Geo-Spatial Trend Detection Through Twitter Data Feed Mining

Maarten Wijnants(✉), Adam Blazejczak, Peter Quax, and Wim Lamotte

Expertise Centre for Digital Media, Hasselt University – tUL – iMinds,
Wetenschapspark 2, 3590 Diepenbeek, Belgium
maarten.wijnants@uhasselt.be

Abstract. Present-day Social Networking Sites are steadily progressing towards becoming representative data providers. This paper proposes *TweetPos*, a versatile web-based tool that facilitates the analytical study of geographic tendencies in crowd-sourced Twitter data feeds. To accommodate the cognitive strengths of the human mind, TweetPos predominantly resorts to graphical data structures such as intensity maps and diagrams to visualize (geo-spatial) tweet metadata. The web service's asset set encompasses a hybrid tweet compilation engine that allows for the investigation of both historic and real-time tweet posting attitudes, temporal trend highlighting via an integrated animation system, and a layered visualization scheme to support tweet topic differentiation. TweetPos' data mining features and the (geo-spatial) intelligence they can amount to are comprehensively demonstrated via the discussion of two representative use cases. Courtesy of its generic design, the TweetPos service might prove valuable to an interdisciplinary customer audience including social scientists and market analysts.

Keywords: Twitter · Social networking sites · Social media · Geographic trends · Investigative tool · Data mining · TweetPos

1 Introduction

Social Networking Sites (SNSs) were conceived as a means to virtually connect users and to offer them an intuitive forum to ubiquitously contribute and disseminate information in real time. As their number of subscribers rose over time, so did the amount of content that is managed by SNSs. As a result, they nowadays host a wealth of user-generated data that is highly heterogeneous in nature.

Over the years, SNSs have also evolved functionality-wise. While many such services were purely text-based upon their inception, they nowadays typically grant users the option to attach multimedia items like pictures and video clips to their contributions. Another feature that has become nearly commonplace in the SNS landscape, is geotagging (i.e., attaching geographic coordinates as metadata to messages). It is apparent that such novel facilities embellish the core SNS content and further extend its value.

V. Monfort and K.-H. Krempels (Eds.): WEBIST 2014, LNBIP 226, pp. 212–227, 2015.
DOI: 10.1007/978-3-319-27030-2_14

Given their popularity and broad adoption, it is becoming evermore valid to regard SNSs as real-life, real-time and crowd-sourced sensor systems that "monitor" a varied spectrum of (physical) properties and topics (see, for example, [1]). By intelligently exploiting the data feeds that can be accumulated from SNSs, innovative and value-added services can be conceived. In addition, mining and analyzing the information that is shared by end-users through social media can lead to valuable insights and knowledge. Possible application domains include consumer behavior modeling, consumer profiling, intelligent recommendation systems, and population sentiment assessment. Extracting such kinds of intelligence from SNSs however typically requires external tools, as profound mining and analysis mechanisms by default are lacking from their feature set.

In this paper, we tend to Twitter, the authoritative microblogging platform in the western world, and we focus on investigating the data that is hosted by this SNS from a geo-spatial perspective. In particular, we introduce the web-based *TweetPos* tool, a convenient means to display and study the geographic origin of tweets, and to uncover the geographical evolution of the popularity of tweet topics. A hybrid visualization method encompassing both heatmap- and chart-based data representation allows for thorough analysis and mining with regard to the geo-spatial distribution of tweeted material over time. The TweetPos web service affords keyword-based topic selection and includes a layering system that allows for easy comparison of the geographical trends of multiple subjects. Furthermore, our tool is able to compile data sets that integrate a representative sample of tweets from the recent past with present-day tweet messages that are captured in real time, in order to grant insight in both historical and current tweet posting behavior. Finally, the accumulated data collections can be aggregated and studied on either a per-day or per-hour basis to provide some degree of analytical granularity. We argue that, combined, these features offer all necessary measures to perform significant research about the geographical sources of Twitter data. We will back this claim by presenting the results of two prototypical analyses that illustrate the versatility, effectiveness and comprehensiveness of the proposed instrument. At the same time, the provided demonstrations serve as prove of the extensive applicability of TweetPos: courtesy of its generic methodology, it may one way or another cater to the demands of a variety of human consumer profiles, including social researchers, marketeers, advertisers, analysts and journalists.

A primordial aspect of the TweetPos solution is its emphasis on providing graphical representations of the crawled Twitter data. Contrary to computers, the typical human mind does not excel at handling large quantities of raw data. On the other hand, our cognitive features make us more adept than computers at interpreting visual data structures [2] like heatmaps and charts, which are exactly the output modalities that are supported by our platform. The TweetPos tool is hence intended to offer human operators an adequate graphical workspace that allows them to readily and conveniently assess geo-spatial trends in social media contributions.

