Automatic Identification and Presentation of Twitter Content for Planned Events

Hila Becker*[†], Feiyang Chen[†], Dan Iter[†], Mor Naaman[‡], Luis Gravano[†]
[†] Columbia University, [‡] Rutgers University

Abstract

We demonstrate a system for augmenting information about planned events with Twitter messages, using a set of automatic query building strategies. We present two alternative interfaces to our system, namely, a browser plug-in and a customizable Web interface.

1 Introduction

Event-based information sharing and seeking are common user interaction scenarios on the Web today. The bulk of information from events is contributed by individuals through social media sites such as Twitter, Facebook, and others. These sites provide event information in anticipation of, during, and following scheduled events. In this demo, we focus on automatically augmenting structured Web pages for planned events with related messages from Twitter.

We consider a Twitter message, or *tweet*, to be related to an event if it provides a reflection on the event before, during, or after the event occurs. Consider the "Yoga at the Great Lawn" event that took place on June 22, 2010 in New York City's Central Park. The event's related tweets can reflect anticipation of the event (e.g., "We are giving away 10 FREE tickets to Yoga at the Great Lawn on 6/22"), participation in the event (e.g., "Waiting on line... have your on-time friends save you a spot"), and post-event reflections (e.g., "Yoga at the Great Lawn may have rained out..."). All of these tweets may be relevant to a person seeking information about this event at different times.

Existing tools to find, organize, and present social media content associated with events are limited. Most related efforts focus on detecting unknown events in social media (Becker, Naaman, and Gravano 2010; Sankaranarayanan et al. 2009), while others need manually selected terms describing the event (Sakaki, Okazaki, and Matsuo 2010; Yardi and boyd 2010). Our demo relies on a variety of strategies for *automatically* constructing queries for events based on structured event content from sites such as Upcoming (`upcoming.org`).

A key challenge is that the (user-contributed) structured content about events cannot generally be used as queries directly. Specifically, this content is often either ambiguous and does not uniquely identify the event (e.g., "Yoga in the Park"), or too specific (e.g., "10,000-person Yoga practice at the Great Lawn in Central Park"), making the extraction of essential event information difficult. We developed a system that returns Twitter messages related to an event using a combination of simple rules and advanced (learned) query building strategies. At the core of the demo are two interfaces to this system, namely, a browser plug-in and a customizable Web interface, which automatically display tweets related to Upcoming events.

2 Strategies for Identifying Event Tweets

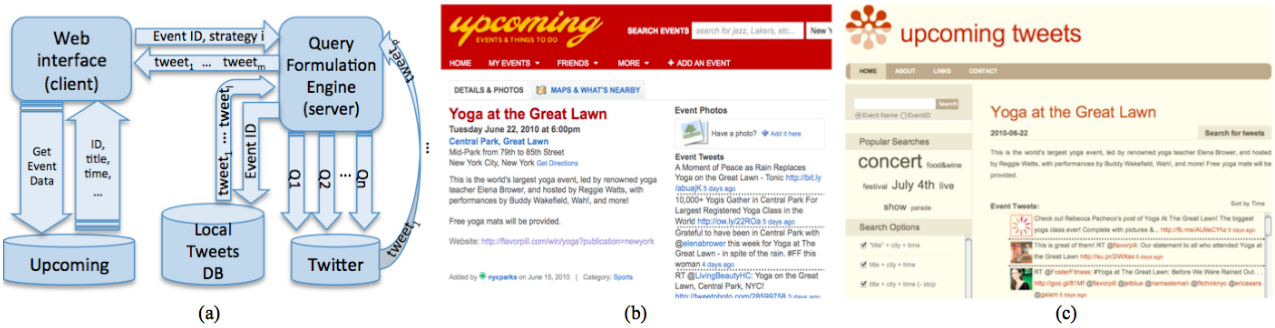
To identify Twitter messages for an event, we begin with simple query building strategies aimed at achieving high-precision results. These strategies derive queries from the structured description of an event, and touch on various aspects of the event (e.g., combining title and venue: ["yoga at the great lawn"+"central park"]), following the intuition that these highly restrictive queries should only result in messages that relate to the intended event. To identify high-precision strategies, we used an annotator to label the results returned by each compelling strategy for over 50 events. The resulting final set of strategies, by design, generally offers high precision, though often at the expense of recall.

We use the high-precision strategies to retrieve tweets for each event. To improve recall, we employ term-frequency analysis and co-location techniques on the high-precision tweets to identify descriptive event terms and phrases, which we use, in turn, to define new queries. Additionally, we build queries using URL and hashtag statistics from the high-precision tweets for an event. We build a rule-based classifier to select among this new set of queries, and then use the selected queries to retrieve additional event messages.

3 System and User Interfaces

We created a system that, given a set of Upcoming event features and a strategy, builds the appropriate queries as outlined by the strategy and returns all matching event messages retrieved via the Twitter API (`http://dev.twitter.com`). Our system is useful as a stand-alone service, but it can also be incorporated into, and extended by, a variety of applications. In this demo, we present two applications that build on this query formulation system to create two user experiences for interacting with tweets for planned events.

[*] Contact author: Hila Becker, hila@cs.columbia.edu



(a)

upcoming EVENTS & THINGS TO DO

SEARCH EVENTS search for jazz, Lakers, etc... New Y

HOME MY EVENTS FRIENDS MORE + ADD AN EVENT

DETAILS & PHOTOS MAPS & WHAT'S NEARBY

Yoga at the Great Lawn
Tuesday June 22, 2010 at 6:00pm
Central Park, Great Lawn
Mid-Park from 79th to 85th Street
New York City, New York Get Directions

This is the world's largest yoga event, led by renowned yoga teacher Elena Brower, and hosted by Reggie Watts, with performances by Buddy Wakefield, Wah!, and more!

