A Semantic-Aware Data Management System for Seismic Engineering Research Projects and Experiments

Md. Rashedul Hasan
Department of Civil, Environment
and
Mechanical Engineering
University of Trento, Italy

Feroz Farazi
Department of Information
Engineering and
Computer Science
University of Trento, Italy

Oreste Bursi
Department of Civil, Environment
and
Mechanical Engineering
University of Trento, Italy

Md. Shahin Reza
Department of Civil, Environment and
Mechanical Engineering
University of Trento, Italy

Ernesto D'Avanzo
Department of Political, Social and
Communication Sciences
University of Salerno, Italy

Abstract—**The invention of the Semantic Web and related technologies is fostering a computing paradigm that entails a shift from databases to Knowledge Bases (KBs). There the core is the ontology that plays a main role in enabling reasoning power that can make implicit facts explicit; in order to produce better results for users. In addition, KB-based systems provide mechanisms to manage information and semantics thereof, that can make systems semantically interoperable and as such can exchange and share data between them. In order to overcome the interoperability issues and to exploit the benefits offered by state of the art technologies, we moved to KB-based system. This paper presents the development of an earthquake engineering ontology with a focus on research project management and experiments. The developed ontology was validated by domain experts, published in RDF and integrated into WordNet. Data originating from scientific experiments such as cyclic and pseudo dynamic tests were also published in RDF. We exploited the power of Semantic Web technologies, namely Jena, Virtuoso and VirtGraph tools in order to publish, storage and manage RDF data, respectively. Finally, a system was developed with the full integration of ontology, experimental data and tools, to evaluate the effectiveness of the KB-based approach; it yielded favorable outcomes.**

Keywords—*Ontology; Knowledge Base; Earthquake Engineering; Semantic Web; Virtuoso*

I. INTRODUCTION

This is an extended version of the following paper: Hasan et al. 2013. The inventor of the Web, Tim Berners-Lee, envisioned a more organized, well connected and well integrated form of the Web data that are suitable for humans to read and for machines to understand. This new form of the Web is called the Semantic Web (T. Berners-Lee, 1999; T. Berners-Lee et al., 2001). On the Semantic Web data can be published using Resource Description Framework (RDF) and Web Ontology Language (OWL). Traditional databases are a persistent storage mechanism that enables large scale of data; however, they were not originally designed for managing RDF and OWL data or ontologies. KBs can do this job effectively. Ontologies are intended to be stored in the KBs, which can offer better user experience by supporting reasoning over ontological data and semantics. Moreover, KB-based systems provide mechanism to manage information and semantics thereof that can make systems semantically interoperable and as such can exchange and share data between them. To overcome the interoperability issues and to exploit the benefits offered by the state of the art technologies, we moved to the KB-based system.

In fact, we have developed an ontology named as Earthquake Engineering Research Projects and Experiments using a faceted approach that gives emphasis on research project management and experiments. Following the validation of the ontology by domain experts, it was published in the knowledge representation language RDF and integrated into the generic ontology WordNet[1]. The experimental data coming from, inter alia, the cyclic and pseudo-dynamic tests were also published in RDF. We used Jena[2], Virtuoso[3] and VirtGraph[4] tools for ontology and data publishing, storage and management, respectively. Finally, a system was developed to verify the effectiveness of the approach through the integration of the aforementioned tools, ontologies and data.

The rest of the paper is organized as follows. Section II depicts an ontology based information management system development approach. Section III describes the ontology development steps and the created ontology (partially). In Section IV, we provide a brief description of the ontology representation languages RDF and OWL. In Section V, we present existing ontology/thesaurus that are relevant for this work and as such worth discussing them. Section VI provides

---

[1] http://wordnet.princeton.edu
[2] https://jena.apache.org
[3] http://virtuoso.openlinksw.com
[4] http://docs.openlinksw.com/jena/virtuoso/jena/driver/
VirtGraph.html

the ontology integration approach and Section VII describes experimental data collection procedure. While Section VIII demonstrates the architecture of the final system that was built on top of the integrated ontology, Section IX reports evaluation results that show the effectiveness of the ontology. In section X we briefly describe related work and in Section XI we conclude the paper.

