A Novel Ontology Design and Comparative Analysis of Various Retrieval Schemes on Education Domain in Protégé

Neera Chaudhary, Suresh Kumar, and Sachi Gupta

Abstract The goal of Semantic Web ensures that the computer is able to recognize the Web data and thus can offer people with a variety of intelligence service. Linked data is the data of today's web. As time passes, the world is changing and the technologies are developing in the same way. The computers are developing, and from the isolated version, they have been entered in the field of networks of information exchange. As the data available on the Web is increasing, it is becoming difficult to handle it and retrieve it relevantly, so Semantic Web helps to search semantically as data on it is available in form of linked data (ontology).In this paper, we have proposed an ontology on education domain with the help of Protégé tool.

Keywords Semantic Web · Ontology · Query

1 Introduction

In the present environment, we can transform a massive library of interlinked documents, via computers, and present to people. Basically it has grown from hypertext system so everyone can contribute or share to it. The drawback is that liability and the quality of information cannot be same or guaranteed [1]. So, there is a need to extent the version of HTML. This was the main reason for defining another version called XML (extensible markup language) which has the task-specific extension and arbitrary domains. Semantic Web (SW) is an XML application, whose prime objective is to develop and make the present Web more semantically richer [1]. From the time of

N. Chaudhary (✉) · S. Gupta
Raj Kumar Goel Institute of Technology Ghaziabad, Uttar Pradesh, India
e-mail: neerachaudhar18@gmail.com

S. Gupta
e-mail: sachifcs@rkgit.edu.in

S. Kumar
Ambedkar Institute of Advanced Communication Technologies & Research Government of NCT of Delhi, New Delhi, India
e-mail: sureshpoonia@aiactr.ac.in

S. Fong et al. (eds.), *ICT Analysis and Applications*, Lecture Notes in Networks and Systems 154, https://doi.org/10.1007/978-981-15-8354-4_48

the 1990s, AI researcher communities have investigated that ontology has become a thrust area of Semantic Web research [1] that helps to provide shared and common understanding of some specific domain and make deliberate communication between people and application system. This reason made ontology more popular [1].

Semantic Web consists of formal Resource Description Framework (RDF), and it helps to overcome the limitation of understanding information that is faced by the current web. It also consists of a set of languages known as OWL, and it is based on RDF models. Vocabularies are used in Semantic Web which describes and explains the concepts and relationship that are related to a specific area of concern. A vocabulary can be considered as a special form of ontology, sometimes also merely as a collection of URIs with a described meaning. Link data is current Web data. It is publicly available in RDF data, identified by URI. In linked data, data is linked to one another.

The entire idea of ontology may sound alike to the concept of RDF. Every ontology is an RDF graph [2]. Ontology can be defined in several ways. There are several definitions of ontology. One of the most popular definition is ontology as "An explicit specification of conceptualization" [2, 3]. The SW depends heavily on the formal ontologies that structure underlined data for the purpose of comprehensive and transportable machine understanding. Ontology is the branch of philosophy that studies the nature of existence and the structure of reality [4]. The term ontology is the backbone of Semantic Web. It gives a concise and systematic means for defining the semantics of Web resources and describes the relevant domain concepts and properties of those concepts [2]. The Semantic Web understands the content of Web resources and combines and relates the content of other resources so we need a system which should be able to interpret the semantics of each resource so that it can represent the content of those resource [3, 4].

2 Designing Proposed Ontology

Education domain is one of the major areas of research today. It consists of a group of institutions such as schools, colleges, universities, ministries of education and teacher training institutions. Its prime objective is to provide right to education to all children as well as young people. It covers a wide range of people, for example, students, teachers of school and professors of university, colleges, etc. [4, 5].

The main objective to design this ontology is to integrate society which interacts a large number of stakeholders like schools, colleges and universities. Moreover, it can be extended to include parents, local communities like MCD schools, RWA institutions and pre-nursery and primary schools [6].

Designing an ontology depends on various components such as class, sub-class, individuals, attributes, relations, functions and restriction terms, rules, axioms and events. The common ontology engineering methodology [7] containing several steps is mentioned below which helps to design and share structural similarities of an effective ontology of any domain.

