# Making Sense of the Immense: An Application of Big Data Visual Analytics for Spatiotemporal Sensemaking in Social Media

IGERT Trainee:
Joshua E. Stevens[1]
Project Team (alphabetical order):
Justine Blanford[1], Ying Chen[2], Sujatha Das[3], Wenyi Huang[3], Frank Hardisty[1], Anuj Jaiswal[3], Yiting Ju[1], Morteza Karimzadeh[1], Alan M. MacEachren[1], Prasenjit Mitra[3], Scott Pezanowski[1], Anthony C. Robinson[1], Alexander Savelyev[1], and Joshua E. Stevens[1]
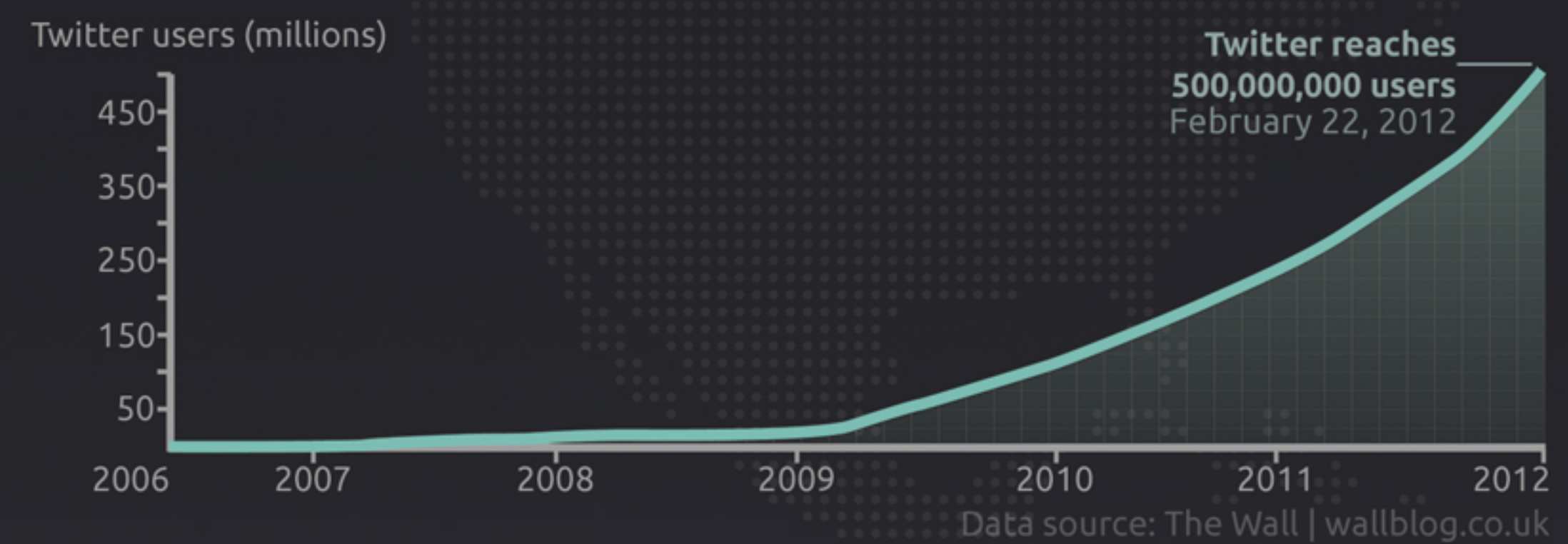
[1]Department of Geography, [2]Department of Computer Sciences and Engineering, [3]College of Information Sciences and Technology