II. APPROACH

Figure 1 describes an ontology based information management system development approach that involves standard three-tier architecture. KB works as a backend of the system hosting ontologies represented in RDF, while query processing, inference mechanism and reasoning are incorporated in the business logic layer. Issuing queries and showing the corresponding results are supported by the User Interface (presentation) layer. However, for ontology development (see Section III) we follow the DERA (Domain, Entity, Relation, Attribute) methodology (Giunchiglia and Dutta, 2011), for ontology representation (see Section IV) in RDF we use Jena and for ontology integration (see Section VI) we implemented a facet based algorithm.
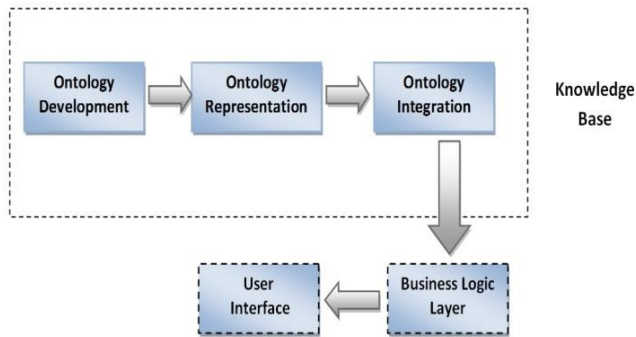

Fig. 1.   Ontology based development Approach

III. ONTOLOGY DEVELOPMENT

We use the DERA methodology (Giunchiglia and Dutta, 2011) for ontology development in fact it is known to be extendable and scalable and some ontologies including GeoWordNet were developed following this approach (Giunchiglia, 2010).

DERA methodology allows for building domain specific ontologies. Domain is an area of knowledge in which users are interested in. For example, earthquake engineering, oceanography, mathematics and computer science can be considered as domains. In DERA, a domain is represented as a 3-tuple $D = < E, R, A >$, where E is a set of entity-classes that consists of concepts (e.g., device and experiment) and entities (e.g., an instance of device and an instance of experiment); R is a set of relations that can be held between entity-classes (e.g., IS_A and $PART\_OF$) and A is a set of attributes of the entities (e.g., *number of devices and name of the experiment*).

In this three basic components concepts, relations and attributes are organized into facets; hence, the ontology is based on faceted methodology. Facet is a hierarchy of homogeneous concepts describing an aspect of a domain. S. R. Ranganathan, who was an Indian mathematician-librarian, was the first to introduce faceted approach capable of categorizing books in the libraries (Ranganathan, 1967).

Note, however, that a domain can alternatively be called as domain ontology. Henceforth in this paper it will be referred to as domain ontology. Among the macro-steps to develop each component of a domain ontology, we used the following ones.