2.1 Ontology Engineering Process

Ontology engineering process basically contains the following five steps in ontology development [7].

Step-1: Selection of domain.
Step-2: Define classes and sub-class related to the domain.
Step-3: Define properties (Object/Data) of each class and sub-class as much possible.
Step-4: Define Individual or instance of each of the objects.
Step-5: Make relationship among classes and sub-classes.

2.2 Ontology for Education Domain

All the above steps are linked to each other. Protégé tool provides a flexibility to design an ontology. Follow the above steps. We design an ontology for education domain. The proposed ontology is mentioned in the following Fig. 1.

The designed ontology is saved in OWL extension file format. For this, it uses two main Semantic Web standards RDF and OWL. Resource Description Framework (RDF) specifies the metadata and describes Web resources which is processed by machines. RDF has many applications in search engines, pattern recognition and retrievals. There are three parameters in RDF, i.e., subject, predicate and object called triplets. OWL is very rich with respect to properties. It is the extension of RDF and has additional properties also to design ontologies in more effective manner [1, 2].

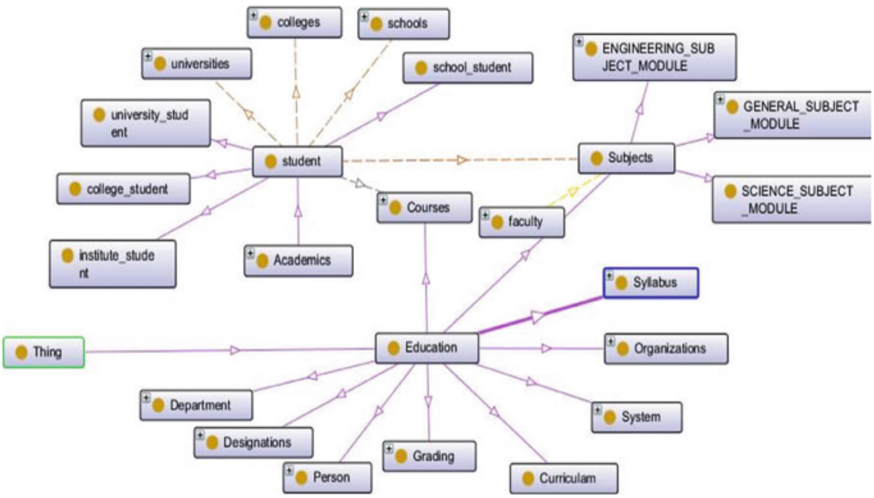

Fig. 1 Proposed ontology for education domain

3 Ontology Analysis

The proposed ontology has number of objects as well as data properties used with its domain and ranges. The following Table 1 represents the limited version of property, domain and ranges which is mentioned below.

Object properties map instance to other instance. Data properties are related to instance to literal data/constant value. In the case of the domain property, when we map a subject to an object using a property with the associated attribute, then the subject qualifies as a type of thing which specifies in the domain. The range works exactly like the domain, but it is applied to the object of the statement and not the subject [8, 9].

After designing the proposed ontology, there is an ontology matrices between the % of count of each matrices components and matrices are mentioned in the following Fig. 2. The figure shows that in our proposed ontology, class axiom is used almost 50.17%, Logical Axioms—25.41%, Class Count—15.84%, Object Property—1.82%, Data Property—1.82% and Individuals—4.95% in education domain.

Table 1 Property-domain-range description

Object property			Data property
Property	Domain	Range	
admitted_in	Student	School, university, college	topDataProperty course empno fcontact fdepartment fdesignation fname saddreess scontact sname ssn ssubject
affiliated_to	Student	School, university, college	
has	Faculty	Department	
has_department	Student, faculty	Department	
has_designation	Faculty	School, university, college	
has_faculty	Department	School, university, college	
opt	Faculty	Subject, course	
regiters_in	Student	Course	
studies	Student	Subjects	
teaches	Faculty	Courses, subjects	

Fig. 2 Ontology matrices

4 Empirical Results

Query language used for making queries in search engine search index is used to retrieve information. Formally Query language is defined in a context-free grammar (CFG). Query language tools [10] are used by users in textual, visual/UI or speech form. Query language mostly used forms to search are keyword and sentence based. In this paper, we have used keyword search mechanism.

