

# Taqreer: A System for Spatio-temporal Analysis on Microblogs\*

Amr Magdy<sup>2</sup>      Mashaal Musleh<sup>1</sup>      Kareem Tarek<sup>1</sup>      Louai Alarabi<sup>2</sup>  
Saif Al-Harhi<sup>1</sup>      Hicham G. Elmongui<sup>1,4</sup>      Thanaa M. Ghanem<sup>3</sup>  
Sohaib Ghani<sup>1</sup>      Mohamed F. Mokbel<sup>2</sup>

<sup>1</sup>KACST GIS Technology Innovation Center, Umm Al-Qura University, Makkah, KSA

<sup>2</sup>Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA

<sup>3</sup>Department of Information and Computer Sciences, Metropolitan State University, Saint Paul, MN, USA

<sup>4</sup>Department of Computer and Systems Engineering, Alexandria University, Alexandria, Egypt

{amr,louai,mokbel}@cs.umn.edu, {mmusleh,ktarek,sharhi,elmongui,sghani}@gistic.org,

thanaa.ghanem@metrostate.edu

## Abstract

*Motivated by the wide popularity of microblog services (e.g., Twitter and Facebook) along with the sheer sizes of microblogs issued every second, this article introduces Taqreer as a scalable and efficient system for auto-generation of spatio-temporal analysis reports on microblogs. Taqreer is composed of two main modules: The Taghreed query engine, which is a scalable and efficient query processing engine for spatio-temporal keyword queries on microblogs and a Report Generation Tool, which receives the user analytic report request and divides it into a set of queries sent to the Taghreed engine, and a set of analysis tasks executed on top of the returned query answers. As of now, Taqreer is able to produce three analysis report types, namely comparative reports, categorical reports, and image gallery reports. Other report types will be added in the future.*

## 1 Introduction

Microblogs, e.g., tweets and Facebook comments, have become incredibly popular in the last few years. Everyday, over a billion of users post about four billions microblogs on Twitter and Facebook [5, 20]. As user-generated data, microblogs are associated with various types of rich contents, including user locations, used language, event updates, news items, opinions, reviews, and/or discussions. With the importance of temporal aspect in microblogs [4, 11] (i.e., more recent microblogs are more important than older ones), and the wide availability of location information of microblogs [3, 12], a high fraction of analysis applications on microblogs rely on *spatio-temporal* analysis. Examples of such analysis include user analysis for geo-targeted advertising [14], event detection [1, 7, 13, 16, 21], news extraction [2, 15, 17], and analysis [6, 18, 19]. Unfortunately, existing systems cannot manage microblogs data efficiently as they are designed for managing either fast or large data. Thus, none of existing systems provide indexing-based data management for data that is fast and large simultaneously, like microblogs.

In this article, we report on our work for the *Taqreer* system. *Taqreer* is a system for generic spatio-temporal analysis and report generation on large numbers of microblogs. *Taqreer* users can generate an analysis report that tracks the appearance of a certain keyword over space and time. *Taqreer* is so scalable that it can generate such report for the whole world over a period of more than a year. Such reporting functionality is important for tracking interest in diseases (e.g., Ebola) or natural disasters (e.g., Nepal earthquake). *Taqreer* users can also

