Making Sense of Sensor Data Using Ontology: A Discussion for Residential Building Monitoring

Markus Stocker, Mauno Rönkkö, and Mikko Kolehmainen

University of Eastern Finland,
P.O. Box 1627, Kuopio, Finland
`{markus.stocker,mauno.ronkko,mikko.kolehmainen}@uef.fi`

Abstract. We illustrate the application of automated representation of knowledge acquired from sensor network data to quality of life services. Specifically, for a sensor network used to monitor a residential building we acquire knowledge about events of interest to occupants and represent such knowledge in ontology. An event of particular interest to quality of life which we discuss is 'unhealthy' exposure to carbon monoxide. Hence, we aim at reducing the considerable gap between raw sensor data and abstract domain terminology. Our results support the claim that computational techniques in signal processing, machine learning, and ontology engineering are important elements to systems that make use of environmental sensing, including systems for quality of life information services.

Key words: Sensor data, ontology, knowledge representation, residential building monitoring

1 Introduction

We discuss our preliminary work on automated acquisition and representation of knowledge about events of interest observed by a sensor network used to monitor variables such as temperature, humidity, carbon monoxide, carbon dioxide, particulate matter, water consumption, and energy consumption of a residential building.

Independently of the domain, information systems and services that build on sensor networks face a common challenge: to bridge the considerable gap between raw sensor data and abstract domain terminology [7] that services in interaction with users should 'understand'. An example for such an information system is the Measurement, Monitoring and Environment Assessment (MMEA) platform currently developed within the Finnish Cluster for Energy and Environment (CLEEN).[1] Among other aims, a key feature of the MMEA platform is to manage

---

[1] http://www.cleen.fi/en/program_overviews/mmea

observation and measurement data. Users and client applications may, however, want to interact with the platform using more abstract domain terminology, such as in a query for 'unhealthy exposure to carbon monoxide' which requires the system to understand how 'unhealthy exposure' translates into concentration levels and exposure duration.

Terminologies to describe the characteristics of sensors and sensor networks, aimed at improving the integration and communication among sensors [14], have received considerable attention. Compton et al. [6] review eleven sensor ontologies for the range and expressive power of their concepts, used to describe, for instance, the deployment, configuration, components, data, observations, location, accuracy, or sampling frequency of sensors.

Rather than to describe sensors and sensor networks with metadata, our aim is to represent knowledge about events of interest observed by a sensor network. Our work relates to Conroy et al. [8, 7] who extract from sensor data various biological and physiological properties of athletes during exercise, and propose the EventSense architecture as a solution to sensor data management. EventSense is a matured architecture compared to the system presented here, though the underlying aim is similar. Contrary to EventSense, which is based on XML, we suggest using an expressive ontology language to formalize acquired knowledge.

Gaglio et al. [9] propose a generic architecture to extract information from an environment sensed by a wireless sensor network and discuss it for a case study on wildfire detection. The three-layered architecture aims at bridging the sub symbolic layer of raw measurement data with the symbolic knowledge layer via a conceptual layer that implements a metric space and a notion of similarity for perceived objects, described as vectors with components corresponding to dimensions of the domain of interest.

Liu and Zhao [11] and Whitehouse et al. [22] discuss the architecture of a system that can be queried for high-level events without requiring handling of raw magnetometer data, specifically for a parking garage case study. The authors elaborate on a programming model called Semantic Streams which rests on two fundamental elements, *event streams* being one of them. Hence, at the base of Semantic Streams are detected world events such as objects, people, or vehicles.

The methodology adopted in this paper had previously been discussed by Stocker et al. [16] in a case study for road vehicle classification by means of machine learning and measurement of road pavement vibration. Contrary to Whitehouse et al. [22], the authors of [16] discuss the entire processing chain from sensor measurement data to knowledge represented in ontology.

