Application-Based Ontology Engineering

Noryusliza Abdullah and Rosziati Ibrahim

Abstract—**Ontology as a knowledge representation has addressed the evolution of Semantic Web in Web 3.0. Data representation is more structured and easily interpreted by machine. In this paper, University Tun Hussein Onn Malaysia (UTHM) Ontology, also known as UTHM Onto is built by referring HERO (Higher Education Reference Ontology) with modifications. This ontology engineering is using Protégé 4.1 and OWL/XML language. In retrieving the data, Jena Inference and SPARQL query language have been considered. UTHM Onto presented in this research can be utilized in other application. This work has been part of Semantic Web Search Engine using ontology, clustering and user profiling techniques. The developed ontology will be utilized in the user profiling part whereby obtained depth and superclasses is used for the Similarity Measurement.**

Index Terms—**Ontology, ontology engineering, Protégé, OWL**

I. Introduction

Ontology is rapidly evolved and utilized in current technologies. In computer science, generally ontologies are divided into Upper and Domain ontologies with each of them is responsible to certain role. Yet, certain researcher use Application ontologies term as a subset of domain ontologies to reflect specific usage for a certain application. In this research, application ontology named UTHM Onto is built by referring domain specific ontology.

Researchers are motivated to facilitate ontology in several field for instance semantic web, information retrieval and knowledge representation. Among those fields, semantic web has been using ontology extensively in research and application development. A promising bright future in the development and usage has encourages researchers to build ontologies and make them reusable. Reusing factor has assist in the expansion of research in this area. Ontology mapping is performs to integrates more than one ontologies in application.

From the existing available ontologies, HERO (Higher Education Reference Ontology) is chosen as reference and it is modified to fulfill a particular need. HERO is ontology developed using OWL API. It consists of higher educational structure. Classes are added and removed to allow representation of specific organization in this case, UTHM. It can be utilized in Hybrid Semantic Web Search Engine using ontology, clustering and user profiling techniques proposed in [1]. The developed ontology will be used in the user profiling part.

Manuscript received August 20, 2013; revised November 1, 2013.

The authors are with the Faculty of Computer Science and Information Technology, University Tun Hussein Onn Malaysia (UTHM) (e-mail: {yusliza, rosziati} @uthm.edu.my).

Several ontologies editor are used in editing them including Protégé, OntoEdit, Topbraid, SWOOP, Neon Toolkit, Knoodl and Ontolingua. Protégé [2] is chosen in this research to implement UTHM Onto based on certain advantages. Modification of HERO is permitted using this editor. In retrieving the data, Jena Inference and SPARQL query language have been considered.

The remainder of this paper is organized as follows. Section II discusses the related work. Section III describes ontology engineering. In this section, it is divided into sub topics: Reusing Ontologies, Ontology Editor and Ontology Language. Section IV contains description of UTHM Ontology implementation. The ontology integration is explained in Section V. Finally, Section VI provides the conclusion and future work.

II. Related Works

Ontology has different interpretations. Derived from Greek philosophical study, it represents the nature of being, existence, or reality, as well as the basic categories of being and their relations [3]. This term is eventually widely used within computer science to describe the world consisting set of types, properties, and relationship. Referring to widely cited Gruber's research paper [4], ontology is an explicit specification of a conceptualization. Additionally, ontology is a domain and knowledge representation [5], [6]. In consonance with Hepp [7], ontologies are the vocabulary that can be used to express a knowledge base while Diez-Rodriguez et al. [8] discusses that the intention to represent concepts in ontologies is to improve knowledge searching and discovery mechanisms.

Upper, domain and application ontologies have signifies great progress. The development of them begins with upper ontology, a top-level or foundation ontology. Jos de Bru ijn [9] describes upper ontology is a general concepts and independent from particular task or domain. Certain number of upper ontologies are developed such as Base Formal Ontology (BFO), Cyc (or OpenCyc), Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE), General Formal Ontology (GFO), PROTON (PROTo ONtology), Sowa's ontology, WordNet and Suggested Upper Merged Ontology (SUMO) [10]. These ontologies facilitate capturing general, domain independent knowledge (e.g. space, time) [9].