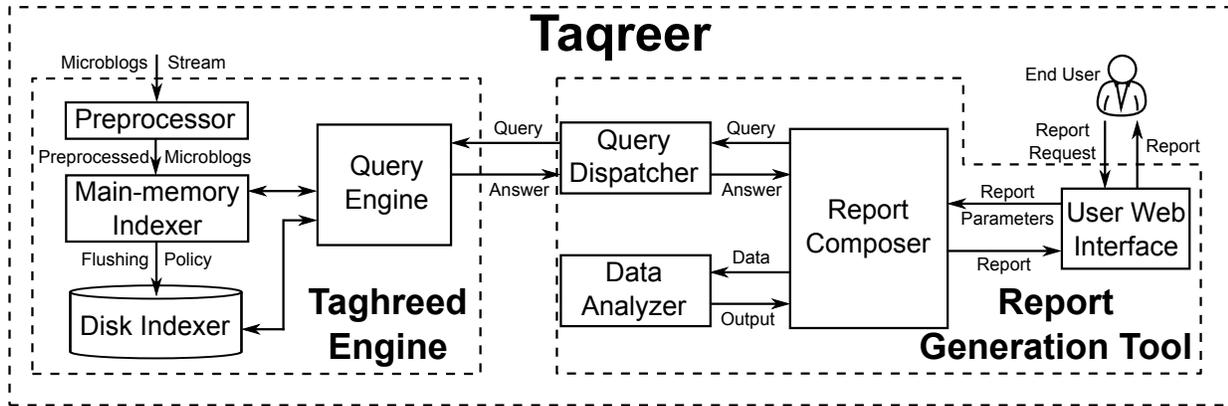


Figure 1: Taqreer Architecture.

generate scalable and interactive comparison reports that compares the trending of various keywords over space and time, which can be used to analyze the status of election candidates, the interest in various soccer teams, or other comparisons based on social media discussions. Other *Taqreer* reports include categorical reports that study the used languages over space and time, which is a powerful report in understanding the diversity of various countries and cities worldwide. In general, *Taqreer* is an extensible system as it provides a rich platform that allows adding various reports easily.

## 2 System Overview of Taqreer

Figure 1 gives *Taqreer* system architecture, which is composed of two main system components, the *Taghreed* query engine [8, 9] (detailed in Section 3) and a *Report Generation Tool* (detailed in Section 4). *Taqreer* users submit their report generation requests to a *User Web Interface* module. Then, the report parameters are forwarded to a *Report Composer* module, which parses the parameters and divides the report into: (a) a set of spatio-temporal keyword queries that retrieve the necessary data to generate the requested report, and (b) a set of analysis tasks to run on the retrieved answer of the spatio-temporal keyword queries. All spatio-temporal keyword queries are sent to the *Taghreed* query engine through a *Query Dispatcher* module.

Meanwhile, the *Taghreed* query engine continuously receives an incoming stream of microblogs with high arrival rates of up to 5,000 microblogs per second. The incoming stream is processed and digested in a highly scalable and efficient main-memory index structure. Once the memory becomes full, a flushing policy is invoked to move a portion of memory contents to the disk storage in another disk-based highly scalable and efficient index structure. The *Taghreed* query engine answers its incoming spatio-temporal queries from both in-memory and disk-based index structures, based on where the data needed for the query answer reside. Once the query answer is collected, it is sent back to the *Report Generation Tool*, which invokes the *Data Analyzer* module to perform the required data analysis tasks. Finally, the report composer gets the analysis output, composes the report in its final form as an interactive web page and sends it back to the requesting user.

---

*Copyright 0000 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.*

**Bulletin of the IEEE Computer Society Technical Committee on Data Engineering**

---

\*This work is supported by KACST GIS Technology Innovation Center at Umm Al-Qura University (GISTIC), under project GISTIC-13-06, and was all done when all the authors were only affiliated with GISTIC.

