A Validation Tool for the W3C SSN Ontology Based Sensory Semantic Knowledge

Şefki Kolozali, Tarek Elsaleh, and Payam Barnaghi
e-mail: {s.kolozali, s.elsaleh, p.barnaghi}@surrey.ac.uk

Centre for Communication Systems Research (CCSR)
University of Surrey, Guildford, United Kingdom

Abstract

This paper describes an ontology validation tool that is designed for the W3C Semantic Sensor Networks Ontology (W3C SSN). The tool allows ontologies and linked-data descriptions to be validated against the concepts and properties used in the W3C SSN model. It generates validation reports and collects statistics regarding the most commonly used terms and concepts within the ontologies. An online version of the tool is available at: (http://iot.ee.surrey.ac.uk/SSNValidation). This tool can be used as a checking and validation service for new ontology developments in the IoT domain. It can also be used to give feedback to W3C SSN and other similar ontology developers regarding the most commonly used concepts and properties from the reference ontology and this information can be used to create core ontologies that have higher level interoperability across different systems and various application domains.

1 Introduction

With the advancement of the Internet of Things (IoT) vast amounts of devices will report data based on new applications and services in diverse application domains such as factory optimisation, transport, smart homes and smart cities. According to a report published by Cisco [2] it is predicted that in the next 5-10 years there will be around 50 billion Internet connected devices that will produce 20% of non-video traffic on the Internet. In order to process the IoT data, information management tools that allow effective organisation of data and knowledge representation tools which provide a frame of reference and enable the representation of abstract concepts in a machine-processable way are vital. While IoT devices provide information that are beneficial to diverse application domains, semantic web technologies allow to represent the domain knowledge as a way to handle various forms of heterogeneity and multi-modality by providing semantic models and interoperable data representation forms. Utilisation of semantic technologies for IoT advances interoperability among IoT resources, information models, data providers and consumers. In an effort to agree on a common consensus on a standardisation towards semantic descriptions of sensor networks an ontology has been developed by the W3C Semantic Sensor Network Incubator group (i.e. W3C SSN Ontology) [1].

Most of the current ontology development methods still require tremendous effort and subjective judgments from the ontology developers to acquire, maintain and validate the ontology. On the one hand, the ability to design and maintain ontologies requires expertise in both the domain of the application and the ontology language used for modelling. However, with their growing utilisation, not only the number of available ontologies increased considerably, but they are also becoming larger and more complex to manage. On the other hand, although there have been numerous work on publishing linked-data on the semantic web and ontology development methodologies in order to transform the art of building ontologies into an engineering activity; ontology and linked-data validation process is another crucial problem since developers need to tackle a wide range of difficulties when modelling and validation ontologies. These difficulties, such as the appearance of anomalies in ontologies or the technical quality of an ontology against a frame reference plays a key role in the ontology engineering projects.

The purpose of this study is to describe and examine the validation issues of sensory information in the IoT domain, and analyse various terminologies in order to provide assistance in the ontology development process. Thus, we propose the W3C SSN ontology validation service, which is based on Eyeball validator to check the RDF descriptions, to enable a user to validate an ontology or linked data on various common problems including use of undefined properties and classes, poorly formed namespaces, problematic prefixes, literal syntax validation and other optional heuristics. Moreover, enabling validation of Linked IoT data descriptions against W3C SSN ontology, we allow users to detect domain specific semantic and factual mistakes that may need an overview of a domain expert. This can help an effective integration of domain specific ontologies into linked-data models. We also collect and present information regarding the popularity of terms that are used by ontologies and the IoT data submitted for validation. This work is developed in the context of the CityPulse project[3][1]. The remainder of this paper is organised as follows. Section 2 describes the SSN validation web tool. Section 3 details the evaluations in which we investigate the most popular terms and modules that have been used within the W3C SSN ontology by examining various SSN related available ontologies. Finally, in the section 4, we note on further challenges of the semantic modelling for the IoT data and outline our future work.

2 The SSN Validation Tool

Fig. 1 presents the architecture of the SSN Validation Tool. The W3C SSN validation web application integrates the ontology and data validation functionalities in a web-based client-server architecture. The client runs in most popular

[1] This work is supported by the EU FP7 CityPulse Project under grant No.603095. http://www.ict-citypulse.eu

web browsers, and provides an easy to use interface. It allows the user to interact with the application and perform the following actions: *i*) enter an RDF document into text box or upload via a browse button to be validated against a reference ontology (i.e. in this case the W3C SSN ontology) *ii*) retrieve a list of evaluation results, *iii*) select and see namespaces of a term from the tag clouds as one would from a search engine and also visualise the most common terms and concepts used in the ontologies.

