The Essence of Ontological Knowledge Representation in building a Semantic Web – A Comprehensive study

E.Arul[1], C.Murale[2], M.Maranco[3]

[1]*Assistant Professor (CSE& IT) Coimbatore Institute of Technology Coimbatore Tamilnadu India* earulcse@yahoo.com

[2]*Assistant Professor (CSE& IT) Coimbatore Institute of Technology Coimbatore Tamilnadu India* cmurale.chellamani@gmail.com

[3]*Assistant Professor (CSE& IT) Coimbatore Institute of Technology Coimbatore Tamilnadu India* marancit@gmail.com

Abstract–**With the advancements in retrieving efficient information, retrieval of Web pages which is of user intent has become the huge task. Although the keyword search provides a easier way, it suffers primarily due to the lack of precision and recall. It is practically always not possible to get the required Web results which the user is interested in, since the query may not match all the keywords and henceforth to overcome these disadvantages, a semantic web is always sought which retrieves the information based on interesting events. In order to build a semantic web we must first define the language representation which is necessary for understanding the relationship between entities, properties as well as other intrinsic ontological properties in developing a system structure. This paper discusses about the need of ontology for knowledge extraction, the language representation for ontology and the tools which are helpful to develop domain ontology.**

Index Terms— **Domain Ontology, Knowledge Representation, Ontology Web Language, Semantic Web.**

INTRODUCTION

With the developments in technology, computation has become increasingly complex and the need to serve the end user has become the predominant factor .For serving the user by means of retrieving the Web results for the query, there arises a possibility that no results can match to the intended query or uninterested results can find its place. In order to avoid ambiguity and fuzziness, ontology can always be built which represents the relationship among the intrinsic properties which in turn will help to serve better Web results to the user when the user makes use of semantic web.

A. ONTOLOGY

The word "ontology" is used with different meaning in different communications. Computational ontology are formally a means to analyze a system structure. The ontology engineer analyzes entities and organizes into concepts and relations. More precisely, ontology can be described as the formal specification of a shared conceptualization. For example the sentence "universe is made up of atoms and molecules" in which the term "universe" can be treated as entities or objects and the term "atoms and molecules" can be treated as a relationship to an entity. In other words, ontology describes the interconnection of several entities which are linked by a shared relationship which exists among them.

For the proper understanding of conceptualization, let us consider the ESPN Cricinfo website. There is always a manual annotation for each ball being bowled irrespective of the runs scored or wickets taken. There is always going to be more than a single commentator who does the annotation. The first commentator may refer the ball being bowled as a "ball". The second

commentator may refer to the ball as a "delivery". Humans may be able to know both these concepts but same is not the case with Web search engines or with the machines.

Therefore a concept has to be built for each entity which in this case has to be built for both the "ball" and the "delivery". After building the concept we can define the relationship between each entities, the properties of an entity, change of a state and other intrinsic properties like class, subclass etc. which may be helpful for building an ontology.

B.CHALLENGES WITH WORLD WIDE WEB (WWW)

In today's Web, meaning of Web content is not easily accessible due to the lack of semantics. In order to have a deeper understanding, consider these two sentences indicated by S1 and S2.

S1: Iam a bowler in Indian cricket.

S2: Iam a bowler in Indian cricket, you may think!!!!Well.....

The above two statements are not in a form which is easily machine processable. Hence there arises a need in which the information which needs to be processed has to represented in a machine understandable form.

In order to have a better understanding, consider the following problem: Is Prof. Natraj available for meeting on Thursday at 9 AM on Nov 22nd? The program which runs on World Wide Web would resolve the problem by asking the following questions:

Where does Prof. Natraj work?

Is Nov 22nd a holiday for the institute?

What are the published slots available on that day?

Which are the slots scheduled around 9 AM on that day?

Is Prof. Natraj engaged in any of these slots?

In order to avoid the above mentioned problems there is a need to use the Semantic Web[3].

C. SEMANTIC WEB

The Semantic Web represents the Web content which is more easily accessible in a machine readable form. It also makes use of intelligent techniques for representation of information with the notion that the Semantic Web has got nothing to do with Artificial Intelligence except it makes use of Artificial Intelligence techniques for knowledge representation. The Semantic Web is not a competitor to World Wide Web and can be considered as a deeper web for extracting meaningful results.