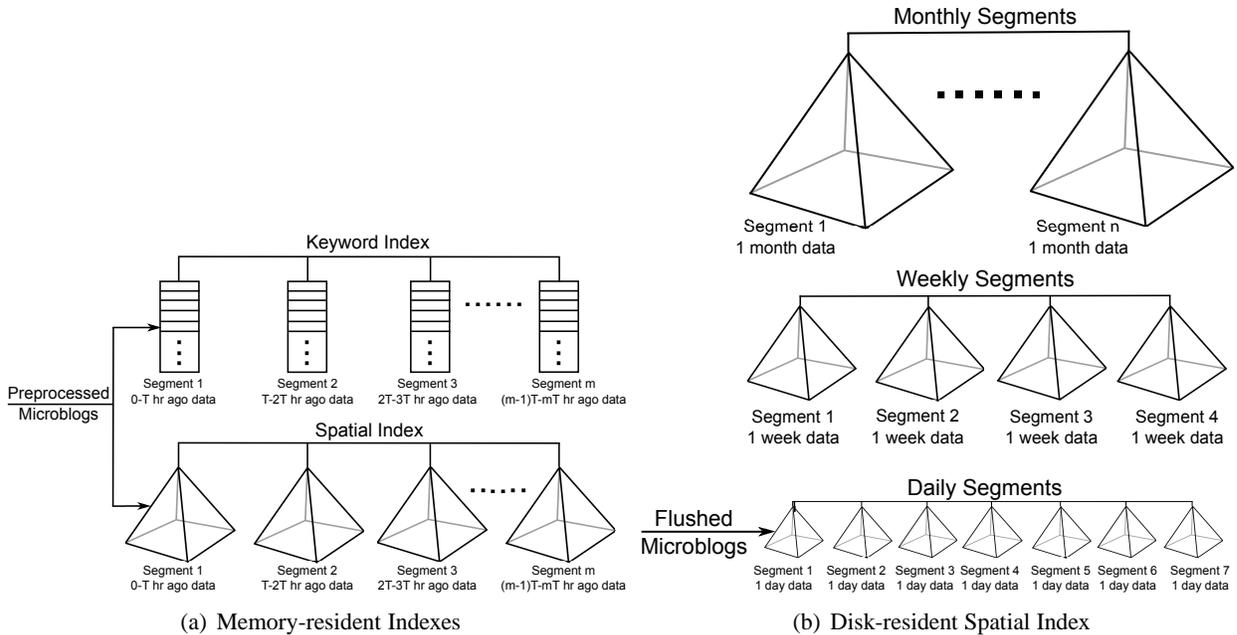


Figure 2: Taghreed Indexes Organization.

### 3 Taghreed Query Engine

*Taghreed* [8, 9] is the query engine behind *Taqreer*, which has two main responsibilities: (1) Digesting incoming stream of microblogs with high arrival rates, and (2) Efficient support for spatio-temporal keyword queries over large set of microblogs. *Taghreed* is composed of five main components, namely, *in-memory index structure*, *disk-based index structure*, *flushing policy*, *query optimizer*, and *recovery manager*, described below.

**In-memory index structure.** *Taghreed* employs two in-memory index structures; a keyword index and a spatial index. Both indexes are segmented into temporal segments that partition data based on arrival timestamp. Figure 2(a) gives the organization of the memory indexes. Each index is segmented into disjoint segments, where each segment includes the data of the last  $T$  hours, where  $T$  is a system parameter. Incoming microblogs are digested in the most recent segment. Once the segment spans  $T$  hours of data, it is concluded and a new empty segment is introduced to digest the new data. Index segmentation has two main advantages: (a) new microblogs are digested in a smaller index, which is the most recent one, and hence we can support higher digestion rates, and (b) it makes it easier to flush data from memory to disk.

**Disk-based index structure.** Similar to main-memory index structures, disk-based indexing supports both spatial and keywords attributes, where each index embeds the temporal aspect in its organization. However, the disk-based index structures are a bit different from the main-memory ones. Figure 2(b) gives the organization of *Taghreed* disk-based spatial index structure. The disk-based keyword index has a similar structure. The index is organized in temporally partitioned segments. The temporal segments are replicated in a hierarchy of three levels, namely, daily segments, weekly segments, and monthly segments. The daily segments level stores data of each calendar day in a separate segment. Each weekly segment level consolidates the data in daily segments of one calendar week. Similarly, each monthly segment level consolidates data of weekly segments of a whole calendar month. The main reason behind the three levels replication is to minimize the number of accessed index segments while processing queries for different temporal periods. For example, for an incoming query asking about data of two months, if only daily segments are stored, then the query processor would access sixty index structures to answer the query. On the contrary, the query processor would access only two index structures of

the two months time horizon of the query. This significantly reduces the query processing time for queries on relatively long periods.