To find the relevancy of retrieved result in protégé, we have done the analysis of three (03) Query languages which are OntoGraf Query, DL Query and SPARQL Query. These languages are like in-built tool in protégé that empower protégé users to Query the ontology data set and retrieve information. With the help of these Query tools, precise queries can be constructed using different searching criteria (searching keywords). In this paper, the table of all the three Query languages (QL) retrieved result using the same five (05) search keywords that are Student, faculty, Course, Subject, Department and Organization, in correspondence to the criteria used for searching by each QL is explained.

4.1 *OntoGraf Query Analysis*

OntoGraf [11] supports OWL ontologies for interactively navigating relationships. Table 2 describes the OntoGraf Query result description. Table 2 represents five search term or keywords used for searching. OntoGraf Query search is done based on the following five criteria mentioned below:

- Contains—In this, information containing the search terms are retrieved.
- Start With—In this, information starting with the search terms are retrieved.
- Ends With—In this, information ending with the search terms are retrieved.
- Exact Match—In this, information exactly matching with the search term are retrieved.

Table 2 OntoGraf Query result description

Search term	Criteria for OntoGraf Query search					Result
	Contains	Start with	Ends with	Exact match	Regular expression	
Student	8	4	5	1	8	26
Faculty	7	3	5	1	7	23
Course	7	3	4	0	7	21
Subject	4	1	0	0	4	09
Department	5	1	5	1	5	17
Organization	1	1	1	1	1	05

Table 3 DL Query result description

Search term	Criteria for DL Query search						Result
	Super classes	Ancestor classes	Equivalent classes	Sub-classes	Descendant classes	Individuals	
Student	1	4	1	4	4	2	16
Faculty	1	4	1	4	4	1	15
Course	1	2	1	4	4	1	13
Subject	1	2	1	3	10	5	22
Department	1	2	1	4	4	0	12
Organization	1	2	1	8	31	3	46

- Regular Expression—In this, information in the form of sequence of characters matching with the search terms are retrieved.

Based on the above explained criteria, a number of result retrieved from the novel designed ontology on education domain are mentioned in Table 2.

4.2 DL Query Analysis

DL Query [12] in protégé is used for searching a classified ontology. Table 3 describes the DL Query result description. Table 3 also represents the same five search terms or keywords used for searching. DL Query search is done based on the following six criteria explained and described below:

- Super classes—Retrieves the super class related to search term.
- Ancestor classes—Retrieves all the super classes related to search term.
- Equivalent classes—All the classes equivalent to search term are retrieved.
- Sub-classes—Sub-class that directly relates to the search term is retrieved.
- Descendant classes—All the sub-classes that relates to the search term are retrieved.
- Individuals—Individuals related to class belonging to search term are retrieved.

Based on the above explained criteria, a number of result retrieved from the novel designed ontology on education domain are mentioned in Table 3.

4.3 SPARQL Analysis

SPARQL [13, 14] is the Query language of the Semantic Web. It is an in-built tool for querying from ontology. It retrieves the user Query result in triple—subject, predicate and object. Based on the proposed education ontology, we apply command

Fig. 3 Query result in response to the user Query

Table 4 SPARQL Query result description

Search Term	Subjects	Objects	Results
Student	4	1	5
Faculty	4	1	5
Course	4	1	5
Subject	3	1	4
Department	4	1	5
Organization	6	1	7

and generated the Query results. The sample of Query in SPARQL is mentioned in the following Fig. 3 which represents the Query result output.

The following Table 3 mentioned the result based on the search terms in SPARQL. Here, based on the search terms, subjects' frequencies are presented by individual objects. The results are retrieved on the bases of one to one, one to many and many to many relationships (Table 4).