In the first step (*identification*) towards building an ontology, we identified the atomic concepts of terms collected from research papers, books, existing ontological resources and experts belonging to Earthquake Engineering domain giving emphasis on research projects and experiments aspects. We found terms such as device, shaker, experiment, dynamic test, etc., and identified the atomic concept for each of them. We bootstrapped our Knowledge Base with the concepts and relations of WordNet. The term device has 5 different concepts in it. In our case, we selected the one that has following description: *device -- (an instrumentality invented for a particular purpose)*. We have found 193 atomic concepts. In the second step (analysis) we analyzed the concepts, i.e., we studied their characteristics to understand the similarity and differences between them. Once the analysis was completed, in the third step (*synthesis*) we organized them into some facets according to their characteristics. For example, shaker is more specific than device, actuator is more specific than device, motor is a part of electric actuator and we assigned the following relationships between them: shaker *IS_A* device, actuator *IS_A* device, motor *PART_OF* electric actuator. This is how we built device fact. In this way, we built 11 facets. A partial list of the facets is as follows: device, experiment, specimen, experimental computation facility, project, project person and organization. Device and experiment facets are shown in Fig. 2. In the fourth step (*standardization*), we marked concepts with a preferred name in the cases of availability of synonymous terms. For example, while experiment and test are used to refer to the same concept, we assigned the former term as the preferred one. Finally, the ontology was validated by domain experts. They suggested a number of changes, e.g., the inclusion of the concepts *shaker-based test* and *hammer-based test* in the experiment facet, the exclusion of the concept *simulation* from the same facet.
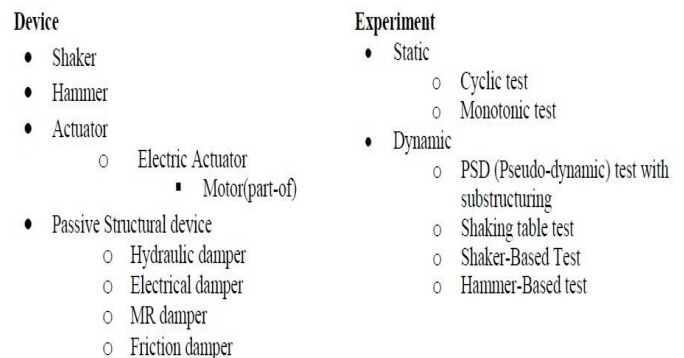

Fig. 2.   The device and experiment facets

Note that in Fig. 2, concepts which are connected by PART_OF relation with the concepts one level above in the hierarchy are explicitly written, for example, motor is PART_OF electric actuator. In the other cases, IS_A relation

holds between them, for example, electric actuator IS_A actuator.

IV. ONTOLOGY PRESENTATION

In the following subsections, we describe the Knowledge Representation Languages RDF and OWL in terms of their capacity in representing ontologies of varied kinds.

A. RDF

The Resource Description Framework (RDF) is a data model used to represent information about resources in the World Wide Web (WWW) and can be used to describe the relationships between concepts and entities. It is a framework to describe metadata on the web. Three types of things are in RDF: resources (entities or concepts) that exist in the real world, global names for resources (i.e. URIs) that identify entire web sites as well as web pages, and RDF statements (triples, or rows in a table) (Klyne, 2004). Each triple includes a subject, an object and a predicate. RDF is designed to represent knowledge in a distributed way particularly concerned with meaning. The following RDF statements describe the resources *Hammer* and *Damper.*

```
<rdf:Description rdf:about="http://earthquake.linkeddata.it/resource/Hammer">
    <rdfs:subClassOf rdf:resource="http://earthquake.linkeddata.it/resource/Device"/>
    <ontology:desciption  rdf:datatype="http://www.w3.org/2001/XMLSchema#string">A  hand
    tool with a heavy rigid head and a handle; used to deliver an impulsive force by strik-
    ing</ontology:desciption>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
</rdf:Description>

<rdf:Description rdf:about="http://earthquake.linkeddata.it/resource/Damper">
    <rdfs:subClassOf rdf:resource="http://earthquake.linkeddata.it/resource/Device"/>
    <ontology:desciption rdf:datatype="http://www.w3.org/2001/XMLSchema#string">A de-
    vice that decreases the amplitude of electronic, mechanical, acoustical or aerodynamic oscil-
    lations</ontology:desciption>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
</rdf:Description>
```

Fig. 3.   RDF statements describe the resources Hammer and Damper

The above example represented relationship between Hammer and Device concepts; and the rdfs: sub Class Of property is used to relate the former concept to its more generic later concept.