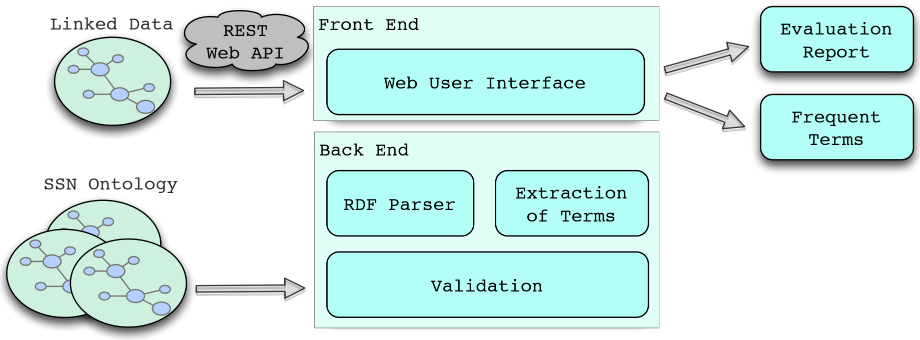

Fig. 1: The architecture of the SSN Validator web application

2.1 Front-End

The validator application is developed using Java EE, HTML, JSP technologies. It takes an RDF input or an ontology in order to produce a set of evaluation results. The web user interface consists of a single view where the user enters the RDF data into text box or uploads via a browse button. In response to user interaction, the server performs the core functionalities as shown in Fig 1. Concerning client-server communication, the validator follows the Representational State Transfer (REST) style web application design. In this architecture web pages form a virtual state machine, allowing a user to progress through the application by entering or uploading an RDF document which results in a transition to the next state of the application by transferring a representation of that state to the user.

2.2 Back-End

There are three main functionalities in the main system, namely RDF Parser, Extraction of Terms, and Validation. Initially, the RDF document describing the ontology or RDF document is parsed using the Jena API to obtain RDF triples as an input to the validation system. The server side of the SSN Validation web application builds on the Eyeball validator, which is a Java library for validating RDF documents. This is extended with modules to domain specific analysis for the W3C SSN ontology. It scans for errors from those available in the Eyeball list regarding RDF, Turtle and N3 syntax and some modelling suggestions are also generated.

The validation results are displayed by means of the web user interface showing a list of errors, and explanations regarding the ontology elements affected. The application also reports the recurrence of terms that are not present within the W3C SSN ontology. We have developed a server-side JavaScript code which interacts with the embedded Tag Cloud to display extracted terms that are not present in the W3C SSN ontology. The terms requested when the user starts the validation process, and returned using JavaScript Object Notation with evaluation results, which is presented at the same time with extracted terms as tag cloud. The term recurrence tags are displayed using an adapted version of the WP-Cumulus Flash-based Tag Cloud plug-in. This plug-in utilises XML data generated on the server side from the extracted terms. The light-weight client uses a combination of standard web technologies (HTML, CSS, JavaScript) and uses a Java library to dynamically load content from an object oriented database (i.e. DB4o).

3 The Ontology Validations

We collected a set of available ontologies and semantic description models that report using and/or extending the W3C SSN ontologies. The ontology dataset includes the Smart Product Ontology[2], the SPITFIRE project ontology[3], The IoT.est project service model[4], The SemSorGrid4Env project ontology[5], The OntoSensor ontology[6], The WSML Event Observation ontology[7], The WSML Environment Observation Ontology[8].

We evaluated these ontologies using our validation tool to find out the noise, inconsistency and syntax errors along with the similarity between the W3C SSN concepts and the terms and concepts used in these ontologies. It can be difficult for an ontology engineer to identify some errors and unexpected incorrect inferences in RDF. In the RDF data model, terms are typically named by Web URIs, which may be dereferenced to access more information such as vocabulary definitions about their meaning. However, while the principal notion behind the Semantic Web is to experience a machine-oriented world of Linked Data,

---

[2] http://www.w3.org/2005/Incubator/ssn/wiki/SSN_Smart_product
[3] http://spitfire-project.eu/ontology.owl
[4] http://personal.ee.surrey.ac.uk/Personal/P.Barnaghi/ontology/OWL-IoT-S.owl
[5] http://www.semsorgrid4env.eu/ontologies/CoastalDefences.owl
[6] https://www.memphis.edu/eece/cas/onto_sensor/OntoSensor.txt
[7] https://code.google.com/p/wsmls/source/browse/trunk/global/Observations/0.2/Observation.n3?spec=svn70&r=70
[8] https://code.google.com/p/wsmls/source/browse/trunk/global/Event-observation/0.2/EventObservation.n3?spec=svn207&r=207

ontology engineers should be very cautious to prevent broken links as well as make URIs dereferencable in order to empower automatic data access for Semantic Web applications. In accordance with the use of HTTP URIs, we found in our validations that in some instances (i.e. WSML event and WSML environment) different URIs were utilised rather than primary resources. As a result, it redirects the application user to their local directory instead of original locations such as SSN Ontology. Some of other errors were identified in IoT.est model and SemSorGrid4Env, in which multiple prefixes were defined (i.e. `owl` and `CoastalDefences`, respectively), in addition to utilisation of upper-case in namespaces of IoT.est model (i.e. `http://www.loa-cnr.it/ontologies/DUL.owl#`). It is interesting to see that while the latter is not actually wrong, it is accepted as unconventional and pointless for eyeball tool.

