Five challenges for the Semantic Sensor Web

Oscar Corcho and Raúl García-Castro
Ontology Engineering Group, Departamento de Inteligencia Artificial, Facultad de Informática,
Universidad Politécnica de Madrid, Campus de Montegancedo, Boadilla del Monte, 28660, Madrid, Spain

Abstract. The combination of sensor networks with the Web, web services and database technologies, was named some years ago as the Sensor Web or the Sensor Internet. Most efforts in this area focused on the provision of platforms that could be used to build sensor-based applications more efficiently, considering some of the most important challenges in sensor-based data management and sensor network configuration. The introduction of semantics into these platforms provides the opportunity of going a step forward into the understanding, management and use of sensor-based data sources, and this is a topic being explored by ongoing initiatives. In this paper we go through some of the most relevant challenges of the current Sensor Web, and describe some ongoing work and open opportunities for the introduction of semantics in this context.

Keywords: sensor, ontology, query language

1. Introduction

The combination of sensor networks with the Web, web services and database technologies, was named some years ago as the Sensor Web or the Sensor Internet [1, 2, 3, 4, 5]. Most of the work done on this topic, performed in some cases under the umbrella of the OGC Sensor Web Enablement Working Group[1], focused on the creation of specifications for different functionalities related to the management of sensor-based data (observations, measurements, sensor network descriptions, transducers, data streaming, etc.), and for the different types of services that may handle these data sources (planning, alert, observation and measurement collection and management, etc.).

Some additional work has focused on the provision of platforms that provide the services needed to develop sensor-based applications. These platforms include libraries for common domain-independent data management tasks, such as data cleaning, storage, aggregation, query processing, etc., and they are used to provide domain-specific aggregated services (e.g., coastal imaging [1], patient care [2]).

Finally, centralized registries for sensor-based data have appeared (e.g., Pachube[2], SensorMap[3]), focused on the registration of sensor-based data sources, and on the provision of access to them in multiple ways, by means of REST-based interfaces, web services, or ad-hoc query languages, to name a few.

Figure 1 presents a general architecture of Sensor Web applications; which can be characterised by:
- variability and heterogeneity of data, devices and networks (including unreliable nodes and links, noise, uncertainty, etc.);
- use of rich data sources (sensors, images, GIS, etc.) in different settings (live, streaming, historical, and processed);
- existence of multiple administrative domains; and
- need for managing multiple, concurrent, and uncoordinated queries to sensors.

We will now review some of the most relevant challenges in this area, for which we will later propose descriptions of how semantic-based approaches could be applied.

---

[1] http://www.opengeospatial.org/projects/groups/sensorweb

[2] http://www.pachube.com/
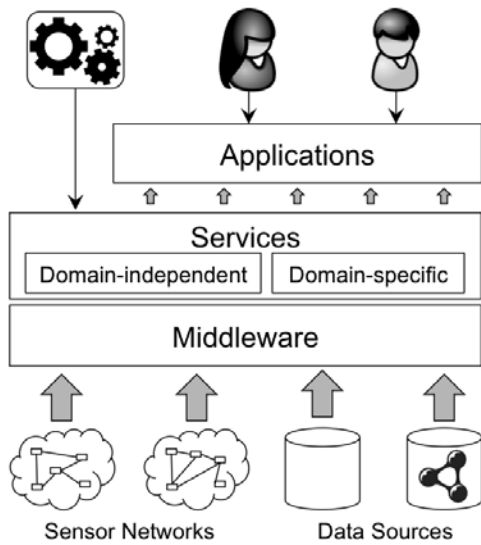[3] http://atom.research.microsoft.com/sensewebv3/sensormap/

Figure 1. Architecture of a Sensor Web application

2. Five Challenges for Sensor Web Applications

This section starts presenting those challenges in the area of the Sensor Web that have to do with the characteristics of the data sources that are handled in typical Sensor Web applications, and then moves into those challenges that have to do with the creation of applications based on these data sources. We do not aim at being exhaustive on the identification of challenges, but we hope that this categorization is useful to understand some of the open problems in this area.

One of the first challenges is related to the **abstraction level** in which sensor data can be obtained, processed and managed in general. Sensor data can be managed at a very low level, at the device- and network-centric levels, generally by means of using low-level programming languages and operating systems. But it can be also managed through higher-level formalisms (e.g., via declarative continuous queries over streams), thereby insulating clients and users from the infrastructural and syntactic heterogeneities of autonomously-deployed sensor networks.

Another challenge is related to the adequate characterisation and management of the **quality** (and **quality of service**) of sensor data. Issues like the unavailability of a piece of data over a period of time may have different meanings when seen from an application perspective: the sensor was not available, there was no event to trigger the data generation during that time, the communication with the sensor was broken, etc. Other issues like the accuracy of the

sensed data may depend on a number of internal and external conditions to the sensor network. In summary, there are a number of quality characteristics that are relevant to the quality of service and that may affect the results obtained from a data observation process, normally with important trade-offs among each other (e.g., longevity vs. latency or completeness vs. throughput).

