A Torrent of Tweets: Managing Information Overload in Online Social Streams

Michael Bernstein
MIT CSAIL
32 Vassar Street
Cambridge, MA 02139
msbernst@mit.edu

Lichan Hong
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
lichan.hong@parc.com

Sanjay Kairam
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
sanjay.kairam@parc.com

Ed H. Chi
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
chi@parc.com

Bongwon Suh
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
bongwon.suh@parc.com

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract
Twitter streams are on overload: active users receive hundreds of items per day and existing interfaces force us to march through a chronologically-ordered morass to find tweets of interest. We propose that the research community engage with microblogging feed consumption practice: how do users manage the incoming flood of updates, and how can we help them do it? We have pursued these questions through formative studies and prototype development. An online survey and semi-structured interviews with information stream users revealed that users assign different value to feed items according to tie strength, topical relevance and serendipitous discovery. Faced with an intractable number of these streams, users either manicure follow lists for manageability or give up on reading everything and subscribe to many more individuals. We introduce an alternative Twitter client called Eddi that groups tweets in a user's feed into topics mentioned explicitly or implicitly. Using Eddi, Twitterers can browse for tweets of interest by topic, filtering out undesired topics.

Introduction

Social status streams threaten to become a torrent. As microblogging services like Twitter have evolved into a powerful means for sharing ideas and opinions [2, 3], highly active users now regularly receive over a thousand tweets each day. This flood of tweets covers an impressive variety of topics, many of interest to the user and many not. Users must choose to reduce the amount of information they subscribe to, or stop reading everything and resign themselves to missing interesting or entertaining tweets.

In contrast to other media such as e-mail [5], we know relatively little about how individuals actually manage and consume their feed. We propose that the research community examine feed management in services like Twitter: How do users deal with the incoming stream of updates? How can we support or improve their management strategies?

Studies of Current Consumption Patterns

Here we outline our investigations of microblog management via surveys and semi-structured interviews.

We began with an informal survey, recruiting participants via Amazon Mechanical Turk who were users of social streams such as Twitter, Facebook and MySpace. We inquired as to personal/professional boundaries, tools used, what role the streams played in users' lives, and collaboration opportunities using the services. 78 participants responded (42 female), aged 14-47 (μ=25.6, σ=8.0), mostly college graduates.

During a later study of a prototype Twitter client, we had the chance to investigate feed consumption patterns in greater depth via semi-structured interviews. 15 users answered our call on Twitter in exchange for a small gratuity. Their ages ranged from 19 to 49 (median 29). These users had an average of 786 tweets per day in their feed (σ=658), with a large right skew: the maximum was 2,840.

The combined output of both studies taught us much about current strategies:

Scale: Turn It Down, or Go "Twitter Zen"?
Some users pruned their follow list to ensure that they could read every tweet. Others went "Twitter Zen" and let go of the urge to read everything in their feed. These users resorted to temporal sampling: dipping into the stream in a few moments of downtime to read recent tweets. The sampling process leads to a high degree of serendipity in the encountered information, but large contiguous portions of the feed are never seen and the user has little control over what is read.

Valued Tweets: People, Topics and Serendipity.
We asked participants in the semi-structured interviews to identify tweets they were particularly glad to have seen. Enjoyment was largely driven by three factors:

- **Topic relevance:** Users particularly appreciated tweets relevant to their interests. We empowered this approach in the Eddi prototype, described below.
- **Tie strength:** Users made sure to read tweets by strong ties and individuals of particular interest. This approach lends credence to the tie-strength design taken by Eric Gilbert in WeMeddle (wemeddle.com).
- **Serendipity:** Some users followed accounts that were far from their usual social circles or interests, simply because the person's tweets were interesting. These users deliberately increased the amount of "noise" in their feed for the reciprocal opportunity to find a needle in the haystack.

