Conversational Tagging in Twitter

Jeff Huang, Katherine M. Thornton, Efthimis N. Efthimiadis
University of Washington
Information School
Seattle WA 98195
ht@jeffhuang.com, thornt@uw.edu, efthimis@uw.edu

ABSTRACT
Users on Twitter, a microblogging service, started the phenomenon of adding tags to their messages sometime around February 2008. These tags are distinct from those in other Web 2.0 systems because users are less likely to index messages for later retrieval. We compare tagging patterns in Twitter with those in Delicious to show that tagging behavior in Twitter is different because of its conversational, rather than organizational nature. We use a mixed method of statistical analysis and an interpretive approach to study the phenomenon. We find that tagging in Twitter is more about filtering and directing content so that it appears in certain streams. The most illustrative example of how tagging in Twitter differs is the phenomenon of the Twitter micro-meme: emergent topics for which a tag is created, used widely for a few days, then disappears. We describe the micro-meme phenomenon and discuss the importance of this new tagging practice for the larger real-time search context.

Categories and Subject Descriptors
H.5.4 [**Information Interfaces and Presentation**]: Hypertext/Hypermedia – *User issues, architecture*

General Terms
Measurement, Human Factors

Keywords
Twitter, tagging, trends, memes.

1. INTRODUCTION
Tag selection in social tagging sites is often *a posteriori*, the key concepts are distilled into short strings that are then attached to a document, image, or resource to facilitate retrieval. In contrast, tagging has emerged as a method for filtering and promoting content in Twitter, rather than as a tool for recall. Users on Twitter have developed a tagging culture of placing a hash symbol (#) in front of short strings, called *hashtags*, on their posted messages, called *tweets*. Since then, a phenomenon of tagging which we call *micro-meme* has arisen. Participation in micro-memes is *a priori*; an individual user is unlikely to have composed a tweet on the topic in question if they had not observed the micro-meme tag in use by other Twitter users.

HT'10, June 13–16, 2010, Toronto, Ontario, Canada.

The concept of *a priori* tagging may seem counterintuitive, but tagging in Twitter is harnessed to achieve goals different from those of other social tagging platforms. A user who observes the rise of a compelling trending topic micro-meme may be inclined to take the tag associated with the meme and compose his or her own tweet on the subject. Thus, it is overwhelmingly likely that they might never have written the tweet if they had not been inspired to participate in the micro-meme phenomenon. This is supported by observations from our study, where the tags associated with micro-memes generally had not been used before they were embraced as micro-memes.

The differences between tagging practices in Twitter and tagging practices in traditional tagging platforms, such as Delicious, are partly due to structural differences between the two platforms. Tags in Delicious can be used to browse and discover new information about a topic, while tags in Twitter are primarily used to find messages from other users about a topic. The relationship between tag usage surrounding trending topics in Twitter differs from the tag usage surrounding trending topics in Delicious. Through the analysis of corpuses of tagging data from Twitter and Delicious, we will demonstrate that the lifecycles of tags related to trending topics have distinct trajectories in each of the two systems.

In this paper we explore a subset of hashtags used in Twitter from December 2008 to October 2009, which is the period when hashtags became widely adopted on the site. This data was collected by a third party by sampling tweets continuously over several years. We use a mixed method of statistical analysis as well as an interpretive approach to study the data. From looking at newly coined tags, we will discuss how statistical metrics differentiate micro-memes from other newly coined tags in Twitter. These results are then compared with those from Delicious data, where we do not observe similar phenomenon.

Our research question is: how are tags used in Twitter compared to Delicious; specifically, what can we tell about the adoption of these tags from analyzing their trends? Our contributions in this paper include being the first large-scale analysis of Twitter hashtags and introduction of the notions of conversational tagging and micro-memes.

1.1 Terminology
To accurately describe the tagging practices in Twitter, we use the following terminology consistently throughout the paper.

A *hashtag* is the specific name for a tag in Twitter. Hashtags derive their name from the fact that they are preceded by the symbol '#', also known as a hash mark, e.g., #nowplaying.

Adoption is the process by which a newly coined hashtag is embraced by a critical mass of users and disseminated through Twitter. For the purposes of this paper, we define adoption as tags representing less than 0.001% of the tags appearing per-day for the first 10 days data is collected, but later becomes widely-used.

Abandonment is when a critical mass of users stop attaching a specific hashtag to their tweets. Essentially, this is a decline in use over time until appearances of the tag become infrequent.