**Flushing policy.** The main goal of *Taghreed* flushing policy is to determine which microblogs should be flushed from main-memory indexes to disk indexes when the memory becomes full. *Taghreed* flushing manager allows the system administrator to employ one of multiple available flushing policies. The default flushing policy is *flush-temporal*, which depends only on timestamps to select its victim microblogs to be flushed. In particular, we expel a portion of the oldest microblogs to empty a room for the newly real-time incoming microblogs. To reduce the flushing overhead, this policy requires the main-memory indexing to partition the data into segments with the same flushing unit. Referring to the main-memory index organization in Figure 2(a), the flushing unit is defined as  $T$  hours, i.e., the oldest  $T$  hours of data are flushed periodically.  $T$  is adjusted by a system administrator based on the available memory resources, the rate of incoming microblogs, and the desired frequency of flushing.

**Query optimizer.** *Taghreed* query optimizer selects which index segment(s) should be accessed to retrieve the query answer. Specifically, *Taghreed* provides two index structures in both main-memory and disk: a keyword index and a spatial index. In addition, disk-resident data is replicated in three temporal levels, daily, weekly, and monthly index segments. Consequently, the query processor may have different ways to process the same query based on: (1) the order of hitting keyword or spatial indexes, and (2) the number of disk index segments to hit. For example, a query that spans from May 1 to May 9 can be answered from disk indexes by either accessing nine daily index segments, or accessing one weekly and two daily index segments. Each of these is called a query plan. The costs of different query plans are different. The main task of the query optimizer is to generate a plan to execute so that the estimated cost is minimal. To this end, the query optimizer employs two cost estimation models, one for estimating the cost of accessing a keyword index segment and the other for estimating the cost of accessing a spatial index segment. Using these cost models, the query optimizer proposes two selection criteria, one for the main-memory index, where each index has a single level of disjoint segments, and one for the disk index, where each index has three levels of overlapped segments.

**Recovery manager.** With millions of microblogs managed in main-memory, *Taghreed* system accounts for memory failures that may lead to data loss. *Taghreed* employs a simple, yet effective, triple-redundancy model where the main-memory data is replicated three times over different machines. The core idea of this model is similar to Hadoop redundancy model that replicates the data three times. In particular, when *Taghreed* is launched, all the main-memory modules are initiated on three different machines. Each machine is fed with exactly the same stream of microblogs, thus they form two backup copies of the main-memory system status. Any flushing from memory to disk leads to throwing the data out from the memory of backup machines. On memory failure, the backup machines continue to digest the real-time microblogs and one of them work as a replacement to the system memory contents. Replicating the data three times significantly reduces the probability of having the three machines down simultaneously and lose all the main-memory data.

## 4 Taqreer Report Generation Tool

*Taqreer* report generation tool is composed of four modules, namely, *User Web Interface*, *Report Composer*, *Query Dispatcher*, and *Data Analyzer*. The user web interface is the interfacing module between *Taqreer* and its end users. The user input parameters are forwarded to the report composer module to sync the work among other modules. In particular, the composer goes through the following four steps: (1) Based on the report type and parameters, the composer determines the set of queries that retrieve the required data and a set of analysis tasks to be performed on that data, (2) The report composer calls the query dispatcher module to submit spatio-temporal keyword queries to *Taghreed* query engine, (3) The retrieved query answers are forwarded to the data analyzer module to perform the required analysis, (4) The report composer adds all the output to an interactive web page and sends it as the final report to the user.

