

Semantic Annotation and Reasoning for Sensor Data

Wang Wei¹ and Payam Barnaghi²

¹ School of Computer Science, University of Nottingham (Malaysia Campus)
Jalan Broga, 43500, Semenyih, Selangor Darul Ehsan, Malaysia
eyx6ww@nottingham.edu.my

² Centre for Communication Systems Research, University of Surrey
Guildford, Surrey, GU2 7XH, United Kingdom
p.barnaghi@surrey.ac.uk

Abstract. Developments in (wireless) sensor and actuator networks and the capabilities to manufacture low cost and energy efficient networked embedded devices have lead to considerable interest in adding real world sense to the Internet and the Web. Recent work has raised the idea towards combining the Internet of Things (i.e. real world resources) with semantic Web technologies to design future service and applications for the Web. In this paper we focus on the current developments and discussions on designing Semantic Sensor Web, particularly, we advocate the idea of semantic annotation with the existing authoritative data published on the semantic Web. Through illustrative examples, we demonstrate how rule-based reasoning can be performed over the sensor observation and measurement data and linked data to derive additional or approximate knowledge. Furthermore, we discuss the association between sensor data, the semantic Web, and the social Web which enable construction of context-aware applications and services, and contribute to construction of a networked knowledge framework.

Keywords: Sensor data modelling, Semantic annotation, Linked data, Reasoning, Semantic Web.

1 Introduction

The Internet and the Web are now the most influential communication medium through the world. With the infrastructure of the current Web, most of the information exchange is related to human-created and logical data such as text, image, audio, and video. The current Internet has been less concerned on providing a global solution to sense the physical world incidents, which could facilitate more natural and unobtrusive human-machine interaction. To this end, this vision is closely related to the grand aim of Mark Weiser's "Ubiquitous Computing" [1] which stated that "the most profound technologies are those that disappear, they weave themselves into the fabric of everyday life until they are indistinguishable from it", as well as the recent proposal of "Internet of Things" [2].

Recent developments in sensor technology has made manufacturing low-cost while energy-efficient hardware for sensing devices possible. It in turn leads to a potential interest in adding real world sense to the Internet and various applications and services on the Web. Deployment of (wireless) sensor and actuator networks makes it possible to observe and measure physical phenomena and the obtained information can be further processed to be used in different services and applications.

Universal access to sensor observation and measurement data provides a platform for a wide range of applications in different domains such as geographical information systems, health care, and smart homes. Over the past few years, considerable research effort has been devoted into developing large scale sensor networks such as SENSEI project¹ and SensorWeb², and also designing industrial standards for sensor data description such as the Sensor Model Language (SensorML) proposed by the Open Geospatial Consortium (OGC)³, the Sensor Web Enablement (SWE) activity [3]. The forms of sensor data representation are mostly based on XML data, which has significant limitations in supporting semantic interoperability and linking the described resources to the existing knowledge.

To realise the vision of ambient intelligence, i.e. creating a global (wireless) sensor network and service environment, data collected from heterogeneous wireless sensor and actuator networks have to be integrated and made available to different services and applications. The observation has recently encouraged development of sensor and actuator data representation using semantic Web [4] technologies, for example, the Semantic Sensor Web [5], OntoSensor [6], the SensorData Ontology proposed in [7], and other research works on creating service layers such as [8,9].

In this paper, we continue the discussion on adding semantic Web technologies into the sensor network research. Given the fact that sensor and actuator data representation models using semantic Web technologies have been relatively well developed, we particularly focus on issues related to semantic annotation using domain ontologies based on the “linked data” principle⁴. We discuss associations between the emerging data from sensor networks, the semantic Web and Social Web [10], and describe how semantic reasoning can be further exploited to help discover new knowledge from the annotated sensor data. This will support developments of low-level (i.e. sensor observation, discovery, and retrieval) and high-level services (i.e. service planning and recommendation) which enable construction of a networked knowledge platform [11].