A *micro-meme* is a small-scale meme emerging around a Twitter hashtag. As more users adopt the hashtag, they add to an asynchronous, massively multi-person conversation by tweeting their thoughts about the topic prompted by the hashtag.

2. RELATED WORK

A number of previous studies have focused on time-series trends in search queries on the web. These studies of query patterns are highly relevant because of our argument that tagging practices in Twitter are shaped by the real-time communication dimension of the system. Adar *et al.* [1] studied the general trends for queries in several datasets of queries, blog posts, and news articles. Like our study, they looked at frequency of terms over time; they found that distinct trends could be explained by different user behaviors relating to the topic of the terms. Vlachos *et al.* [11] focused on burst and periodic queries, representing them concisely using coefficients in a Fourier transform. Identifying trends in queries has real applications; for example, Ginsberg *et al.* [6] were able to predict flu epidemics using search engine data before institutions such as the US Centers for Disease Control and Prevention. Extending this further with the same search data, Shimshoni *et al.* [10] were able to predict trends for categories of seasonal queries. Chien and Immorlica [3] presented an efficient method for finding related queries by correlating queries with similar time-series distributions. Carman *et al.* [2] analyzed queries, tags, and web content and found that queries and web content had a stronger overlap than tags and queries or tags and content. This illustrates that tags are not necessarily used exclusively for retrieval.

Other studies have explored tagging activity for websites where the goal is information organization. Santos-Neto *et al.* [9] studied tagging vocabulary used for scientific publications, for the purpose of classification and retrieval. They concluded that tagging is used for organizational purposes more than for collaboration with others. Wu and Zhou [13] uncovered a social aspect to tags in Delicious through various visualizations, showing that tags and users connected in a network have high levels of semantic and social relatedness. Marlow *et al.* [8] proposed a framework for the analysis of social tagging systems. We have found that Twitter can be classified as a social tagging system under their definition, but that the patterns of tagging behavior within Twitter are distinct. We also provide in-depth analysis of the tag entry mechanism Marlow *et al.* call "blind tagging", the category for which they have the least data. Ding *et al.* [5] looked at tagging practices in Delicious, YouTube, and Flickr. We expand upon their taxonomy of the tagging features of popular social networking sites to include Twitter. Twitter provides a useful set of distinctions because the ways tags can be used are differently constrained. Dellschaft and Staab [4] looked at tag stream data and ranked tags by frequency of use to understand how individual users made tag selections. It is uncommon for a tweet to be assigned more than one tag, so we do not compare tag co-occurrence; instead, we look at tag selection from a social context since Twitter users are influenced by the tags used by people in their network or from lists of trending topics, when they choose tags according to the 'Imitation' model proposed by Dellschaft and Staab.

3. METHOD

3.1 Data

Our data comes from 2 different sources, both of which are online services which allow users to tag content (Table 1). We have a sample of 42 million hashtags used in the microblogging website Twitter, inserted in messages posted by users. We have a sample of 378 million tags from the online bookmarking service Delicious, created by users to organize their bookmarks. Both of these datasets contain the tag along with the timestamp of when that tag was attached.

Dataset	Start Date	End Date	# Tags
Twitter[1]	Dec 2008	Oct 2009	42M
Delicious [12]	Jan 2006	Dec 2007	378M

Table 1: Summary of datasets used in the analysis.

While it would have been ideal for the datasets to cover the same time periods to control for temporal variations, obtaining large amounts of user data from different sources is a challenging task; we were constrained by the availability of data.

3.2 Processing

To explore this data, we constructed time-series charts of the tags, where the elements of the time-series are the number of times a tag is issued per day. We used only tags which appeared at least 10,000 times in the data to obtain sufficient data points to explore. Next, we removed tags which appeared more than 0.001% of the time on average for the first 10 days. This eliminated recurring tags so we could focus on newly coined tags. We normalized the frequency of each tag per day by dividing by the total number of tags sampled that day. This eliminated the weekly variation in online activity (since there is significantly more web traffic on Wednesday than Sunday) as well as different sample sizes used throughout the datasets. From looking at the data, we removed obvious cases of spam, in which a single website tagged their link on Delicious repeatedly over several months, and cases of data noise, for which it was obvious that the tag was invalid. Finally, we calculated the standard deviation, skew, and kurtosis.