The remainder of this article is organized as follows. Section 2 presents an overview of the functional features of the TweetPos web service. Next, Sect. 3

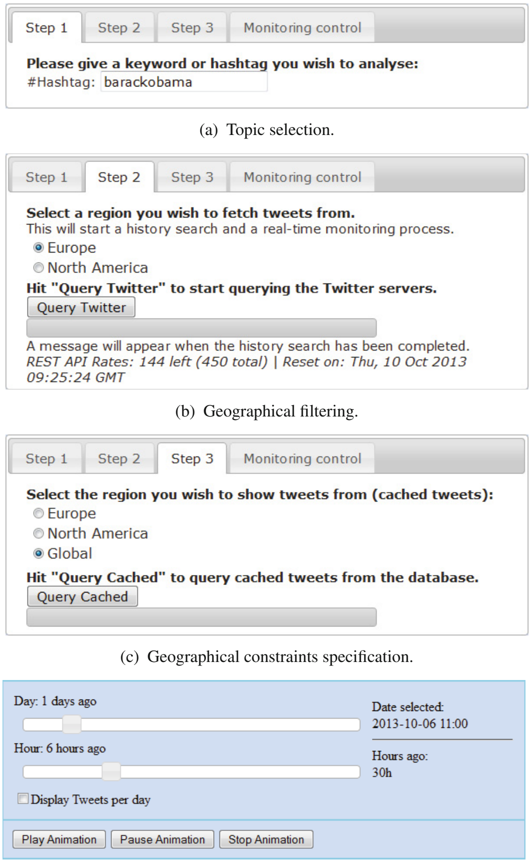
handles the architectural design and implementation of the tool. We then evaluate our work in Sect. 4 by discussing some representative examples of investigations into the geographical evolution of recently trending Twitter themes that have been produced with the proposed tool. Section 5 briefly reviews related work on the analysis and mining of information that has been shared via social networks, and at the same time highlights our scientific contributions. Finally, we draw our conclusions and suggest potential future research directions in Sect. 6.

2 TweetPos

The TweetPos instrument is implemented as a web service that is accessible via a standard web browser. Screenshots of the tool's input widgets are bundled in Fig. 1. As these images illustrate, keywords or so-called Twitter *hashtags* are the service's essential ingress parameters. Based on the specified topic of interest, the tool will compile a corpus of tweets that deal with this subject. This corpus will encompass a representative sample of historical messages as well as a completely accurate set of current and future tweets on the topic at hand. The user is hereby granted the option to apply geographical filtering by limiting the tweet compilation to either Europe or North America, if so desired (see Fig. 1(b)). An identical filtering option is included in the input pane that controls the visualization of the accumulated data (see Fig. 1(c)). Finally, a number of standard HTML input elements allow for controlling the temporal constraints and the animation of the result set. In particular, via two HTML sliders and a checkbox, users can enforce the discrete time interval with which (the timestamps of) gathered tweets need to comply for them to be included in the output. Two fixed levels of granularity are supported for the specification of the temporal constraints, which cause TweetPos to aggregate filtered tweets per hour and per day, respectively. An animation engine that utilizes either hourly or daily increments allows for the animated, video-like presentation of the tweet data set and as such might yield valuable insights into the geo-spatial trends that are exhibited by tweet topics over time.

On the output front, the principal GUI element consists of a topographic map that scaffolds heatmap-based visualization of the geo-spatial provenances of filtered Twitter messages. Stated differently, this output component displays the intensity, from a geographic point of view, of tweets that encompass the specified input keyword. Besides a map, two additional output widgets are included in the tool. The first is a line chart that visualizes the quantitative volume of the compiled tweet archive, aggregated either on a per-hour or a per-day basis, while the second enumerates the textual contents of the collected tweets. Figure 2 illustrates the TweetPos output interface.

An important feature of TweetPos is its keyword layering functionality. The tool allows multiple keyword filters to be active simultaneously, by conceptually associating (the results of) each concurrent hashtag search with an individual layer. Figure 2(a) and (b) for instance illustrate a setup in which two queries are



Please give a keyword or hashtag you wish to analyse:
#Hashtag: barackobama

(a) Topic selection.

Select a region you wish to fetch tweets from.
This will start a history search and a real-time monitoring process.
● Europe
○ North America
Hit "Query Twitter" to start querying the Twitter servers.

A message will appear when the history search has been completed.
REST API Rates: 144 left (450 total) | Reset on: Thu, 10 Oct 2013 09:25:24 GMT

(b) Geographical filtering.

Select the region you wish to show tweets from (cached tweets):
○ Europe
○ North America
● Global
Hit "Query Cached" to query cached tweets from the database.

(c) Geographical constraints specification.