Report Type	Parameters	Queries	Analysis Tasks
Comparative Reports	<ul style="list-style-type: none"> <li>• <math>n</math> spatial regions <math>R_i, 1 \leq i \leq n</math></li> <li>• <math>m</math> keywords (topics/entities) <math>W_j, 1 \leq j \leq m</math></li> <li>• Time range <math>[T_s, T_e]</math></li> </ul>	$n \times m$ queries, each takes: <ul style="list-style-type: none"> <li>• Spatial region <math>R_i</math></li> <li>• Keyword <math>W_j</math></li> <li>• Time range <math>[T_s, T_e]</math></li> </ul>	None
Categorical Reports	<ul style="list-style-type: none"> <li>• Spatial region <math>R</math>, auto divided into <math>n</math> sub-regions of fixed default area</li> <li>• Time range <math>[T_s, T_e]</math></li> <li>• Categorical attribute <math>A</math></li> <li>• Optional <math>m</math> keywords <math>W_j, 1 \leq j \leq m</math></li> </ul>	$n$ queries, each takes: <ul style="list-style-type: none"> <li>• Spatial region <math>R_i \subset R</math></li> <li>• Time range <math>[T_s, T_e]</math></li> <li>• Optional <math>m</math> keywords <math>W_j, 1 \leq j \leq m</math></li> </ul>	<ul style="list-style-type: none"> <li>• Count categories of attribute <math>A</math> for each query microblogs</li> <li>• Aggregate counts over less granular spatial levels</li> </ul>
Image Gallery Reports	<ul style="list-style-type: none"> <li>• <math>m</math> keywords <math>W_j, 1 \leq j \leq m</math></li> <li>• Time range <math>[T_s, T_e]</math></li> <li>• Optional spatial region <math>R</math></li> </ul>	One query that takes: <ul style="list-style-type: none"> <li>• <math>m</math> keywords <math>W_j, 1 \leq j \leq m</math></li> <li>• Time range <math>[T_s, T_e]</math></li> <li>• Optional spatial region <math>R</math></li> </ul>	<ul style="list-style-type: none"> <li>• Extract photos</li> </ul>

Table 1: Parameters, Queries, and Analysis of Different Report Types

*Taqreer* supports three types of spatio-temporal reports, namely, *Comparative Reports*, *Categorical Reports*, and *Image Gallery Reports*. Table 1 gives the parameters, queries, and analysis tasks for each of the three reports. Details of these reports are described in the rest of this section.

#### 4.1 Comparative Reports

Comparative reports aim to compare individual microblogs that are related to different topics (or entities) in different spatial regions within a certain time range. Topics (or entities) are defined by a set of keywords/hashtags. An example of these reports is to compare tweets about the two Spanish soccer teams *Real Madrid C.F.* and *FC Barcelona* in different cities in Spain during the week of their soccer game. This can also include analysis related to presidential candidates, political parties, product trademarks, or events. The first row in Table 1 gives the parameters, queries, and analysis tasks of comparative reports. The user inputs  $n$  spatial regions of interest  $R_i, 1 \leq i \leq n$ ,  $m$  entities or topics (identified by keywords  $W_j, 1 \leq j \leq m$ ), and an arbitrary time range  $[T_s, T_e]$ . A set of  $n \times m$  queries are submitted to *Taghreed* query engine to retrieve the report data, each query takes a spatial region  $R_i$ , a keyword  $W_j$ , and the time range  $[T_s, T_e]$ . The retrieved data is displayed in an interactive web page that allows arbitrarily inclusion/exclusion of microblogs of certain spatial regions and navigation along the whole report timeline, either for a single point of time or on a time range. The analysis of these reports include creating a heatmap for the microblogs, optional pie charts that show percentage analysis for the displayed microblogs, optional tag cloud that shows popular topics in the displayed microblogs, and locating and displaying individual microblogs on a geographical map with full text and user information. In the rest of this section, we present using comparative reports for two purposes: (1) analyzing event-specific tweets, and (2) general-purpose social media analysis.