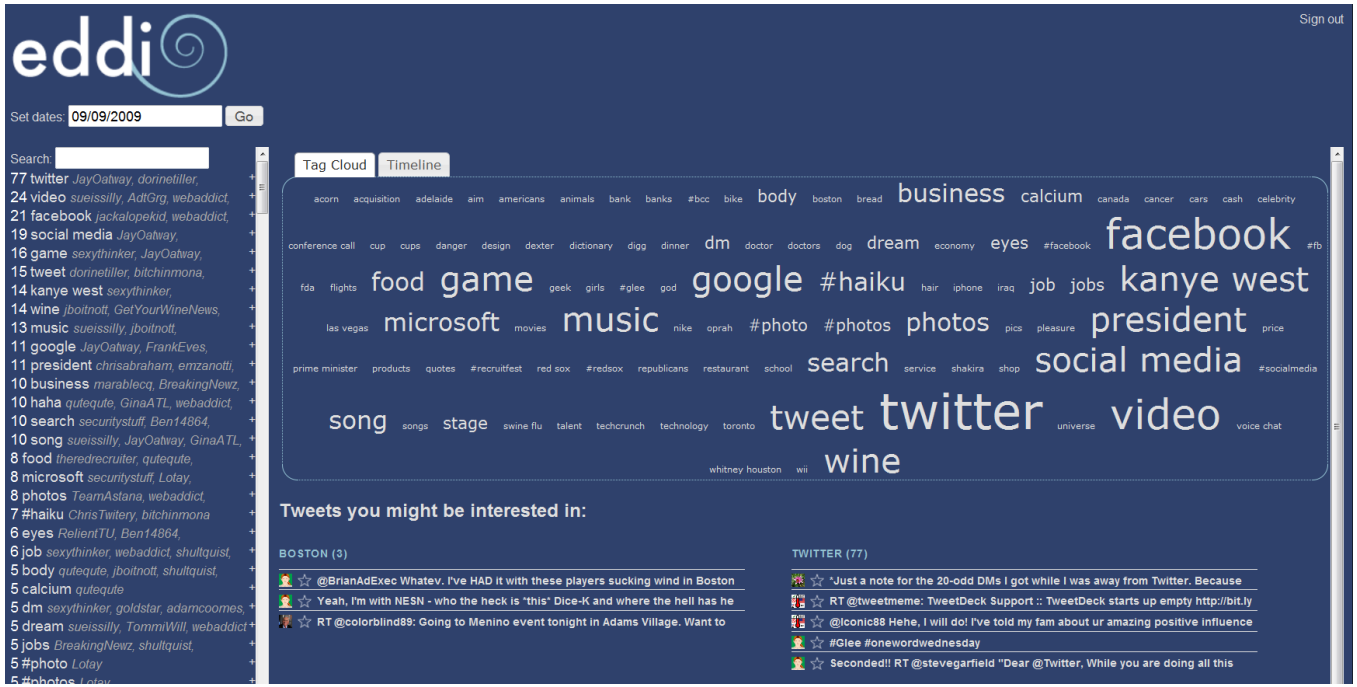

Figure 1. Eddi is a topic-oriented browsing interface for Twitter. Clockwise from upper right is the tag cloud, timeline (hidden in another tab), the topic dashboard, and the navigational list.

Many users attempted to balance these goals by separating the "must read" tweets from the ambient stream. For example, one interview participant used four accounts to manage readership expectations; others used filtered views in clients like TweetDeck.

Twitter Professional Enterprise Edition 2010 – or not. Of our 78 Mechanical Turk participants, 70 reported using microblogging sites for personal reasons. 16 said they used the sites for professional work. It seems that microblogging has not yet achieved saturation for professional uses.

Eddi: Topic Browsing For Your Feed

Our exploratory studies have led us to believe that a topic-oriented interface for an information stream like Twitter can help users manage an overwhelming feed. Topic categories allow a user to focus on updates on topics of interest, giving more control over a stream that cannot be read wholesale. However, 140 characters is too short for typical topic identification algorithms to succeed. We have developed a novel technique for solving the topic identification problem by combining approaches from the information retrieval and web research communities [1, 4]. In our approach,

the tweet "macbook died, but the Genius guys gave me a new one!" will be associated with topics like *MacBook, Apple, Apple Store* and *Genius Bar*. We utilize this algorithm in the interface.

We developed an alternate Twitter browsing interface, called Eddi (Figure 1), that visualizes a user's Twitter feed using topic clusters constructed using our novel topic identification algorithm. It is a trending topics interface for your own feed. Topic browsing enables behaviors that were very difficult in Twitter before: The overview allows users to get an at-a-glance view of their stream. In Figure 1, the user's friends started tweeting about *President* (Obama) as well as *Kanye West*. Then, users choose a topic to dive into detail on a topically coherent subset of their Twitter feed. For example, a user might choose to browse tweets relating to *research*, then *design* or *Kanye West*.

In our semi-structured interview study, we introduced participants to Eddi and gave them 3 minutes to consume 24 hours' worth of their Twitter feeds from 4 separate days. Participants found Eddi to be an efficient way to browse their feeds. "Eddi helps me find things that I'm interested in, faster," one participant explained. "It gives me a very quick way to have a first pass and to keep me from needing to read 140 characters about something I don't care about," another participant offered. The main drawback of the interface was that it was more difficult for users to see every tweet if so desired, since individual tweets were buried under a layer of hierarchy. The chronological interface was simply "less enjoyable but more comprehensive." Other users, however, didn't feel a need to see every tweet, so the abstraction was a positive trait: "With the serial feed I feel like I need to see everything. Here, I can quickly get a gestalt."

Conclusion

Our work has driven us to explore the role of the consumer in the microblogging ecosystem. We found that users struggle to balance the promise of interesting content with the sheer volume of the incoming tweets. Several users value tweets on particular topics, but topic metadata is not available in Twitter. Therefore, we have developed a system called Eddi that empowers topic-based browsing.

References

[1] Bendersky, M. and Croft, W.B. Discovering key concepts in verbose queries. ACM Conference on Research and Development in Information Retrieval, ACM Press (2008), 491.

[2] Java, A., Song, X., Finin, T., et al. Why We Twitter: Understanding Microblogging Usage and Communities. Proc. WebKDD '07, ACM Press (2007), 56-65.

[3] Naaman, M., Boase, J., and Lai, C. Is it Really About Me? Message Content in Social Awareness Streams. Proc. CSCW '10, ACM Press (2010), 0-3.

[4] Sahami, M. and Heilman, T.D. A web-based kernel function for measuring the similarity of short text snippets. Proc. WWW '06, ACM Press (2006), 377.

[5] Whittaker, S. and Sidner, C. Email overload: exploring personal information management of email. Proc. CHI '96, ACM Press, 276-283.