2   Materials and Methods

The residential building under consideration is located in the city of Kuopio, Finland. A total of 38 sensors are installed to monitor a number of indoor and outdoor variables, including temperature, relative humidity (RH), carbon dioxide, carbon monoxide (CO), air pressure difference, volatile organic compounds,

motion, particulate matter (0.5 µm and 2.5 µm), electricity and water consumption, as well as the use of district-distributed heat energy. Consumption sensors compute integrated values for the building while sensors for, e.g., temperature and relative humidity compute values specifically for 6 areas of the building, namely the entrance, living room, bathroom, and 3 bedrooms. Outdoor temperature is also measured. The data collected by the sensors are transmitted over (wired or wireless) data communication links and centrally stored to a relational database management system.

In this case study we acquire and represent knowledge for (1) observations by the relative humidity sensor in the bathroom for the event 'person taking a shower' and (2) observations by the carbon monoxide sensor for the event 'beyond average carbon monoxide' for homes without gas stoves (0.5-5 ppm).[2] Hence, we focus on two sensors, namely the carbon monoxide [ppm] sensor installed in the living room with identifier S-CO-L and the relative humidity [%] sensor installed in the bathroom with identifier S-RH-B. Both sensors sample at frequency below 1 Hz, officially one measurement value every 5-7 seconds. However, the actual number of measurement values stored in the database per minute can vary considerably.

Observations are extracted for the month of February 2012. With respect to S-RH-B, we detect maximally one observation per hour if measured relative humidity exceeds 40 % within the 60 minutes time interval.[3] For each observation we compute the start time, end time, and the duration. Assuming that relative humidity drops after ending a shower, the end time corresponds to the time of measured maximum relative humidity. Start time corresponds to the time at which the first derivative of the time series exceeds 2 %. Duration is simply end time minus start time. Figure 1(a) shows relative humidity as measured by sensor S-RH-B on February 5, 2012. Visible are two observations for relative humidity above 40 %.

With respect to S-CO-L, we detect maximally one observation per hour if measured carbon monoxide concentration exceeds 5 ppm within the 60 minutes time interval. For each observation we compute the start time, end time, peak time, duration, and peak concentration. Figure 1(b) shows carbon monoxide as measured by sensor S-CO-L on February 5, 2012. Visible is one observation for carbon monoxide above 5 ppm.

To represent domain knowledge extracted from sensor data we extend the Semantic Sensor Network (SSN) ontology [19, 5]. (Note that the SSN ontology imports the DOLCE+DnS Ultralite ontology[4] (DUL). Hence we use both SSN and DUL terminology.) Specifically, our ontology imports the SSN ontology and asserts the two sensor types for carbon monoxide and relative humidity measurement as subclasses of SSN `SensingDevice`. The sensors S-CO-L and S-RH-B are modelled as individuals of these sensing device classes, respectively. In addition,

---

[2] http://www.epa.gov/iaq/co.html

[3] Note that thresholds are chosen experimentally based on data sampled from the database for individual sensors.

[4] http://www.loa.istc.cnr.it/ontologies/DUL.owl

(a) Relative humidity as measured by sensor S-RH-B on February 5, 2012. Visible are two observations for relative humidity above 40 %.

(b) Carbon monoxide as measured by sensor S-CO-L on February 5, 2012. Visible is one observation for carbon monoxide above 5 ppm.
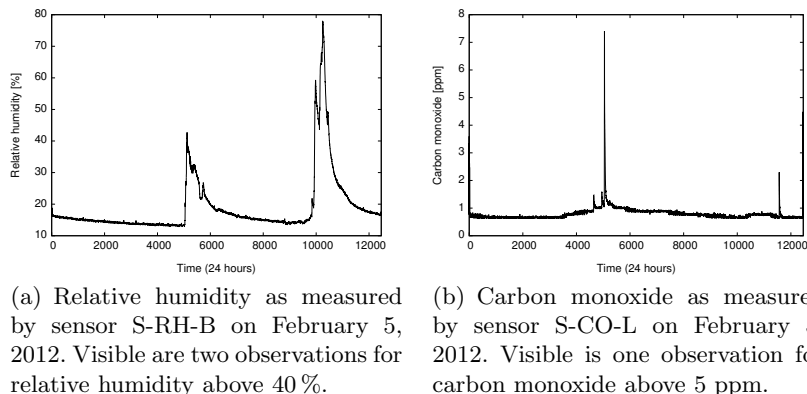
Fig. 1. Relative humidity and carbon monoxide as measured on February 5, 2012 by sensor S-RH-B and sensor S-CO-L, respectively.