In parallel, sometimes properties or classes are used without any formal definition. In SPITFIRE, for instance, it has been defined that `:savedEnergyOf rdfs:domain :SavedEnergy`, even though `:SavedEnergy` is not defined as a class. Nevertheless, although such practice is not prohibited, such ad-hoc undefined classes and properties make automatic integration of data less efficient and prevent the possibility of making inferences through reasoning. An additional error that has been found for SPITFIRE via our validation tool is a syntax error where `ssn:subPropertyOf` was used instead of `rdfs:subPropertyOf`. Finally, we discovered in OntoSensor that there was clearly a misuse of functional property syntax along with a data property. It needs to be updated in line with OWL-2 guidelines using `FunctionalDataProperty` that describes properties for each individual allowing for at most one distinct literal.

Table 1: Summary of ontology evaluations against the W3C SSN ontology. Similar terms: s-terms; Dissimilar terms: d-terms; Similar properties: s-prop; Dissimilar properties: d-prop; Similar concepts: s-concept; Dissimilar Concepts: d-concept

	s-terms	d-terms	s-prop	d-prop	s-concept	d-concept
Smart Product	12	25	11	5	10	11
SPITFIRE	2	94	0	67	3	26
IoT.est model	0	12	0	10	0	2
SemSorGrid4Env	2	31	1	3	2	27
OntoSensor	0	331	0	226	0	105
WSML event	0	7	0	0	0	7
WSML environment	0	7	0	0	0	7

Table 1 summarises the results of similarity of terms and shows statistics using the W3C SSN ontology concepts in these ontologies. We found that the most frequently used SSN terms are as follows: `Property`, `Device`, `Observation`, `FeatureOfInterest`, and `ObservationValue`. Based on the validation results, we also created a Tag Cloud that shows the most common concepts that are used in the validated ontologies. We also checked linked ontologies and other common description models that can be used in the form of linked-data. Considering the most common concepts and properties that are used from the W3C SSN ontology can help to create an optimum core ontology in which the main concepts and properties are used in several other related ontologies. This can also give an indication of which parts of the SSN ontology are used more than others. The latter can provide feedback to the ontology developers to help them focus on the most commonly used features and create automated tools and software that can enhance and increase interoperability across different applications and systems in the IoT domain.

4 Conclusions

This paper describes a validation tool that is mainly designed for the W3C SSN ontology. However, it can be also used with other base line ontologies to validate linked-data descriptions and ontologies against reference ontologies. As the number of semantic models and description frameworks in the IoT domain increases, interoperability between the various models becomes an issue. The W3C SSN ontology is designed to describe sensor networks and device related features. This ontology has been used and extended in several projects and applications. We have created an online tool to validate semantic models and linked-data descriptions against the W3C SSN ontology. We used the tool to validate a set of ontologies that are available online in which the W3C SSN ontology was used as a base ontology. We created a Tag Cloud and presented the most common terms and concepts that are used from the W3C SSN ontology and provided statistics regarding the number of concepts and properties that are adapted from the SSN model in each of the ontologies.

The validation service can be used for checking and evaluating the new ontologies against a base line ontology; i.e. the W3C SSN ontology. It can be also used to collect information and statistics about the use of the W3C SSN ontology and provide feedback to the ontology developers. Future work will focus on automated matching and ontology alignment to improve interoperability between different ontologies that are developed in the IoT domain.

References

1. Michael Compton, Payam Barnaghi, Luis Bermudez, Raul García-Castro, Oscar Corcho, Simon Cox, John Graybeal, Manfred Hauswirth, Cory Henson, Arthur Herzog, et al. The ssn ontology of the w3c semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web*, 17:25–32, 2012.
2. Dave Evans. The internet of things: How the next evolution of the internet is changing everything. Technical report, Cisco Internet Business Solutions Group (IBSG), April 2011.
3. R. Tonjes, M. I. Ali, P. Barnaghi, S. Ganea, F. Ganz, M. Haushwirth, B. Kjargaard, D. Kumper, A. Mileo, S. Nechifor, A. Sheth, V. Tsiatsis, and L. Vestergaard. Real time iot stream processing and large-scale data analytics for smart city applications. *European Conference on Networks and Communications*, 2014.