Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal

PEER REVIEWED PAPERS
NEW TRENDS, CHALLENGES AND PERSPECTIVES ON HEALTHCARE COGNITIVE COMPUTING: FROM INFORMATION EXTRACTION TO HEALTHCARE ANALYTICS
Vol. 14, n.1, 2018
ISSN: 1826-6223 | eISSN: 1971-8829

INTEGRATING HETEROGENEOUS DATA OF HEALTHCARE DEVICES TO ENABLE DOMAIN DATA MANAGEMENT

Antonella Carbonaro[1]
Filippo Piccinini[2]
Roberto Reda[3]

[1] Department of Computer Science and Engineering, University of Bologna, Italy - antonella.carbonaro@unibo.it
[2] Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST) S.r.l., IRCCS, Oncology Research Hospital, Meldola (FC), Italy - filippo.piccinini@irst.emr.it
[3] Master's Degree in Computer Science and Engineering, University of Bologna, Italy - roberto.reda@unibo.it

Keywords: Internet of Things; Cognitive computing; Healthcare Data; Ontologies; Semantic Web Technologies

The growth of data produced for example by IoT devices has playing a major role in developing healthcare applications able to handle vast amount of information. The challenge lies in representing volumes of data, integrating and understanding their various formats and sources. Cognitive computing systems offer promise for analysing, accessing, integrating, and investigating data in order to improve outcomes across many domains, including healthcare. This paper presents an ontology-based system for the eHealth domain. It provides semantic interoperability among heterogeneous IoT fitness and wellness devices and facilitates data integration and sharing. The novelty of the proposed approach lies in exploiting semantic web technologies to explicitly describe the meaning of sensor data and define a

for citations:
Carbonaro A., Piccinini F., Reda R. (2018), *Integrating Heterogeneous Data of Healthcare Devices to enable Domain Data Management,* Journal of e-Learning and Knowledge Society, v.14, n.1, 45-56. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1450

common communication strategy for information representation and exchange.

1 Introduction

The term eHealth refers to systems which leverage information and communications technologies to support healthcare and health-related fields including wellness and fitness industry.

Due to the ever-increasing usage of mobile health apps and wearable devices in our day-to-day lives, Internet of Things (IoT) technology is becoming one of the most popular trends for assisted eHealth (Islam *et al.*, 2015; Kim *et al.*, 2014). For instance, wrist-worn fitness bands such as Fitbit or Jawbone UP are wellness devices which can monitor 24/7 users' physical activity and health by keeping track of steps taken, distance travelled, stairs climbed and sleep hours. More advanced wearable fitness devices such as smartwatches, along with a GPS for outdoor sport tracking include also sensors for keeping track of basic physiological parameters of the wearer with an emphasis on heart rate and body temperature. Consumer IoT fitness trackers can even be used to monitor patient with health problems, for instance cardiologic and oncologic patients (Mendoza *et al.*, 2017).

Nowadays, thanks to the aforementioned smart wearable devices we can revolutionize the way we manage our physical well-being and training sessions by analyzing the large volumes of structured and unstructured produced health data.

However, collected data are often stored and managed within separate repositories that are disconnected and isolated from other contextual and external systems (i.e., data silos) preventing users and health practitioners from having an integrated view of the whole knowledge.

Cognitive computing eHealth systems (i.e., the emulation of human thought process in a computerized model applied to healthcare) can deal with all the amount of data available from IoTs to address specific issues (Riccucci *et al.*, 2005) which involve the combination of that data with users' medical history or environment information to produce actionable insights for a more targeted medical care. Moreover, eHealth can also take advantage of cognitive computing systems when services and applications use different vocabularies and models for the same concepts, thus solving issues related to interoperability and knowledge sharing (Fox, 2017). Within eHealth knowledge systems, the employment of shared ontologies constitutes a good strategy to deal with diversity. From a Semantic Web (SW) perspective, the term "ontology" denotes "a shared conceptualization of an explicit machine-readable specification". Ontologies provide a generic infrastructure for exchanging and integrating structured data and promote a creative reuse of the data which can help to respond to some of the key challenges that eHealth systems are nowadays facing.