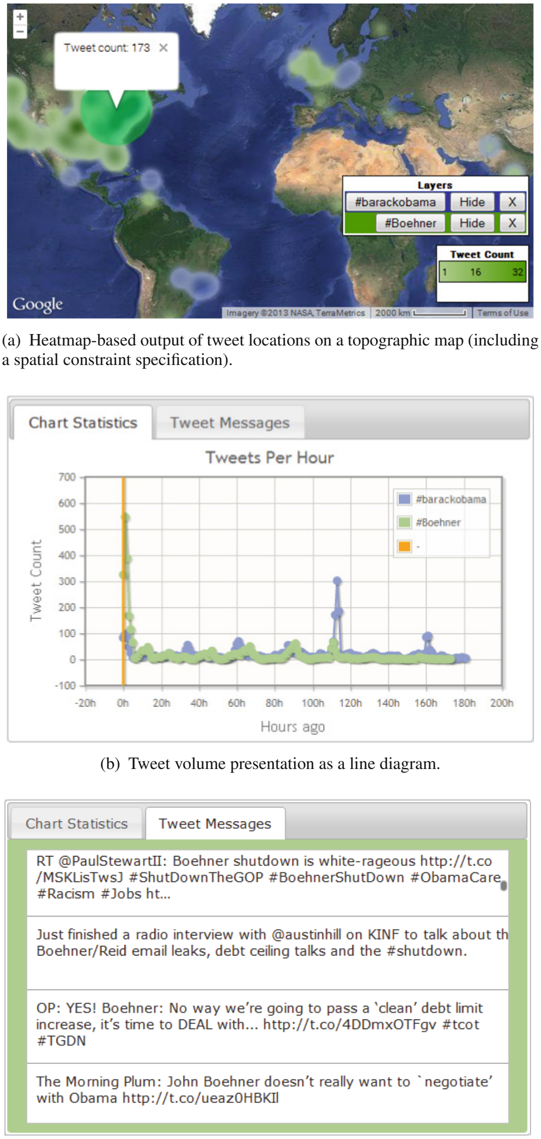
Day: 1 days ago
Hour: 6 hours ago
□ Display Tweets per day
Date selected: 2013-10-06 11:00
Hours ago: 30h

(d) Temporal constraints specification and animation control.

Fig. 1. TweetPos input GUI.

involved. Layers are rendered on top of the topographic map as uniquely colored overlays, whose visualization can be independently toggled on and off. Analogously, distinct tweet volumes are plotted in the line graph for each currently deployed keyword filter. A layer can be eliminated from the visualization process via the legend that is incorporated in the geographic map. The layering system provides a powerful means to investigate (the geo-spatial evolution of) multiple subjects concurrently, to offset them against each other, to reveal potential correlations between them, and so on.

Apart from temporal filtering parameters, the TweetPos service also supports the specification of spatial constraints. This type of constraint is deployed by clicking on the topographic map, which causes a circular area to be drawn around the selected location (see Fig. 2(a)). The map's zoom level and the stretch of the marked geographical region have been designed to be inversely proportional properties, which implies that the spatial extent of the highlighted area is controllable by zooming the map in and out. In effect, installing a spatial



(a) Heatmap-based output of tweet locations on a topographic map (including a spatial constraint specification).

(b) Tweet volume presentation as a line diagram.

RT @PaulStewartII: Boehner shutdown is white-rageous http://t.co/MSKLisTwsJ #ShutDownTheGOP #BoehnerShutDown #ObamaCare #Racism #Jobs ht...

Just finished a radio interview with @austinhill on KINF to talk about th Boehner/Reid email leaks, debt ceiling talks and the #shutdown.

OP: YES! Boehner: No way we're going to pass a 'clean' debt limit increase, it's time to DEAL with... http://t.co/4DDmxOTFgv #tcot #TGDN

The Morning Plum: John Boehner doesn't really want to `negotiate' with Obama http://t.co/ueaz0HBKIl

(c) Textual contents of the filtered tweets.

Fig. 2. TweetPos output GUI.

constraint under a relatively high zoom level will result in the selection of a relatively tight geographical region, while the opposite holds true when the map is heavily zoomed out.

All output components are dynamic, in the sense that their content is updated on-the-fly when the user modifies one or more input parameters. Obviously this applies to the keywords or hashtags that are searched for. In particular,

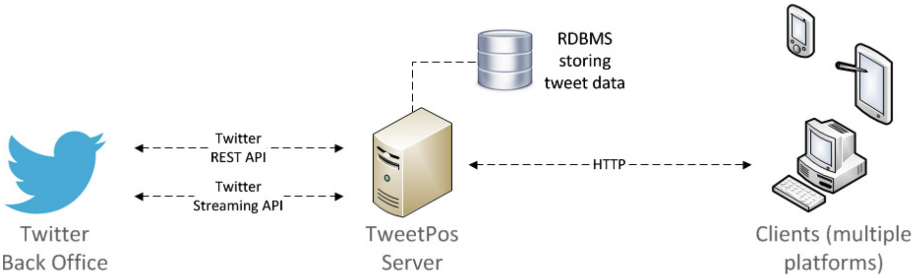


Fig. 3. High-level system architecture.

initiating a new search operation causes an additional layer to be introduced in both the 2D map and the line chart. Responding to less profound input settings however also occurs in real time. For example, exploiting the HTML sliders to modify the time constraints causes the map, the line chart as well as the list of tweet message to be updated instantaneously. The map will be adjusted to draw the geographic intensity that applied at the specified timestamp, the volume plot will be updated so that it correctly marks the currently selected time, and the textual list will only display tweet messages that satisfy the installed temporal restrictions. Analogous actions are dynamically undertaken in reaction to the definition of a spatial constraint. More precisely, the volume plot and textual message list only reckon with tweets that originated from the designated spatial area, if any. This feature allows human operators to zoom in on certain geographic regions and to perform fine-grained, localized analyses. As a final example of the dynamism of the output GUI, switching between layers via the legend in the topographic map causes the contents of the textual tweet enumeration widget to be updated so that it only displays those messages that apply to the keyword that corresponds with the currently selected layer.

