Volume 2, Issue 8, August 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Knowledge Representation of 'Published Articles' in Semantic Web using Upper Ontology

Ms. Swaminarayan Priya R.*
Head & Asso. Prof.,
Department of MCA,
ISTAR, India

Miss Nehal Daulatjada,
Lecturer,
SEMCOM College,
India

Dr.P V Virparia,
Associate Professor,
GDCST,
India

Dr.V R Rathod,
Retired Prof & Head,
Dept. of CSE, Bhavnagar
University, India

***Abstract* - The current searches require keywords to be entered in the search engines where irrelevant searches are also displayed. If there is a mechanism by which the computer can understand the information, then that information can be searchable in the relevant context. The task of the end user can be reduced. Various technologies such as XML Schema, Resource Description Framework (RDF), Resource Description Framework Schema (RDFS), Web Ontology Language (OWL) etc. have been developed for knowledge representation on the semantic web.**

Ontology is used to represent Knowledge of any domain as a set of concepts and the relationships between those concepts. An upper ontology describes very general concepts that are the same in all knowledge domains. One of the important functions of an upper ontology is to support broad Semantic Interoperability among large number of Ontologies. In this paper we have developed ontology for the articles published by faculty members. Each publication is described using Dublin core upper ontology. This enables the well-formed information to be semantically searchable on the web environment. For our purpose, we have used Protégé software to develop the ontology. The testing of the ontology is done using SPARQL querying language. This paper is an attempt to represent knowledge of published article on Semantic Web and try to accomplish the vision of semantic web.

Keywords: **Semantic web, Ontology, Upper ontology, Protégé, SPARQL, Metadata**

I. INTRODUCTION

1.1 The Semantic Web

The Semantic Web is the vision of World Wide Web Consortium (W3C). The Semantic Web offers a common framework which permits data to be shared and reused by enterprise, community boundaries and application [1].

The W3C defines the main goals of the Semantic Web as follows: "The Semantic Web is a Web of Data. The vision of the Semantic Web is to extend the principles of the Web from documents to data. Data should be accessed using the general Web architecture e.g., URIs; data should be related to one another just as documents (or portions of documents) are already. This also means creation of a common framework that allows data to be shared and reused across application, enterprise, and community boundaries, to be processed automatically by tools as well as manually, including revealing possible new relationships among pieces of data. In order to achieve these goals, it is necessary to define and describe the relations among data on the Web. This is not unlike the usage of hyperlinks that connect the current page with another page: the hyperlinks define a relationship between the current page and the target. On the Semantic Web, such relationships can be established between any two named resources or values and the relationship itself (i.e, the link) is also named. By contrast, a link on the traditional Web is not named, which means that the significance or meaning of that link needs to be deduced by the human reader. The naming and defining those relations explicitly enables better and automatic interchange of data. RDF, which is one of the fundamental building blocks of the Semantic Web, gives a formal definition for that interchange."[2]

In the Semantic Web, complex and evolving concepts, resources and relationships could have a deep impact on how we access data and use data [3]. Much efforts have been made to accomplish the vision of Semantic Web i.e. machine-readability. Some of the industries are shifting from a "wait-and-see" approach [4] to the real-world deployment of applications which will give them a competitive advantage [5].

1.2 Ontology

An Ontology is a formal description of important concepts in a specific domain [6]. An ontology is a formal explicit representation of concepts in a domain, characteristics of concepts can be described by the properties of concepts. Concept is also known as class, properties are also known as roles [7]. An ontology defines a common vocabulary for sharing of information in a domain. It contains machine-interpretable definitions of basic concepts in the domain and relations among concepts. An ontology offers a means to share a common understanding of the structure of information among software agents as well as among people. An ontology is a way of separating operational knowledge from domain knowledge [8,9].