Ontologies, as cognitive computing approach to better access knowledge, provides a common semantic description for eHealth domain concepts. Moreover, specific imaging ontologies can also be used utilized in cooperation with computer-vision techniques (Carbonaro, 2009; Carbonaro 2010) when data collected by IoT are images. Semantic technologies provide a promising way also for image analysis and interpretation, in fact, qualitative aspects of an image (e.g. low-level visual features) can be easily mapped to high-level concepts formally described by an ontology.

In this paper, we propose an ontology-based cognitive computing eHealth system which aims to provide semantic interoperability among heterogeneous IoT fitness devices and wellness appliances in order to facilitate data integration, sharing and analysis. The original contribution of our works lies in exploiting SW technologies to explicitly and formally describe the meaning of sensors data, and to facilitate the interoperability and data integration among eHealth systems.

The paper is organized as follows: the next section introduces the main previous works describing SW technologies used in the healthcare context and explores research efforts related to manage different IoT data, highlighting the main open-issues in the field. Section 3 describes the technological aspects of the context and shows the overall architecture of the proposed framework. Section 4 proposes eHealth ontologies used to describe domain concepts. Section 5 reviews our development process in order to design the ontology and describes its main characteristics. Section 6 introduces the mapping process by which data are semantically annotated according to our ontology. Finally, Section 7 provides some considerations on a case study.

2 Related works

One of the main challenging problems connected with the existing IoT applications is that devices are not (or little) interoperable with each other since their data are based on proprietary formats and do not use common terms or vocabulary. Moreover, promote interoperability, reusability, and resource sharing among IoT applications is more complicated if IoT solutions are designed by considering a single domain. SW technologies can be employed in IoT (SWoT) to overcome these challenges.

Recently, Patel *et al.* (2017) created SWoTSuite, which is an infrastructure that enables SWoT applications. It takes high-level specifications as input, parses them and generates code that can be deployed on IoT sensors (e.g., temperature sensors, transportation devices) at the physical layer and IoT actuators (e.g., heaters) and user interface devices (e.g., smartphones, dashboards) at the application layer.

Ruta *et al.* (2012) proposed a general framework for the SWoT evolved over the basic knowledge base model. An ontology along with a set of asserted facts build a knowledge base from which further knowledge can be inferred. By sharing the system infrastructure, several object classes, described using different ontologies, can co-exist in a physical environment. Resources belonging to the same domain will be described by means of the same ontology, while objects of different categories may refer to different ontologies. The main values from the use of semantic technologies in IoT can be derived through cross domain or horizontal applications rather than focusing on domain specific of vertical IoT application development.

CREDO (Fox, 2017) is a long-term research program on reasoning, decision-making, planning, and autonomy. Its primary goal is a theoretical foundation for high-level cognition understanding; the program has been inspired by clinical expertise and medicine has been central for validating the results. The paper highlighted the increased understanding of the need to capture and sharing the meaning of concepts in complex domains like medicine where great efforts are being made to establish standard terminologies and ontologies. It suggested improving CREDO system using knowledge representation and ontology design as are domain terminologies and ontological content.

From the above-mentioned papers, it is possible to underline some main issues: the semantic interoperability of eHealth connected objects and their data is crucial but still poorly widespread. Often, data concerning the connected objects (e.g., characteristics, state, properties) and data concerning the patient (e.g. vital signs, activity) are represented in a disjointed context resulting into vertical application development.

3 Semantic Web approach to eHealth

By offering a generic infrastructure for interchange, integration, and creative reuse of structured data, SW can help to cross some of the boundaries that Web 2.0 is facing. Currently, Web 2.0 offers just basic query possibilities like searching by keywords or tags. There has been a great deal of interest in the development of semantic-based systems to facilitate knowledge representation and extraction and content integration (Carbonaro, 2010; Henze *et al.*, 2004; Andronico *et al.*, 2003; 2004). We can consider semantic information representation as an important step towards a wider efficient manipulation and knowledge representation (Carbonaro, 2006). In the digital library community, a flat list of attribute/value pairs is often assumed to be available (Carbonaro, 2010). In the SW community, annotations are often assumed to be an ontology instance. Through the ontologies, the system will express key entities and relationships describing resources in a formal machine-processable representation.