## Introduction

Since Twitter launched in 2006, the amount of new users and total posts has grown precipitously. Twitter currently boasts more than **517 million total users**, accounting for over **500 million tweets each day** [1,2]. With several billion tweets posted each month, an enormous opportunity awaits emergency managers and analysts.
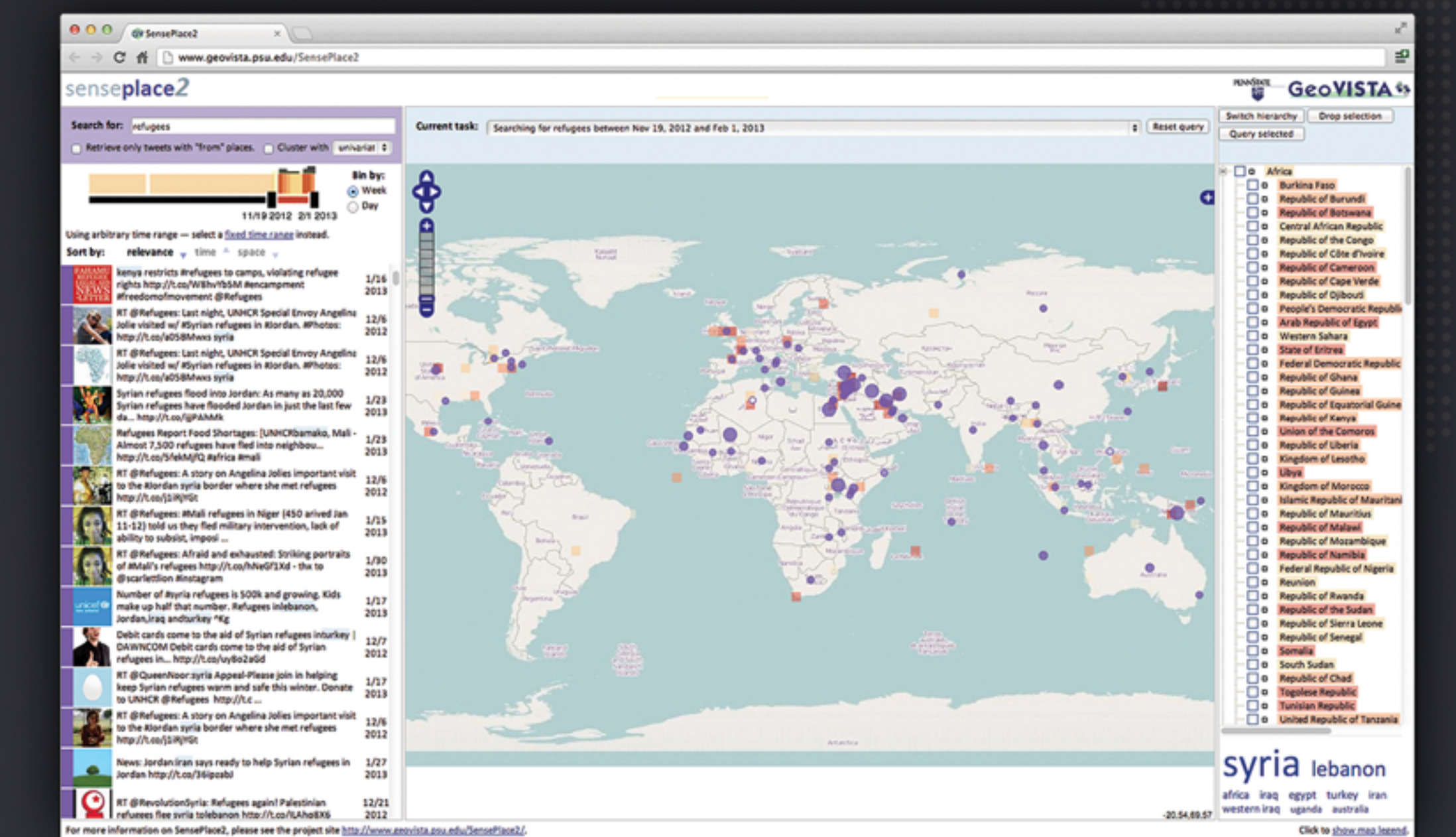
Despite the ability to associate tweets with geographic locations (via geotagging), **fewer than 1% of tweets use this feature** [3], though **in times of crisis up to 15% of tweets may be geotagged** [4].

Our work pairs the torrential stream of tweets with entity extraction and geocoding algorithms to determine *what is being tweeted about*, *where tweets are referring to*, and *how topics vary across space and time*.