The Semantic Web also retrieves the interesting events which may differ from person to person. For example when a ball by ball annotation is being annotated on sites like ESPN Cricinfo. Person A may be interested in knowing the details about the player i.e., the person may be interested in knowing about Sachin Tendulkar irrespective of whether he is batting or bowling. He may be interested in knowing the batting statistics or bowling statistics of that player. Person B may be interested in knowing the number of runs scored in an innings or in an entire match. Person C may be interested in knowing the number of wickets which felt during that session or in an entire match. Hence the Semantic Web is required which is helpful to retrieve the interesting events for each person based on interestingness measure.

IJSER

II. EXISTING SYSTEM

The ontology which is now being developed can be classified into upper ontology and lower ontology in which the former deals with generalization while the latter deals with specialization. Bungean Ontology is the first adopted ontology which was more considered to be a general systems theory in which the postulates are widely accepted statements about real world phenomenon. They are everyday experiences, facts and observations. Wand and Weber adopted it to Software Systems later.

III. RELATED WORK

In order to build ontology, there is always a need to identify the intrinsic ontological categories and the primitive relational categories as well as the rules which are needed for ontological categorization which would uniquely identify the intrinsic ontological categories like thing, event, property and state. For developing ontology we also require the language representations which in this case can be Extended Markup Language (XML), for providing the service from the service provider to the service consumer and can act as a communication bridge between the provider and the consumer, Ontology Web Language (OWL) which is meant especially for building ontologies. We can also make use of Resource Description Framework (RDF) and a schema for RDF in developing ontology.

A. ROLE OF XML

XML is a mark up language like HTML with the XML tags comprising the XML header and body for incorporating the information within the tags. When the information needs to be transmitted from the sender to the receiver there is always a need to go for XML. Let us consider the client at Make my Trip.com wants to book the ticket to Jet Airways. In this case if all the programming paradigms are similar then the transaction can happen easily. If on the other hand, if the client uses .NET as the programming language and Apache as the server and if the server uses Java as the programming language and IIS as the server, interoperability issues may arise. To resolve those issues we need a common platform on which client and server can communicate with each other despite interoperability issues. XML plays the role of solving the interoperability issues and providing the service to the consumer by means of transport protocol like Simple Object Access Protocol (SOAP) and making use of distributed component technologies like Remote Procedure Call (RPC).

A.THE SERVICES OF RDF AND RDFS IN WEB XML

RDF is a framework for describing resources on the Web. It is a part of W3C's recommendation on Semantic Web activity. The resource can be anything. i.e., a printer or a Web page.RDF is not designed to be used by humans and it will not be displayed on the screen. RDF is more like XML which will have a Resource, Property and a Property Value. A Resource can be URI, Property defines the name of the resource. Property Value defines the value of the property. The collection of all gives a Statement.

For example consider the statement: The author of http://www.xyz.com/rdf is Jan Egil. In this statement, the resource is the URI, the property is the "author" and the property value is "Jan Egil".RDF Schema organizes the vocabulary for RDF. It organizes the vocabulary in a hierarchical format like Class of, Subclass of, Type, Property of, Sub

Property of and the schema for RDF can be written as

rdf: Type

rdfs: Sub Class of /Sub Property of (read as rdfs: Sub Class of or rdfs: Sub Property of)

IV. A DETAILED LOOK ON ONTOLOGIES

A.INTRINSIC ONTOLOGIES

The most significant intrinsic ontologies consist of Thing, Property and State. The Thing has identity or existence in reality and it possess properties. The book can be considered as an example of Thing. The Property characterizes the object and they do not have independent existence on their own. The color "White" can be considered as a Property to an object "Book". Every object is in some state or other at a given time[2].

The representational categories for the intrinsic categories can be Schema, Attribute or a State Variable. The Schema is a representation of a Thing. For example the Thing "Book" may have the schema of book name, author etc.. The Attribute is a representation of the Property. For example for the Schema "House", the "Address" may be considered as a Property and for the address we can have multi valued attributes like house number, street, city etc.. The State variable indicates the representation of a state[2].

The primitive relational categories relate two intrinsic categories. They are classified into Possess, Precedes and Event. Possess is a relational category between Thing and Property. We can consider Book as a Thing which possess properties like book price, author etc… Precedes is a relational category between Property and Property. The following statement can be considered as an example of precedes relation: Being a person precedes being a student. The Event is a change of two states of a thing[2].

B.RULES FOR ONTOLOGICAL CATEGORIZATION

If the entity satisfies a particular relation then we can easily categorize into Thing, Property or an Event. For example consider the relation which has no Run method, no Main method, no Methods and if it has only attributes then we can infer it as a Property. Similarly if the relation does not have a Run Method, no Main method, if the members are not static and if it has Constructors and Interactions between them, we can easily classify them as a Thing[2].