In eHealth domain, there are massive information resources in which the knowledge formation process is associated with multiple data sources. However, the systems, grammar, structure and semantics of resources are heterogeneous. The idea behind SW approaches is using the Web to allow exposing, connecting and sharing data through dereferenceable Uniform Resource Identifiers (URIs). The goal of SW is to extend the Web by publishing various open datasets as Resource Description Framework (RDF) triples and by setting RDF links between data items from several sources. Using URIs, everything can be referred to and looked up both by people and by software agents. Using URIs, RDF language and Ontology Web Language (OWL) ontologies, SW technologies easily allow users to connect pieces of data, share information and knowledge on the web. Ontologies constitute the backbone of the SW. Ontologies are a means to express concepts of a given domain and the relationships among the concepts; they specify complex constraints on the types of resources and their properties.

Interoperability is the ability to interconnect and create communications between different systems and along with data; integration is one of the vital issues still unsettled in IoT. However, interoperability can be solved if communicating smart systems are semantically interoperable. This challenge is particular relevant in the eHealth where a multitude of diverse devices collect the same type of data but store and exchange them in many different ways so generating semantic and syntactic conflicts. Semantic Web of Things (SWoT) is a research field which aims to combine SW technologies to IoT by providing interoperability among ontologies and data (Jara *et al.* 2014; Pfisterer *et al.*, 2011).

Semantic data annotation is the key step for every SW project. Our framework aims to semantically annotate heterogeneous IoT fitness and life logging data collected by wearable devices and wellness appliances in order to make it integrated and machine-understandable.

Figure 1 shows the overall architecture of the proposed framework. The two core components of the entire system are the IoT Fitness Ontology (IFO) and the mapping process (i.e., the RML processor and the mapping specifications).

The primary role of the IFO ontology is to provide a formal representation of the main concepts within the IoT fitness domain. The RML processor, supplied with the mapping specifications for the various sources, consumes the IoT raw data and transforms it into an RDF graph, which is the same input data semantically annotated according to the IFO ontology.

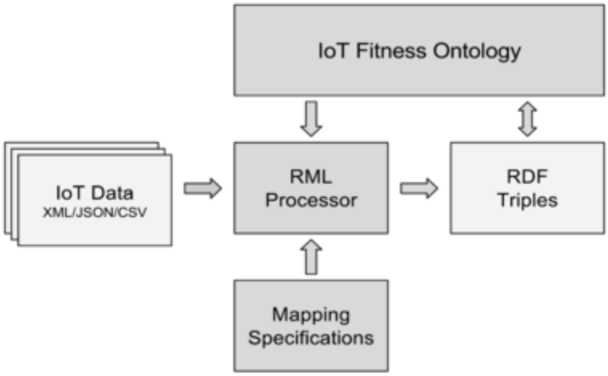

Fig. 1 - Proposed system architecture: how to build ontology using mapping and IoT raw data.

4 eHealth ontologies

Due to the extreme complexity of eHealth terminology systems, ontologies play a central role for the representation, management, and sharing of knowledge and data. In the past years, a plethora of eHealth domain ontologies has been created. Such representations are used to systematically denote, categorize, and relate eHealth concepts, allowing fitting handling of the data [46]. The number of ontologies in eHealth is constantly increasing; BioPortal provides access to a library of biomedical ontologies and terminologies developed in RDF, OWL, Open Biological and Biomedical Ontologies (OBO) formats (Smith, Ashburner *et al.*, 2007).

Below the main characteristics of SNOMED-CT and LOINC ontologies are briefly reviewed.

4.1 SNOMED-CT

The Systematized Nomenclature of Medicine-Clinical Term (SNOMED-CT) is considered as the main ontology for a standardized representation and automatic interpretation of clinical concepts, terms and relationships in the field of eHealth. The ontology covers most of the areas that are used in medical practice, including clinical findings, symptoms, diagnoses, pharmaceuticals, body structures, medical devices, social contexts, and so on. SNOMED-CT has hierarchy structure with a set of top level general concepts. All other concepts are subtypes of one these top concepts. Each concept is assigned a unique ConceptID and a Fully Specified Name (FSD). SNOMED-CT provides a consistent way to represent data that can enhance the interoperability between different systems.

4.2 LOINC

The Logical Observation Identifiers Names and Codes (LOINC) is a universal code system for laboratory test and other clinical observations. For each observation it provides a code, a short name, a long formal name and synonyms. The primary purpose of LOINC is to provide common codes and terminology to receive clinical observations from multiple sources, so that they can automatically fill the data in the right slots of their medical records, research, or public health systems.