4. INTERPRETIVE ANALYSIS

We begin our study with an interpretive analysis of the tags used in Twitter and Delicious. The authors went through the 224 most common tags in the Twitter dataset and the 304 most common tags in the Delicious dataset. The reason we first chose a qualitative approach was to make observations that would take into account the meaning of the tag itself. This allowed us to inform our interpretation with our knowledge of historical events and gave us a better understanding of the tag by seeing it used in the context of the message. One author had participated in Twitter for several years, allowing us to combine observed data with historical familiarity of the Twitter ecosystem to better understand phenomenon.

4.1 Following the Trend

Once users began to understand that tweets, in the aggregate, provided rich real-time information about specific issues, they began to build tools to help filter and highlight trending topics.

[1] Collected by Infochimps. http://infochimps.org/datasets/twitter-census-::-conversation-metrics-one-year-of-urls-hashtags

Many third-party Twitter clients, such as Brizzly[2] and Tweetdeck,[3] have automatically refreshing sidebars to display the topics currently the most popular. Later, Twitter itself began to display trending topic information on their front page (see Figure 1). These trending topic lists are individually linked to the current set of tweets composed on that topic. While tweets without hashtags were also displayed in trending topic lists, the act of tagging a tweet increased the likelihood of a tweet being displayed in a group of tweets on a trending topic.

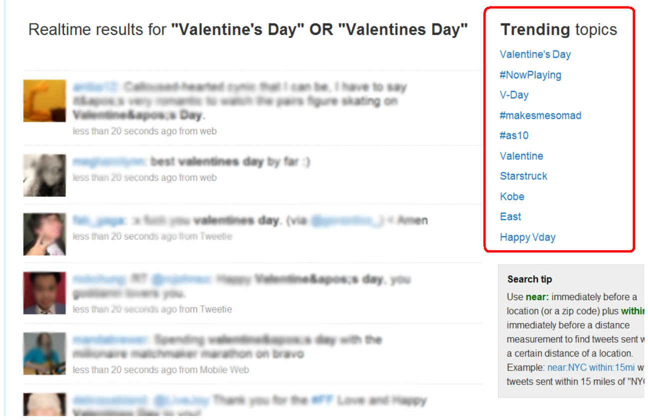

Figure 1: A screenshot of a Twitter search page, showing the Trending topics on the right sidebar, circled in red. Actual tweets are blurred for privacy.

4.2 Conversational vs. Organizational Tagging

Twitter is often characterized as an ephemeral stream of status updates. In one-hundred-forty characters or less, members of the Twitter community answer the question "What's happening?" back and forth among themselves. One year after Twitter went live, members of the community, without involvement or support from Twitter administrators, began tagging their tweets. If Twitter is a temporal stream of the aggregate thoughts of its users, why would anyone want to go to the effort of tagging their tweets? It would seem that there would be no use for attaching metadata to this evaporating pool of thoughts. It would also seem that in an online environment purposely constrained by space, the characters required to tag a tweet, which count toward the one-hundred-forty-character-per-tweet limit, tagging would be slow to catch on. In reality, many members of the Twitter community do use hashtags in their tweets.

Tagging practices differ between communities in which they are used due to the design of the individual systems as well as patterns of behavior which develop in response to those systems. In our analysis, we saw tagging practices in Delicious as examples of organizational tagging. In Delicious, tags are attached to resources in order to facilitate access to the resources at a later date. Tags in Delicious also facilitate discovery through browsing as each tag is a link to a list of all resources to which that tag has been attached. Thus, a user who tags a resource immediately places it in the context of all content that has been similarly tagged. Delicious was designed as a metaphorical library of bookmarks for users to access easily on the web. Tags are the metadata that provide organizational structure in the system.

[2] http://brizzly.com/
[3] http://www.tweetdeck.com/

Tagging practices in Twitter are an example of a new type of tagging, which we have chosen to call 'conversational' tagging. In conversational tagging, the tag itself is an important piece of the message. The tag can either serve as a label in the traditional sense of a tag, or it can serve as a prompt for user comment. In many trending topics, Twitter tags sometimes serve as prompts, and the resulting content is an asynchronous massively-multi-person conversation. While these are not the only types of tags used in Twitter, we argue that this is a type of tagging behavior that emerged due to the structure of the Twitter system.

4.3 Participating in Micro-memes

#igrewupon, #liesmentell, #igottacrushon and #90stweet are examples of hashtags we observed in Twitter associated with emergent micro-memes. These hashtags are rarely used to retrieve old tweets; instead, they provide synchronic metadata used to funnel related tweets into common streams. Figure 2 shows a time-series chart for the hashtag #itsabouttime. At its peak popularity, 3% of the tweets that day contained this micro-meme.