**Analyzing event-specific tweets.** One of the most important and consistent behavior of Twitter users is posting a plethora of tweets about *events* of different types, e.g., Oscars ceremony, soccer games, and natural disasters. For such kind of event-specific tweets, geotagged tweets grab a high attention as all events, by nature, have a spatial extent. For example, while Boston Marathon explosions were going on in April 2013, users rush to Twitter seeking tweets from the marathon location [2]. An event is generally defined by a temporal horizon, a set of hashtags, and a spatial extent. Events fall in one of two categories: (a) multi-side events, e.g., sports games or elections, and (b) independent events, e.g., Oscars ceremony or New Year Eve. Figure 3 gives two examples of event-specific reports. Figure 3(a) gives a heatmap for New Year Eve tweets over different points of time. The figure shows popular hashtags in a tag cloud and enables going through exact tweets with full text and user information. Figure 3(b) gives tweets of a soccer game in Saudi Arabia, where an extra option is added



(a) New Year Eve 2015 tweets

(b) Soccer game tweets

Figure 3: Spatio-temporal Analysis of Event-specific Tweets.



Figure 4: Analyzing tweets mentioning different car brands in Saudi cities.

to classify visualized tweets based on local cities and show percentage of tweets that support each team.

**General-purpose social media analysis.** With its generic usage, comparative reports can be used as a powerful tool to analyze social media contents. Figure 4 gives an example for a generated report by *Taqreer* that has: (1) January-March 2015 as the temporal horizon, (2) Riyadh, West, East, and North districts as sub-regions within Saudi Arabia, and (3) Toyota, Nissan, and Ford as entities to analyze. The figure shows heatmap as well as individual tweets of each car brand on a geographical map, a time line that allows navigation in different time instances and/or ranges, percentage of tweets that mention each car brand in a pie chart, and spatial filters to include/exclude tweets of each district. Such a generic tool for analyzing social media contents is very helpful in getting insights from the public discussions in different contexts and applications.

## 4.2 Categorical Reports

The plethora of social media active users enables meaningful analysis tasks that can deduce fruitful conclusions for actual population. One of the underutilized attributes are the *categorical attributes*: the attributes that can take one of multiple discrete values. Prime examples of important categorical attributes in Twitter data include the language attribute that indicates the language used in each tweet and the tweet source attribute, which determines from which operating system, device, or application the tweet is posted. Categorical reports in *Taqreer* perform spatial aggregate analysis over a categorical attribute for microblogs that lie within certain spatial and temporal ranges. As the second row in Table 1 shows, the user inputs a spatial region of interest  $R$ , an arbitrary time range  $[T_s, T_e]$ , a categorical attribute  $A$ , and an optional set of keywords. The report composer divides the space into  $n$  small spatial regions of default fixed size. Then, a set of  $n$  queries are submitted to *Taghreed*, each query

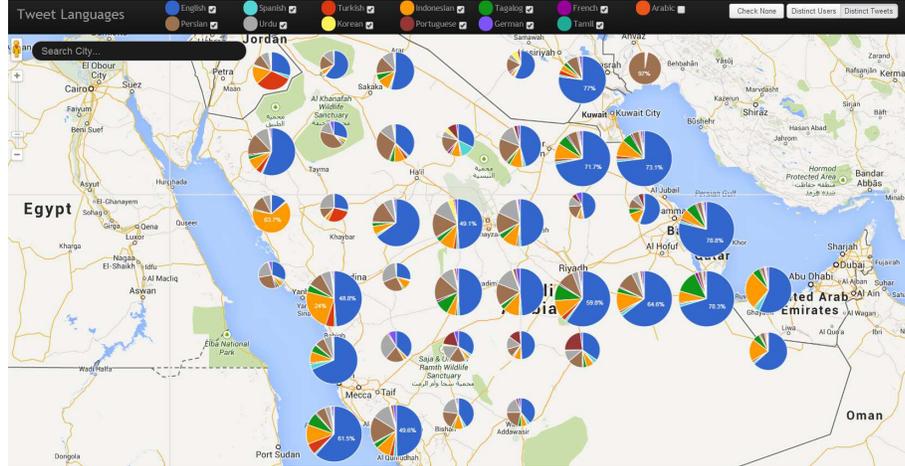


Figure 5: Tweets Languages Spatial Analysis in Arab Gulf Countries.

takes one of the small regions, the time range  $[T_s, T_e]$ , and the set of keywords. Each query retrieves individual microblogs that lie within the query parameters. The retrieved data is forwarded to the data analyzer module to count microblogs in different categories of attribute  $A$ . After counting is performed for all queries, the counts are then aggregated at higher levels of spatial granularity to support zoom in/out analysis in the final report. Finally, the report composer puts all the aggregates on pie charts aligned with latitude/longitude coordinates of a geographical map and embed all of this in a web page. This forms an interactive web page that is sent as the final report to the user.