5 Proposed ontology

The proposed IFO ontology is a lightweight extensible domain-specific ontology which aims to provide a formal representation of the most common concepts and their relationships within the IoT fitness and wellness devices, including mobile health applications.

This section briefly reviews our development process in order to design the IFO ontology and describes the main characteristics of the developed ontology.

5.1 Development process

For the development process of the IFO ontology, we followed the well-known methodology proposed by Noy and McGuiness which is a quick but complete approach for building ontologies (Noy & McGuiness, 2001). We wrote the IFO ontology in OWL, which is a W3C recommendation and the de facto standard language for publishing and sharing ontologies in the SW and we validated it using the ontology reasoner HermiT (Shearer *et al.*, 2008) to check for inconsistencies.

To identify the concepts described in the IFO ontology we considered and carefully analysed the characteristics and functionalities provided by several IoT wearable devices and wellness appliances as well as health mobile applications available in the market. The list of products and vendors that were taken in consideration during the design process includes: Apple HealthKit, Microsoft HealthVault, Google Fit, Fitbit, Jawbone, Strava, Runtastic, iHealth and Nokia Health. The result is a harmonised ontology of the most important common concepts in the domain considered. The first version of the IFO ontology consists of 93 classes, 16 object properties, 7 data properties, and 47 individuals.

5.2 Ontology Structure

The root concept of the IFO ontology is modelled by the class Episode, representing the set of the all possible events that can be measured by the IoT fitness and wellness systems. To each episode, is associated by means of OWL properties, a time reference (i.e., start time and end time of the event) and a numeric measurement value with the unit. These information are essential because they allow us to numerical quantify the object of the episode and give it a temporal collocation and duration. An example of episode could be a running activity or a treadmill session at the gym as well as a vital sign measurement such as the heart rate or the blood pressure.

The IFO ontology organizes the episodes in a hierarchical structure based on single inheritance. We consider physical activities and body measurements as important episode categories. Physical activities comprehend any kind of activity involving body movement such as running, swimming or steps taken. Body measurements, on the other hand, are relative to the physiological parameters of a person such as the body weight or the heart rate. Other categories that our ontology defines concern the sleep and the meditation. Furthermore, the IFO ontology describes complementary concepts to specify additional information regarding a single episode such as geospatial locations (i.e., acquired via GPS sensor), device characteristics, personal annotations, temporal relationships with respect to other person's life activities (e.g., temporal relationship of an episode with respect to a meal), physical activity intensity and person's information such as the gender or the date of birth.

To achieve a better integration with other systems and better specify the meaning of each class, we referred to other standardized ontologies such as SNOMED-CT. Personal information (e.g., date of birth) are based on FOAF ontology and the Basic geo (WGS84 lat/long) vocabulary was used for the geospatial locations. In order to keep the ontology simple, we avoided to import other OWL top level ontologies for concepts related to the measurements (e.g., units of measure) or for the time information (e.g., time intervals). Figure 2 shows the IFO ontology classes hierarchy as can be seen within the Protegè editor.

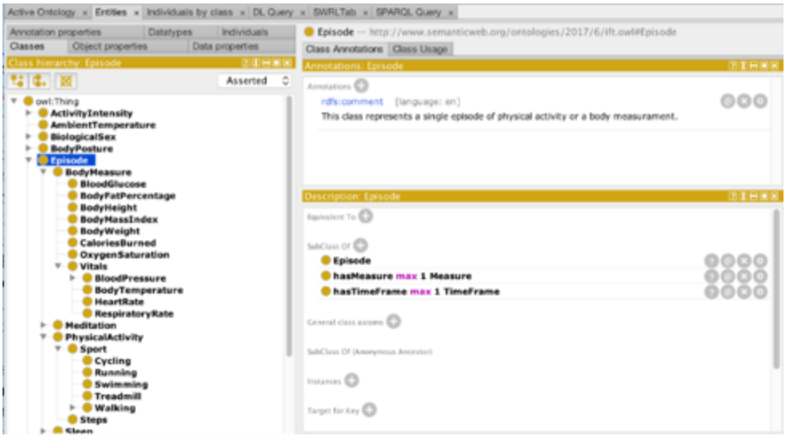

Fig. 2 - Developed ontology structure in Protegè editor.

