Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-30 (2016) 257–262

Using Big Data and Real-Time Analytics to Support Smart City Initiatives

Arthur Souza * **Mickael Figueredo** ** **Nélio Cacho** *
Daniel Araújo * **Carlos A. Prolo** *

* *Department of Informatics and Applied Mathematics ,Federal University of Rio Grande do Norte ,Natal, Brazil*
** *School of Science and Techology Federal University of Rio Grande do Norte, Natal,Brazil*

Abstract:
A central issue in the context of smart cities is related to the capability to acquire timely information about city events. This paper describes a platform which focuses on processing messages posted in Twitter social network. Key issues here are the high throughput a large volume of data per second that needs to be processed, and the need to process ill formed natural language texts. With these in mind the platform has pipelined modules for robust, fast, real time tweet acquisition and storage, filtering of several kinds, natural language processing and sentiment analysis, that feed a final analysis and visualization module. A case study of sentiment analysis during the 2014 FIFA World Cup in Brazil is used to validate the effort made so far.

Keywords: Smart cities, social networks, tweets, natural language processing.

1. INTRODUCTION

The city's growth adds complexity and management challenges to the government authorities in dealing with problems related to water supply, local waste disposal, urban traffic management system, health, education, public safety, economy, environment and tourism. In this sense, the great challenge to be faced is to ensure sustainable urbanization associated with socioeconomic progress.

Politicians around the world are seeking for answers and ways to deal with these challenges. One of the strategies proposed encompasses the creation of smart cities. The work of Caragliu et al. (2011) argue that a city can be defined as "smart", when there is investment in human and social capital, as well as in information and communication technology (ICT) infrastructure.

Smart city incorporates a large number of systems, which represent the most basic infrastructure for integrating the real and virtual worlds. One of the great challenges of deployment of smart cities is the extraction of relevant information from the ICT infrastructure of cities. For Komninos et al. (2013), such extraction usually relies on the use of sensors that are installed to capture the flow of vehicles, water and energy consumption, thus requiring high public investment for the development of smart cities.

To overcome such difficulty, some studies (Doran et al. (2013); Anantharam et al. (2015)) suggest using social media to identify the perception of residents and visitors about a particular city. For example, social media can be used to obtain relevant information about the situation of public transport, traffic and environmental conditions, public safety and general events in cities. In this sense, the purpose of this paper is to present a platform that uses social media as a data source to support the decisions of policymakers in the context of a smart city.

The proposed platform was implemented and tested by means of analyzing 7.5 million Tweets. The results show that it is possible to identify relevant information about points of agglomeration and concentration of users, movement of users and the user perception about city events. Overall the initial results suggest that data collected from social media can be applicable to the effective management of smart city initiatives. The remainder of this paper is organized as follows. Section 2 describes some details of the Natal Smart City initiative. Section 3 presents the architecture and the implementation details of the platform. Section 4 describes the evaluation process. Section 5 presents some related works. Finally, Section 6 provides some concluding remarks.

2. NATAL SMART CITY INITIATIVE

Natal is located on the northeast of Brazil by the Atlantic Ocean. The capital city of the state of Rio Grande do Norte is home of approximately 862.000 people. Natal was a host city of the 2014 FIFA World Cup. Although Natal was not the location for the World Cup knockout stage, Natal hosted 4 games in the group stage with an average attendance of 40,000 fans at each game. In total, Natal received around 173,000 tourists during the World Cup period. According to study performed by ForwadKeys and Pires&Associados (2014), Natal presented the highest growing number of bookings among all host cities when compared to the same period in 2013, for which bookings did have a increase of 1000%.

The high number of tourists puts severe pressure on the urban infrastructure and services related to transportation,

10.1016/j.ifacol.2016.11.121

safety and water consumption. In order to handle such pressure, the Natal city council in partnership with public and private sector have engaged in an initiative to transform itself into smart city. The development of Smart City facilitates seamless access to value-added services such as access to real-time information on public transportation network, enriches tourist experiences and enhances city competitiveness (Buhalis and Amaranggana (2013)).

The purpose of the Natal smart city approach is to accelerate and enable the delivery of outcomes across various sectors, through a truly integrated approach. For instance, the plan creates a network infrastructure, named Giga Metropole, that has an optical backbone of approximately 160km, as well as a passive network of approximately 300km to interconnect public institutions in the state of Rio Grande do Norte. More precisely, the Giga Metropole will benefit around 650 public and private institutions in Natal's metropolitan area, including: 350 state and municipal public basic education schools, police stations, universities and technical schools, teaching laboratories, and 10 hospitals.