However, the restriction of upper type ontology is the inability to satisfy user's need. Due to this, researchers are trying to produce domain ontologies. It models a specific domain, which represents part of the world [3]. Various domain-specific ontologies are developed including domain in biomedical, pharmaceutical, university, currency, family relationship and others. Although these ontologies are able to

DOI: 10.7763/IJFCC.2014.V3.291

overcome constraints carried by the first type of ontology, something has to be done to resolve specific requirement in application. Accordingly, application ontologies are implemented.

Researchers are developing their ontologies with intention to share and allow reusing concept. This idea has lead to the formation of ontology libraries. D'Aquin, Natalya F.Noy [11] lists 11 new generation ontology libraries. They are BioPortal, CupBoard, The OBO Foundry, oeGov, OLS, Ontology Design Patterns, OntoSelect, OntoSearch, The ONKI ontology server, The TONES repository and Schema-Cache.

Rapid emergence of ontologies has encourages reusing concept which involve modification. Creating new ontology is done when there is no one exists that satisfy the requirement. Occasionally, ontology mapping is used if two or more ontologies are needed. Section III describes about the ontology engineering.

III. Ontology Engineering

In this section we discuss the details of reusing ontologies and ontologies editor.

A. *Reusing Ontology*

Ontologies allow reuse and this is the ultimate advantage. According to Cardoso [12], building ontology is more complex in terms of logic and structure compared to building software. Equally important, reuse existing, well-established and well-tested ontologies to describe data is more cost-effective than develop from scratch [11]. Therefore, researchers and developers prefer to reuse it. In reusing ontologies, several steps and techniques are used. Simperl E. [13] presents eleven most prominent case studies on ontology reuse. The studies show activities such as ontology assessment, integration, translation and customization are applied. In this research HERO is used to produce specific ontology for application usage. In HERO, there are domain and classes with attributes and instances. These are then studied, modified and reengineer.

B. *Ontology Editor*

Ontology development needed an editor. There are several editors including Protégé, SWOOP, OntoEdit, OntoStudio, NeOn Toolkit, KAON and many more. Among all, Protégé is the most used editor due to the support of wide variety of plugin and import formats. Furthermore, it is free open source.

Many researchers and developers use Web Ontology Language (OWL) in Protégé. It is a semantic markup language for publishing and sharing ontologies on the World Wide Web and used to describe the classes and relations between them [14].

C. *Ontology Language*

Ontology is implemented using ontology language. Resource Description Framework (RDF) is a framework for representing information in the web. It initiates in annotating web pages to enable machine to process metadata. Specifically, it is developed for the Semantic Web. Only few things can be inferred from this language. Therefore, RDF Schema (RDFS) is introduced. It is a semantic extension of RDF and more expressive compared to the former type. It stands between light-weight and heavy-weight ontologies. Dealing with heavy-weight ontologies need more advance language that is Web Ontology Language (OWL).

RDF and its family are developed from triples consisting of subject-predicate-object <s-p-o>. It generates RDF Graph, g also known as Directed Graph as shown in Fig. 1.

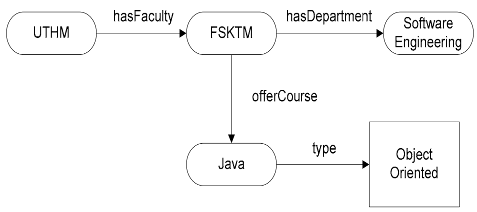

Fig. 1. Directed graph.

Ontologies can be described in another way of s-p-o as below:

(has Faculty, UTHM, FSKTM)
(has Department, FSKTM, Software Engineering)
(offer Course, FSKTM, Java)
(type, Java, Object Oriented)

Section IV describes the details of implementing UTHM Ontology using Protege.

IV. Implementation of UTHM Ontology

In this section, the proposed approach to implement UTHM ontology is presented.

A. *Tools*

In the implementation phase, Protégé 4.1 is used as the ontology editor. The ontologies can be exported into a variety of formats including RDF/XML, OWL/XML, OWL Functional Syntax, Turtle and N-Triple.