B. OWL

Web Ontology Language is designed to represent comparatively complex ontological relationships and to overcome some of the limitations of RDF such as representation of specific cardinality values and disjointness relationship between classes (Giunchiglia et al. 2010). The language is characterized by formal semantics and RDF/XML based serializations for the web. As an ontology representation language, OWL is essentially concerned with defining terms that can be used in RDF documents, i.e., classes, properties and instances (Antoniou et al. 2004). It serves two purposes: first, it helps identifying current document as an ontology and second it serves as a container of metadata about the ontology. This language focuses on reasoning techniques, formal foundations and language extensions. OWL uses URI references as names

and constructs these URI references in the same manner as that used by RDF. The W3C allows OWL specification to include the definition of three variants of OWL, with different levels of expressiveness. These are OWL Lite, OWL DL and OWL Full ordered by increasing expressiveness.

V. EXISTING ONTOLOGY/ THESAURUS

Ontologies and thesaurus, which are germane to our Earthquake Engineering ontology, are described in terms of the amount of concepts they have and the types of relations that exist between concepts.

A. WordNet

WordNet (Miller et al. 1990) is an ontology that consists of more than 100 thousand concepts and 26 different kinds of relations, e.g., hyponym, synonym, antonym, hypernyms and meronyms. It was created and is being maintained at the Cognitive Science Laboratory of Princeton University. The most obvious difference between WordNet and a standard dictionary is that its concepts are organized into hierarchies, like professor *IS_A* kind of person and person *IS_A* kind of living thing. It can be used for knowledge-based applications. It is a generic knowledge base and as such does not have good coverage for domain specific applications. It has been widely used for a number of different purposes in information systems including word sense disambiguation, information retrieval and automatic text summarization.

B. NEES Thesaurus

The Network for Earthquake Engineering Simulation NEES is one of the leading organizations for Earthquake Engineering in the USA.

TABLE I.   NEES EARTHQUAKE ENGINEERING THESAURUS

Term	Broader Term	Narrower Term
AASHTO_LRFD_Bridge_Design_Specifications	AASHTO_2001	
Peak_Base_Acceleration	Acceleration	
Dynamic_Actuator	Actuator	
Static_Actuator	Actuator	
Cyclic_Axial_Load	Axial_Load	
Preformed_Fabric_Pads	Bearings	Cotton_Duck_Bearing_Pads

(HTTPS://NEES.ORG/)

They developed an earthquake engineering thesaurus, which is based on Narrower and Broader terms. It contains around 300 concepts and we have integrated in our ontology 75 of them. Table I reports a small portion of NEES thesaurus.

VI. ONTOLOGY INTEGRATION

Developed facets include concepts that were selected from NEES thesaurus to be incorporated into our ontology. This integration was accomplished in fact when we built the facets. In this Section, we describe how we integrated our developed ontology with Wordnet. Basically, we applied the semi-automatic ontology integration algorithm proposed in Farazi et al. (2011). In particular, we implemented the following macro steps:

1) Facet concept identification: For each facet, the concept of its root node is manually mapped to WordNet, in the case of availability.

2) Concept Identification: For each atomic concept C of the faceted ontology, it checks if the concept label is available in WordNet. In the case of availability, it retrieves all the concepts connected to it and maps with the one residing in the sub-tree rooted at the concept that corresponds to the facet root concept.

3) Parent Identification: In the case of unavailability of a concept it tries to identify the parent. For each multiword concept label it checks the presence of the header, and if it is found within the given facet, it identifies it as a parent. For instance, in WordNet it does not find hydraulic damper for which damper is the header and that is available there in the hierarchy of device facet. Therefore, it recognizes the damper with the description damper, muffler -- (a device that decreases the amplitude of electronic, mechanical, acoustical, or aerodynamic oscillations), as the parent of the hydraulic damper.