we assert two subclasses of SSN `Observation` for carbon monoxide and relative humidity observations. Hence, an observation made by the relative humidity sensing device is modelled as individual of the relative humidity observation class.

In SSN, observations made by sensing devices relate to a so-called feature-of-interest. Thus, we assert our two domain-specific events of interest 'beyond average carbon monoxide' and 'person taking a shower' as subclasses of SSN `FeatureOfInterest`. An observation made by the carbon monoxide sensing device is, hence, for a 'beyond average carbon monoxide' feature-of-interest. A SSN `FeatureOfInterest` may have one or more related properties, each of which may relate to a data value. Hence, we assert the property classes for start time, end time, duration, carbon monoxide concentration, relative humidity, and temperature as subclasses of SSN `Property`.

We implemented the software to process sensor data, extract domain knowledge, and represent knowledge in Java. Knowledge is represented by means of Web Ontology Language (OWL) [18] and Resource Description Framework (RDF) [12] technologies. We used Protégé[5] (version 4.1) and Jena[6] (version 2.7.0) to manage the ontology.

3   Results

We evaluated our methodology and implementation for observations occurring during the month of February 2012. During this period, 59 observations of type relative humidity were made for the feature-of-interest 'person taking a shower' and 2 observations of type carbon monoxide were made for the feature-of-interest 'beyond average carbon monoxide'.

---

[5] http://protege.stanford.edu
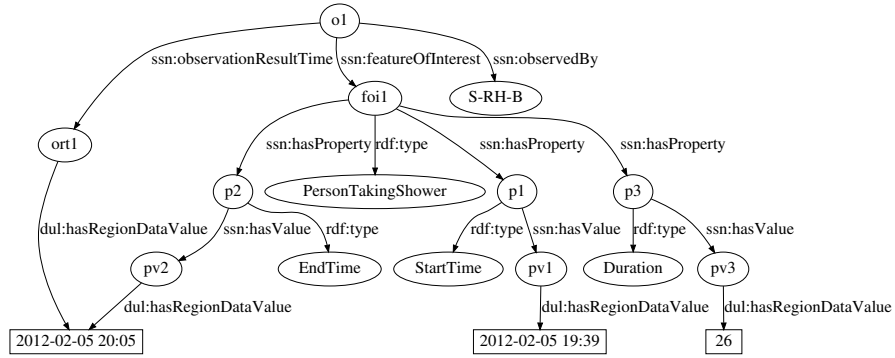[6] http://jena.apache.org

Fig. 2. Sub-graph describing the observation (`o1`) made by S-RH-B at 20:05 on February 05, 2012 for a person taking a shower of duration 26 minutes.

Observations of type relative humidity are observed by S-RH-B and relate via 'observation result time' to the time of the observation. Further, they relate to the 'person taking a shower' feature-of-interest. This feature-of-interest holds 3 property relations, namely for start time, end time, and duration. Figure 2 corresponds to a RDF sub-graph that represents knowledge for the taller peak in Figure 1(a). The observation was made at 2012-02-05 20:05 by S-RH-B for the observed event 'person taking a shower'. The feature-of-interest holds 3 property relations for start time 2012-02-05 19:39, end time 2012-02-05 20:05, and duration 26 minutes.