Free yoga mats will be provided.

Website: http://flavorpill.com/win/yoga?publication=newyork

Added by nycparks on June 15, 2010 | Category: Sports

Event Photos

Have a photo? Add it here

Event Tweets

A Moment of Peace as Rain Replaces Yoga on the Great Lawn - Tonic http://bit.ly/abuajK 5 days ago

10,000+ Yogis Gather in Central Park For Largest Registered Yoga Class in the World http://ow.ly/22ROa 5 days ago

Grateful to have been in Central Park with @elenabrower this week for Yoga at The Great Lawn - in spite of the rain. #FF this woman 4 days ago

RT @LivingBeautyHC: Yoga on the Great Lawn, Central Park, NYC! http://tweetphoto.com/28599758 3 days ago

(b)

upcoming tweets

HOME ABOUT LINKS CONTACT

Search
◉ Event Name ○ EventID

Popular Searches
concert food&wine festival July 4th live show parade

Search Options
☑ "title" + city + time
☑ title + city + time
☑ title + city + time (- stop

Yoga at the Great Lawn
2010-06-22
Search for tweets

This is the world's largest yoga event, led by renowned yoga teacher Elena Brower, and hosted by Reggie Watts, with performances by Buddy Wakefield, Wah!, and more! Free yoga mats will be provided.

Event Tweets:
Sort by Time

Check out Rebecca Pacheco's post of Yoga At The Great Lawn! The biggest yoga class ever! Complete with pictures &... http://fb.me/AU9eCYhz 5 days ago

This is great of them! RT @flavorpill: Our statement to all who attended Yoga at the Great Lawn http://su.pr/2iWXax 5 days ago

RT @FosterFitness: #Yoga at The Great Lawn: Before We Were Rained Out. . . http://goo.gl/81Mf @flavorpill @jetblue @namastemari @fitchicknyc @ericasara @gaiam 5 days ago

(c)

Figure 1: System diagram (a), browser plug-in (b), and customizable interface (c).

3.1 Browser Plug-In

Our first sample application uses a browser plug-in script that enables seamless embedding of tweets related to planned events on Upcoming (Figure 1(b)). This plug-in script calls a query formulation engine with a set of pre-configured querying strategies (Figure 1(a)), which can be modified via plug-in settings, or dynamically selected based on event type. When a user navigates to an Upcoming event page, the plug-in script collects the event ID and associated event features, and sends them to the query formulation engine, along with the selected query building strategies. The query formulation engine automatically constructs appropriate event-specific queries using the event's features, as required by each query building strategy. The engine then collects all of the matching messages from Twitter, and finally sends the results to the plug-in script.

For efficiency, we issue asynchronous requests to the query formulation engine, where each strategy corresponds to one request, and post the tweets in the order in which they are returned. We use a hash-map to keep track of all tweets that are already displayed on the page, in order to avoid displaying duplicate messages, which may be returned by the different strategies. Additional performance improvement is gained from issuing a request for any locally cached tweets (see local tweets database in Figure 1(a)) that were previously retrieved for this event by any user of our system. We dynamically append the resulting tweets to the Upcoming page, alongside the event description.

3.2 Customizable Web-based Interface

Our customizable interface enables users to select specific strategies for automatically retrieving tweets for any given Upcoming event. Through this interface, users can either search for events or select from a list of recent Upcoming events (Figure 1(c)). On the sidebar, we display the list of query building strategies that, if checked, will be used in the retrieval process. When a user selects the "search for tweets" link for an event, our interface issues simultaneous, asynchronous calls to our query formulation engine, to retrieve tweets for this event according to each selected strategy.

We display the Twitter messages dynamically, as soon as they are retrieved, for efficiency reasons. However, we also include options for ranking the tweets according to various criteria. One such ranking criterion is to order the tweets according to the time at which they were posted, in case a user is interested in the most up-to-date information about an event. Another ranking option orders tweets according to the number of strategies for which they were retrieved. In the future, we plan to experiment with richer ranking functions.

This interface provides flexibility by allowing users to dynamically modify the set of retrieval strategies, which is particularly useful when the high-precision strategies have insufficient recall. Another feature is a user-driven option to remove generally high-precision strategies when they are expected to introduce noise for a specific event: for example, this scenario might happen with a "title-only" query strategy when an event title is ambiguous (e.g., "4th of July Celebration," referring to an event in Charleston, SC).

4 Conclusion

We present a system that uses a variety of query building strategies, aimed at automatically augmenting user-contributed information for planned events with dynamically generated Twitter content. Our demo system enables experimentation and exploration of substantially different query formulation strategies for the important task of identifying relevant Twitter content for planned events.

5 Acknowledgments

This material is based on work supported by NSF Grants IIS-0811038, IIS-1017845, and IIS-1017389, and by two Google Research Awards. In accordance with Columbia Univ. reporting requirements, Prof. Gravano acknowledges ownership of Google stock as of the writing of this paper.

References

Becker, H.; Naaman, M.; and Gravano, L. 2010. Learning similarity metrics for event identification in social media. In *WSDM'10*.

Sakaki, T.; Okazaki, M.; and Matsuo, Y. 2010. Earthquake shakes Twitter users: Real-time event detection by social sensors. In *WWW'10*.

Sankaranarayanan, J.; Samet, H.; Teitler, B. E.; Lieberman, M. D.; and Sperling, J. 2009. Twitterstand: News in tweets. In *GIS'09*.

Yardi, S., and boyd, d. 2010. Tweeting from the town square: Measuring geographic local networks. In *ICWSM'10*.