3 Implementation

The TweetPos implementation is completely web-compliant. HTML and CSS are used for rendering the GUI and for handling page layout and style, while all programmatic logic is scripted in PHP and JavaScript (at server and client side, respectively).

Our motivations for realizing the TweetPos application as a web service are manifold. First of all, selecting the web as deployment platform acknowledges the pervasiveness of the Internet in modern society. At the same time, it renders the TweetPos functionality available on all environments and devices that support widespread and standardized web technologies, which maximizes the portability of our implementation. Finally, numerous utility libraries and supportive tools exist for the web, which we have gladly leveraged to expedite the development process.

3.1 Architectural Design

A schematic overview of TweetPos' architectural setup is given in Fig. 3. TweetPos adopts a client/server network topology. The back-end HTTP server forms the heart of the system; it interfaces with Twitter, implements the data filtering and compilation, hosts a relational database (RDBMS) for data persistence purposes, and responds to incoming HTTP requests. The client on the other hand is very lightweight, as its responsibilities are limited to user interfacing and data visualization. As such, the server (and the RDBMS which it encapsulates) forms a level of abstraction in the TweetPos system architecture between respectively the external information source (i.e., Twitter) and the client-side presentation of the disclosed data.

3.2 Twitter Data Collection

Twitter provides multiple HTTP-based APIs to enable third-party software developers to interface with the platform and to build socially-inspired applications. The TweetPos tool exploits two of these APIs in order to harvest both historical and up-to-date (public) Twitter data. First of all, the Twitter Search API (which is embedded in the Twitter REST API as of version 1.1) is leveraged to compose a non-exhaustive yet representative sample of tweets from the past 7 days that dealt with a particular subject. The quantitative incompleteness is intrinsic to Twitter and represents a deliberate strategy in the platform's design [3]. In effect, the Search API has been designed for relevance and not completeness, which implies that it is not intended to deliver a rigorous index of past tweets. The second Twitter interface that fuels TweetPos' data collection procedure is a low-latency gateway to the global stream of tweets, called the Streaming API. This particular API allows developers to set up a long-lived HTTP connection to the Twitter back office, over which tweets from that moment on will then be streamed incrementally. In combination with extensive filtering and querying mechanisms, applications in this way obtain near-real-time and exhaustive access to exactly the type of tweets they are interested in. To facilitate the interaction with the Twitter Streaming API, the TweetPos tool integrates the 140dev Streaming API framework [4].

For the sake of comprehensiveness, we will now describe the complete set of actions and operations that constitute TweetPos' data ingestion pipeline. When a user initiates a new data collection process by transmitting a keyword-based query to the TweetPos server, the latter will spawn a total of seven PHP daemons. Each of these background processes utilize the Twitter Search API to jointly compile a pool of relevant historical tweets that were contributed during the past week (i.e., one process per day). At the same time, the back-end server manages a (PHP-based) daemon that permanently monitors the Twitter Streaming API. As an end-point is only allowed to set up a single connection to the Streaming API, this background process runs a cumulative filter to guarantee that all present and future tweets that satisfy one of the currently active queries are captured. In contrast to the Search API daemons, which have a finite execution time and are query-specific, the Streaming API process runs indefinitely and is shared by queries. A dedicated widget in the client-side GUI empowers users to stop the real-time monitoring of a particular topic (which is enforced by updating the cumulative filter of the Streaming API daemon).

3.3 Data Storage and Processing

Fetched tweets are persisted at server side in a MySQL database. To streamline the integration of the 140dev framework in the TweetPos tool, we have opted to integrally adopt its cache architecture and accompanying database schema. The caching mechanism of the 140dev framework applies a two-step approach. An aggregation step continuously filters JSON-encoded tweet data (including the actual message and all sorts of metadata) from the Twitter Streaming API and inserts the resulting data directly into a designated caching table in the back-end database. In effect, this task is fulfilled by the Streaming API daemon that was mentioned in Sect. 3.2. Simultaneously, an independent background process successively pulls single raw JSON items from this table, parses and conveniently formats the composing entities of the corresponding tweets (i.e., the textual message itself, the encapsulated hashtags and mentions, etcetera), and distributes the outcome across dedicated database tables. By isolating the aggregation from the parsing of relevant tweets, real-time and lossless data ingestion is guaranteed (the Twitter Streaming API might yield tremendous quantities of data, whose sheer volume might prohibit on-the-fly parsing and processing).

Besides leveraging the 140dev caching methodology and database schema for the Streaming API context of the TweetPos tool, we have decided to extend their application to the Twitter Search API component of our implementation. This entails that historical tweets that are harvested by the Search API daemons are just as well cached in raw JSON format and then parsed by the same process that also handles Streaming API contributions. The beneficial implications of this design are that it yields a clean software architecture, ensures uniform treatment of tweets originating from heterogeneous sources, and enables the elimination of data duplication in an integrated manner (i.e., without requiring an exogenous control loop).

Once the data collection procedure for a particular keyword-based query has been initiated, all client requests that are related to this query are handled at server side by means of pure RDBMS interactions. As an example, the execution of adequate SQL statements suffices for the server to be able to forward an up-to-date overview of Twitter data pertaining to the queried topic to the client.

