
Making sense of Twitter Search

Gene Golovchinsky

FX Palo Alto Laboratory, Inc.
3400 Hillview Ave, Bldg 4
Palo Alto, CA 94304 USA
gene@fxpal.com

Miles Efron

Graduate School of Library and
Information Science
University of Illinois at Urbana-
Champaign
Champaign, IL 61820 USA
mefron@illinois.edu

Abstract

Twitter provides a search interface to its data, along the lines of traditional search engines. But the single ranked list is a poor way to represent the richly-structured Twitter data. A more structured approach that recognizes original messages, re-tweets, people, and documents as interesting constructs is more appropriate for this kind of data. In this paper, we describe a prototype for exploring search results delivered by Twitter. The design is based on our own experience with using Twitter search, and as well as on the results of a small online questionnaire.

Keywords

Twitter, information seeking, HCIR

ACM Classification Keywords

H5.2. User Interfaces: User-centered design, H.5.4 Hypertext/Hypermedia: Navigation.

General Terms

Human Factors

Introduction

Twitter is fast gaining popularity as an effective way to communicate for a variety of purposes, including sharing and disseminating news, keeping in touch, marketing or promoting products, etc. [8]. While much interaction with Twitter happens by passive monitoring

Copyright is held by the author/owner(s).

Workshop on Microblogging: What and How Can We Learn From It?

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

ACM 978-1-60558-930-5/10/04.

of followers punctuated by more focused conversations, search is becoming more important as the volume of information on Twitter grows.

This importance has been reflected in the recent incorporation of Twitter results into search engines such as Google and Bing, that can now retrieve trending Twitter topics related to a traditional web search. This incorporation of near-real time information is meant to improve the recency of the retrieved information.

Twitter offers its own search interface, patterned on more traditional search engines. Each posting is formatted to show the icon and screen name of the tweep, the tweet text, and the time it was sent; the list is ordered showing the most recent tweet first (**figure 1**). In addition, the right margin shows some general information about trending topics, nifty queries, and language filtering and translation.



figure 1 Twitter search interface

The problem with this interface is that it is optimized for showing the most recent results, underscoring Twitter's interest in what's going on *right now*. In fact, this focus is so pervasive that Twitter search only retrieves tweets of up to a couple of weeks in age for

any query; older material, even though it is accessible by browsing or by direct linking, is not returned.

But people search for many reasons, not just to find the latest and greatest, and they search for many kinds of information. In the following discussion, we briefly characterize some information needs that people bring to Twitter, and then describe an interface designed to facilitate the exploration of search results.

Reasons for searching Twitter

A number of studies have appeared recently examining a variety of Twitter-related phenomena, including re-tweeting [2], twitter trends [3], collaboration [4], real-time commentary on events [5], and disaster-related communication [6]. In addition, there is considerable work in mining the Twitter stream for sentiment analysis, product mentions, etc. Little attention, however, has been paid to how people search Twitter, and to how they explore returned search result sets. One study [1] did explore the use of timelines to cluster search results, but did not look at other available metadata.

We conducted a short online questionnaire (a link to a Google form distributed via Twitter) that asked people to indicate how often they searched Twitter, and what kinds of information they were looking for. We also asked for optional free-form comments that describe a specific recent search experience.

More than half of the respondents use Twitter search at least once a week (**table 2**), and most searched for events (such as conferences), for people, and for trending topics (**table 2**). Trending topics queries can be handled the same as other kinds of queries, and so will not be discussed further..

| Frequency of use | # ppl |
|-----------------------|-------|
| Daily | 6 |
| At least once a week | 7 |
| At least once a month | 8 |
| At least once a year | 2 |
| Never | 0 |

table 2 Frequencies of Twitter Search Use (23 respondents)

| Purpose | # ppl |
|-----------|-------|
| Events | 17 |
| Trending | 12 |
| People | 13 |
| Documents | 3 |
| Other | 10 |

table 2 Types of information sought (23 respondents)

Hash tags are user-generated metadata intended to represent a particular topic. We argue that using hash tags in Twitter search can improve users' experience. For example, events such as trade shows (#ces), conferences (#chi2010), elections (#debate08), geo-political events (#iran), etc. are often represented by hash tags in Twitter. Searching on such a hash tag can retrieve hundreds or thousands of tweets, including discussion, re-tweets, links to documents, etc. The small size of each tweet makes it hard to estimate the relevance of that tweet that underlies traditional ranking algorithms. Furthermore, the usefulness of a tweet may lie in who sent it, when it was sent, or what document it referred to. Twitter search's temporal linear presentation may make it difficult to extract key memes or documents that characterize an event.

figure 2 Screenshot showing tweeters and re-tweeters on a search on 'sfusgs'. The list on the left shows tweets by @ayman, the list on the right re-tweets by @ayman

A novel search results browser

We built an initial version of a browser for exploring Twitter search results. The system issues a user-specified query to Twitter, parses the results, groups tweets by users, extracts document references, and organizes the results in terms of people, tweets, and documents. It also gives an overview of the statistics associated with the results.