VII.   EXPERIMENTAL DATA COLLECTION

In this section, an experimental test on a piping system under earthquake loading carried out by Reza et al. (2013) is briefly discussed to provide the reader with an overview of experimental Data Acquisition (DAQ) procedure.
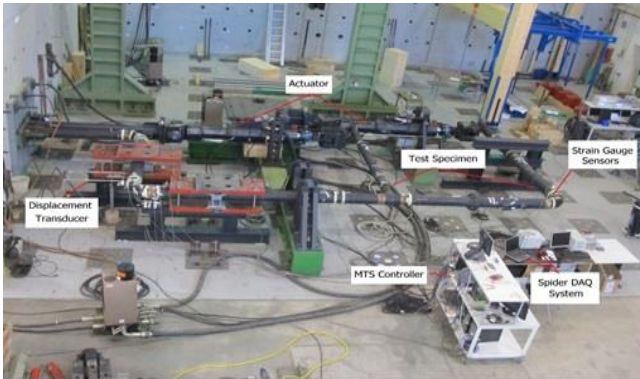

Fig. 4.   Experimental set-up of a piping system tested under earthquake loading (Reza et al., 2013)

Fig. 4 illustrates the relevant set-up of the experiment. As can be seen in this figure, the test specimen, i.e. the piping system, is excited with earthquake loading by means of two actuators which are controlled via an MTS controller. The test specimen is mounted with several sensors, such as strain gauges and displacement transducers, in order to observe its responses under applied seismic loading. In this particular experiment, four Spider8 DAQ systems were used to collect data from the sensors. Generally, output from a sensor, e.g. displacement transducer, is found in voltage, which is then transformed in another unit, such as mm, through a predefined calibration made in the DAQ measurement software. This data are then stored in a computer in an easily manageable format, such as Matlab (.mat) excel or ASCII, which are published in the ontology.

VIII.   EXPERIMENTAL SET-UP

In Fig. 5, we describe the process of creating the KB. The domain specific ontology that we developed was published into RDF by means of Jena (a Semantic Web tool for publishing and managing ontologies) and integrated with WordNet RDF using the approach described in Section VI. In order to increase the coverage of the background knowledge in the KB, we performed the integration of the two ontologies. The outcome of the ontology integration was put in Virtuoso triple store.
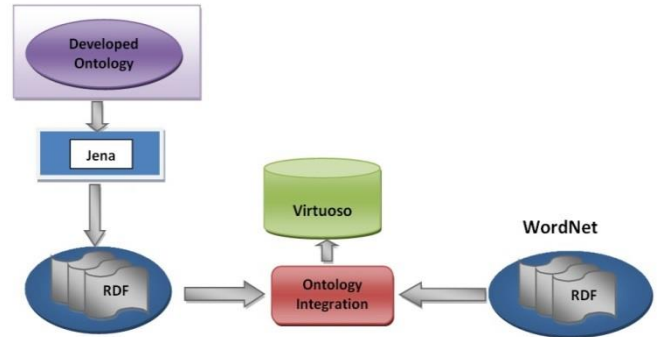

Fig. 5.   Ontology Integration and Population to KB

Fig.6 illustrates the architecture of our KB-based information management system that uses Semantic Web tools and technologies. As presented in the figure, the system is organized into three layers, which are User Interface (UI), Middleware and KB.

To execute any user request, for example, visualizing the whole ontology or part of it, the corresponding service is called from the middleware. Each service communicates with the KB using SPARQL query. SPARQL is a query language especially designed to query RDF representations. It allows add, update and delete of RDF data.

User Interface: Developed user interface allows people to perform the following operations on the ontological TBoxes: edit, search, integration and visualization, which are shown in the upper-most layer of Fig. 6 alongside the following operations defined to be performed on the ABoxes: edit entity, entity navigation and experimental result visualization. With the edit ontology operation, concepts and relations can be created, deleted and updated. With the search ontology operation, concepts can be queried with their natural language labels. For the aggregation of an external ontology with the ones already present in the KB we perform the integration operation. In order to view and surf any of the ontologies, we employ (ontology) visualization operation. Note that in the KB until now we have two ontologies, WordNet and EERPE.

Edit entity operation is designed to help perform create, delete and update entities. Existing entities can be viewed and browsed with the entity navigation operation and experimental results can be shown with the corresponding visualization operation.