Figure 5 gives an example of analyzing tweets languages in Arab Gulf countries. The figure gives a pie chart for each sub-region/city. Each pie chart shows the distribution of tweets languages in its region. Zooming in/out gives a finer/coarser granular analysis for language distributions up to the street level. Users can arbitrarily include/exclude languages from the top bar to focus on a subset of all languages. This language analysis, combined with ground truth data enables a full study on language diversity and minorities in local communities [10].

### 4.3 Image Gallery Reports

Image gallery reports exploit the availability of many photos on the social media to summarize certain topics or entities through creating a photo gallery for their microblogs. An example of such reports is to extract and organize photos that are posted in response to a certain event, e.g., human crisis, terrorist attack, elections, or sports game. Topics and entities are defined by a set of keywords/hashtags. Analyzed microblogs should lie within a certain time range and can optionally be filtered based on a spatial region of interest. As described in the third row in Table 1, users input  $m$  keywords  $W_j$ ,  $1 \leq j \leq m$ , a time range  $[T_s, T_e]$ , and an optional spatial region  $R$ . A single query with the input parameters is submitted to *Taghreed* to retrieve the report data. The retrieved microblogs are scanned to extract their images. Extracted images are organized and displayed in an interactive web page that allow users to navigate, enlarge, and share portions of the report on social media websites. Figure 6 gives an image gallery for the event of 2015 Chapel Hill Shooting. The shown images are extracted for the hashtag #ChapelHillShooting for 11 days after the accident happened. Thus, *Taghreed* is queried with time range of February 10 to February 20, 2015 and hashtag #ChapelHillShooting. The returned tweets are analyzed to extract their images and organized them as the figure shows.



Figure 6: Tweets Image Gallery for 2015 Chapel Hill Shooting.

## 5 Conclusion

This article presented *Taqreer*; a scalable and efficient system for auto-generation of spatio-temporal analysis reports on microblogs. *Taqreer* is composed of two main modules, the *Taghreed* query engine, which is responsible for efficiently supporting spatio-temporal keyword queries on microblogs, and a *Report Generation* tool, which is responsible for receiving the user requests, extracting the required queries for the report, sending them to the *Taghreed* query engine, and performing a set of analysis and visualization tasks on top of the returned query results. We have presented three report types as example of what *Taqreer* can generate, namely, comparative reports, categorical reports, and image gallery reports. For each report type, we show the user input parameters, the queries that will be sent to the *Taghreed* query engine, and the set of analysis tasks that will be performed on top of the returned query answers. Other report types can be defined within the *Taqreer* system in a similar way.

## References

- [1] Hamed Abdelhaq, Christian Sengstock, and Michael Gertz. EvenTweet: Online Localized Event Detection from Twitter. In *Proceedings of the International Conference on Very Large Data Bases, VLDB*, 2013.
- [2] After Boston Explosions, People Rush to Twitter for Breaking News. <http://www.latimes.com/business/technology/la-fi-tn-after-boston-explosions-people-rush-to-twitter-for-breaking-news-20130415,0,3729783.story>, 2013.
- [3] Ceren Budak, Theodore Georgiou, Divyakant Agrawal, and Amr El Abbadi. GeoScope: Online Detection of Geo-Correlated Information Trends in Social Networks. In *Proceedings of the International Conference on Very Large Data Bases, VLDB*, 2014.
- [4] Michael Busch, Krishna Gade, Brian Larson, Patrick Lok, Samuel Luckenbill, and Jimmy Lin. Earlybird: Real-Time Search at Twitter. In *Proceedings of the IEEE International Conference on Data Engineering, ICDE*, 2012.
- [5] Facebook Statistics. <http://newsroom.fb.com/company-info/>, 2015.
- [6] Harvard Tweet Map. [worldmap.harvard.edu/tweetmap/](http://worldmap.harvard.edu/tweetmap/), 2013.