Twitter users (millions)

Twitter reaches 500,000,000 users February 22, 2012
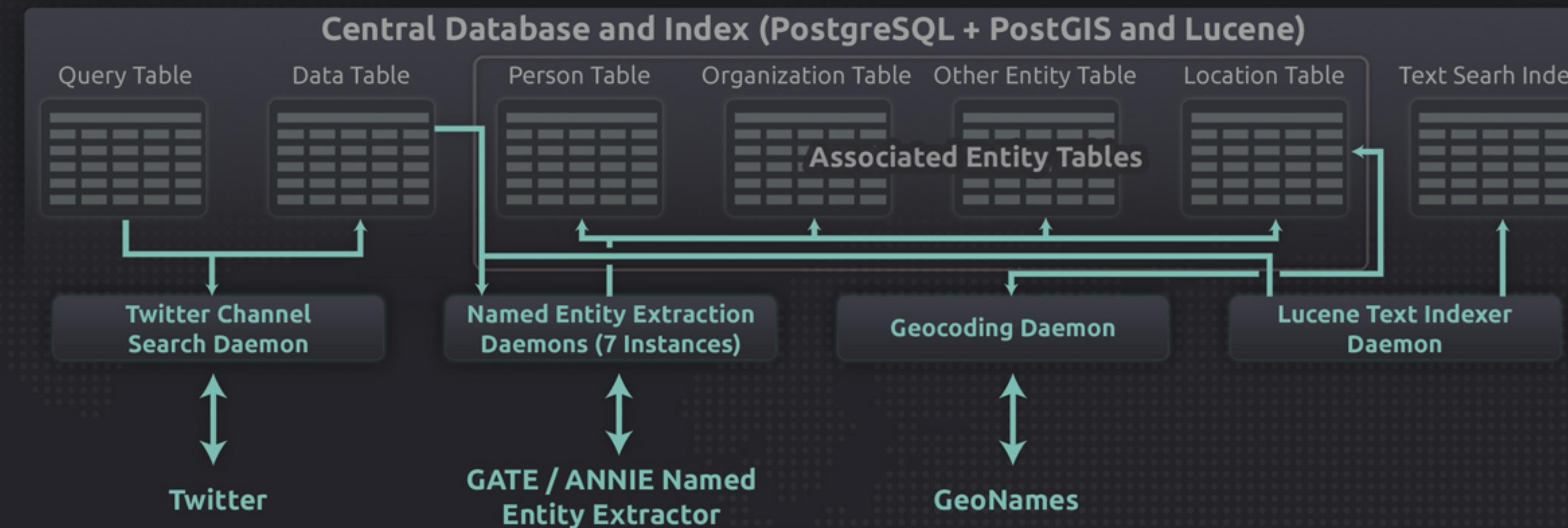
Data source: The Wall | wallblog.co.uk

## Enabling Analytics: *SensePlace2*

*SensePlace2* is a web-accessible geovisual analytics tool that allows analysts to **achieve greater situational awareness about key topics of interest**. *SensePlace2* collects, analyzes, and visualizes millions of tweets, which analysts can then **explore through ad hoc queries and a rich set of user interactions** [5].

## System Architecture

**Central Database and Index (PostgreSQL + PostGIS and Lucene)**

Query Table | Data Table | Person Table | Organization Table | Other Entity Table | Location Table | Text Searh Index

**Associated Entity Tables**

Twitter Channel Search Daemon | Named Entity Extraction Daemons (7 Instances) | Geocoding Daemon | Lucene Text Indexer Daemon

Twitter | GATE / ANNIE Named Entity Extractor | GeoNames

## The UI: Leveraging Multiple Coordinated Views

The user interface of *SensePlace2* employs Shneiderman's Visual Information Seeking Mantra [6]. In this way, users are provided with an overview that can then be explored to reveal additional details as the analyst forages and interacts with data of interest. The multiple coordinated views of *SensePlace2* include ① a temporal filter, ② the tweet list, ③ the map view, ④ a place-tree hierarchy, and ⑤ a word cloud. The screenshot and text below highlights and explains each of these components.