Middleware: All the functionalities germane to the operations that can be requested and eventually be performed from the user interface are implemented as services and deployed on a web server.

Each service is basically communicating with the KB to execute one or more of the CRUD (create, read, update and delete) operations on its knowledge objects.
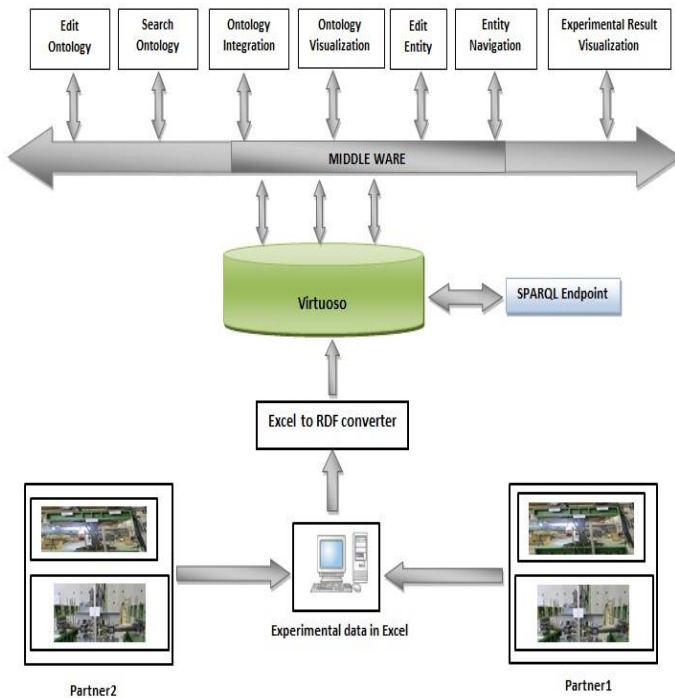

Fig. 6.   KB-based System Architecture

KB: This is our Knowledge Base hosting the ontologies consists of concepts and relations thereof, entities and their attributes and relations, and exogenous data from our own experimental setup and the one of our partner university, the University of Napoli.

IX.   RESULTS

In Table II, we report the detailed statistics about EERPE ontology. This ontology consists of 11 facets, 193 entity classes, 6 relations and 13 attributes. Note that each of the entity classes, relations and attributes represents an atomic concept. Hence, in total we found 212 atomic concepts in the ontology and out of them 100 concepts are available in WordNet.

TABLE II.   STATISTICS ABOUT EERPE ONTOLOGY

Object	Quantity
Facet	11
Entity class	193
Relation	6
Attribute	13
Concept	212
Concepts found in WordNet	100

Moreover, we describe basically what sort of advantages users can get with KB-based systems over traditional DB systems. In particular, we performed synonym search and more specific concept search.

Synonym Search: when a concept is represented with two or more terms, they are essentially synonymous and can be represented in RDF with *owl:equivalentClass*. For example, test and experiment represent the same concept and in the ontology they are encoded accordingly with equivalent relation. Therefore, as can be seen in Fig. 7, user query for test can also return experiment because they are semantically equivalent.
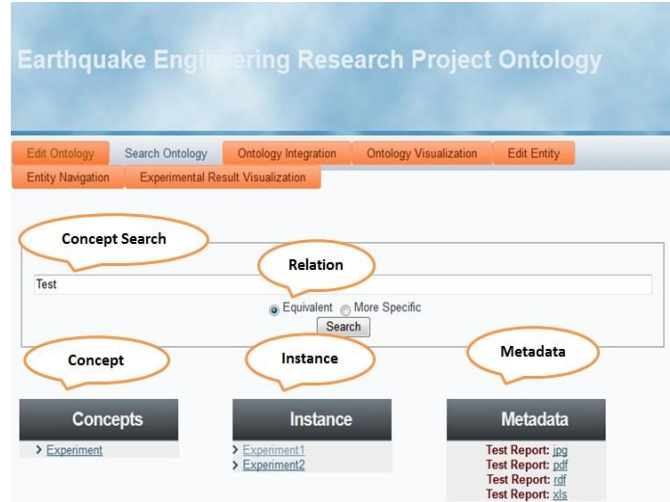

