Enabling Query Technologies for the Semantic Sensor Web

Jean-Paul Calbimonte, Universidad Politécnica de Madrid, Spain
Hoyoung Jeung, SAP Research, Australia
Oscar Corcho, Universidad Politécnica de Madrid, Spain
Karl Aberer, Ecole Polytechnique Fédérale de Lausanne, Switzerland

ABSTRACT

Sensor networks are increasingly being deployed in the environment for many different purposes. The observations that they produce are made available with heterogeneous schemas, vocabularies and data formats, making it difficult to share and reuse this data, for other purposes than those for which they were originally set up. The authors propose an ontology-based approach for providing data access and query capabilities to streaming data sources, allowing users to express their needs at a conceptual level, independent of implementation and language-specific details. In this article, the authors describe the theoretical foundations and technologies that enable exposing semantically enriched sensor metadata, and querying sensor observations through SPARQL extensions, using query rewriting and data translation techniques according to mapping languages, and managing both pull and push delivery modes.

Keywords: Query Processing, Semantic Sensor Web, Semantic Web, Sensor Networks, Sensor Technologies

INTRODUCTION

Every second, massive amounts of data are being produced by sensors all around the world. From environmental measurement devices to smartphones, the sources of sensor data continue to proliferate, increasing the possibility of blending the diverse sources to collaboratively detect and identify a multitude of observations, from simple phenomena to complex events and situations. As these sensors become more accessible, due to lower costs and simpler configuration and maintenance, they can be deployed not only by companies and government institutions, but also by enthusiasts and citizen scientists. Therefore the volume of data produced is extremely large and highly heterogeneous, making it complex to discover and use.

The heterogeneity of data as well as sensing environments is a key obstacle for realizing a connected sensor world. Different sensor network deployments usually represent the information that they capture in different ways. The data models and schemas are different, the data types and structures are not always compatible, and even the data values often

DOI: 10.4018/jswis.2012010103
use different representations. For example, consider multiple sensor networks measuring the same type of physical phenomenon. Each sensor deployment may have its own way to represent semantically identical information, e.g., “wind speed” vs. “average wind speed,” or “temperature” vs. “thermometer.” If a user wants to obtain the latest wind speed or temperature data values over the region where all the sensor networks are deployed, the user must employ a mechanism for letting the system understand the semantically equivalent but different representations of data, in order to fully answer the query.

One of the solutions to deal with heterogeneity is through the semantic annotation of sensor data (Sheth, Henson, & Sahoo, 2008), and the provision of ontology-based access to it (Calbimonte, Corcho, & Gray, 2010; Taylor & Leidinger, 2011). However, there is a lack of evidence of how this approach scales, especially with high data rates, and in push-based delivery of streaming data.

In this article we focus on two problems in this context: (i) how to find relevant heterogeneous sensor data sources based on their metadata, and (ii) how to query streaming sensor data from these sources. We summarize our contributions as follows:

• Our main contribution to the first problem is the use of the SSN ontology (Compton et al., in press), along with domain-specific vocabularies, for modeling sensor metadata and observations, augmented with mappings to the original sensor schemas. To this end, we use R2RML (Das, Sundara, & Cyganiak, 2012) (RDB-to-RDF mapping language) for mapping relational streams -instead of tables- to ontologies. Thus we use ontologies as a common model for representing sensor data and metadata, to make it possible to search for data sources and to access them through ontological schemas.

• For the second problem, we propose a query rewriting and data translation approach that allows querying virtual RDF streams using the SPARQL language with streaming extensions. This approach exploits the R2RML mappings to provide access to the sensor streaming data, not only the metadata. Furthermore, we show that our query rewriting and execution mechanisms are applicable for both pull and push delivery modes, and also for various state-of-the-art stream processing engines, such as SNEE (Galpin, Brenninkmeijer, Jabeen, Fernandes, & Paton, 2009), GSN (Aberer, Hauswirth, & Salehi, 2006), Pachube (https://pachube.com/), and Esper (http://esper.codehaus.org/). We provide empirical evidence of performance with respect to sampling rates and delivery latency in both pull and push-based modes.

As an illustrative example, we show how the GSN implementation of this approach was used in a federated sensor network in the Swiss-Experiment (http://www.swiss-experiment.ch) project, a collaborative platform for sharing real-time sensor data across various institutions to improve environmental hazard forecasting and warning.

The paper is organized as follows: first we introduce the fundamental concepts of ontology-based access to streams. We then discuss the modeling of sensor data, metadata and mappings. We also provide the theoretical foundations of query translation and how it can be implemented. An experimental evaluation is also presented, before our conclusions.

PRELIMINARIES

Our approach is based on the fundamental concepts detailed below: streaming data querying, and semantic data access using ontology-based queries and mappings.

Queries over Streaming Data

A data stream consists of an unbounded sequence of values continuously appended, each of which carries a timestamp that typically indicates when it has been produced. Examples of
On Using Wiki as a Tool for Collaborative Online Blended Learning
www.igi-global.com/chapter/using-wiki-tool-collaborative-online/39188?camid=4v1a

Property-based Semantic Similarity and Relatedness for Improving Recommendation Accuracy and Diversity
www.igi-global.com/article/property-based-semantic-similarity-and-relatedness-for-improving-recommendation-accuracy-and-diversity/145854?camid=4v1a