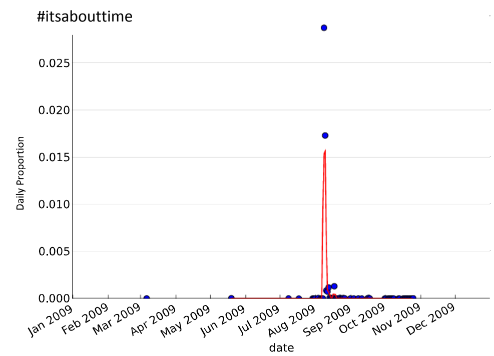

Figure 2: Use of the Twitter hashtag #itsabouttime over time. The value on the y-axis represents the proportion of this tag on that day. The red line is a 3-day moving average.

This type of tagging behavior s not observed in the Delicious data because the tags used in Delicious are attached to resources rather than to user-contributed data. In Twitter, a massively multi-party conversation can emerge as increasing numbers of users adopt the hashtag around which a micro-meme arises. Each user contributes his or her own commentary to the stream of conversation.

The motivation for an individual to participate in an episode of micro-meme tagging is to see their tweets displayed in the filtered stream of messages with that tag attached. Many of the micro-memes are constructed in topic-comment format, so people who use Twitter might be interested in skimming a few dozen to a few hundred tweets offering individual (often humorous or insightful) responses to the micro-meme. For example, in December 2009, the hashtag #willgetyouslapped was a trending topic. A Twitter user saw this hashtag and tweeted, "Taking online courses without owning a computer #willgetyouslapped. You can not take yo class on yo iphone". In this instance, the hashtag serves as the topic and the tweet can be interpreted as commentary on that topic.

5. STATISTICAL ANALYSIS

Next, we analyze tag and timestamp data using the second, third, and fourth moments: standard deviation, skew, and kurtosis. This allows us to quantitatively describe the adoption of tags over time. Our analysis covered both Twitter and Delicious; however, we focus our statistical report on Twitter hashtags because the patterns of Delicious data were unsurprising and generally regular.

5.1 Standard Deviation

We can calculate the standard deviation of the timestamps for each tag in Twitter. This measures the spread of activity of the tag, representing how long a tag remained in use. The standard deviation of our sample is calculated by the equation,

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}$$

From this, we find that a low standard deviation is a good indicator that the tag is used for conversational (i.e. social) rather than organizational purposes. Table 2 illustrates the tags extracted from Twitter along with the standard deviation of their post date. The tags are sorted by standard deviation, but only tags with the lowest and highest standard deviations are shown for brevity.

Tag	Std Dev
#ladiespleasestop	0.22
#ruleofrelationships	0.22
#clubrules	0.36
#helooksgoodbut	0.37
#thingssaidb4sex	0.38
#doyoumind	0.39
#oneletteroffmovies	0.51
#anybodyseen	0.54
#shelooksgoodbut	0.60
#excusemebut	0.68
Skip 214 tags…	
#goodnight	70.53
#fml	71.16
#gov20	71.62
#nsfw	72.76
#energy	73.52
#p2	75.03
#radio	78.22
#dollhouse	78.70
#bbcqt	82.74
#contest	90.91

Table 2: Tags from the Twitter dataset with their post timestamp sorted by standard deviation.

The tags in Table 2 with small standard deviations are what we earlier referred to as micro-memes. These micro-memes can be quantified in terms of their measures of standard deviation and kurtosis. Their measures of standard deviation are low compared to the measures of standard deviation of hashtags which are newly coined but not associated with micro-memes. This means they are both adopted and abandoned in a short period of time.