The implantation of smart cities initiatives usually requires the use of a variety of sensors and applications that are connected and can interact with each other to create controlled environments that can be adjusted in real time (Hancke et al. (2012)). In many cases the potential of smart cities depends on the density and integration of the sensors and their applications. Examples of applications for smart cities include: implementation of vibration sensors to monitor the flow of vehicles and the integrity of the pavement (López-Higuera et al. (2011)), the use of sensors to predict traffic conditions and optimize the public lighting in Lyon, France (Perchet (2013)), use of cameras around the city to identify incidents in Liverpool, England (Coleman and Sim (2000)), etc.

However, the use of a variety of sensors poses a number of challenges that are related to device management with restricted physical capabilities (as energy, processing, and memory), or are related to security and privacy (particularly of citizens, critical infrastructure and systems information), or to dependability requirements (reliability and availability), since defects can lead to system failures, resulting in financial losses and environment or people damage. In addition to all these challenges, Brazilian cities face the very limited budget to spend on ICT.

In this context, some researches (Doran et al. (2013); Anantharam et al. (2015)), suggest to use social media to identify in real-time the tourists or resident perception about a particular event in a city. According to Doran et al. (2013), the use of sensors can be useful for identifying "what" is happening, but is unable to identify "why" and "how" such an incident occurs. In this case, Doran et al. (2013) suggests using social media to capture the human perception of events. Such suggestion is based on the fact that the perceptions of a city event are often described through comments on social media. One of the main social media networks currently available is Twitter. Twitter has over 500 million users and generates around 500 million posts (tweets) per day (Telegraph (2013)).

The use of social media as a source of information does not require the deployment and maintenance of a comprehensive IT infrastructure spread throughout the city. It needs only a computer connected to the Internet to make use of the API provided by the social network used. Despite these facilities, the use of social media in the context of smart cities poses several challenges. The first is related to the capability to process in real-time all posts (about 6,000 per second Telegraph (2013)) sent on various subjects. The second challenge is related to the capability to interpret such posts. The sentiment expressed by a tweet can be used to identify the occurrence of events. The third challenge is to store such data in a way that it is possible to generate appropriate views. In order to address these challenges, the next section describes a platform developed and implemented to support smart cities initiatives.

3. PROPOSED PLATFORM

Peoples perceptions about events and issues they encounter in their cities are often embodied in the words, terms and phrases that form their spoken language, and now also in their social media posts (Doran et al. (2013)). Social media produce millions of posts being broadcast over time. These posts need to be analyzed in order to extract situation awareness about people behavior in a smart city. This work has adapted and optimized a set of machine learning and natural language processing techniques to deal with real-time and high-volume text streams. These techniques are packaged in a proposed software platform.

3.1 Real-time Stream Infrastructure

The proposed platform uses real-time processing infrastructure as a central component. This infrastructure is implemented through the use of Apache Storm (Apache (2015a)). Storm is a free and open source stream-processing framework capable of processing one million 100 byte messages per second per node (Apache (2015a)). A Storm cluster is formed by a distributed network of processing nodes that process a set of data compartmentalized in tuples. For this, three components are defined: (i) Zookeeper, (ii) Nimbus and (iii) the Supervisor. Zookeeper (Apache (2015b)) is a high-performance service that coordinates distributed applications through configuration management, appointment and work services group synchronization. On storms architecture, it stores the synchronization of data and the processing state of tuples that will be performed at the nodes supervisor. Supervisor represents the node(s) of the cluster Storm responsible for the data processing. Finally, Nimbus is the primary node of the cluster Storm, responsible for distribution of code to be processed, assigning tasks to nodes supervisors and fault monitoring.

The workflow on Storm architecture is structured by Topology concept. Topologies as mechanisms for computational organization and are defined as a processing graph where each node in a topology contains a logic processing and links between them. The links indicate that data can be exchanged between nodes. The data stream is represented by Streams elements. Stream is a sequence of tuples that can be affected by spouts components and bolts. Spouts and bolts have interfaces that can be used by developers to implement the program logic. Spouts are elements that receive a data stream and organize them in