1.3. Upper Ontology

Upper ontologies are quickly becoming a key technology for integrating heterogeneous knowledge coming from different sources. In fact, they may be used by different parties involved in a knowledge integration and exchange process as a reference, a common model of the reality. The definition of upper ontology, (also named top-level ontology, or foundation ontology) given by Wikipedia [10] is "an attempt to create an ontology which describes very general concepts that are the same across all domains. The aim is to have a large number on ontologies accessible under this upper ontology"[11].

There are two kinds of benefits from using these formal upper ontologies. First, they provide a rich vocabulary for describing the systems we are building and working with, so they can assist us in their design and in understanding them. Second, they form the basis of a suite of abstract data types which can be provided as libraries to the designers of agents which will automatically interoperate with our systems and with each other [12]. The potential advantages of ontology for the purposes of information management are obvious. Each group of data analysts would need to perform the task of making its terms and concepts compatible with those of other such groups only once – by calibrating its results in the terms of the single canonical backbone language. If all databases were calibrated in terms of just one common ontology (a single consistent,stable and highly expressive set of category labels), then the prospect would arise of leveraging the thousands of person-years of effort that have been invested in creating separate database resources in such a way as to create, in more or less automatic fashion, a single integrated knowledge base of a scale hitherto unimagined, thus fulfilling an ancient philosophical dream of a Great Encyclopedia comprehending all knowledge within a single system.

The obstacles standing in the way of the construction of a single shared ontology in the sense described are unfortunately prodigious. Consider the task of establishing a common ontology of world history. This would require a neutral and common framework for all descriptions of historical facts, which would require in turn that all legal and political systems, rights, beliefs, powers, and so forth, be comprehended within a single, perspicuous list of categories.

The top-level ontology would then be designed to serve as common neutral backbone, which would be supplemented by the work of ontologists working in more specialized domains on, for example, ontologies of geography, or medicine, or ecology, or law, or, still more specifically, ontologies of built environments[13], or of surgical deeds[14,15].

1.4 Dublin Core

Metadata is defined as data providing information about one or more aspects of the data. Each metadata schema has a limited number of elements with their meaning [16]. The Dublin Core (DC) metadata standard is a simple yet effective element set for describing a wide range of networked resources. The Dublin Core standard includes two levels: Simple and Qualified. Simple Dublin Core comprises fifteen elements; Qualified Dublin Core includes three additional elements (Audience, Provenance and RightsHolder), as well as a group of element refinements (also called qualifiers) that refine the semantics of the elements in such a way that they may be useful in resource discovery. The semantics of Dublin Core have been established by an international, cross-disciplinary group of professionals from librarianship, computer science, text encoding, the museum community, and other related fields of scholarship and practice. Another way to look at Dublin Core is as a "small language for making a particular class ofstatements about resources".

Although the Dublin Core was originally developed with an eye to describing document-like objects (because traditional text resources are fairly well understood), DC metadata can be applied to other resources as well. Its suitability for use with particular non-document resources will depend to some extent on how closely their metadata resembles typical document metadata and also what purpose the metadata is intended to serve.

As given in Dublin Core Metadata Initiative, following are the main goals of Dublin Core [17]:

- **Simplicity of creation and maintenance**: The element set of Dublin Core is small and simple which allows a non specialist to create descriptive records for information resources very easily and inexpensively.
- **Commonly understood semantics**: Search of information over the Internet is delayed due to differences in terminology and descriptive practices from one field of knowledge to the other. Element set of upper ontology increases the visibility and accessibility of all resources.
- **International Scope:** The Dublin Core element set is created in English, Finnish, Norwegian, Thai, Japanese, French, Portuguese, German, Greek, Indonesian, and Spanish.
- **Extensibility**: The element set of Dublin Core can be easily extended as the developer of DC provides a mechanism for the extension of DC element set for additional resource discovery needs. Metadata experts of other communities can create and administer additional metadata sets according to their requirement. Metadata elements from these sets could be used in combination with Dublin Core metadata to meet the need for interoperability.

Main advantages of DC are: [18]: Commonly understood semantics, an economical alternative that can accommodate richer description models, improves the prospects for cross-disciplinary search, users retrieve documents sets that represent multiple occurrences of an element or qualifier within a database.