5.2 Skew

The skew of the tag timestamps tells us whether the tag is one which grows slowly and becomes adopted more over time, or one that catches on instantly and decays in use. In other words, it compares the rates of adoption against abandonment. Examples of positive and negative skew are shown in Figure 3. Skew is also known as the third moment, calculated by,

$$g_1 = \sqrt{\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^3}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{3/2}}}$$

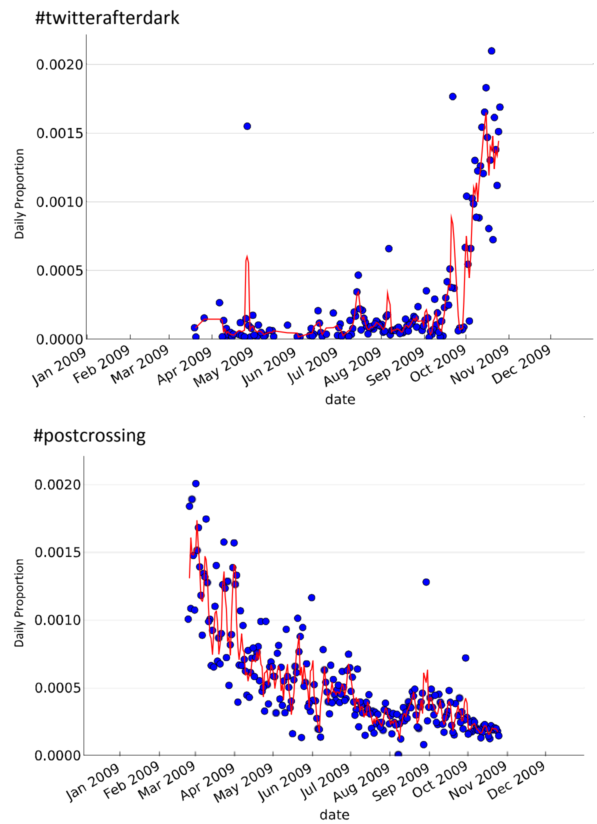

Figure 3: Time-series charts for #twitterafterdark (above), which has negative skew representing a large gain in adoption, and #postcrossing (below), where the positive skew indicates a slow abandonment of the tag.

5.3 Kurtosis

The fourth moment, kurtosis, represents the staying power of a tag around its peak popularity. The equation we used for kurtosis is,

$$g_2 = \sqrt{\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{2}}} - 3$$

This is a metric that concisely represents whether a tag maintained continued use for some period of time. It can be used to differentiate between micro-memes, recurring tags, or spam. Examples of tags with positive and negative kurtosis are presented in Figure 4.

6. DISCUSSION

The metrics we have presented can support automatic detection of tagging behavior associated with emergent micro-memes. This work is a first step towards building classifiers that can track the short-lived participatory phenomena which we expect to proliferate in other real-time aggregations of user-created content.

We have found a relationship between certain statistical metrics and the adoption of specific tags over time. This indicates that metrics like standard deviation and kurtosis have potential applications for automatically classifying the type of tag if its usage pattern is known. Perhaps micro-memes can be tracked as they are occurring to enhance social communication in Twitter. Skew can be used to measure the viral nature of a tag: whether a tag slowly gains traction up to a peak, or is adopted rapidly, but abandoned shortly after.

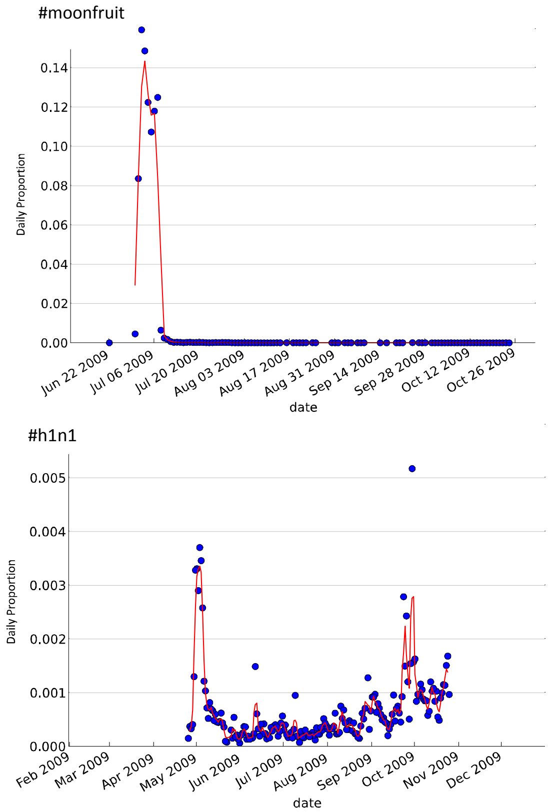

Figure 4: Charts for #moonfruit (above) where the positive kurtosis represents short temporal activity, and #h1n1 (below) with negative kurtosis, emphasizing the tag's staying power.