B. *Design*

Ontology creation is using certain methodology. Uschold and King Skeletal Model, Gruniger and Fox, Methontology and On-To-Knowledge (OTK) are several methods commonly used. OTK has been used throughout the UTHM Onto development since it is more suitable in application specific ontologies.

In designing UTHM Onto, several things have been taken into consideration. Domain and scope of this ontology is defined. This research only apply small scope in ontology development since we only intend to prove the usage of ontology as user profiling concept in the Semantic Web Search mentioned in [1]. Therefore, it is only concentrates on IT faculty (FSKTM) in Universiti Tun Hussein Onn Malaysia (UTHM). The process is then followed with class definition and class hierarchy. From various approach in class hierarchy development, top-down is chosen. The object properties between classes are developed to link classes involved. They are hasDepartment, hasFaculty, offerCourse and type.

Corroboration in Instances also known as Individuals are included in the ontology. Examples of the instances are stated below:

Faculty, I_f = {FSKTM}

Organization, O_f = {UTHM}
Department, D_f= {software Engineering, Web Technology, multimedia, Information Security}
Object Oriented, B_f= {java, php, visual Basic}
Non Object Oriented, N_f= {c, cobol, fortran, pascal}

Entire UTHM Ontology is shown in hierarchy as Fig. 2.

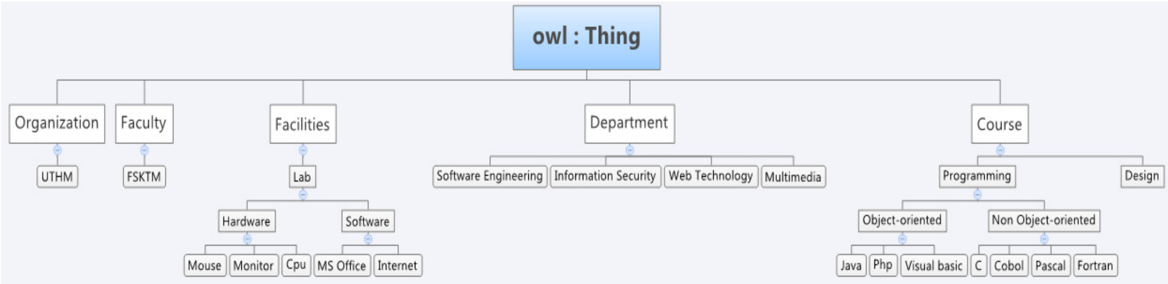

Fig. 2. Hierarchy for UTHM onto.

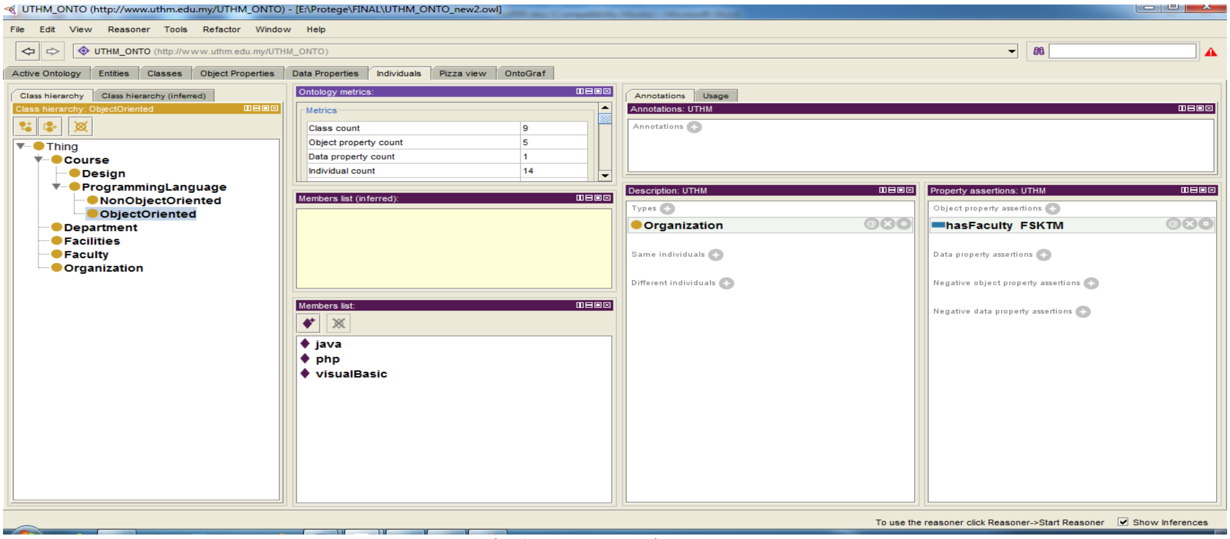

Fig. 3. UTHM onto in Protégé.