3.4 Geocoding

As the TweetPos tool is chiefly concerned with the geo-spatial provenance of tweets, it is clear that geographic metadata plays a primordial role in its operation. To be more precise, geographic coordinates are needed in order to pinpoint a tweet on a topographic map. Some Twitter users include these coordinates directly in their posts (e.g., users with smartphones with built-in GPS receivers), yet the majority only inserts a descriptive representation of the involved location (e.g., in the form of a textual address), or even leave out all geographic references altogether.

TweetPos' data accumulation procedure is agnostic of the presence of geo-spatial metadata in tweets. Stated differently, tweets that lack any trace of geographical metadata are not filtered out by either the Streaming API or Search API data compiler. Tweets holding exact geographic footprints are directly cached, as they can be readily localized on a map. In case the tweet only incorporates a descriptive geo-spatial reference, the data processing daemon described in Sect. 3.3 will invoke the Google Geocoding API [5] to translate the description into geographic coordinates prior to database insertion. Finally, although non-localized contributions are not exploitable in the current implementation, they are still recorded in the database "as is" for the sake of completeness (i.e., they may hold some value in future extensions of the tool).

3.5 Visualization

All visualization and GUI interaction operations are performed at client side by means of HTML and JavaScript.

Heatmap-Based Geolocation Clustering. The topographic output map has been implemented by means of the JavaScript variant of the Google Maps API [6]. Tweets are positioned on this map on the basis of the geographic location from which they were posted. Instead of marking (the location of) individual tweets on the map, a heatmap-based design has been adopted. Heatmaps are a general-purpose data visualization technique in which the intensity of data points is plotted in relative comparison to the absolute maximum value of the data set. Typically, data point intensity is indicated by means of a color coding scheme. Compared to mashups of discrete markers (which might easily clutter the map in the case of voluminous data sets), heatmaps hold the perceptual advantage that, without sacrificing much detail, they are naturally surveyable and interpretable. The Google Maps JavaScript API has built-in support for heatmap rendering.

Line Graph. While the heatmap at a glance provides users with an impression of the spatial characteristics of a particular Twitter topic, it fails to communicate exact quantitative figures concerning the tweet volume. To counter this deficiency, the TweetPos tool includes a line graph visualization that discretely plots, either per hour or per day, the number of tweets that address the queried subject(s). As such, it visualizes a precise overview of the temporal evolution of the popularity of themes (expressed in tweet quantity). The line diagram is implemented via jqPlot, a plotting and charting plug-in for the jQuery JavaScript framework (http://www.jqplot.com/). The data values that compose the graph are interactive in the sense that they can be clicked to leap the date selection sliders (see Fig. 1(d)) to the corresponding timestamp.

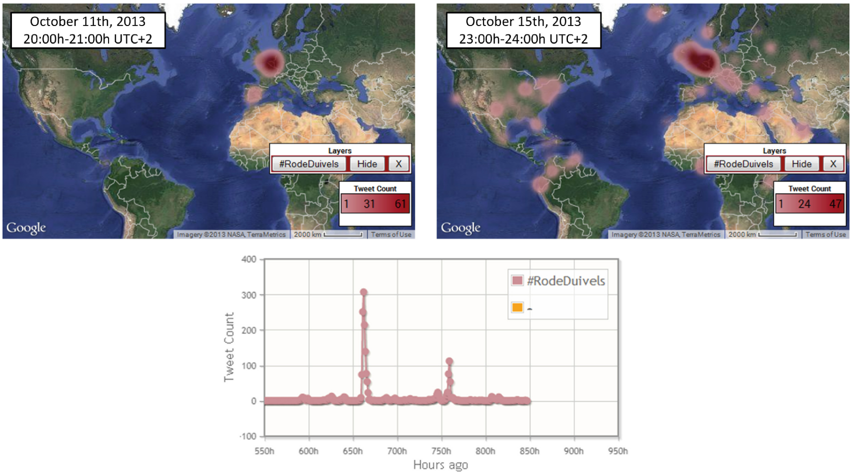


Fig. 4. Results of the 2014 FIFA World Cup qualifiers experiment.

Tweet Message Enumeration. The TweetPos tool is also able to output the textual contents of filtered tweets. This output method has been realized by means of the MegaList jQuery plug-in (http://triceam.github.io/MegaList/). Like the other output widgets, it is adaptive in the sense that it dynamically adjusts its contents to imposed spatiotemporal constraints. This widget is intended to provide users insight into the context in which the queried topic is referenced. As such, it allows for accurate, context-aware classification of tweets based on the messages they carry. For instance, a tweet about a certain incident might plead for or, conversely, against it; by inspecting the textual context, the stance of the tweet publisher becomes apparent.

4 Evaluation

This section serves to showcase the capabilities of the TweetPos instrument by presenting two representative examples of (geo-spatial) analyses of Twitter content that have been produced with it. The first test case is intended to rigorously demonstrate TweetPos' overall practicalities and to generally exemplify the data mining options which the tool scaffolds, while the second example focuses on TweetPos' layering functionality and the analytical features it entails. Space limitations force us to be brief in our discussion, and prevent us from including additional demonstrations.

