Joining Spatial Visualization Tools with Social Media Data Using Free and Open Source Software

M. Zurbarán**, D. Oxoli*, C. E. Kilsedar*, P. Wightman**, M. A. Brovelli*

* Department of Civil and Environmental Engineering, Politecnico di Milano, Como Campus, Via Valleggio 11, 22100 Como
** Department of Systems Engineering, Universidad del Norte, Km 5 Via Pto. Colombia, Barranquilla, Colombia

Extended Abstract

The development of mobile technologies and its fast adoption by users promote virtual social activities and a growing dependence on location-based services for everyday tasks, i.e. accessing local news, finding a near shop, consulting transportation, etc. This phenomenon provides valuable digital footprint in a geographical context that was not possible in years before. In this work, the authors propose the use of crowdsourced georeferenced data from different social platforms along with a Web-based Geographic Information System (GIS) that visually communicates what is reported by social media users.

In the presented demo, data has been collected from Twitter and Foursquare since the beginning of 2016 focusing specifically on the Lombardy Region. Both, data collection and visualization are performed using only open data and Free and Open Source Software (FOSS), with the intention to keep it openly available for interested parties.

By combining different resources available, it is now possible to identify concentrations of Points of Interest (POIs) in real time and understand how these change over time using date ranges for filtering. This provides valuable insights related to important phenomena in the region (i.e. comparison of the behavior on

transportation terminals, social media usage during natural disasters or sport events, etc.).

For this spatio-temporal Web-based GIS client-server architecture is used, depicted in detail in *Figure 1*. On the server-side, the Web application that collects and serves the data is developed using the Django Web framework built with Python, which handles the daily collection of georeferenced data on a PostgreSQL database with the PostGIS plugin enabled. For Twitter the *Twitter Streaming Application Programming Interface* (API) is used that permanently stores the new tweets with a continuous streaming connection, whereas for Foursquare the *Venues API* is used that queries the Lombardy Region three times a day and stores the total check-ins each time for all the reported POIs. This data is subjected to restrictions that social media vendors may apply, but nevertheless represents a valuable source of users' location information.
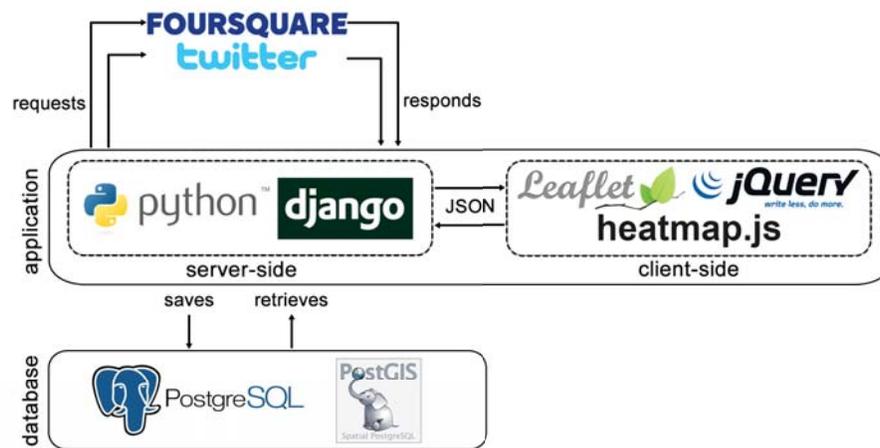

Figure 1. Architecture of the Web-based GIS

A Python script is developed on the server-side in order to aggregate the POIs stored in the PostgreSQL database using a predefined square grid. For this demo, a 200 m grid is selected, which allowed to optimize both the performance of the system and the visualization of the map. For the data aggregation process JavaScript Object Notation (JSON) is used with the list of the coordinates of the cell centers as keys and the associated count of POIs within the cell area as values. This JSON is passed as input to the client-side for the heatmap generation. The user is able to

make requests filtering by date range and a keyword. While the grid is always static, the POIs are selected dynamically using the filters to display only the locations registered during the selected period and that include the keyword (if specified). The heatmap.js library and Leaflet are used to create a dynamic heatmap that is responsive to the zoom level over the map. Higher density locations are emphasized using a gradient color scheme depending on the amount of points included on screen, by using local maximum computation. Moreover transparency is added to the Twitter and Foursquare layers using heatmap.js, so that the overlaid layers can be visually interpreted. The users can also enable or disable data sources to be displayed on the heatmap using a legend. A sample use, where the start and end date is set to 1 August 2016 and 6 October 2016, Twitter tag is set to "hike" and Foursquare keyword is set to "Mountain" is given in *Figure 2*. On the map both the Twitter and Foursquare data is shown, where density variations, intersection and difference areas between two sources are visible. The demo application is available at http://geomobile.como.polimi.it/heatmap.
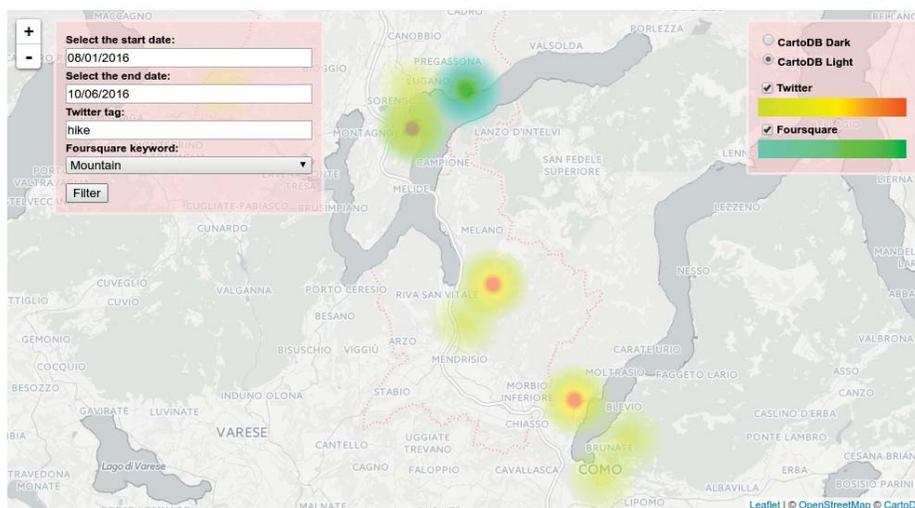

Figure 2. A sample use of the Web-based GIS

In conclusion, bringing all the data sources together is a useful novelty for anyone who is interested in the area for several purposes. Furthermore, presenting a large amount of data through a

heatmap is an intuitive method where being able to filter it according to time and a keyword is an additional dimension to usefulness. This helps also to understand for what purposes different social media platforms are used by easing data analysis as well as inferring correlation between user activity and their geographical context. Moreover, the possibility of customizing queries to the database (e.g. by specifying tags) enables to display thematic heatmaps according to the user's needs.

References

Django: the web framework for perfectionists with deadlines. https://www.djangoproject.com. Accessed 23 September 2016

Heatmap.js: dynamic heatmaps for the Web. https://www.patrickwied.at/static/heatmapjs. Accessed 23 September 2016

Leaflet,js: an open-source JavaScript library for mobile-friendly interactive maps. http://leafletjs.com. Accessed 23 September 2016