Contribution ID: **24**   Type: **Poster**

Distributed databases to study human mobility through social networks

Monday, 16 September 2013 09:00 (8h 30m)

Description of Work

To study mobility in Zurich, London and Barcelona we use geolocalized tweets. One can get 1% of all the tweets by using the stream API provided by Twitter, but in this case, less than 12 % of the collected tweets are geolocalized and only a small fraction of them are located in the cities we are focusing
in. To get the network of users, we identified the users that have tweets geolocalized in the cities considered. Then, in addition to the stream data, we download the Twitter timeline of these specific users.
The data type to be stored and managed was suited to be stored in a database and there are basically two kinds of databases: relational and non relational or no-SQL. Two conditions have been taken into account when choosing a database: 1) Streaming retrieval volume is not constant and 2) the information contained in the tweets and the format of the fields can change over time. Attending these needs we focus on non-SQL databases but comparing MySQL and MongoDB performance.

Printable Summary

As the amount of digital data has increased in the last years, storage systems are evolving to fit users' needs and traditional relational databases are currently competing with non-relational databases whose main goals are horizontal scalability and dynamic schemas support. For large amounts of data, non-relational databases leverages sharding or distribution of data to different computers over the network, to speed up query results. In the framework of EUNOIA project we are currently gathering information related to mobility. The aim is to study mobility behaviour of people in urban areas , such as attitudes and lifestyle, which are particularly important, e.g., for developing demand management concepts aiming to influence mobility decisions. In this context we are using geolocalized tweets as a source of information for mobility patterns and we relied on non-relational databases to get efficient data storage and fast access to the data.

Primary author: TUGORES, Antonia (IFISC (UIB-CSIC))

Co-author: Prof. COLET, Pere (IFISC (UIB-CSIC))

Presenter: TUGORES, Antonia (IFISC (UIB-CSIC))

Session Classification: Posters display