2 Semantic Modelling of Sensor Data

Providing universal descriptions and high-level interfaces for sensor and actuator data significantly ease the tasks for application developers without forcing

¹ <http://www.ict-sensei.org/>

² <http://research.microsoft.com/en-us/projects/senseweb/>

³ <http://www.opengeospatial.org/standards/sensorml/>

⁴ <http://www.w3.org/DesignIssues/LinkedData.html>

them to be involved in complex issues related to heterogeneity of the underlying technologies in different platforms. Research in the semantic Web community in the past decade has developed well established standards, i.e. formal knowledge representation framework (i.e. RDF⁵, RDFS⁶, and OWL⁷) and ontologies. These standard representations are an ideal choice for modelling various sensor data (sensor node and interface descriptions, and also observation and measurement data collected from sensors) at different levels to support interoperability. More importantly, enriched descriptions can be easily integrated with sensor data in forms of semantic annotation, and logical reasoning can be performed over the sensor data to support advanced query and retrieval tasks.

There have been a number of works focusing on the development of representation models for sensor data using ontologies [6,9]. OntoSensor [6] constructs an ontology-based descriptive specification model for sensors by excerpting parts of SensorML descriptions and extending the IEEE Suggested Upper Merged Ontology (SUMO)⁸. However, it does not provide a descriptive model for observation and measurement data. The work presented in [9] proposes an ontology-based model for service oriented sensor data and networks, but how to represent and interpret complex sensor data is not specified. Eid *et al.* [8] propose a universal ontology which includes three sub-ontologies, i.e. extension plug-in ontology, sensor data ontology, and sensor hierarchy ontology. Similarly, the model does not provide details of sensor data specification and relationships between complex sensor data. The SensorData Ontology developed in [7] is built based on “Observations & Measurements” and SensorML specifications defined by the Sensor Web Enablement (SWE) [3].

Sheth *et al* take the research one step further and bring forward the concept of “Semantic Sensor Web” [5]. The idea behind is that with the semantic representation of the sensor observation and measurement model, one could add semantic annotations in terms of time, location, and thematic data into the actual sensor data to facilitate advanced query and reasoning. Collectively, these annotations represent the context data collected from the spatial, temporal, and thematic dimensions. Some application scenarios with reasoning over the semantically annotated sensor data with rules are described in [5,12]. The sensor data is annotated with concepts from the OWL-time domain ontology [12], as such, videos can be retrieved by using (semantic) temporal concepts such as “within”, “contains”, and “overlaps” when querying with an interval of time. The Semantic Sensor Web also promotes the creation of the so-called “event Web” [13], and supports travelling the semantic Web through space, time, and theme [14].

In the following we focus on the idea of Semantic Sensor Web by extending the discussion of semantic annotation using concepts taken from various domain ontologies. In particular, we propose using “linked data” principle to connect the sensor data to existing knowledge represented in different ontologies. We

⁵ <http://www.w3.org/RDF/>

⁶ <http://www.w3.org/TR/rdf-schema/>

⁷ <http://www.w3.org/TR/owl-features/>

⁸ <http://www.ontologyportal.org/>

demonstrate the idea with examples in which semantic reasoning over the sensor data and the existing “linked data” can be performed to deduce new knowledge. Although the sensor data is intended to be primarily utilised by machines, sensor node descriptions and the metadata about the observations and measurement and surrounding environment will be potentially valuable for sensors discovery and their trustworthiness evaluation. With the “collective intelligence”, (wireless) sensor and actuator networks can also be linked to the “Social Web” [10], harmonising the Web of data and facilitating creation of real world applications on top of the “Internet of Things” [2].

3 Annotation of Sensor Data Using Linked Data

Utilising metadata and semantic annotations to describe sensor/actuator and in general real world and logical world resources (i.e. “things”) in a scalable and heterogeneous platform will enable different communities to exploit the emerging data and exchange information and knowledge in a collaborative environment. Semantic annotations, no matter inferred from the sensor data or provided by users, represent the context data which can be utilised to create context-aware applications.

Previous works in computer science have discussed the representation and modelling of context. Schmidt *et al* emphasises the importance of roles of users in representing context information [15]. They define context as information of users (e.g. knowledge of habits, emotional state, bio-physiological conditions), user’s social environment (e.g. co-location of others, social interaction, group dynamics), and user’s tasks (e.g. spontaneous activity, engaged tasks, general goals). In service oriented computing context-awareness can refer to including functionalities that make services more dynamic, i.e. to be able to adapt and respond to environment and situation changes [16]. On the other hand, context can be also defined as attributes related to situation of a resource related to physical environment such as location (e.g. absolute position, relative position, co-location), infrastructure (e.g. surrounding resources for computation, communication, task performance), and physical conditions (e.g. noise, light, pressure). An example is Sheth *et al*’s work on Semantic Sensor Web in which the context is modelled as spatial, temporal, and thematic data [5]. In their work, sensor data is annotated with concepts of time, location, and also other domain ontologies. However, the source of these domain ontologies are not specified and the relationship between the Semantic Sensor Web and the existing knowledge is not discussed in detail.