Observations of type carbon monoxide are all observed by S-CO-L. Here, the feature-of-interest is for 'beyond average carbon monoxide' and holds 5 property relations, namely for carbon monoxide concentration, start time, peak time, end time, and duration. Figure 3 corresponds to a RDF sub-graph that represents knowledge for the taller peak in Figure 1(b). Due to lack of space we show only some of the key properties of a carbon monoxide observation made by S-CO-L. Specifically, Figure 3 includes the 3 property relations for concentration, peak time, and duration related to the feature-of-interest. Note that the duration for a 'beyond average carbon monoxide' feature-of-interest is the time interval during which measured carbon monoxide concentration is beyond average, here above 5 ppm.

The RDF graphs shown in Figure 2 and Figure 3 for represented knowledge about observations visible in Figure 1(a) and Figure 1(b) are two examples used to demonstrate how knowledge extracted from sensor network data is automatically represented. Naturally, but not discussed here, similar representations are generated for all 61 observations made during February 2012.
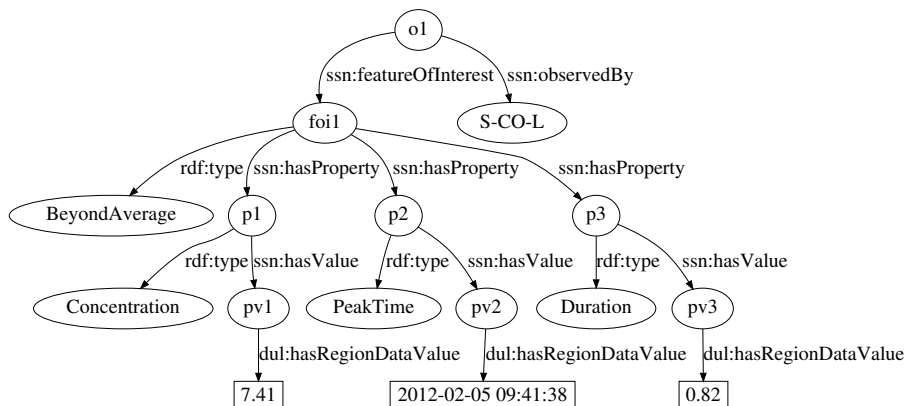
Fig. 3. Sub-graph describing the observation (`o1`) made by S-CO-L at 09:41:38 on February 05, 2012 for carbon monoxide concentration beyond average with peak concentration 7.41 ppm and duration approximately 49 seconds.

4   Discussion

Given a sensor network and data for the monitoring of (mostly) indoor air (quality) and consumption in energy and water of a residential building, we have demonstrated the methods used to automatically represent domain knowledge acquired from sensor network data. For the relative humidity sensing device S-RH-B we acquired knowledge about observations with feature-of-interest 'person taking a shower'. For the carbon monoxide sensing device S-CO-L we acquired knowledge about observations with feature-of-interest 'beyond average carbon monoxide'. Hence, we have demonstrated how knowledge about observations made by sensors can automatically be represented in ontology. We have evaluated our implementation for the month of February 2012 and represented a total of 61 observations. In particular we presented and discussed in details the representation of a relative humidity observation and a carbon monoxide observation made on February 5, 2012.

Representing knowledge using de-facto standards such as OWL and RDF comes with a number of benefits, in particular the standard set of reasoning services (e.g. consistency checking and classification) typically implemented in reasoning engines, rule-based inference, as well as a wide array of storage systems, query engines and query languages (e.g. SPARQL [13]), and modelling tools that are readily available.

For a sensor network the size and diversity presented here it may be expected that knowledge about a multitude of events of interest could be acquired. In this experiment we limited our evaluation to univariate analysis, i.e. knowledge is acquired for individual time series. We are currently investigating multivariate analysis. Moreover, we need to clarify the exact interpretation of the data for consumption sensors before we can take those into account. However, we have

explored the data for other events of interest, for instance significant drops in room temperature, which may signal prolonged opening of windows, or elevated readings in 2.5 µm particulate matter, which may be a health concern. With respect to the former, we found that room temperature is rather constant. For example, during February 2012 the data for the temperature sensor in one of the bedrooms has a mean and standard deviation of 23.3 ± 0.4 °C and a minimum value of 20.3 °C. With respect to particulate matter, the sensor was installed only in April 2012. Hence, there was little data available during the evaluation phase for this paper, and none for February 2012.