1.5 Protégé

Protégé is a java based free, open source ontology editor and knowledge-base framework. It is an extensible and provides a plug and play environment which make it more flexible for prototyping and application development. In the Protégé platform ontology can be modelling by two ways - the Protege-Frames and Protege-OWL editors. The ontology which is developed by Protege can be exported into a variety of formats including Resource Description Framework Schema (RDFS), Web Ontology Language (OWL) and XML Schema [19].As per survey done by Jorge Cardoso, more than 75% ontology developer used Protégé as Ontology development tool[20] Another online survey done by M. Rahamatullah Khondoker, Paul Mueller, they asked question to the Ontology developers that "Which Ontology development tool did you try most?" They permitted only single response for this question, then 24 out of 32 Ontology developers used Protégé as Ontology Development tool. Due to this online survey, we also used Protégé as Ontology development tool for the development of ontology for published articles on Web [21].

II. CURRENT PROBLEM

A lot of efforts have been done by the scientists and researchers in the area of Computer Science to achieve the Semantic Web vision. In this paper, we have developed ontology for published research articles on Internet using Protege tool and searching them semantically using SPARQL query language. Current Web supports only text-based search, so the major problem of the World Wide Web is the keywords used in searching have different meanings in different context which cannot be interpreted by the machine. Hence additional efforts are required by human users to retrieve exact information from the Web.

Solution towards this problem is to develop a common framework which would allow the search process to infer knowledge from it. In our context, if an article search is required, the keywords may bring many other irrelevant links in the result. One of the solutions is to represent knowledge using Semantic Web technologies like Resourve Description Framework, Resource Description Framework Schema or Web Ontology Language. Here we can develop Ontology for published articles but major issue of this Ontology is Ontology mapping problem – i.e. one entity may be represented differently by different Ontology developers. A solution to this would be to develop Ontology for published articles by adding an upper ontology which describes what is published, which in turn can be used by the search engines as additional search parameters. This would help the number of results generated based on the context of the search.

III. PROTOTYPE MODEL TOWARDS SOLUTION

The problem of searching for publication details by machines was undertaken. We have developed a prototype which includes two entities:

Faculty: firstName, lastName, birthDate, designation, orgName

Article: articleType, category, keywords, publishedIn

Metadata for each Article: (Using Dublin Core metadata) Creator, date, description, publisher, title

In the **Faculty** class, information regarding the faculty members is stored. In the **Article** class, information regarding articles published is added. The main details of the Articles are stored using Dublin Core metadata in the annotations. These annotations are machine-readable metadata. A faculty can publish one or more articles.

If all the information is available in Article class while searching the current search engines would display the results which may include additional results with partial keywords matches. With our prototype, we are associating metadata or semantics using Dublin Core which contains additional information (metadata) about the Article. The advantage is that in addition to the normal searching by humans, even an automated application can search for information without human help since semantics is associated with the information.

In the following query, we are using two concepts. The first is using an inverse relationship and secondly, we are using metadata through this relationship.

Query - 1: Display creator name and organization name for all articles.

```
SELECT ?article ?oname ?creator
WHERE
{
    ?article a:isPublishedBy ?faculty.
   ?article dc:creator ?creator.
    ?faculty a:orgName ?oname.
}
```

Fig. – 1: SPARQL Code for Query - 1

TABLE – 1: RESULT OF QUERY – 1

Article	**OrgName**	**Creator**
Mining	SVIT	Amisha Shingala
Development of Institute Ontology	SEMCOM	Nehal
Development of Institute Ontology	SEMCOM	Nehal
University Ontology	SEMCOM	Nehal
University Ontology	Institute of Science and Technology for Advanced Studies and Research	Swaminarayan Priya R.