Fig. 7.   Synonymus relationship of Test

More specific concept search: In our ontology concept hierarchies are represented using *rdfs:subClassOf*. For example, hammer and damper are more specific concepts of device, hence, they are represented as follows: hammer *rdfs:subClassOf* device; and damper *rdfs:subClassOf* device.
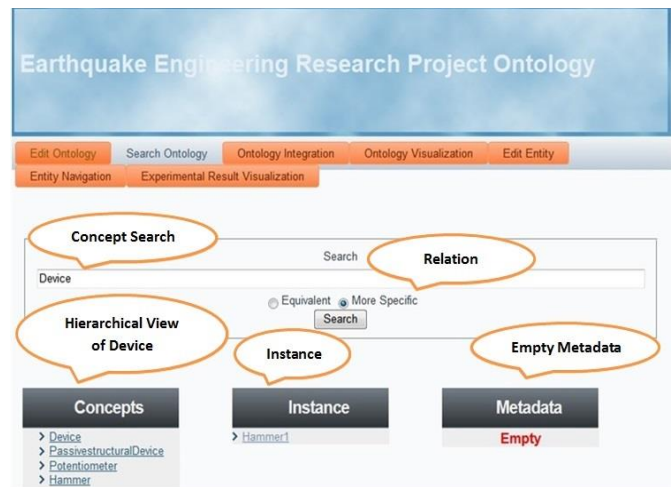

Fig. 8.   Transitive Relationship of Device

Moreover, hydraulic damper is more specific than damper and it is encoded as hydraulic damper *rdfs:subClassOf* damper. Note that *rdfs:subClassOf* is a transitive relation. Using OWL inference engine, we can utilize the power of transitivity and for a given concept we can retrieve all the more specific concepts that are directly or indirectly connected by *rdfs:subClassOf*. Therefore, a search for device retrieved all of its more specific concepts as shown in Fig. 8.

In addition to the search facility, we have implemented ontology editing, integration and visualization, entity editing and navigation and experimental result visualization functionalities. We have tested them with the help of a number of users. Their feedbacks were satisfactory.

X. RELATED WORK

We have classified the related works into two kinds. One covers the earthquake engineering ontology topic and another focuses on the faceted approach for developing ontologies.

Earthquake Engineering Ontology: NEES ontology has been developed in the domain of earthquake engineering. However, it is mainly a thesaurus encoding broader and narrower relations that cannot capture ontological details. For instance, it cannot be clarified in thesaurus whether a relation between two concepts is *IS_A* or *PART_OF*. As a result ontologies represents as thesaurus might lead to some unexpected results. DB Pedia is an example that uses broader/narrower relations and ended up establishing connection between Telecommunication, and Flora and Fauna. In contrast, the ontology developed in this paper does not suffer from this issue; rather it provides better clarification because it exploits ontological relations.

Faceted ontology development: This approach was followed in developing Geo WordNet, a faceted ontology aimed at building geospatial Semantic Web and enhancing interoperability among numerous information systems developed in isolations dealing with data of the geographic domain (Giunchiglia et al. 2010). By taking into account the advantages offered by this approach, such as easy to follow and linear time requirement, it was employed in the creation of some other ontologies including the one for the Autonomous Province of Trento for developing their semantic geo-catalogue (Farazi et al. 2011).

XI. CONCLUSION

In this paper, we provided a detailed description of the development of Earthquake Engineering Projects and Experiments ontology. We followed DERA methodology for building this domain specific ontology. We exploited an ontology integration algorithm that was employed to incorporate our ontology into WordNet. It helped to increase the coverage of the Knowledge Base. On top of the integrated ontology that is kept in an instance of Vrituoso, we experimented the semantic and ontological capabilities of the developed system and interesting results were found.