> " *Overview first, zoom and filter, then details on demand.* "
>
> **Ben Shneiderman (1996)**
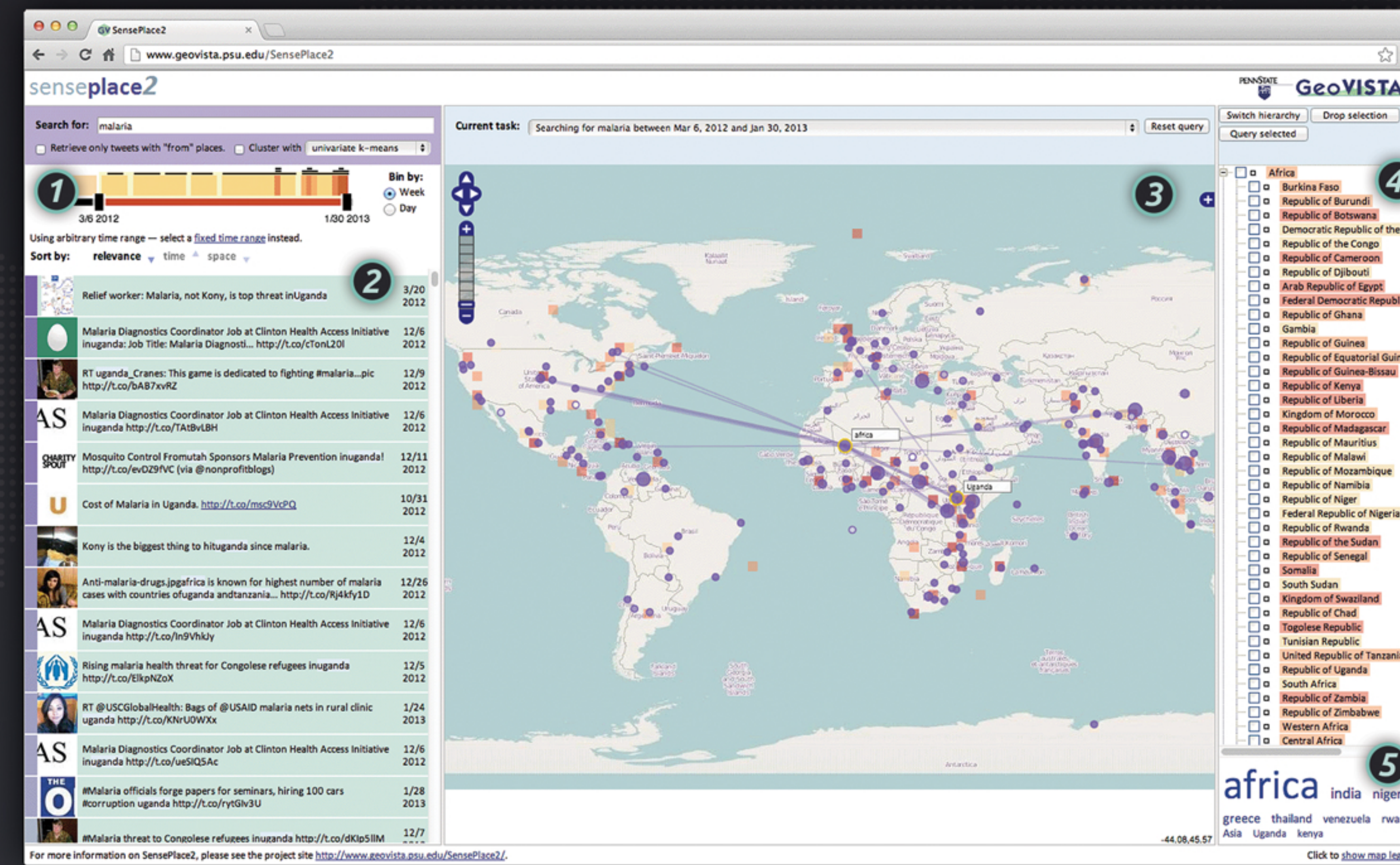
The screenshot above demonstrates a query on the term 'malaria.' The analyst has constrained the date range between 3/6/2012 and 1/30/2013. After selecting tweets on the map that mention Africa and Uganda, **the associated tweets are highlighted and brought to the top of the tweet list**. Notice the connections that are drawn when a single tweet contains multiple place mentions.
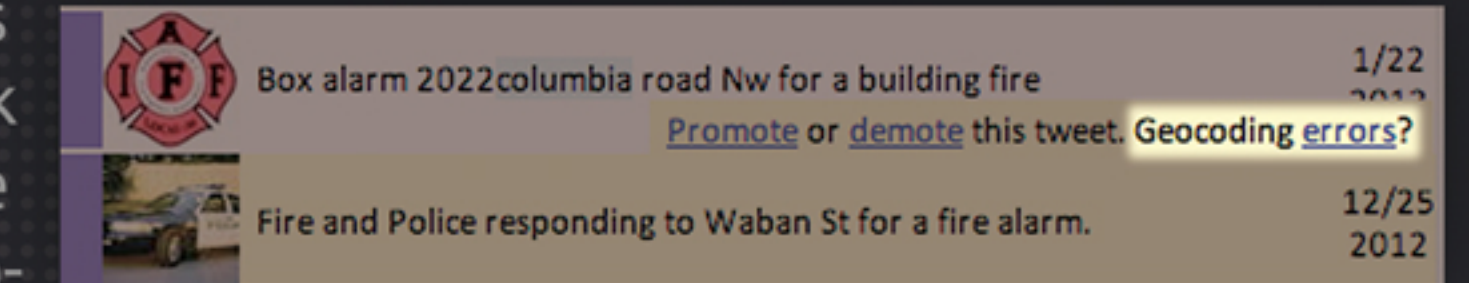
## Spatiotemporal Comparisons