Query-2: Display all articles and faculty details of all articles.

```
SELECT ?atype ?cate ?key ?pubin ?fname ?lname
?bdate ?desg ?oname
WHERE
{
   ?article a:isPublishedBy ?faculty.
   ?article a:articleType ?atype.
```

Fig. 2: SPARQL Code for Query - 2

```
   ?article a:category ?cate.
   ?article a:keywords ?key.
   ?article a:publishedIn ?pubin.
  ?faculty a:firstName ?fname.
  ?faculty a:lastName ?lname.
  ?faculty a:birthDate ?bdate.
 ?faculty a:designation ?desg.
  ?faculty a:orgName ?oname.
}
```

Fig.3: SPAQL Code for Query - 3

TABLE – 2: RESULT OF QUERY – 2

atype	cate	key	pubin	fname	lname	bdate	desg	oname
Survey Paper	Computer Science	Data Mining	International Journal of Computer Science	Amisha	Shingala	1982-04-02	Assistant Professor	SVIT
White Paper	Computer Science	Ontology, Mapping	IEEE Journal	Priya	Swaminarayan	1978-04-10	Associate Professor	ISTAR
White Paper	Computer Science	Ontology, Mapping	IEEE Journal	Nehal	Daulatjada	1980-04-09	Assistant Professor	SEMCOM
Research Paper	Computer Science	Ontology, Protege	National Journal of Information Science	Nehal	Daulatjada	1980-04-09	Assistant Professor	SEMCOM

Query – 3: Display all articles along with its metadata and faculty details of all articles.

```
SELECT ?atype ?pubin ?title ?creator ?date ?descr
?pub ?desg ?oname
WHERE
{
   ?article a:isPublishedBy ?faculty.
   ?article a:articleType ?atype.
```

Fig. 3 : SPAQL Code for Query – 3

```
?article a:publishedIn ?pubin.
   ?article dc:title ?title.
   ?article dc:creator ?creator.
   ?article dc:date ?date.
   ?article dc:description ?descr.
   ?article dc:publisher ?pub.
  ?faculty a:designation ?desg.
  ?faculty a:orgName ?oname.
}
```

Fig. 4 : SPAQL Code for Query - 3

TABLE – 3: RESULT OF QUERY – 3

Atype	Pubin	Title	Creator	Date	Descr	Pub	Desg	oname
Survey Paper	International Journal of Computer Science	Study of Data Mining techniques for decision making	Amisha Shingala	2012-01-01	Related to Data Mining in Industry	ACM	Assistant Professor	SVIT
White Paper	IEEE Journal	Analysis of Upeer Ontology	Nehal Daulatjada	2011-02-02	Study of Upper Ontology	IEEE Journal	Associate Professor	ISTAR
White Paper	IEEE Journal	Analysis of Upeer Ontology	Nehal Daulatjada	2011-02-02	Study of Upper Ontology	IEEE Journal	Assistant Professor	SEMCOM
Research Paper	National Journal of Information Science	Study and survey of Ontology Mapping	Nehal Daulatjada	2010-04-04	Paper on Ontology Mapping	Springer Journal	Assistant Professor	SEMCOM
Research Paper	National Journal of Information Science	Study and survey of Ontology Mapping	Swaminarayan Priya R.	2010-04-04	Paper on Ontology Mapping	Springer Journal	Assistant Professor	SEMCOM

Following is the screen shot of the implemented system which shows list of entities under subject section and two relationships between entities under association section. It also shows ten individuals of entities as well as relationships.