The design of this ontology is then developed using Protégé. It is stored in OWL/XML format with OWL Lite, OWL DL and OWL Full as the sublanguages. Compared to other formats, OWL is better in giving more information about the data. Advance reasoner and useful annotation has been the attractive features in describing ontology using this language.

Although this research use simple data and querying process, preparing for advance data acquisition can ensure in proper retrieval in later stage. Data regarding Class, Subclass, Instance and Properties are transformed into ontology file in Fig. 3.

Viewing this file in the form of Web Ontology Language (OWL) is shown in Fig. 4. Stored data in this format will be accessed online or downloaded in a server before manipulated in the application.

Fig. 5 shows visualization of the finalized UTHM Onto. OntoGraph features in Protégé is able to give a clearer picture of ontology. Additional information can be obtained by hovering mouse on each relationship.

V. Implementation of UTHM Ontology

Reusing concept appearing in the ontology has made it so widely accepted in the semantic world. Applications can easily exploit the features. Reading or engaging to UTHM Onto needs Jena API. Fig. 6 is the snippet used to call the ontology.

This ontology will be used in the hybrid Semantic Web Search. In order to manipulate Similarity Measurement, depth or superclasses is counted. There are two methods to fulfill the task. First is using built-in Jena inference, OWL_DL_MEM_RULE_INF. It is shown in Fig. 7.

```
<Ontology
xmlns="http://www.w3.org/2002/07/owl#"

xml:base="http://www.uthm.edu.my/UTHM_ONTO"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-sche
ma#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-sy
ntax-ns#"
xmlns:xml="http://www.w3.org/XML/1998/namespac
e"

ontologyIRI="http://www.uthm.edu.my/UTHM_ONTO"
>
    <Prefix name="xsd"
IRI="http://www.w3.org/2001/XMLSchema#"/>
    <Prefix name="swrlb"
IRI="http://www.w3.org/2003/11/swrlb#"/>
    <Prefix name="owl"
```

Fig. 4. UTHM onto in OWL syntax.

