive terms which are also likely to be re-used by other users when tweeting about the same resource.

7 Conclusions

In this paper we have investigated the similarity (and in some cases dissimilarity) between terms derived from two different forms of social data: tags from Delicious and tweets from Twitter. We surmised that if it could be shown that terms drawn from tweets are similar to Delicious tags then they could be used as a cheap and up-to-date source of metadata in a similar way. We collected tags, tweets and content for a number of URLs and performed a number of statistical analyses.

We found that, provided a sufficient number of tweets are made regarding a particular URL, the top terms drawn from these tweets are similar to Delicious tags and have very comparable similarity with the content of the web pages they are discussing. For both tags and tweets terms where dense data is available we found that the overlap with the keywords and description of the web page is quite high, indicating strongly that they serve as a good proxy for such metadata and for summarising the web page. We demonstrated that as more tweets mentioning a given URL are added the top terms tend to converge, displaying similar behaviour to that originally found in tags. Finally, we analysed the top terms for emotional content and found that a large proportion were not sentiment-bearing, further contributing to the hypothesis that they are good descriptive keywords.

Our results are important because they offer a more complete understanding of how the vast amounts of data available from microblogging sites such as Twitter can be used. By showing that the converged top terms are similar to Delicious tags we open up the possibility of using the data for similar purposes such as improving web search and browsing interfaces. This allows the wealth of algorithms and techniques developed for social tagging data to be used with tweet data instead of or in concert with tags. From our analysis we identified that there are several reasons why people tweet about URLs and in future work we would like to classify these and conduct analysis on each class separately. We intend to investigate tweets where people are responding to requests from other users as we believe these may yield good descriptive terms [16]. Furthermore we wish to analyse whether or not the emotive terms uncovered by our analysis can be useful as a means to determine the quality of web pages or to build more accurate user interest profiles.

References

1. S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC'10*, pages –1–1, 2010.
2. S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. Optimizing web search using social annotations. In *WWW '07*, pages 501–510, New York, NY, USA, 2007.
3. T. Berners-Lee, R. Cailliau, A. Luotonen, H. F. Nielsen, and Arthur Secret. The world-wide web. *Commun. ACM*, 37:76–82, August 1994.
4. M. James Carman, M. Baillie, R. Gwadera, and F. Crestani. A statistical comparison of tag and query logs. pages 123–130, 2009.
5. eMarketer. US twitter usage surpasses earlier estimates.
6. S. Golder and B. Huberman. Usage patterns of collaborative tagging systems. In *Journal of Information Science*, volume 32(2), pages 198–208, 2006.
7. S. Golder and B. A. Huberman. The structure of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, 2005.
8. H. Halpin, V. Robu, and H. Shepherd. The complex dynamics of collaborative tagging. In *WWW '07* pages 211–220, New York, NY, USA, 2007.
9. M. Harvey, I. Ruthven, and M. J. Carman. Improving social bookmark search using personalised latent variable language models. *WSDM 2011*, pages 485–494, 2011.
10. P. Heymann, G. Koutrika, and G. Garcia-Molina. Can social bookmarking improve web search? In *WSDM 2008*, February 2008.
11. J. Huang, K. M Thornton, and E. N Efthimiadis. Conversational tagging in twitter. In *HT 10*, page 173, 2010.
12. J. Hurlock and M. L. Wilson. Searching twitter: Separating the tweet from the chaff. *ICWSM*, 2011.
13. A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *WebKDD/SNA-KDD '07*, pages 56–65, 2007.
14. C. D. Manning and H. Schütze. *Foundations of statistical natural language processing*. MIT Press, Cambridge, MA, USA, 1999.
15. P. McFedries. Technically speaking: All a-twitter. *Spectrum, IEEE*, 44:84–84, 2007.
16. M. R. Morris, K. Panovich and J. Teevan What do people ask their social networks, and why? In *CHI 2010*, 1739–1748, 2010.