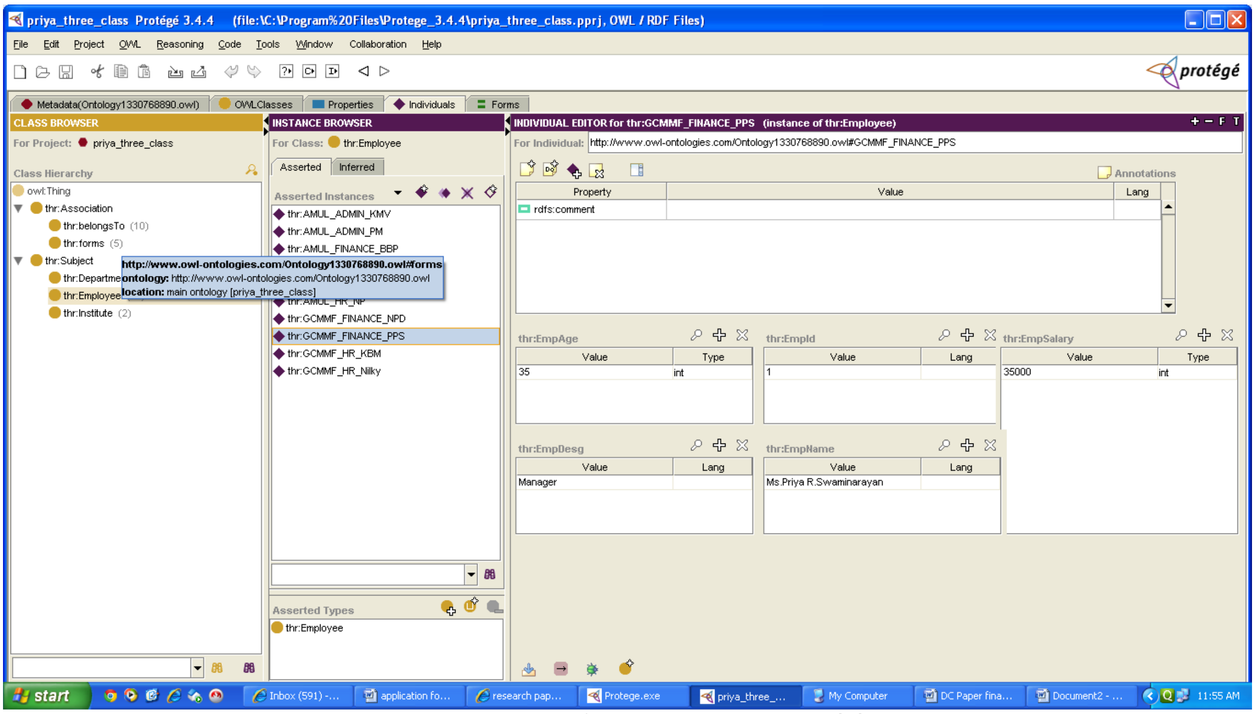

Fig. – 5: Screen shot of Implemented system in Protégé

IV. CONCLUSION AND FUTURE WORK

The sample ontology which we developed was tested by executing semantic queries against it. All the nodes as well as each instance metadata could be searched through the SPARQL query. This is sufficient to prove that the ontology is semantically searchable. If such ontology is available on the Semantic Web, then a semantic search engine would be able to search for it. The main advantages are reusability and semantically searchable.

In the era of Semantic Web, the ontologies have become a powerful tool for knowledge sharing and it also supports the semantic interoperability among heterogeneous distributed systems. Ontologies and agent technologies are essential part of the semantic web, and their combined use will make possible the sharing of heterogeneous, autonomous knowledge sources in a capable, adaptable and extensible manner. In the multi-agent system, Ontology is used to assist the interactions among different agents and improve the quality of the service provided by each agent.

Our future plan is to develop multi-agent system using FIPA Specification, so there is no need to fetch knowledge from Ontology by using semantic web query languages such as SPARQL, etc, but the data from Ontology can retrieved by the agents, so it can reduce human intervention and it can serve the goal of Semantic Web. Another future task would be to include another upper ontology FOAF, so that it is more semantically expressive and searchable.