The "people" view (**Error! Reference source not found.**) is divided into tweeps who tweeted and those who re-tweeted messages. A tweep can appear in both lists. A tooltip for each tweep shows the person's name and the number of tweets in the result set. Clicking on the icon associated with a tweeter shows messages associated with that person.

Re-tweets are identified by looking at the metadata provided by the Twitter API, and by applying some heuristics to the text of the messages in the absence of re-tweet metadata. Heuristics look for textual similarity, for proper temporal sequence, for the mention of a sender's screen name, etc.

The "tweets" view shows a hierarchical grouping of tweets with their re-tweets. Tweets can be ordered by the number of "descendant" re-tweets or by time.

The "documents" view lists the documents mentioned in the results sets, shows all the tweets that mention the document, and allows the searcher to read the document without leaving the search results.

Ongoing work

We have just started on this process of exploring how best to support people in their explorations of Twitter

result sets. We are in the process of extending the software to organize tweets by content. Hash tags provide an obvious point of leverage here. That is, organizing search results into groups based on shared hash tags that were found in the search results offers an avenue for presenting topically clustered results.

Another avenue of exploration is how to incorporate the contents of documents referred to by the tweets. Finally, social network analysis may be applied to the tweeks identified in search results to help understand the discussion from that perspective. Tweeks can be ordered by metrics computed from the social graph, such as by in- or out-degree, or by aggregate measures such as TunkRank [7].

As the number of tweeks in a search result set goes up, in addition to sorting people by a variety of attributes, filtering on those (or other) attributes may also be useful. The exact configuration of capabilities and interfaces will have to be determined through an iterative, user-centered design process.

One limitation of Twitter search is the two or so week horizon for tweets returned in search results. To compensate for this limitation, tools such as www.twapperkeeper.com allow users to create archives of tweets that use a particular hashtag. These archives collect tweets on an on-going basis, and can be exported in a CSV file. Our system will ingest these files to browse these archived collections.

While this research has focused pragmatically on Twitter, it has broader implications. We are exploring how to help people make sense of moderate-sized corpora of messages and related information. This kind

of analysis is useful not only for Twitter and Facebook status updates, but also for understanding e-mail and other message traffic. Our approach is to create an interface structure that enables people to use their skills and knowledge to make sense of the data, while providing scaffolding for incorporating automated approaches to data analysis.

References

- [1] Alonso, O., Gertz, M., and Baeza-Yates, R. 2009. Clustering and exploring search results using timeline constructions. In *Proc. CIKM'09*. ACM 97-106. DOI= <http://doi.acm.org/10.1145/1645953.1645968>.
- [2] boyd, d., Golder, S., and Lotan, G. 2009. "Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter," In *Proc. HICSS-43*.
- [3] Cheong, M. and Lee, V. (2009). Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base. In *Proc. SSM2009*.
- [4] Honeycutt, C. and Herring, S.C. (2009). Beyond microblogging: Conversation and collaboration via Twitter. In *Proc. HICSS-42*. IEEE Press.
- [5] Shamma, D.A., Kennedy, L., and Churchill, E.F. (2009). Tweet the Debates: Understanding Community Annotation of Uncollected Sources. In *Proc. ACM Multimedia*.
- [6] Sutton, J., Palen, L., and Shlovski, I. (2008). Back-Channels on the Front Lines: Emerging Use of Social Media in the 2007 Southern California Wildfires. In *Proc. ISCRAM 2008*.
- [7] Weng, J., Lim, E-P., Jiang, J. and He, Q. 2010. TwitterRank: Finding Topic-sensitive Influential Twitterers. In *Proc WSDM 2010*.
- [8] Zhao, D. and Rosson, M. 2009. How and why people Twitter: the role that micro-blogging plays in informal communication at work. In *Proc. GROUP '09*. ACM Press, pp. 243-252. DOI= <http://doi.acm.org/10.1145/1531674.1531710>