Another challenge has to do with the **integration and fusion of data** coming from autonomously-deployed sensor networks, with varying qualities of service and different throughput rates, geographical scales, etc. This is related not only with the integration of data coming from different sensor networks, but also with the combination of such data with data persisted in other sources, such as static data or archived sensor data.

Another challenge of utmost importance, related to the previous one, is the **identification and location of relevant sensor-based data sources** with which data integration and fusion tasks can be performed. The number of sensor networks being deployed in the real world is growing continuously, given the fact that the prices of hardware are decreasing. As a result, more experiments and initiatives deploy sensor networks in different (sometimes overlapping) areas, and finding the right information to be used in integration and fusion tasks is highly relevant.

Finally, another important challenge has to do with the need to enable the **rapid development of applications** that are able to handle sensor data, taking into account the aforementioned characteristics and challenges. This includes dealing with data integrity and validation issues as well as the need for common interfaces and formats between applications, databases, sensor networks, etc. This challenge requires enabling the development of applications with different resource models and qualities of service (e.g., energy, bandwidth, processing, storage) and facilitating the interaction with sensor data from the developer and user points of view.

3. Applying Semantic-Based Approaches to Sensor Web Challenges

In this section we provide brief descriptions of how the aforementioned challenges are being addressed in existing initiatives and projects, by means of semantic-based methods, techniques and technologies.

We start with the characterization of the abstraction level at which sensor data can be obtained, processed and managed. A number of **sensor network ontologies** have been defined in the literature [6], which aim at describing different aspects of sensor-based data, from the device point of view (focusing on the hardware that is being used in order to generate the data) to the domain point of view (focusing on the types of data that can be generated from sensors and sensor networks in the context of specific domains such as Health or Environment). Several aspects are relevant in the development of most of these ontologies, such as the distinction between raw observed data and derived data, the representation of aspects like accuracy, or the consideration of observations and measurements according to the relevant OGC models; the ontological representation of this last aspect has received attention on its own [7, 8]. The development of an ontology in this area is one of the main tasks being performed in the W3C Incubator Group on Semantic Sensor Networks[4].

The aforementioned work on sensor network ontologies also takes into account the **quality of the data sources**, although it is not central to the work being performed in the context of the Incubator Group. Data quality is a large research area that is not only applicable to sensor-based data, but to any type of data that can be managed in an application. It is common to talk about data quality in relational databases, in semi-structured data sources, in user generated content, etc. Therefore, it is a property of data sources in general, and not of sensor-based data in particular. However, sensor-based data depends largely on the context of the sensor network, such as the network physical infrastructure, deployment strategy, or surrounding environmental conditions. This context may influence the quality of data (e.g., the accuracy of measurements) and has to be taken into account to correctly interpret them (e.g., to interpret the meaning of data gaps). Early work is being done on the definition of data quality models for this type of data, by categorising existing approaches for other types of sources and selecting and adapting them to the context of sensor networks. The same applies to the quality of service of sensor network sources, in terms of parameters that are also applied to other types of sources (e.g., reliability) and are specialized for sensor networks (e.g., reading rate, battery levels).

With respect to the integration and fusion of data, work has been done in the context of integrating and fusing heterogeneous data streams. Some of this work uses semantic techniques, and some does not. A recent research trend is focused on the generation of **Linked Data from sensor network data streams** [9, 10] by means of transforming sensor-based data into RDF and making it available using HTTP by means of sensor-related URIs. This will allow the seamless navigation across sensor-based (and other types of) data. Other work is being done on the provision of **semantic queries** that are adapted to sensor-based data. They leverage declarative querying infrastructure to define logical views over sensor network data and open the way for view- and ontology-based techniques to be used. These approaches extend query languages like SPARQL with constructors normally applied to stream-based sources (e.g., time and tuple-based windows). Examples of such extensions are the C-SPARQL [11] or the Streaming SPARQL [12] languages, and an example of approaches that provide transformations between sensor data sources and these languages is the work described in [13].

In the context of identifying and locating relevant sensor-based data in the real world, work is being done on the definition of **sensor data registry interfaces**, and in the development of the appropriate infrastructure that can cope with the types of queries that are usually handled in sensor-based applications. These registries should provide support for spatio-temporal queries (e.g., "get sensor data sources that contain information about the temperature in this region for the last two days") and for metadata queries related to existing sensor network ontologies. Some work in this context can be found at [14].

Finally, another identified challenge is related to the development of **high-level application programming interfaces** (APIs) that ease **the rapid development of thin applications** (e.g., mashups) that use data from sensor networks and legacy databases. These programming interfaces should cope in a homogeneous way with the different types of data (persisted and streamed), support the use of the semantic extensions already identified (e.g., semantic-based descriptions of data, linked sensor data, semantic-based registries), and help users interact with and make sense of the potentially enormous and heterogeneous amounts of data coming from the Sensor Web. Examples of these interfaces are already available, although without much semantic support (e.g., SensorMap [15]) and

---

[4] http://www.w3.org/2005/Incubator/ssn/

some early work is also done to develop decision support systems for environmental management.