We envision semantic annotations with the linked data principle which is capable of connecting the emerging data from (wireless) sensor and actuator networks to the existing semantic Web. The idea behind the linked data is that the value and usefulness of data increases as it is interlinked with large amount of other data. On one hand, the connection enables qualitative annotations to promote interoperability, avoiding creating repetitive data. Data already published on the semantic Web and widely used by communities can be reused for the sensor data annotation purpose. When annotating sensor observation and

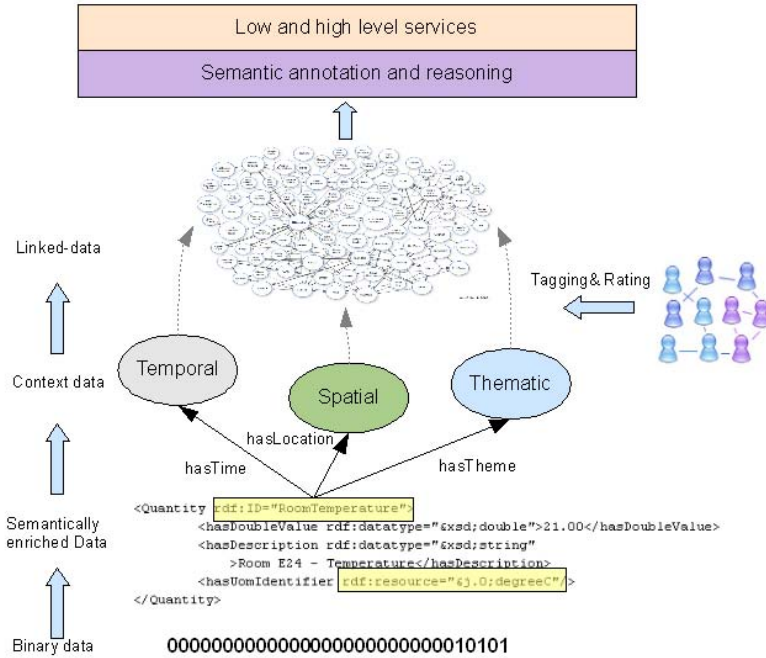


Fig. 1. Integrating sensor data with the semantic Web and linked data

measurement data, or inferred concepts (i.e. a “phenomenon” in Semantic Sensor Web [5]), instead of creating new concepts or instances, one could annotate the sensor data with those concepts on the semantic Web by creating RDF links, i.e. data published by authoritative sources using the linked data principle such as DBpedia⁹. On the other hand, making connections to the linked data also binds the sensor data to potentially endless data or knowledge (provided by authoritative sources). It enables reasoning over the sensor data and the linked data to provide advanced sensor data query and retrieval functions. Figure 1 demonstrates the proposed framework for integrating the sensor network with the semantic Web (and possibly the social Web).

Through the semantic annotation, the real world resources and sensor data can be connected to the existing semantic Web. This associates abundant data and knowledge to the original and inferred sensor data, improving the ways that sensor data is utilised; e.g. the data can be exploited to design enhanced services through the use of reasoning and rules (see Section 4).

4 Reasoning Sensor Data with Context Data

Logical reasoning is a powerful mechanism to derive new and implicit knowledge from semantically annotated sensor data, and to answer complex user queries.