6 Mapping

Mapping is the process by which values within the data sources are semantically annotated according to an ontology (Amardeilh, 2008). We use RML, a language to specify mappings for heterogeneous and hierarchical serializations into RDF, according to an RDF schema or ontology of the user's choice. This solution leads to mapping on the semantic level, provides a solution applicable to a broader domain and can be extended to cover a great number of file serialization. The mapping process constitutes the second core component of our framework. Essentially, a map processor consumes the input sources along with the mapping specifications to produce the RDF graph. It is important that the semantic annotation process adheres to a common standard to guarantee interoperability between different systems. RDF, the cornerstone of the SW provides a standardized means for adding metadata annotations to resources (Manola *et al.*, 2004), in order to reach semantic interoperability and integration and querying of data having heterogeneous formats. These heterogeneous data can be following leveraged in Linked Data applications. Therefore, generating RDF triples from sources having various formats is a key step for our system.

This section presents a brief overview of how data can be retrieved and the most common data serializations used by the IoT fitness devices, and then it discusses the technologies we used to implement the mapping process.

6.1 IoT data

Raw data collected by IoT devices can be manually retrieve when systems are provided with data export functionalities. IoT raw data can normally be exported in eXtensible Markup Language (XML) or Comma Separated Value (CSV) serialization formats. On the other hand, when a data export function is not directly available within the device or the mobile application, data collected by IoT systems can be downloaded from the Cloud, usually in JavaScript Object Notation (JSON) format, through RESTful APIs provided by the device vendor.

6.2 Mapping process

From a data perspective, the context of IoT, is characterized by a high heterogeneity of data representation and serialization formats. Among different vendors of IoT fitness devices the same concepts are represented using different types and stored in different formats. To address this issue, we chose to implement our mapping system using the RDF Mapping language (RML) along with the RML Processor.

RML is a mapping specification language based on RDF syntax; it derives from the R2RML language which is the W3C standard for mapping relational databases into RDF. RML is specifically designed for dealing with heterogeneous data sources. To refer to specific values within the input data sources, RML relies on a target expression language relevant to the source format. For instance, given a data source serialized in XML format, XPath can be used as an expression target language to extract the specified values, in a similar way values in a JSON source can be referenced using JSONPath.

We defined the mapping specifications for three IoT systems among the ones we used to construct the ontology (i.e., Fitbit, Apple Health and Nokia Health). In particular, the mapping rules are relative to some shared concepts among these systems (e.g., the heart rate). As an evidence of the flexibility of the mapping language, we selected the IoTs devices aforementioned because they use different formats to store the data collected. Even though we mapped only a limited number of devices, mapping definitions can be easily reused across different sources that provide similar information. As a mapping process executor, we opted for RML Mapper which is a Java implementation of an RML mapping processor. RDF Mapper already supports XML, JSON and CVS data formats, and therefore we did not need to extend or modify the existing software.

Conclusions

The cognitive computing approach proposed in this paper, allows establishing a common semantic for addressing the eHealth concepts. Ontology representation provides semantic interoperability among heterogeneous IoT fitness and life log data collected by wearable fitness devices and wellness appliances and facilitates data integration and sharing. The two core components of the framework are the IoT Fitness Ontology and the mapping system based on the RML language. SW technologies have been used also to enable advanced analytics over the IoT data. The reasoning tests performed using OWL and SWRL allow the automatic classification of classes using description logic expressions. The inferences were revised by an expert, who confirmed that they were valid based on his analysis of the available information. We are carrying out some more tests to verify that our model can efficiently implement the classification process.

The system addresses the issue of the dimension and heterogeneity in source and format of data captured by eHealth IoT devices by representing the semantics of both connected objects, the domain and their relationship to each other. Two are the main contributions of this work. First, we propose a semantic-based approach that starts with data collection, followed by knowledge extraction and semantic modelling of this knowledge, in order to explicitly describe the meaning of the sensors data. Second, we describe an eHealth system that integrates ontologies to facilitate the interoperability and data integration among different devices by illustrating the effectiveness of the proposed approach for ontology building and evaluation.

REFERENCES

Amardeilh F. (2008), *Semantic annotation and ontology population*. Semantic Web Engineering in the Knowledge Society, 424.