- [7] Rui Li, Kin Hou Lei, Ravi Khadiwala, and Kevin Chen-Chuan Chang. TEDAS: A Twitter-based Event Detection and Analysis System. In *Proceedings of the IEEE International Conference on Data Engineering, ICDE*, 2012.
- [8] Amr Magdy, Louai Alarabi, Saif Al-Harathi, Mashaal Musleh, Thanaa Ghanem, Sohaib Ghani, Saleh Basalamah, and Mohamed Mokbel. Demonstration of Taghreed: A System for Querying, Analyzing, and Visualizing Geotagged Microblogs. In *Proceedings of the IEEE International Conference on Data Engineering, ICDE*, 2015.
- [9] Amr Magdy, Louai Alarabi, Saif Al-Harathi, Mashaal Musleh, Thanaa Ghanem, Sohaib Ghani, and Mohamed Mokbel. Taghreed: A System for Querying, Analyzing, and Visualizing Geotagged Microblogs. In *Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM GIS*, 2014.
- [10] Amr Magdy, Thanaa M. Ghanem, Mashaal Musleh, and Mohamed F. Mokbel. Exploiting Geo-tagged Tweets to Understand Localized Language Diversity. In *the International ACM Workshop on Managing and Mining Enriched Geo-spatial Data, GeoRich. In conjunction with SIGMOD*, 2014.
- [11] Amr Magdy and Mohamed Mokbel. Towards a Microblogs Data Management System. In *Proceedings of the International Conference on Mobile Data Management, MDM*, 2015.
- [12] Amr Magdy, Mohamed F. Mokbel, Sameh Elnikety, Suman Nath, and Yuxiong He. Mercury: A Memory-Constrained Spatio-temporal Real-time Search on Microblogs. In *Proceedings of the IEEE International Conference on Data Engineering, ICDE*, 2014.
- [13] Adam Marcus, Michael S. Bernstein, Osama Badar, David R. Karger, Samuel Madden, and Robert C. Miller. Twitinfo: Aggregating and Visualizing Microblogs for Event Exploration. In *Proceedings of the International Conference on Human Factors in Computing Systems, CHI*, 2011.
- [14] New Enhanced Geo-targeting for Marketers. <https://blog.twitter.com/2012/new-enhanced-geo-targeting-for-marketers>.
- [15] Owen Phelan, Kevin McCarthy, and Barry Smyth. Using Twitter to Recommend Real-Time Topical News. In *Proceedings of the ACM Conference on Recommender Systems, RecSys*, 2009.
- [16] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. In *Proceedings of the International Conference on World Wide Web, WWW*, 2010.
- [17] Jagan Sankaranarayanan, Hanan Samet, Benjamin E. Teitler, Michael D. Lieberman, and Jon Sperling. TwitterStand: News in Tweets. In *Proceedings of the ACM Symposium on Advances in Geographic Information Systems, ACM GIS*, 2009.
- [18] Topsy Analytics: Find the insights that matter. [www.topsy.com](http://www.topsy.com), 2014.
- [19] TweetTracker: track, analyze, and understand activity on Twitter. [tweettracker.fulton.asu.edu/](http://tweettracker.fulton.asu.edu/), 2014.
- [20] Twitter Statistics. <https://about.twitter.com/company>, 2015.
- [21] Kazufumi Watanabe, Masanao Ochi, Makoto Okabe, and Rikio Onai. Jasmine: A Real-time Local-event Detection System based on Geolocation Information Propagated to Microblogs. In *Proceedings of the ACM International Conference on Information and Knowledge Management, CIKM*, 2011.