Telematics and Informatics 35 (2018) 68–81

Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

An approach based on semantic stream reasoning to support decision processes in smart cities

Giuseppe D'Aniello[a], Matteo Gaeta[a], Francesco Orciuoli[b,*]

[a] Dipartimento di Ingegneria dell'Informazione ed Elettrica e Matematica Applicata, University of Salerno, 84084 Fisciano, Salerno, Italy
[b] Dipartimento di Scienze Aziendali – Management & Innovation Systems, University of Salerno, 84084 Fisciano, Salerno, Italy

ARTICLE INFO

Keywords:
Stream reasoning
Smart City
Semantic Web
Decision Support Systems

ABSTRACT

This paper discusses the use of stream reasoning models and techniques to provide a stream reasoning-based architecture to represent, manage and process data streams produced in the Smart City context, to extract useful knowledge for a better understanding of city phenomena and to support the decision making processes in both the city governance and the citizens. The proposed architecture, taking into account the need for processing heterogeneous data/information across several and different domains, is able to sustain decision-making processes deployed at operational, tactical and strategical levels. Such architecture is distributed and adopts a meet-in-the-middle configuration logic that is really effective and scalable in a complex environment like a (smart) city. The applicability of semantic technologies to implement all the aforementioned features is demonstrated by means of a complex case study realized by using a dataset, related to the city of Aarhus, provided by the CityPulse EU Project.

1. Introduction

Nowadays, the volume of data produced worldwide has extremely increased with respect to the previous years and this trend is constantly growing. First of all, recent developments in ICT have significantly fostered data production processes with respect to the common users' perspective: every 60 s an extremely huge volume of data is uploaded via Internet. This is due to the availability of numerous and cheap ways final users can use to share contents through social networks, instant messaging systems, e-mails, etc. Moreover, the rise of new paradigms, like Internet of Things (Atzori et al., 2010) and wearable computing (Ugulino et al., 2012), provides a further boost to the aforementioned phenomenon that acquires even more relevance in the urban areas. In fact, since most of the world population lives in urban areas, the main environment where data are produced is undoubtedly the city. Thus, the question is *why do not exploit these data in order to improve the quality of life?* This is one of the mission of the so called Smart City (Zanella et al., 2014), i.e., an high-tech intensive and advanced city which connects people, information and various entities using new and advanced technologies, with the aim to create a more sustainable and competitive model of city. With respect to data, IBM defines a smarter city as: *one that makes optimal use of all the interconnected information available today to better understand and control its operations and optimize the use of limited resources* (Kehoe et al., 2011).

The above described context introduces serious challenges related to the processing of heterogeneous data produced by different sources. The problem can be framed in terms of Big Data. According to Chen et al. (2014), the term *Big Data* points out to enormous datasets (compared to traditional ones) including masses of unstructured data needing more real-time analysis. The opportunities

* Corresponding author at: Dipartimento di Scienze Aziendali – Management & Innovation Systems, University of Salerno, via Giovanni Paolo II, 132, 84084 Fisciano, Salerno, Italy.

E-mail addresses: gidaniello@unisa.it (G. D'Aniello), mgaeta@unisa.it (M. Gaeta), forciuoli@unisa.it (F. Orciuoli).

http://dx.doi.org/10.1016/j.tele.2017.09.019
Received 21 July 2017; Received in revised form 29 September 2017; Accepted 30 September 2017
Available online 03 October 2017
0736-5853/

provided by Big Data, especially in the Smart City, are related to the value, hidden in these data, that has to be discovered. The challenges related to Big Data are mainly represented by the definition of methodologies and technologies enabling the effective organization and management of the aforementioned datasets. Typically, the concept of Big Data is clarified by considering five Vs (Batty, 2013):

- Volume: refers to the size of data that has been created from all the sources.
- Velocity: refers to the speed at which data is generated, stored, analyzed and processed. An emphasis is being put recently on supporting real-time big data analysis.
- Variety: refers to the different types of data being generated. It is common now that most data is unstructured and cannot be easily categorized or tabulated.
- Variability: refers to how the structure and meaning of data constantly change, especially when dealing with data generated from natural language analysis.
- Value: refers to the possible advantage big data can offer a business based on good big data collection, management and analysis.

In particular, with respect to velocity, data are continuously generated with high frequencies producing what are called *data streams*. Real-time analysis of these data streams enable the definition of effective monitoring, alerting, decision-support and fault detection systems supporting scenarios in which it is possible to rapidly react to what is occurring within the monitored environment.

Real-time analysis of data streams is not a trivial task. Although existing solutions of Data Stream Management Systems (DSMSs) (Geisler, 2013) and Complex Event Processing Systems (CEPs) (Cugola et al., 2012) are already available, they suffer in scenarios characterized by high heterogeneity of data. Unfortunately, Smart Cities are environments characterized by distributed sources of heterogeneous data streams that need to be processed in order to be analysed in real-time to enable cross-domains applications.

The general goal of this paper is to demonstrate that semantic technologies can be fruitfully applied to real-time data stream analysis in the Smart City, i.e., by considering huge volumes of heterogeneous data. More in detail, this work proposes a semantic-driven architecture to represent, manage and process data streams produced in the Smart City in order to extract useful actionable knowledge for a better understanding of city phenomena and to support decision making processes at the city administration level. In particular, the proposed solution is based on: i) the adoption of the Semantic Sensor Network Ontology (SSN) ontology (Compton et al., 2012) to conceptualize and semantically enrich data coming from the different and heterogeneous sensors deployed on the urban area, ii) the application of C-SPARQL (Barbieri et al., 2010) that is useful for filtering, querying, integrating data and executing a first level of analysis on such data, and iii) the use of engines for *Semantic Stream Reasoning* able to infer new knowledge from the existing one. The benefits provided by semantic technologies are due to their native capabilities related to interoperability and integration for facing with heterogeneous data reasoning.

In order to evaluate the applicability of the proposed solution in a realistic scenario, a case study has been planned and executed by using datasets provided by the CityPulse[1] EU Project. In particular, in order to support the experimentation activity, data regarding traffic, pollution and parking measurements, taken from August to September 2014, have been selected. These data have been processed by the stack proposed in the solution on a single machine (providing weak computational power). The performance of the stack (considering both the C-SPARQL layer and the Semantic Reasoning layer) has been measured. Such performance is promising and allows to consider semantic technologies as plausible and suitable choices for implementing such kind of large-scale distributed systems.

The paper is structured as follows. Section 2 introduces the related works with a classification of different models of a Smart City. Section 3 describes the overall approach and the proposed architecture. Section 4 details the process of gathering, representing, elaborating and reasoning on data streams produced in the Smart City. Section 5 presents the results of the evaluation in a case study related to the city of Aarhus. Section 6 concludes the paper with some final remarks and future research directions.

2. Related works

Over the last decade, many models of Smart City have been proposed and applications implemented. To clarify and organize what has been mainly proposed so far, we classify such models and applications in three different classes, as reported in Table 1:

- Models: abstract frameworks trying to describe all the different assets and aspects of a Smart City.
- Specific-purpose applications: models and applications related to one domain or asset of the Smart City, which do not consider the city as a complex system made of several interlinked domains.
- Multi-domain practical models: models and applications that describe the Smart City as a complex system and consider more than one assets and domains.

Next subsections provide further details for each class.

[1] http://www.ict-citypulse.eu/page/.