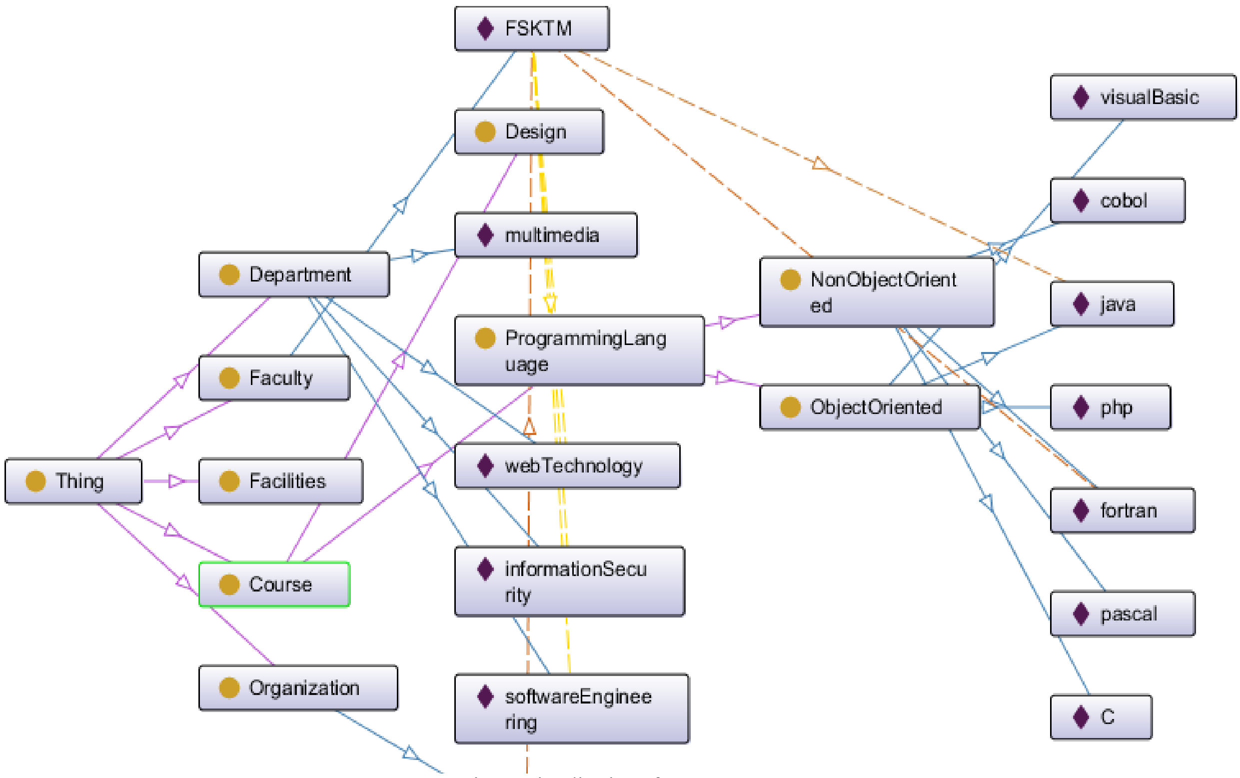

Fig. 5. Visualization of UTHM onto.

```
public class bacaSuperclass {
private static final String URI =
"http://www.uthm.edu.my/UTHM_ONTO#";

public static void main(String[] args) {OntModel
model =
ModelFactory.createOntologyModel(OntModelSpec.
OWL_DL_MEM_RULE_INF);

model.read(FileManager.get().open("data/UTHM_O
NTO.owl"), null);
```

Fig. 6. UTHM onto utilization.

```
public static void main(String[] args) {
    OntModel model =
ModelFactory.createOntologyModel(OntModelSpec.
OWL_DL_MEM_RULE_INF);

model.read(FileManager.get().open("data/UTHM_O
NTO.owl"), null);

System.out.println("\n-- Individual Computer
--");
Individual data = model.getIndividual(URI +
"Computer");
ExtendedIterator<Resource> classes =
data.listRDFTypes(false);

System.out.println(classes.toList().size());
```

Fig. 7. Jena inference.

Other option is to use SPARQL Query, an RDF query language. From the statement below, it will produce the same output. This option has its own advantages compared to the earlier mentioned inference. It has the ability to do complex query. However for simple searching, inference is usually accepted.

```
SELECT ?x
WHERE {
  ?x rdfs :superClassOf uthm:Faculty
}
```

Both techniques will produce same result as in Fig. 8.

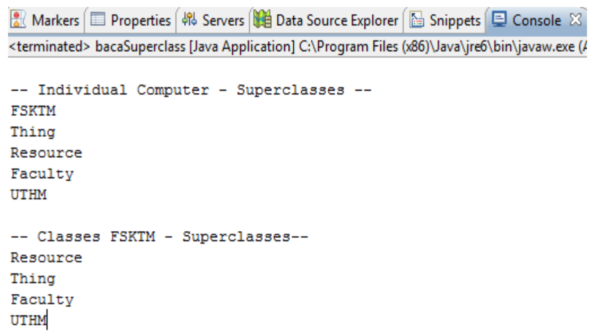

Fig. 8. Results in getting superclasses from UTHM onto.

VI. Conclusion and Future Work

In this paper we have presented an approach to develop ontology, named UTHM Onto. Designing process of the ontology is discussed. Specific editor is utilized throughout the process with certain ontology language. This paper also considers the development and integration part. It uses reusing concept from available online domain ontology. The output is a query to give depth and superclasses. Two methods are evaluated, using reasoner and SPARQL Query. This specific ontology is developed to utilized user profiling method. User profiling is the suggested technique to be merged with lexical ontology in giving more personalized results of web searching. Practically, collaboration of WordNet, ontology and Semantic Search Engine will give better results and it will be implemented in the future work.