4.1 2014 FIFA World Cup Qualifiers

The final two qualifier matches for the 2014 soccer World Cup were played on October 11th and 15th, 2013, respectively. We have exploited the TweetPos service to investigate the (geographic) resonance of these matches on Twitter, specifically for Belgium's national soccer team (which are nicknamed the "Red Devils" or "Rode Duivels" in Dutch). We issued a TweetPos data collection request for the `RodeDuivels` hashtag on October 13th and kept this query active until October 19th. Figure 4 shows the geographic distribution of the tweets that were gathered worldwide in the one hour interval immediately succeeding the end of the two matches, as well as a chart-based representation of the tweet quantity that was harvested during the entire course of the experiment (aggregated per hour). As the query was initiated on October 13th, all tweet data in the result set that precedes this date was acquired via the Search API, while tweets with an older timestamp were filtered from the Streaming API.

Analysis of the experimental results yields four notable observations. First and foremost, the output graph reveals two obvious peaks in tweet volume. These local maxima coincide nicely with the Red Devils' schedule of play. As such, this test case corroborates Twitter's capacity to act as a user-driven distributed sensor system that is able to identify real-world events (see also Sect. 5). As the data collection procedure was started in between the two matches, this capacity applies to both the Search API (for events from the recent past) and Streaming API (for current and future events). Secondly, tweets dealing with the match on October 11th appear to have originated practically exclusively from Belgium and its surrounding countries. In contrast, tweets about the second game exhibit a quasi worldwide distribution, yet again with a strong concentration in Western Europe. As the first set of tweets was ingested via the Twitter Search API, this outcome can likely be attributed to the operational principles of this interface (recall from Sect. 3.2 that the Search API aims for relevance, not comprehensiveness). Thirdly, although their volume is rather marginal, tweets embodying the `RodeDuivels` keyword were found to also emerge from non-Dutch speaking countries like the USA, Spain and Turkey (see the rightmost topographic map in Fig. 4). After inspecting the textual contents of these contributions (by means of the tweet mes-sage enumeration widget described in Sect. 3.5), it became clear that these types of tweets can roughly be classified into two categories:

- tweets written in Dutch by Belgian citizens (temporarily) living abroad; e.g., "Come on #RodeDuivels, I am rooting for you from my hotel room in Barcelona!" (English translation)
- retweets by the local population of English messages that include the (Dutch) `RodeDuivels` hashtag; often, the original messages were posted by Dutch natives who wanted to reach an internal audience; e.g., "Belgium versus Wales qualifier starting in 15 minutes #RodeDuivels #RedDevils #belwal #wc2014"

The fourth and final observation pertains to location-driven personalization of the tweeted contents. For example, a tweet by Toby Alderweireld (a Belgian soccer player who plays for Atletico Madrid in Spain), written in English and communicating Belgium's qualification for the 2014 FIFA World Cup, was actively retweeted by his followers in Spain and amounted to the majority of `RodeDuivels` tweets that originated from that country. A single Spanish Atletico

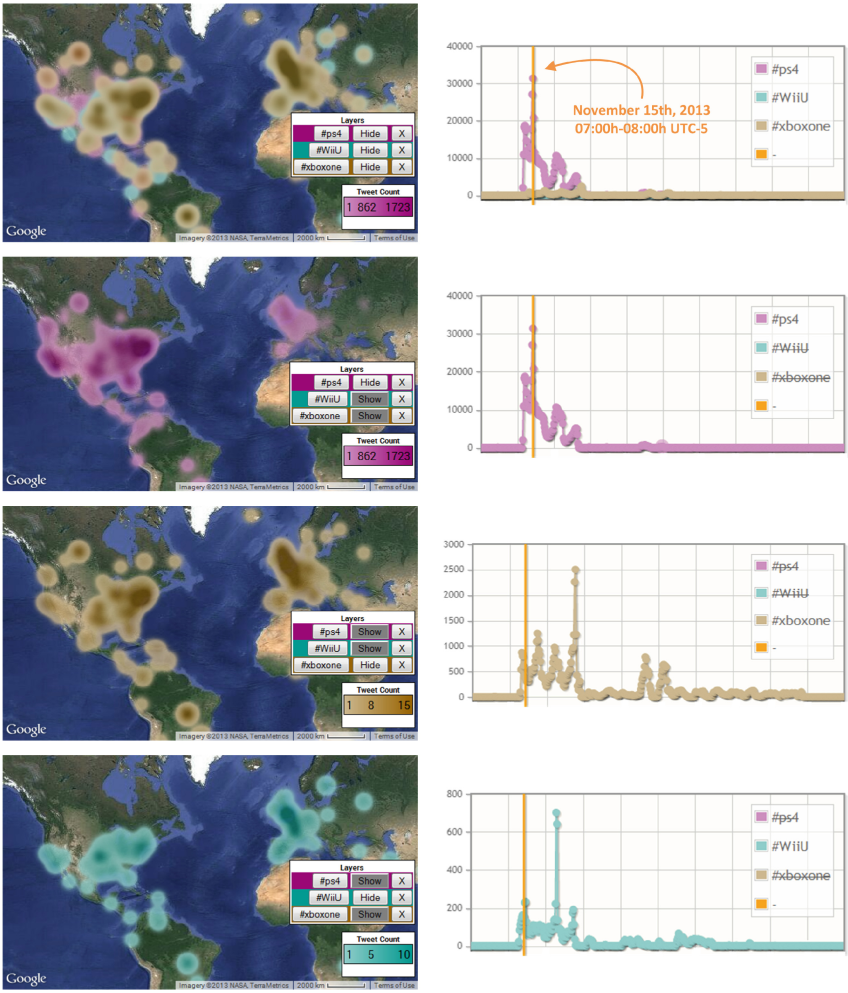


Fig. 5. Heatmap-based as well as quantitative comparison of game console popularity.