In the approach to representation of knowledge acquired from sensor data presented here, the sensor data processing and knowledge extraction steps are executed before the knowledge representation step. In fact, except for the values of certain properties, no sensor measurement data is explicitly represented in the ontology. This approach is akin to the one suggested by Liu and Zhao [11] and Whitehouse et al. [22] for a system that can be queried for high-level events without requiring handling of raw magnetometer or infrared break beam sensor data. For instance, the authors represent the speed of a vehicle, computed from sensor data, without representing, to our understanding, the sensor data. It also relates to the architecture described by Gaglio et al. [9].

However, the approach is in contrast to the practice of representing sensor data explicitly by annotating it with semantic metadata [14]. The semantic annotation of sensor data, and its explicit representation, has received considerable attention. Wanner et al. [20] present an environmental information system in which environmental data, e.g. temperature measurements for a city, are incorporated into a knowledge base [1], using OWL as ontology language. Barnaghi et al. [2] propose a semantic model for (heterogeneous) sensor data representation, and discuss the representation of sensor data using XML and OWL. Wei and Barnaghi [21] annotate sensor data with semantic metadata and relate sensor data with data of knowledge bases available online, such as DBpedia [4], using the linked data principle [3]. Furthermore, eight out of eleven sensor ontologies reviewed by Compton et al. [6] support the representation of measurement data.

An advantage of representing sensor data explicitly is that in doing so systems can leverage ontology and rule-based reasoning to infer new knowledge from sensor data. This advantage is well-documented, for instance by Sheth et al. [14] to infer a 'blizzard condition'; Henson et al. [10] to infer 'high winds' observations; Stocker et al. [17] to infer the nutrient status of lakes; Wei and Barnaghi [21] to infer the approximate temperature for a city neighbouring a city for which the temperature is known. A second advantage of this approach is its general applicability. In fact, given a domain ontology, the fundamental procedure is to enrich sensor data with semantic metadata, define domain rules, and use off-the-shelf reasoners to infer new knowledge.

However, representing sensor data explicitly may not always be practicable, or even feasible. For the Los Angeles Department of Water and Power, Simmhan et al. [15] highlight how information systems for smart grid networks may be unable to "centrally store and manage all the data that is collected" [15]. Stocker

et al. [16] discuss the representation of knowledge about vehicles classified using three accelerometer sensors sampling road-pavement vibration at 2 kHz. Storing 6000 measurement values per second annotated with semantic metadata in an OWL knowledge base poses considerable demands on resources. Barnaghi et al. [2] underscore the significant increase in volume of data incurred by semantic annotation of sensor data and highlight how this can be of concern especially to power-constrained sensor networks. For some applications this problem may be eased by aggregating sensor data, e.g. by computing the root mean square value for the measurement data of a certain time interval. However, the loss of information incurred by aggregation may impede knowledge acquisition. For instance, the neural network based vehicle classification discussed by Stocker et al. [16] relies on the high-frequency raw sensor data.