The need for ontologies in Earthquake Engineering is demonstrated, and it has been shown that ontology can be a useful tool for knowledge codification, management, sharing and reuse. We have planned the following future works. We will improve the query performing capabilities using Natural Language Processing (NLP) techniques. We will also include automatic ontology updating feature employing supervised machine learning approach.

REFERENCES

[1] Grigoris, A. and Harmelen, F. V. Semantic Web Primer. the MIT Press, 2004.

[2] Farazi, F., Maltese, V. , Giunchiglia, F. , Ivanyukovich, A. A faceted ontology for a semantic geo-catalogue, Extended Semantic Web Conference (ESWC), 2011.K. Elissa, "Title of paper if known," unpublished.

[3] Giunchiglia, F., Dutta, B. DERA: A Faceted Knowledge Organization Framework, March 2011

[4] Giunchiglia, F., Dutta, B., Maltese,V., Farazi, F. A Facet-Based Methodology for the Construction of a large-scale GEO-SPATIAL Ontology, 2012.

[5] Giunchiglia, F., Farazi, F., Tanca, L. , R. de Virgilio. The Semantic Web Languages. Semantic Web Information Management, Springer Verlag, Berlin, 2010.

[6] Giunchiglia, F., Maltese, V., Farazi, F., Dutta, B.. GeoWordNet: a resource for geo-spatial applications. In The Semantic Web: Research and Applications. Springer Berlin Heidelberg, 2010.

[7] Miller, G. A., Beckwith, R. , Fellbaum, C. D. , Gross, D., Miller, K. WordNet: An online lexical database. Int. J. Lexicograph. 3, 4, pp. 235–244, . 1990.

[8] Klyne, G., Carroll, J. (Editors). Resource Description Framework (RDF): Concepts and Abstract Syntax, W3C Recommendation, 10 February 2004.

[9] Ranganathan, S. R. Prolegomena to library classification. Asia Publishing House,1967.

[10] Berners-Lee, T. Weaving the Web. Orion Business Books, 1999.

[11] Berners-Lee, T., Hendler, J. A. and Lassila, O. The Semantic Web. In Scientific American Journal, 284(5), 34-43, (2001).

[12] Reza, Md Shahin, et al. "Pseudo-Dynamic Heterogeneous Testing With Dynamic Substructuring of a Piping System Under Earthquake Loading." ASME 2013 Pressure Vessels and Piping Conference. American Society of Mechanical Engineers, 2013.

[13] Hasan, M.R., et al. A Semantic Technology-Based Information Management System for Earthquake Engineering Projects and Experiments. In 5th International Conference on Advances in Experimental Structural Engineering, Taipei, Taiwan: National Center for Research on Earthquake Engineering, 2013.

[14] Sure, Y., Angele, J., Staab, S. OntoEdit: Multifaceted Inferencing for Ontology Engineering, Journal on Data Semantics, LNCS 2800, Springer, 2003, pp. 128-152.

[15] Maedche, S. Staab.: KAON: The Karlsruhe Ontology and Semantic Web Meta Project, Künstliche Intelligenz. Special Issue on Semantic Web 3/2003, pp. 27-30.

[16] Gennari, John H., Mark A. Musen, Ray W. Fergerson, William E. Grosso, Monica Crubézy, Henrik Eriksson, Natalya F. Noy, and Samson W. Tu. "The evolution of Protégé: an environment for knowledge-based systems development." International Journal of Human-computer studies 58, no. 1 (2003): 89-123.

[17] Giunchiglia, F. et al. "GeoWordNet: a resource for geo-spatial applications." The Semantic Web: Research and Applications. Springer Berlin Heidelberg, 2010. 121-136.

[18] Giunchiglia, F., Dutta, B., Maltese, V. "Faceted lightweight ontologies." Conceptual Modeling: Foundations and Applications. Springer Berlin Heidelberg, 2009. 36-51.