4.4 Overall Analysis

Various users are biased to use SPARQL, DL Query, OntoGraf Query on the basis of results extracted by these tools for knowing their education information. It was observed from different empirical result tables that SPARQL has less retrieved results in comparison to other Query tools like OntoGraf and DL Query. The fact that SPARQL could be far appropriately suited in education-based ontology for extracting imperative information rather than the other knowledge builder tools could be concluded by analyzing the analogy in Fig. 4.

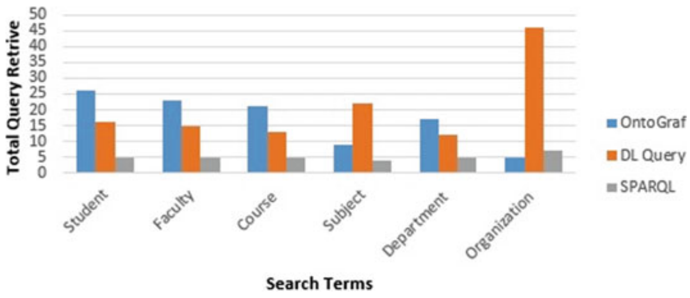

Fig. 4 Overall analysis of OntoGraf, DL and SPARQL Query result

5 Conclusion and Future Scope

In this paper, we have proposed an ontology on education domain. We have performed six queries (keywords) [15] on this ontology using three tools. Then after, we have compared the results of these three tools. In our empirical analysis, we have concluded that the results obtained through SPARQL tool are better than DL Query and OntoGraf Query tool. As the result retrieve through DL Query and OntoGraf Query tool provides semantically rich response but these are not much expressive whereas SPARQL tool provides semantically as well as more expressive power of producing the result.

For future, we have planned to use results of all these tools as an application for ranking. We are developing an effective mathematical method for ranking in information retrieval, which will be published in near future.

References

1. Giri, K. (2011). Role of ontology in semantic web. *DESIDOC Journal of Liberty & Information Technology, 31*(2), 116–120.
2. Synak, M., Dabrowski, M., & Kruk, S. R. Semantic Web and Ontologies. In *Semantic digital libraries,* (pp. 41–54).
3. Bamashmoos, F., Holyer, I., Tryfonas, T., & Woznowski, P. (2017). Towards secure SPARQL queries in semantic web applications using PHP. In *2017 IEEE 11th International Conference on Semantic Computing*.
4. Anand, S., & Verrma, A. (2010). Development of ontology for smart hospital and implementation using UML and RDF. *IJCSI International Journal of Computer Science Issues, 7*(5), 206.
5. Ceron-Figueroa, S., Lopez-Yanez, I., Alhalabi, W., Camacho-Nieto, O., Villuendas-Rey, Y., Aldape-Perez, M., et al. (2017). Instance-Based ontology matching for e -learning material using an associative pattern classifier. *Computers in Human Behavior, 69,* 218–225. https://doi.org/10.1016/j.chb.2016.12.039.
6. Jacob, E. K. Ontologies and the semantic web. In *April/May 2003-Bulletin of the American Society for Information Science and Technology*.

7. Liu, Y., Wang, Y., & Yang, D. (2014). DLOSS: An improved ontology-based semantic web search system. In *Fifth International Conference on Intelligent Control and Information Processing,* (pp. 18–20). China: Dalian, Liaoning.
8. http://protegeproject.github.io/protege/views/object-property-characteristics/.
9. https://www.w3.org/TR/owl-ref/.
10. Goncalves, R. S., Tu, S. W., Nyulas, C. I., Tierney, M. J., & Musen, M. A. (2017). An ontology-driven tool for structured data acquisition using Web forms, Goncalves et al. *Journal Biomedical Semantics, 8,* 26 https://doi.org/10.1186/s13326-017-0133-1.
11. https://protegewiki.stanford.edu/wiki/OntoGraf.
12. https://protegewiki.stanford.edu/wiki/DLQueryTab.
13. https://www.w3.org/TR/rdf-sparql-query/.
14. https://en.wikipedia.org/wiki/SPARQL.
15. Singh, P. K., Agarwal, G., & Gupta, S. (2013). A new ranking technique for ranking phase of search engine: Size based ranking algorithm (SBRA). *International Journal of Computer Applications, 82*(5), 0975–8887.