REFERENCES

1. W3C, http://www.w3.org/2001/sw/
2. W3C, http://www.w3.org/2001/sw/SW-FAQ
3. http://wiki.cetis.ac.uk/images/1/1a/The_Semantic_Web.pdf The Semantic Web, Linked and Open Data - A Briefing PaperBy Lorna M. Campbell and Sheila MacNeill http://jisc.cetis.ac.uk JICS cetis - Centre for educational technology and interoperability standards
4. Cardoso, J., J. Miller, et al. Academic and Industrial Research: Do their Approaches Differ in Adding Semantics to Web Services. Semantic Web Process: powering next generation of processes with Semantics and Web services. Heidelberg, Germany, Springer-Verlag. LCNS 3387: 14-21,2005.
5. https://wiki.nci.nih.gov/download/attachments/19793270/sw-survey-2007.pdf Jorge Cardoso, "The Semantic Web Vision: Where are We?" IEEE Intelligent Systems, September/October 2007, pp.22-26, 2007
6. Fensel D., 2000. The semantic web and its languages. IEEE Computer Society 15, 6 (November /December), 67−73.
7. Noy N.F., McGuinnes D. L. 2001. ‘‘Ontology Development 101: A Guide to Creating Your First Ontology’’. Stanford Knowledge Systems Laboratory Technical Report KSL−01−05 and StanfordMedical Informatics Technical Report SMI−2001−0880, March.
8. Lili Aunimo, HAAGA- HELIA (2009)," Semantic modeling using RDF", DBTech Ext Workshop on Database Modeling and Semantic Modeling retrieved from http://www.dbtechnet.org/labs/sem_lab/SemanticModelingUsingRDF_Tutorial.pdf
9. Fonseca, F. Egenhofer M., Agouris, P., Camara G. 2002. Using Ontologies for Integrated Geographic Information Systems. Transactions in GIS, −(6):3 in print.
10. Wikipedia. Upper ontology – wikipedia, the free encyclopedia, 2006. [Online; accessed 15-December-2006].
11. A Comparison of Upper Ontologies (Technical Report DISI-TR-06-21) Viviana Mascardi, ValentinaCordì, Paolo Rosso http://www.disi.unige.it/person/MascardiV/Download/DISI-TR-06-21.pdf
12. http://www.loa.istc.cnr.it/Papers/ISIB-CNR-TR-20-02.pdf Use of Upper Ontologies for Interoperation of Information Systems: A Tutorial, Robert M. Colomb, Technical Report 20/02, ISIB-CNR, November, 2002
13. Bittner, Thomas 2001 "The Qualitative Structure of Built Environments," Fundamental Informaticae, 46, 97–126. Uses the theory of fiat boundaries to develop anontology of urban environments.
14. Rossi Mori, A., Gangemi, A., Steve, G., Consorti, F., Galeazzi, E. 1997 "AnOntological Analysis of Surgical Deeds", in, C. Garbay, et al. (eds), Proceedingsof Artificial Intelligence in Europe (AIME '97), Berlin: Springer Verlag.
15. http://ontology.buffalo.edu/smith/articles/ontology_pic.pdf Ontology by Barry Smith, Preprint version of chapter "Ontology", in L. Floridi (ed.),Blackwell Guide to the Philosophy of Computing and Information, Oxford: Blackwell, 2003, 155–166.
16. http://www.library.uq.edu.au/iad/ctmeta4.html AN INTRODUCTION TO METADATA Paper written by Chris Taylor Manager, Information Access Service, University of Queensland Library Revised: 29 July 2003
17. Diane Hillmann(2007) ,Dublin core metadata initiative retrieved from http://dublincore.org/documents/usageguide/
18. http://davinci.ohsu.edu/~appleyar/presentations_html/AMIA97Poster/sld011.htm
19. http://protege.stanford.edu/
20. Jorge Cardoso (2007), "The Semantic Web Vision: Where are We?" IEEE Intelligent Systems, September/October 2007, pp.22-26,2007. [Online] Available:citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97...rep...
21. M. Rahamatullah Khondoker, Paul Mueller (2010), "Comparing Ontology Development Tools Based on an Online Survey" Proceedings of the World Congress on Engineering 2010 Vol I WCE 2010, June 30 - July 2, 2010, London, U.K.
22. 20 Ontology Development 101: A Guide to Creating Your First Ontology Natalya F. Noy and Deborah L. McGuinness Stanford University, Stanford, CA, 94305http://protege.stanford.edu/publications/ontology_development/ontology101.pdf