ACKNOWLEDGMENT

This work is supported by Exploratory Research Grant Scheme (ERGS), Vote Number E006, Ministry of Higher Education, Malaysia and Universiti Tun Hussien Onn Malaysia (UTHM).

REFERENCES

[1] N. Abdullah and R. Ibrahim, "Knowledge retrieval using hybrid semantic web search," in *Proc. 2012 International Conference on Computer & Information Sciences*, pp. 61-65, 2012.
[2] Protege. (2013). [Online]. Available: http://protege.stanford.edu/
[3] Ontology. (2013). [Online]. Available: http://en.wikipedia.org/wiki/Ontology
[4] T. R. Gruber, "A translation approach to portable ontology specifications knowledge acquisition," vol. 5, no. 2, pp. 199-220, 1993.
[5] C. Edwards, "Analysis: Semantic web's hidden meanings," *Engineering and Technology*, vol. 5, no. 16, pp. 52-53, 2010.
[6] J. Cardoso, "The semantic web vision: Where are we?" *IEEE Intelligent Systems*, vol. 22, no. 5, pp. 84-88, 2007.
[7] M. Hepp, "Ontologies: State of the art, business potential, and grand challenges," in *Ontology Management. Semantic Web, Semantic Web Services, and Business Applications*, M. Hepp, P. D. Leenheer, A. D. Moor and Y. Sure, Ed., New York: Springer, 2008.
[8] H. D. Rodriguez, G. M. Luna, and J. O. O. Aguirre, "Ontology-Based Knowledge Retrieval," presented at the 2008 Seventh Mexican International Conference on Artificial Intelligence, 2008.
[9] J. D. Bruijn, "Using ontologies," *Enabling Knowledge Sharing and Reuse on the Semantic Web*, vol. 5, no. 2, pp. 199-220, 2003.
[10] V. Mascardi, V. Cordì, and P. Rosso, "A comparison of upper ontologies," *Technical Report*, pp. 1-15, 2007.
[11] D. Aquin and N. F. Noy, "Where to publish and find ontologies? A survey of ontologies libraries," *International Journal of Scientific & Technology Research*, vol. 11, pp .96-111, 2012.
[12] J. Cardoso, "The semantic web vision: Where are we?" *IEEE Intelligent Systems*, vol. 22, no. 5, pp. 84-88, 2007.
[13] E. Simperl, "Reusing ontologies on the semantic web: A feasibility study," *Data and Knowledge Engineering*, vol. 68, no. 10, pp. 905-925, 2009.
[14] K. Wecel, "Towards an ontological representation of knowledge on the web," in *Knowledge-based Information Retrieval and Filtering*, W. Abramowicz, Ed. From the Web, USA, Kluwer Academic Publisher, 2003.

Noryusliza Abdullah was born in Johore, Malaysia on July 20th, 1979. She obtained her first degree in computer science from the University of Science Malaysia (USM) in 2002. She received a masters degree in information technology from University of Technology Malaysia (UTM) in 2006. She is an information technology officer in University Tun Hussein Onn Malaysia (UTHM) and currently a PhD candidate in Faculty of Computer Science and Information Technology in the same university. While working as an Information Technology Officer, she was involved in several projects such as Student Information System, Inventory System and Web-based application. Her current research interests are Semantic search, Semantic web, Information retrieval and Cloud computing. Mrs. Noryusliza is a Program Committee to the First International Conference on Advance Data and Information Engineering (DaEng2013).

Rosziati Ibrahim is with the software engineering Department, Faculty of computer science and information technology, Universiti Tun Hussein Onn Malaysia (UTHM). She obtained her PhD in software specification from the Queensland University of Technology (QUT), Brisbane and her MSc and BSc (Hons) in computer science and mathematics from the University of Adelaide, Australia. Her research area is in Software Engineering that covers Software Specification, Software Testing, Operational Semantics, Formal Methods, Data Mining and Object-Oriented Technology.