To empower analysts even further, **new methods are being explored to better depict spatial and temporal relationships intuitively and seamlessly**. These include surface and isarithmic maps and new temporal filters that allow comparison between individual queries and the rest of the database. **Novel approaches to map symbology (below) will allow *SensePlace2* to depict the number of connections between locations at a glance**, revealing the spatial patterns that exist across terms without unnecessary foraging.
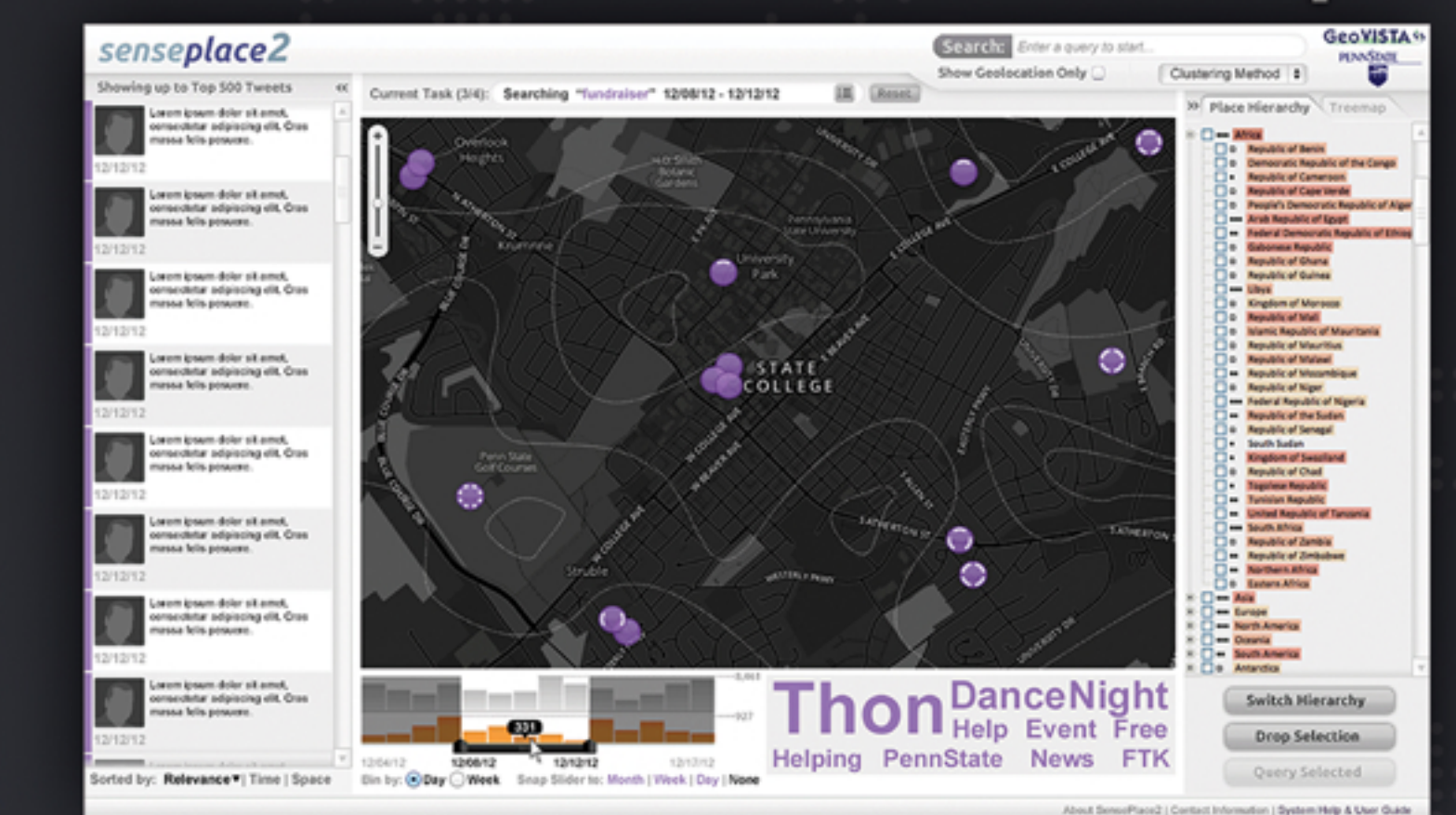
None | One | Two | Three | Four | Five | Many