Andronico, A., Carbonaro, A., Colazzo, L., & Molinari, A. (2004), *Personalisation services for learning management systems in mobile settings*. International Journal of Continuing Engineering Education and Life Long Learning 14.4-5: 353-369.

Andronico, A., Carbonaro, A., Colazzo, L., Molinari, A., Ronchetti, M., & Trifonova, A. (2003), *Designing models and services for learning management systems in mobile settings*. Workshop on Mobile and Ubiquitous Information Access. Springer Berlin Heidelberg.

Carbonaro A. (2006), *Defining personalized learning views of relevant learning objects in a collaborative bookmark management system*, In Z. Ma (Ed.), Web-based Intelligent ELearning Systems: Technologies and Applications (pp. 139-155). Hershey, PA: Information Science Publishing.

Carbonaro A. (2007), *Personalized information retrieval in a semantic-based learning environment*, Social Information Retrieval Systems: Emerging Technologies and Applications for Searching the Web Effectively, Pages 270-288

Carbonaro A. (2010), *WordNet-based summarization to enhance learning interaction tutoring*, Journal of E-Learning and Knowledge Society, Volume 6, Issue 2, Pages 67-74

Carbonaro A. (2009), *Collaborative and semantic information retrieval for technology-enhanced learning*, CEUR Workshop Proceedings, Volume 535, 2009, 8p

Carbonaro A. (2010), *Improving web search and navigation using summarization process*, Communications in Computer and Information Science Volume 111 CCIS, Issue PART 1, Pages 131-138

Carbonaro, A. (2010), *Towards an automatic forum summarization to support tutoring.* Technology Enhanced Learning. Quality of Teaching and Educational Reform. Springer Berlin Heidelberg, 141-147.

Fox J. (2017), *Cognitive systems at the point of care: The CREDO program*, Journal of Biomedical Informatics.

Henze N., Dolog P. & Nejdl W. (2004), *Reasoning and Ontologies for Personalized E-Learning in the Semantic Web*, Educational Technology & Society, 7 (4), 82-97.

Islam S. R., D. Kwak, M. H. Kabir, M. Hossain, & K.-S. Kwak (2015), *The internet of things for health care: a comprehensive survey*. IEEE Access, 3:678–708.

Jara A. J., A. C. Olivieri, Y. Bocchi, M. Jung, W. Kastner, & A. F. Skarmeta (2014), *Semantic web of things: an analysis of the application semantics for the iot moving towards the iot convergence*. International Journal of Web and Grid Services, 10(2-3): 244–272.

Kim J., JW Lee (2014), *OpenIoT: An open service framework for the Internet of Things*, IEEE World Forum on Internet of Things.

Manola F., E. Miller, B. McBride (2004), *Rdf primer. W3C recommendation*, 10(1-107):6.

Mendoza, J.A., Baker, K.S., Moreno, M.A., Whitlock, K., Abbey-Lambertz, M., Waite, A., Colburn, T. & Chow, E.J. (2017), *Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: A pilot study*. Pediatric Blood & Cancer.

Noy, Natalya F., & Deborah L. McGuinness. (2001), *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05.

Patel, A Gyrard, SK Datta & MI Ali (2017), *SWoTSuite: A Toolkit for Prototyping End-to-End Semantic Web of Things Applications*, Proceedings of the 26th International Conference on World Wide Web Companion, pp. 263-267.

Pfisterer, D., Romer, K., Bimschas, D., Kleine, O., Mietz, R., Truong, C., Hasemann, H., Kröller, A., Pagel, M., Hauswirth, M. and Karnstedt, M. (2011), *Spitfire: toward a semantic web of things*. IEEE Communications Magazine, 49(11):40–48.

Riccucci, S., Carbonaro, A. & Casadei, G. (2005), *An architecture for knowledge management in intelligent tutoring system*, IADIS International Conference on Cognition and Exploratory Learning in Digital Age, CELDA, 473-476.

Ruta, M., Scioscia, F., & Di Sciascio, E. (2012), *Enabling the semantic web of things: framework and architecture*. In: 2012 IEEE Sixth International Conference on Semantic Computing. IEEE.

Shearer R., B. Motik, & I. Horrocks (2008), Hermit: A highly-efficient owl reasoner. In OWLED, volume 432, 91.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J. & Leontis, N. (2007), *The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration*, Nature Biotechnology 25, 1251 – 1255.