Madrid fan mentioned not only Toby Alderweireld but also his Belgian teammate Thibaut Courtois in his tweet: "Well done to #Atleti's @thibautcourtois & @AlderweireldTob and their #RodeDuivels teammates. We'll see you in Brazil at #wc2014".

4.2 Game Console Comparison

The market of (next-gen) gaming consoles is (for the time being) dominated by Sony, Microsoft and Nintendo with their PlayStation 4, Xbox One and Wii U hardware, respectively. In this second test case, the TweetPos tool was put to use to compare the attention these three consoles receive on the Twitter network, and to uncover geographic dissimilarities between their respective popularity, if any. Therefore, between November 1st and November 16th, 2013, the `ps4`, `xboxone` and `WiiU` keywords were tracked with TweetPos. An impression of the resulting data set is given in Fig. 5. This figure visualizes the geo-spatial intensities of the three hashtags on the launch day of the PlayStation 4 in the USA (i.e., on November 15th between 07:00h and 08:00h UTC-5), as well as per-hour aggregated overviews of the volumetric magnitudes of the collected data sets.

These experimental results validate that TweetPos succeeds in layering multiple heatmaps, each associated with an independent query, on top of a single topographic map. The same holds true for the tweet volume plotting functionality of the line chart. Notice however from the topmost row of images in Fig. 5 that keyword visualizations might quickly conceal one another in multi-layer scenarios, which in turn is likely to impair analytical efficiency. Courtesy of TweetPos' ability to on-the-fly switch the rendering of individual layers on and off, it nonetheless remains feasible to interactively compare and interpret (the geographic provenance of) tweets in multi-query studies. In effect, the images in the bottom three rows in Fig. 5 communicate exactly the same information as the ones in the upper row, yet in an itemized fashion.

In-depth analysis of the composed data body falls beyond the scope of this article. Instead, we will point out two illustrative insights that we were able to extract from the collected tweets. Firstly, Fig. 5 at a glance reveals the existence of large quantitative differences between the three tracked keywords. In the monitored time interval, the Wii U console garnered only a fraction of the attention that the Xbox One was able to accumulate, whose Twitter coverage in turn was outclassed by that of the PlayStation 4 by an order of magnitude. The fact that the experiment encapsulated the PlayStation 4's USA release date definitely contributed to this outcome. In particular, inspection of the captured tweet messages confirmed considerable hype build-up as the PlayStation 4 release approached. For the same reason, the PlayStation 4 tweets geo-spatially tended towards the USA. Secondly, the volume diagrams show that Microsoft was able to pierce the PlayStation 4's Twitter hegemony exactly once in the course of the experiment. This achievement can be attributed to a clever marketing strategy: by retweeting a message from the official Twitter account of Xbox France, users could reveal the identity of the French Xbox One ambassador, an opportunity that was massively seized by fans. The resulting retweets primarily originated from Western Europe, and France in particular (not shown in Fig. 5).

5 Related Work

The principle of creating map mashups of the geographic sources of tweets has been considered by a number of commercialized web services. Examples include TweepsMap (http://tweepsmap.com/), Trendsmap (http://trendsmap.com/), Tweereal (http://tweereal.com/), Tweetping (http://tweetping.net/) and GlobalTweets (http://globaltweets.com/). The first maps (the home location of) the followers of a particular user's Twitter account, the second provides a real-time, localized mashup of currently trending Twitter themes, and the final three offer real-time geographic visualization of Twitter posts.

The academic literature also holds a number of articles that deal with deriving geo-spatial insights from Twitter data. Stefanidis et al. have proposed a framework to harvest and analyze ambient geographic information (i.e., not specified in terms of explicit coordinates) from tweets [7]. The iScience Maps tool targets behavioral researchers interested in exploiting Twitter for localized social media analysis purposes [8]. The global concept of applying Twitter as a distributed sensor network to identify and locate events in the physical world has been successfully explored by a number of analogous research initiatives [1,9–11]; of particular relevance is the social pixel/images/video approach by Singh et al. that allows for Twitter-powered situation detection and spatio-temporal assessments [12]. Field and O'Brien have investigated the application of cartographic principles to Twitter-powered map mashups [13]. Finally, the software architecture proposed by Oussalah et al. affords the deployment of geolocated services that are fueled by Twitter data [14].