Furthermore, knowledge entailed by an ontology is specified by the expressivity of the ontology language and rule language used. If a knowledge acquisition task on sensor data is within the expressivity of the languages used and, hence, can be formally represented by means of said languages, representing sensor data in ontology is very attractive as it allows us to leverage the powerful reasoning capabilities of inference engines. However, not all knowledge acquisition tasks on sensor data are within the expressivity of state-of-the-art ontology and rule languages. For instance, the knowledge acquisition task discussed by Stocker et al. [16] for the classification of vehicles using road-pavement vibration data and machine learning cannot be formalized in an OWL ontology (the language lacks of a notion for, e.g., Fourier transform). For the use case presented here, it is unclear whether the straightforward computations on time series could be formalized and whether doing so would be computationally efficient. We argue that these examples are hardly isolated cases. For instance, for a hypothetical use case in meteorology, the representation of knowledge for the cloud type using images captured by cameras or satellites relies on image analysis techniques. Naturally, we may implement such knowledge acquisition tasks on top of represented sensor data, i.e. on top of an ontology layer, and feedback the acquired knowledge. This would, however, incur a similar cost and loss of generality as does the approach discussed here – which is arguably the most important limitation, and a concern we aim to address in future work.

Given the acquired and represented knowledge discussed here, we may envision a lightweight mobile phone client application that interacts with the ontology layer to provide users with information. For instance, a mobile phone client application may signal to occupants of a residential building 'unhealthy' exposure to carbon monoxide or particulate matter. By interacting with the ontology layer, client applications can take advantage of high-level domain terminology without having to deal with raw measurement data [11, 22].

5   Conclusions

We have illustrated the automated representation of knowledge acquired from sensor network data for a case study on residential building monitoring of vari-

ables such as relative humidity and carbon monoxide. Specifically, we have acquired and represented knowledge about observations made by a relative humidity sensor for the event of interest of a person taking a shower, and knowledge about observations made by a carbon monoxide sensor for concentrations beyond average. Representing knowledge for carbon monoxide events is of particular interest to quality of life services, given that such knowledge may signal 'unhealthy' exposure to carbon monoxide.

Based on our results and discussion we argue that modern computational techniques allow us to automatically acquire and represent knowledge from sensor network data and, hence, make sense of what often are high-volume numerical time series data. However, different problem classes may require different approaches. Specifically, use cases with low-frequency sensing and knowledge acquisition tasks on sensor data that can be formalized may leverage ontology and rule-based reasoning by representing sensor data explicitly. Use cases with high-frequency sensing and/or knowledge acquisition tasks on sensor data that cannot be formalized may require an approach more akin the one presented here. Information systems may implement a hybrid approach that combines the two methodologies.

References

1. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.: The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, 2nd Edition (2007)
2. Barnaghi, P., Meissner, S., Presser, M., Moessner, K.: Sense and Sens'ability: Semantic Data Modelling for Sensor Networks. Proceedings of the ICT Mobile Summit 2009 (2009)
3. Berners-Lee, T.: Linked Data–Design Issues (2006), `http://www.w3.org/DesignIssues/LinkedData.html`
4. Bizer, C., Auer, S., Kobilarov, G., Lehmann, J., Cyganiak, R.: DBpedia–Querying Wikipedia like a database. In: Developers track presentation at the 16th international conference on World Wide Web, WWW. pp. 8–12 (2007)
5. Compton, M.: What Now and Where Next for the W3C Semantic Sensor Networks Incubator Group Sensor Ontology. In: Proceedings of the 4th International Workshop on Semantic Sensor Networks 2011 (SSN11). pp. 1–8. CEUR-WS (2011)
6. Compton, M., Henson, C., Neuhaus, H., Lefort, L., Sheth, A.: A Survey of the Semantic Specification of Sensors. In: 2nd International Workshop on Semantic Sensor Networks, at 8th International Semantic Web Conference (2009)
7. Conroy, K., May, G., Roantree, M., Warrington, G., Cullen, S.J., McGoldrick, A.: Knowledge acquisition from sensor data in an equine environment. In: Proceedings

of the 13th international conference on data warehousing and knowledge discovery. pp. 432–444. DaWaK'11, Springer-Verlag, Berlin, Heidelberg (2011)

8. Conroy, K., May, G.C., Roantree, M., Warrington, G.: Expanding sensor networks to automate knowledge acquisition. In: Proceedings of the 28th British national conference on Advances in databases. pp. 97–107. BNCOD'11, Springer-Verlag, Berlin, Heidelberg (2011)