4. Conclusions and Future Work

In this paper we have described some challenges in the area of the Sensor Web and how these challenges are being addressed using semantic-based approaches.

We have covered issues that arise from the need to interpret, manage and integrate in a meaningful way data that is coming from heterogeneous sensor networks, with different levels of abstraction, different application areas, and different quality conditions. We have also described how applications that rely heavily on sensor-based data can be more flexibly created, and how they can make use of services to locate data sources that may not have been originally deployed for the specific purpose of the application.

Much work still remains to be done in all these areas, and also in others that have not been covered exhaustively in this position paper, such as event identification and management with sensor data or improved sensor network management using semantic techniques, to name a few.

Furthermore, the achievement of a Semantic Sensor Web is not a task to be made in isolation. We have shown how introducing semantics into the Sensor Web scenario presents new requirements over the Semantic Web specifications and technologies. Even if such requirements are currently being satisfied by extending these specifications and technologies, they can be a valuable input for advancement in the Semantic Web area that will, in turn, benefit the Semantic Sensor Web.

working group on Real World Internet[10] and in the W3C incubator group on Semantic Sensor Networks.

References

[1] P.B. Gibbons, B. Karp, Y. Ke, S. Nath, S. Seshan, IrisNet: An Architecture for a World-Wide Sensor Web. IEEE Pervasive Computing, Volume 2, Number 4, October-December 2003.
[2] J. Shneidman, P. Pietzuch, J. Ledlie, M. Roussopoulos, M. Seltzer, M. Welsh, Hourglass: An Infrastructure for Connecting Sensor Networks and Applications. Harvard Technical Report TR-21-04.
[3] J. Ledlie, J. Shneidman, M. Welsh, M. Roussopoulos, M. Seltzer, Open Problems in Data Collection Networks. SIGOPS European Workshop, Leuven, Belgium, September 2004.
[4] D. Havlik, G. Schimak, R. Denzer, B. Stevenot, Introduction to SANY (Sensors Anywhere) Integrated Project. ENVIROINFO 2006, Shaker, Graz, Austria, p.541-546, 2006.
[5] K. Aberer, M. Hauswirth, and A. Salehi, A middleware for fast and flexible sensor network deployment. In VLDB, pages 1199–1202, 2006.
[6] M. Compton, H. Neuhaus, K. Taylor, K.-N. Tran, A Survey of the Semantic Specification of Sensors. Proceedings of the 2nd Int. Workshop on Semantic Sensor Networks (SSN09), Washington DC, USA, October 26, 2009.
[7] C. Henson, J. Pschorr, A. Sheth, K. Thirunarayan, SemSOS: Semantic Sensor Observation Service. Proceedings of the 2009 Int. Symposium on Collaborative Technologies and Systems (CTS 2009), Baltimore, MD, USA, May 18-22, 2009.
[8] W. Kuhn, A Functional Ontology of Observation and Measurement. 3rd Workshop on Geosemantics (GeoS 2009), Mexico City, Mexico, pages 26-43, 3-4 December 2009.
[9] J. Sequeda, O. Corcho, Linked Stream Data: A Position Paper. Proceedings of the 2nd Int. Workshop on Semantic Sensor Networks (SSN09), Washington DC, USA, October 26, 2009.
[10] K. Page, D. de Roure, K. Martinez, J. Sadler, O. Kit, Linked Sensor Data: RESTfully serving RDF and GML. Proceedings of the 2nd Int. Workshop on Semantic Sensor Networks (SSN09), Washington DC, USA, October 26, 2009.
[11] D.F. Barbieri, D. Braga, S. Ceri, E. Della Valle, M. Grossniklaus. C-SPARQL: SPARQL for Continuous Querying. In Proceedings of the 18th Int. World Wide Web Conference, pages 1061- 1062, 2009.
[12] A. Bolles, M. Grawunder and J. Jacobi. Streaming SPARQL - Extending SPARQL to process data streams. The Semantic Web: Research and Applications, pages 448-462, 2008.
[13] J.P. Calbimonte, O. Corcho, A.J.G. Gray. Ontology-based Access to Streaming Data. Poster at the 7th Extended Semantic Web Conference (ESWC2010). Hersonissos, Greece, June 2010.
[14] K. Kyzirakos, M. Koubarakis, and Z. Kaoudi. Data models and languages for registries in SemsorGrid4Env. Deliverable D3.1 Version 1.0, SemSorGrid4Env, August 2009.
[15] S. Nath, J. Liu, and F. Zhao. SensorMap for Wide-Area Sensor Webs. IEEE Computer, Volume 40, Number 7, Pages 106-109, July 2007.

---

[5] http://www.semsorgrid4env.eu/

[6] http://www.ict-sensei.org/

[7] http://www.cooperating-objects.eu/

[8] http://www.ict-peces.eu/

[9] http://www.fp7-aspire.eu/

---

[10] http://rwi.future-internet.eu/