⁹ <http://dbpedia.org/>

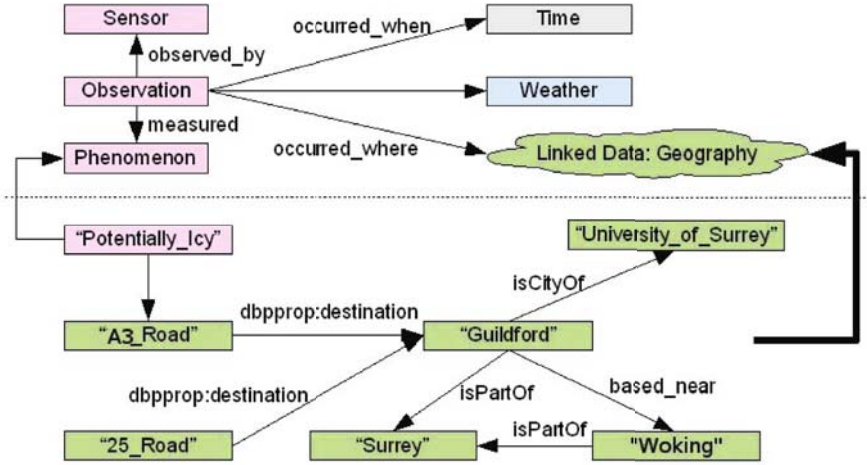


Fig. 2. Associations between sensor data and high-level concepts

The work in Semantic Sensor Web demonstrates some examples (e.g. “Potentially Icy”) of using rule-based reasoning to deduce new ontological assertions from known instances [5]. In one of Sheth *et al*’s example, sensor data is first annotated with temporal data extracted from video with respect to a Time domain ontology. Then they present how videos can be retrieved by using semantic temporal concepts such as “within”, “contains”, and “overlaps”. Here we provide an example as an extension to these examples and show how the semantic annotation with the linked data contribute to answering sensor data queries. In particular, we annotate the observation and measurement data by linking it to geographic data published by DBpedia.

As shown in Figure 2, suppose that a number of sensors installed at “A3_Road” provide information on temperature and precipitation, and using rule-based reasoning one infers that the road condition is “Potentially Icy” which is an instance of the class “Phenomenon” defined in the Semantic Sensor Web ontology. The phenomenon is then linked to an instance called “A3_Road” published by the DBpedia through the property “occurred_where”. In DBpedia, the instance “A3_Road” is related to many of the other objects, such as it has a destination of “Guildford”, which is instance of “City” located in the “Area” of “Surrey” and based near to another instance called “Woking”. The linked data enables advanced queries over sensor data, for example, a user might issue a query for all the roads with the condition “Potentially Icy” located in the area of “Surrey”. Using the following query one is able to obtain desired results (the query is represented in SPARQL¹⁰ form):

¹⁰ <http://www.w3.org/TR/rdf-sparql-query/>

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dc <http://purl.org/dc/terms/>
PREFIX dbpprop: <http://dbpedia.org/property/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
SELECT ?road
WHERE {
  ?road rdf:type dbpedia-owl:Road.
  ?road dbpprop:destinations ?city.
  ?city rdf:type dbpedia-owl:City.
  ?city dc:isPartOf dbpedia:Surrey.
  ?road ssw:described Potentially_Icy.
}

```

Without the domain ontology or the linked data it is impossible for applications to answer this kind of queries automatically. Semantic annotation with the linked data reuses the existing information on the semantic Web and eliminates the needs of creating repetitive and redundant data. Making connections from the real world resources to the existing semantic Web ontologies also naturally promotes the idea of traversal of the semantic Web through the dimensions of time, space, and theme [14].

Rule-based reasoning plays important role in sensor data processing as it is one of the most promising means of deriving new and implicit knowledge. Sheth *et al* show that using rules complex queries (e.g. road conditions that are “Potentially Icy” or “blizzard”) over simple weather readings can be executed. Rule-based reasoning is also effective to deal with missing data or uncertainty. In some situations, precise answers might not exist due to incomplete information or unavailability of data. However, it is meaningful to provide approximate or indirect answers in responding to user queries. Let’s reconsider the previous example as shown in Figure 2: assume that a user is searching the weather condition in “Woking”, however, due to the unavailability of sensor observation the query cannot be answered directly. With a domain ontology, e.g. data published in DBpedia, we know that “Guildford” is a town based near to “Woking” area and their distance is less than 20 KM. Moreover, the weather condition in Guildford has been observed, e.g. temperature is 25 degrees Fahrenheit. Using the following rule one could obtain an approximate temperature range for “Woking”.

Rule: Approximate Temperature:

```

Temperature(?temp_high_Woking)&
Temperature(?temp_low_Woking)&
Temperature(?temp_NearCity)&
City(?city)&
hasTemperature(?city, ?temp_NearCity)&
temp_Value(?temp_NearCity, ?temp_NearCity_Value)&

```

```

City(Woking)&
hasTemperature("Woking", ?temp_high_Woking)&
hasTemperature("Woking", ?temp_low_Woking)&
basedNear(?city, "Woking")&
distance_Value(?dist, ?dist_Value)&
lessThanOrEqual(?dist_Value, 20)&
sourceArea(?dist, ?city)&
destinationArea(?dist, Woking)&
—>temp_Value(?temp_high_Woking, ?temp_NearCity_Value + 2)
or temp_Value(?temp_low_Woking, ?temp_NearCity_Value - 2)

```

In the rule, *distance_Value* is a function for calculating the distance between a source area and destination area. The function could be realised using Web services: it takes two cities or coordinates of the cities and returns the geographical distance. Providing approximate answers to queries, especially for service composition in which intermediary results are essential is more desirable when in typical scenarios it return failure or terminates the service execution prematurely.

5 Connecting Real World Resources, Semantic Web, and Social Web

One of the primary visions of providing enriched data for sensor data is to create an infrastructure for networked knowledge [11] that is aware of physical world incidents and enables construction of new services that remove the boundary between the logical and physical world. Today, (wireless) sensor and actuator networks deployed in numerous locations are able to produce multitude of observation and measurements, from simple phenomena to complex events and situations [5]. Semantic Web technologies are ideal choice to represent, manage, store, analyse, and reason over the observed sensor data, enhancing interoperability among heterogeneous networks to build Semantic Sensor Web, and to bring usefulness of the sensor data to its full potential.

We also envision the potential contribution of the social Web [10] to the Internet of Things. Social Web sites such as blogs and wikis allow their users to generate annotations effortlessly through a process of social tagging. The same techniques can be employed in the sensor networks to exploit the “collaborative intelligence”. Tags and folksonomies contributed by numerous online users can be valuable information to annotate sensor data and associate them to high-level concepts. In the future, quality and trustworthiness of the sensor data will also have direct impact on real world services that make use of them, for example, user-generated tags and quality ratings can be important factors for evaluation of sensor data.

The main issue in this respect is how to orchestrate annotation process and ontology associations to link the resources to high-level concepts and existing knowledge. Another aspect is using rich-descriptions with automated ontology

construction methods to associate the tags to high-level concepts. The annotation of the resources with domain knowledge represented in ontologies facilitates drawing logical reasoning based on the data. This will result in a framework which can “learn” to discover relations between the data and high-level concepts and can “gather” knowledge to perform “adaptable” decision making in dynamic environments. In a recent work, we concentrated on creating ontologies from text using probabilistic methods [17]. In the ongoing work, we also investigate similar approaches to enable (semi-) automated link association to support providing metadata and linked data for the resources on the Internet of Things.

6 Real World Services

The ultimate goal of the Internet of Things is to integrate the physical world into the digital world to support more natural human-computer interaction and to build more intelligent application and services. Real world services utilising sensor data can be generally categorised into two groups: low-level sensor data services (e.g. those provide data integration and semantic reasoning), and high-level services (e.g. context-aware applications, planning, and recommendation).

Typically, low-level services integrate various information from different resources and provide intermediary results or inputs to high-level services and applications (e.g. decision-support systems, planning, and recommendation services). The examples presented earlier in this paper belong to the scope of this service category. In particular, we extend the previous work on Semantic Sensor Web and demonstrate how sensor data can be annotated and enriched using concepts in the spatial, temporal, and thematic dimensions with linked data; i.e. data already exists on the semantic Web. Through rule-based reasoning and using semantically annotated sensor data, additional knowledge and meaningful context data can be derived and used by other applications and services.