## Volunteered Georeferencing Improvement

*SensePlace2* takes advantage of its web-based foundation to enable volunteered feedback on tweet georeferencing. The interface allows users to report and categorize georeferencing errors and this input is being fed back into refining the *SensePlace2* georeferencing algorithms.

Box alarm 2022columbia road Nw for a building fire
Promote or demote this tweet. Geocoding errors?
1/22 2013
Fire and Police responding to Waban St for a fire alarm.
12/25 2012

## Future Work and Next Steps

A **prototype UI on the left demonstrates some of our new ideas**. An updated temporal filter has been designed to show results from both the current query as well as the entire database. A new base map style gives salience to the symbology, which includes an isarithmic surface. **Adjustments across the UI provide consistency and organization**.

## References

[1] Lunden, Ingrid. "Analyst: Twitter Passed 500M Users In June 2012, 140M Of Them In US; Jakarta 'Biggest Tweeting' City." TechCrunch. AOL Tech, 3 February 2013. Web.

[2] Terdiman, Daniel, "Report: Twitter Hits Half a Billion Tweets a Day." CNET. CBS Interactive, 3 February 2013. Web.

[3] MacEachren, A.M.; Jaiswal, A.; Robinson, A.C.; Pezanowski, S.; Savelyev, A.; Mitra, P.; Zhang, X.; Blanford, J., "SensePlace2: GeoTwitter analytics support for situational awareness," Conference on Visual Analytics Science and Technology (VAST), IEEE, 181-190, 23-28 Oct. 2011.

[4] S. Vieweg, A. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: what twitter may contribute to situational awareness," in Proc. of the 28th Inter. Conference on Human Factors in Computing Systems, 1079-1088. 2010.

[5] MacEachren, A. M., Robinson, A. C., Jaiswal, A., Pezanowski, S., Savelyev, A., Blanford, J., & Mitra, P. "Geo-Twitter Analytics: Applications in Crisis Management," In Proceedings, 25th International Cartographic Conference, Paris, France. 2011.

[6] Shneiderman, B., "The eyes have it: a task by data type taxonomy for information visualizations," Symposium on Visual Languages, 1996. Proceedings., IEEE , 336-343, 3-6 Sept. 1996.

**VACCINE**
Visual Analytics for Command, Control, and Interoperability Environments
A U.S. Department of Homeland Security Center of Excellence

**GeoVISTA**

**PennState** 1855

**DEPARTMENT OF GEOGRAPHY**
COLLEGE OF EARTH AND MINERAL SCIENCES