While Twitter is one of the first platforms where tagging has a conversational function, we expect these findings to generalize to other social platforms such as the recently launched Google Buzz. On the other hand, since tagging behavior differs depending on the system's focus on conversation or organization, studying tag behavior on Delicious or Flickr will not necessarily translate into the same findings on Twitter. Similarly, interfaces supporting tagging like the tag cloud have been commonly used for browsing content, but may be less useful for achieving conversational goals.

Our work adds to the discussion of why people use microblogging services. Two studies of Twitter [7,14] report nearly identical reasons: to report daily activities, conversation, sharing information or URLs, and breaking news. Clearly, the short-term conversational nature of hashtags are useful in conversation and breaking news. However, people's use of Twitter changes over time. There has been an increase in services which archive and allow searching of past tweets. This may alter the role of the hashtag in the future.

7. CONCLUSION

We use interpretive and statistical approaches to explore the patterns of Twitter tags over time and compare them with Delicious tags. We find that the differences in tagging practice between the two systems are caused by their design and function. Users add tags to their messages in Twitter to join discussions on existing topics. This leads to the phenomenon of micro-memes, where clever short-lived tags catch on and then die-out quickly. Statistical measures of standard deviation and kurtosis are correlated with patterning of these micro-memes.

This research results in an increased understanding of the motivation for social tagging. The knowledge will also enrich our understanding of how individuals classify their social communication in the aggregate. Through observation of how individuals spontaneously create links within a body of content, we can better address whether to design for the support of searching or browsing, or some combination thereof when elaborating systems for information retrieval. We also better understand the function of tagging in the real-time search environment in addition to the previously documented functions of tagging.

8. ACKNOWLEDGEMENTS

We thank Infochimps and the authors in [12] for collecting and sharing the data used in this study. We also thank the anonymous reviewers for their feedback on earlier drafts.

9. REFERENCES

[1] Adar, E., Weld, D. S., Bershad, B. N., and Gribble, S. S. 2007. Why we search: visualizing and predicting user behavior. In *Proceedings of WWW '07*, 161–170.

[2] Carman, M. J., Baillie, M., Gwadera, R., and Crestani, F. 2009. A statistical comparison of tag and query logs. In *Proceedings of SIGIR '09*, 123–130.

[3] Chien, S. and Immorlica, N. 2005. Semantic similarity between search engine queries using temporal correlation. In *Proceedings of WWW '05*, 2–11.

[4] Dellschaft, Klaas, and Steffen Staab. 2008. An epistemic dynamic model for tagging systems. In *Proceedings of HT '08*, 71–80.

[5] Ding, Y, Jacob, E.K., Zhang, Z., Foo, S., Yan, E., George, N.L., and Guo, L. 2009. Perspectives on social tagging. *JASIST*, 60, 12, 2388–2401.

[6] Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., and Brilliant, L. Detecting influenza epidemics using search engine query data. *Nature*, 457, 1012–1015.

[7] Java, A., Song, X., Finin, T., and Tseng, B. 2007. Why we twitter: understanding microblogging usage and communities. In *SNA-KDD Workshop Proceedings*, KDD 2007, 56–65.

[8] Marlow, C, Naaman, M., Boyd, D., and Davis, M. 2006. HT06, tagging paper, taxonomy, Flickr, academic article, to read. In *Proceedings of HT '06*, 31–40.

[9] Santos-Neto, E., Condon, D., Andrade, N., Iamnitchi, A., and Ripeanu, M. 2009. Individual and social behavior in tagging systems. In *Proceedings of HT '09*, 183–192.

[10] Shimshoni, Y., Efron, N., and Matias, Y. 2009. On the Predictability of Search Trends. Working draft. Retrieved http://research.google.com/archive/google_trends_predictability.pdf on February 2, 2010.

[11] Vlachos, M., Meek, C., Vagena, Z., and Gunopulos, D. 2004. Identifying similarities, periodicities and bursts for online search queries. In *Proceedings of SIGMOD '04*, 131–142.

[12] Wetzker, R., Zimmermann, C., and Bauckhage, C. 2008. Analyzing social bookmarking systems: A del.icio.us cookbook. In *MSoDa Workshop Proceedings*, ECAI 2008, 26–30.

[13] Wu, C. and Zhou, B. 2009. Analysis of tag within online social networks. In *Proceedings of GROUP '09*, 21–30.

[14] Zhao, D. and Rosson, M. 2009. How and why people Twitter: the role that micro-blogging plays in informal communication at work. In *Proceedings of GROUP '09*, 243-252.