9. Gaglio, S., Gatani, L., Lo Re, G., Ortolani, M.: Understanding the Environment Through Wireless Sensor Networks. In: Proceedings of the 10th Congress of the Italian Association for Artificial Intelligence on AI*IA 2007: Artificial Intelligence and Human-Oriented Computing. pp. 72–83. AI*IA '07, Springer-Verlag, Berlin, Heidelberg (2007)

10. Henson, C.A., Pschorr, J.K., Sheth, A.P., Thirunarayan, K.: SemSOS: Semantic Sensor Observation Service. In: Proceedings of the 2009 International Symposium on Collaborative Technologies and Systems (CTS 2009). Baltimore, MD (May 2009)

11. Liu, J., Zhao, F.: Towards semantic services for sensor-rich information systems. In: Broadband Networks, 2005. BroadNets 2005. 2nd International Conference on. vol. 2, pp. 967–974 (October 2005)

12. Manola, F., Miller, E., McBride, B.: RDF Primer. W3C Recommendation, W3C (February 2004)

13. Prud'hommeaux, E., Seaborne, A.: SPARQL Query Language for RDF. Tech. Rep. W3C Recommendation, W3C (2008)

14. Sheth, A., Henson, C., Sahoo, S.: Semantic Sensor Web. Internet Computing, IEEE 12(4), 78–83 (July-August 2008)

15. Simmhan, Y., Aman, S., Cao, B., Giakkoupis, M., Kumbhare, A., Zhou, Q., Paul, D., Fern, C., Sharma, A., Prasanna, V.: An Informatics Approach to Demand Response Optimization in Smart Grids. Tech. rep., University of Southern California (2011)

16. Stocker, M., Rönkkö, M., Kolehmainen, M.: Making Sense of Sensor Data Using Ontology: A Discussion for Road Vehicle Classification. In: Seppelt, R., Voinov, A., Lange, S., Bankamp, D. (eds.) 2012 International Congress on Environmental Modelling and Software. pp. 2387–2394. iEMSs, Leipzig, Germany (July 2012)

17. Stocker, M., Rönkkö, M., Villa, F., Kolehmainen, M.: The Relevance of Measurement Data in Environmental Ontology Learning. In: Environmental Software Systems. Frameworks of eEnvironment, IFIP Advances in Information and Communication Technology, vol. 359, pp. 445–453. Springer Boston (2011)

18. W3C OWL Working Group: OWL 2 Web Ontology Language. Tech. Rep. W3C Recommendation, W3C (2009)

19. W3C Semantic Sensor Network Incubator Group: Semantic Sensor Network Ontology. Tech. rep., W3C (2009)

20. Wanner, L., Vrochidis, S., Tonelli, S., Moßgraber, J., Bosch, H., Karppinen, A., Myllynen, M., Rospocher, M., Bouayad-Agha, N., Bügel, U., Casamayor, G., Ertl, T., Kompatsiaris, I., Koskentalo, T., Mille, S., Moumtzidou, A., Pianta, E., Saggion, H., Serafini, L., Tarvainen, V.: Building an Environmental Information System for Personalized Content Delivery. In: Hrebícek, J., Schimak, G., Denzer, R. (eds.) Environmental Software Systems. Frameworks of eEnvironment, IFIP Advances in Information and Communication Technology, vol. 359, pp. 169–176. Springer Boston (2011)

21. Wei, W., Barnaghi, P.: Semantic Annotation and Reasoning for Sensor Data. Smart Sensing and Context pp. 66–76 (2009)

22. Whitehouse, K., Zhao, F., Liu, J.: Semantic Streams: A Framework for Composable Semantic Interpretation of Sensor Data. In: Römer, K., Karl, H., Mattern, F. (eds.) Wireless Sensor Networks, LNCS, vol. 3868, pp. 5–20. Springer Berlin / Heidelberg (2006)