All systems that have been cited in this section, both commercialized and academic ones, have their specific merits and feature sets. The TweetPos instrument exhibits functional overlaps with all of them. For example, the social pixel approach largely corresponds with our animated heatmap-based visualization solution. Some related tools even provide functionality that is missing in TweetPos. When for instance again looking at the social pixel framework, it incorporates an automated situation detection scheme and exploits domain semantics to autonomously recommend relevant control actions in response to detected events. However, the TweetPos tool exceeds every cited initiative in terms of the variety of analytical means it integrates and the synergistic benefits that stem from this holistic design. As an example, only a minority of the related systems grants insight in both historical and current tweet posting behavior. Also, the combination of a heatmap-based representation of the geographic intensity of topics, a tweet volume diagram, and dynamic means to inspect the textual contents of tweets fosters unprecedented deep mining of (the geo-spatial evolution of) Twitter contributions. A final example of a differentiating TweetPos feature is its layering mechanism and the opportunities in terms of comparative analysis it unlocks. Only the iScience Maps tool provides similar functionality, yet its comparison options are limited to exactly two configurations; in contrast, unlimited numbers of layers can be constructed in TweetPos.

6 Conclusions and Future Work

SNSs have become prominent information channels in present-day society, as is manifested by the massive amounts of information that are shared and communicated through them. Given this quantitative overload, human operators benefit from tools that assist in transforming the constituting raw data into practical knowledge. This article has proposed *TweetPos*, a web service that provides exactly such assistive functions for the Twitter network, hereby allocating elevated attention to the geo-spatial characteristics of tweets. As the human mind is very adept at visual pattern recognition and at interpreting graphical data formats, TweetPos maximally invests in visual output modalities. The tool integrates and blends multiple complementary functions in order to yield a holistic solution for Twitter data analysis. Experimental results collected from two isolated test cases confirm this claim and prove the feasibility, effectiveness and added value of our work. In particular, it has been established that the TweetPos service succeeds in streamlining the ingestion, filtering, processing, analysis and mining of tweeted information, and as such represents a valuable, highly versatile tool with cross-disciplinary application options.

Decision making logic, provisions for automated conclusion drawing and autonomous recommendation systems have deliberately been omitted from the current instantiation of the proposed tool, as we believe these tasks are more suited to human operators than to machines. As part of future research, we nonetheless plan to investigate whether the incorporation of computer-mediated aids might assist users in executing these actions more efficiently and swiftly. Potential supportive technologies include visual pattern recognition and edge detection algorithms to facilitate heatmap analysis, and linguistic processing frameworks to aid human operators in categorizing aggregated tweets on the basis of the textual message they convey. Another trajectory of future work is dynamic data delivery. In the current implementation, all tweet data pertaining to a particular query is transferred from the back-end server to the web browser in bulk. Although this design renders the TweetPos service highly responsive once all data has been downloaded, it also causes start-up delays to be high (i.e., they are directly proportional to the data set size). At the same time, network bandwidth utilization is suboptimal, as the client is likely to end up downloading data which the user will never inspect (or at least not in detail). We will therefore implement a demand-oriented transmission scheme in which relevant data is transmitted just-in-time (i.e., when it becomes needed). By doing so, we will be able to investigate the trade-off between service responsiveness and start-up delay, as well as the impact this balance has on the usage experience.

References

1. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the WWW 2010, Raleigh, NC, USA, pp. 851–860 (2010)
2. Pinto, N., Majaj, N.J., Barhomi, Y., Solomon, E.A., Cox, D.D., DiCarlo, J.J.: Human versus machine: comparing visual object recognition systems on a level playing field. In: Proceedings of the Cosyne 2010, Salt Lake City, UT, USA (2010)
3. Twitter Developers: using the Twitter search API (2013). https://dev.twitter.com/docs/using-search
4. 140dev: 140dev streaming API framework (2013). http://140dev.com/free-twitter-api-source-code-library/
5. Google Developers: The Google Geocoding API (2013). https://developers.google.com/maps/documentation/geocoding/

6. Google Developers: Google Maps JavaScript API v3 (2013). https://developers.google.com/maps/documentation/javascript/
7. Stefanidis, A., Crooks, A., Radzikowski, J.: Harvesting ambient geospatial information from social media feeds. GeoJournal **78**, 319–338 (2013)
8. Reips, U.D., Garaizar, P.: Mining Twitter: a source for psychological wisdom of the crowds. Behav. Res. Methods **43**, 635–642 (2011)
9. Boettcher, A., Lee, D.: EventRadar: a real-time local event detection scheme using Twitter stream. In: Proceedings of GreenCom 2012, Besançon, France, pp. 358–367 (2012)
10. Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski, J.: #Earthquake: Twitter as a distributed sensor system. Trans. GIS **17**, 124–147 (2013)
11. Takahashi, T., Abe, S., Igata, N.: Can Twitter be an alternative of real-world sensors? In: Jacko, J.A. (ed.) Human-Computer Interaction, Part III, HCII 2011. LNCS, vol. 6763, pp. 240–249. Springer, Heidelberg (2011)
12. Singh, V.K., Gao, M., Jain, R.: Situation detection and control using spatio-temporal analysis of microblogs. In: Proceedings of WWW 2010, Raleigh, NC, USA, pp. 1181–1182 (2010)
13. Field, K., O'Brien, J.: Cartoblography: experiments in using and organising the spatial context of micro-blogging. Trans. GIS **14**, 5–23 (2010)
14. Oussalah, M., Bhat, F., Challis, K., Schnier, T.: A software architecture for twitter collection, search and geolocation services. Knowl.-Based Syst. **37**, 105–120 (2013)