High-level services seamlessly integrate the digital world and physical world happenings to create context-aware applications and to support various decision making processes. With the situational knowledge which represented by various sensor data and information gathered from outputs of low-level services, high-level services are able to perceive environment change and autonomously adapt to the new situation. Moreover, they enable personalisation and customisation of contents and services for both human (e.g. planning trips and service recommendation) and machines (e.g. implementation of innovative business applications and semantic Web service composition).

7 Conclusion

Integration of the physical world and the digital world has a wide range of applications such as monitoring, manufacturing, health, tracking and planning. This vision brings forward the issues related to sensor data representation, annotation, sharing, management, and reasoning. Utilising semantic Web technologies, metadata and enriched metadata in particular, to describe sensor data, and more generally real world and logical world resources in heterogeneous platforms enables

different communities to exchange information and knowledge in a collaborative environment. More importantly, it facilitates development of context-aware applications to support effective human-machine and machine-machine interaction, moving forward to the grand aims of ubiquitous computing and the Internet of Future.

In this paper we report some preliminary results of an ongoing research on sensor data annotation using the linked data and automated sensor observation using rule-based reasoning. Examples are provided to demonstrate the feasibility of using reasoning mechanisms to infer additional knowledge from semantically annotated sensor data, even in situations with incomplete information. We also discuss the relationships between resources in the Internet of Things, the semantic Web, and the social Web which are usually seen as separated entities. By annotating sensor observation and measurement data using concepts already published on the semantic Web, and user-generated contents such as tags and ratings we envision a fully connected Internet of Things in the future and the great potential to create real world context-aware applications and services. The underlying services will contribute to collect, represent, process, and reason real world and logical world data in a networked knowledge framework.

References

1. Weiser, M.: The computer for the 21st century. *Scientific American* 265(3), 66–75 (1991)
2. Union, I.C.: The internet of things, iITU Internet Reports, Executive Summary (2005)
3. Botts, M., Percivall, G., Reed, C., Davidson, J.: Sensor web enablement: Overview and high level architecture, pp. 175–190 (2008)
4. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Scientific American* 284(5) (2001)
5. Sheth, A.P., Henson, C., Sahoo, S.S.: Semantic sensor web. *IEEE Internet Computing* 12(4), 78–83 (2008)
6. Russomanno, D.J., Kothari, C., Thomas, O.: Sensor ontologies: from shallow to deep models. In: *Proceedings of the Thirty-Seventh southeastern Symposium on System Theory, SSST 2005, Los Alamitos, CA, USA*, pp. 107–112 (2005)
7. Barnaghi, P.M., Meissner, S., Presser, M., Moessner, K.: Sense and sens'ability: Semantic data modelling for sensor networks. In: *Proceedings of the ICT Mobile Summit 2009 (June 2009)* (to appear)
8. Eid, M., Liscano, R., El Saddik, A.: A universal ontology for sensor networks data. In: *Proceedings of the IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, CIMSA 2007*, pp. 59–62. IEEE Computer Society, Los Alamitos (2007)
9. Kim, J.-H., Kwon, H., Kim, D.-H., Kwak, H.-Y., Lee, S.-J.: Building a service-oriented ontology for wireless sensor networks. In: *Proceedings of the Seventh IEEE/ACIS International Conference on Computer and Information Science, ICIS 2008, Los Alamitos, CA, USA*, pp. 649–654 (2008)
10. Gruber, T.: Collective knowledge systems: Where the social web meets the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web* 6(1), 4–13 (2008)

11. Decker, S., Hauswirth, M.: Enabling networked knowledge. In: Klusch, M., Pěchouček, M., Polleres, A. (eds.) CIA 2008. LNCS (LNAI), vol. 5180, pp. 1–15. Springer, Heidelberg (2008)
12. Henson, C., Sheth, A., Jain, P., Rapoch, T.: Video on the semantic sensor web. In: W3C Video on the Web Workshop (2007)
13. Jain, R.: EventWeb: developing a human-centered computing system. *Computer* 41(2), 42–50 (2008)
14. Sheth, A., Perry, M.: Traveling the semantic web through space, time, and theme. *IEEE Internet Computing* 12(2), 81–86 (2008)
15. Schmidt, A., Beigl, M., Gellersen, H.-W.: There is more to context than location. *Computers & Graphics* 23(6), 893–901 (1999)
16. Vuković, M.: Context aware service composition. UCAM-CL-TR-700, University of Cambridge, Computer Laboratory, Tech. Rep. (October 2007)
17. Wang, W., Barnaghi, P.M., Bargiela, A.: Probabilistic topic models for learning terminological ontologies. *IEEE Transactions on Knowledge and Data Engineering* (to appear, 2009)