C.THE NEED OF ONTOLOGY WEB LANGUAGE (OWL)

There is a possibility that we may think of what is the need of having a separate language for ontology despite using RDF and RDFS. RDF and RDFS allow the representation of some ontological knowledge. The next question which may possibly arise is even when RDF/RDFS provides knowledge modeling for ontology why do we need something like OWL? The main modeling primitives of RDF/RDFS concern the organization of vocabularies: subclass and sub property relationships, domain and range restrictions, and instances of classes. However a number of other features which may be possibly required are missing. rdfs:range defines the range of a property. In RDF Schema we cannot apply the range restriction to a particular class alone like in XML. For example, we cannot say that buffalo eat only plants, while other animals may eat meat, too. Sometimes we may like to

IJSER

indicate that classes are disjoint. For example, mammals and water living species are disjoint. But in RDF Schema we can only state subclass relationships, e.g. water living species is a subclass of animal. In certain cases we may like to place restrictions on how many distinct values a property may take. For example, we would like to say that an industry has exactly two managing authorities in case of two partners for that industry, and that a lecture is taught by at least one instructor. Again such restrictions are impossible to express in RDF Schema[1].

V.ONTOLOGICAL REPRESENTATIONS IN WEB

A.THE ONTOLOGY WEB LANGUAGE (OWL)

OWL builds on RDF and RDF Schema, and uses XML syntax. OWL documents are usually called OWL ontologies, and are RDF documents. So the root element of a OWL ontology is an rdf:RDF element which also specifies a number of namespaces like in RDF/RDFS.OWL ontology may start with a collection of assertions. These assertions are grouped under an owl: Ontologyelement which contains comments, version control and inclusion of other ontologies. The one of these assertions which has any consequences for the logical meaning is owl:imports and this lists other ontologies whose content is assumed to be part of the current ontology. if ontology X imports ontology Y, and ontology Y imports ontology Z, then ontology X also imports ontology Z. Classes are defined using a owl: Class element. The other properties defined by OWL include Object properties and Data type properties. Object properties which relate objects to other objects. Data type properties which relate objects to data type values. The other properties in OWL include Transitive Property, Symmetric Property, Functional Property, Inverse Functional Property[1] .The following segment of code gives an example of OWL syntax.

```
<owl:Ontology rdf:about="">

<rdfs:comment>OWL ontology for Cricket </rdfs:comment>

<owl:priorVersion

rdf:resource="http://www.universe.org/cri-sport"/>

<owl:imports
rdf:resource="http://www.universe.org/cricket"/>

<rdfs:label>Cricket
Ontology</rdfs:label>

</owl:Ontology>
```

B. TRANSITION TO OWL 2.0

OWL 2 is an extension and revision of the OWL Web Ontology Language developed by the W3C Web Ontology Working Group. Like OWL 1, OWL 2 is designed to facilitate ontology development and sharing via the Web, with the ultimate goal of making Web content more accessible to machines. A concrete syntax is needed in order to store OWL 2 ontologies and to exchange them among tools and applications. The primary exchange syntax for OWL 2 is RDF/XML and this is indeed the only syntax that is supported by all OWL 2 tools. While RDF/XML provides for interoperability among OWL 2 tools, other concrete syntaxes may also be used. These include alternative RDF serializations, such as Turtle, an XML serialization and a more readable syntax, called the Manchester Syntax that is used in several ontology editing tools[1][4].

IJSER

C.ONTOLOGY TOOL - PROTEGE

Protege is a free, open-source platform to construct domain models and knowledge-based applications with ontologies. The most recent version of Protege available is Protege 4.3[5] .For the installation and the usage of tool this documentation is recommended.

VI. CONCLUSION

Thus this paper provides insights of various difficulties which are prevalent in key based search engines and it also iterates the need to switch to ontology which provides a more knowledge representation due to its incorporation of semantics. This paper also outlines how RDF is semantically less powerful than OWL. Further this paper also provides an abstract overview of various ontology languages which are suitable for developing a domain and a tool which is more commonly used in today's market.

References

[1] S.Staab and R.Studer "Handbook on Ontologies" Springer Second Edition 2009.

[2] Gómez-Pérez, Asunción, Fernandez-Lopez, Mariano, Corcho, "Ontological Engineering- with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web." Springer First Edition 2004.

[3] Grigoris Antoniou, Frank van Harmelen "A Semantic web Primer" The MIT Press annotated Edition, 2004.

[4] OWL 2 Documentation, http://www.w3.org/TR/owl2-overview/

[5] ProtegeDocumentation, https://dspace.ist.utl.pt/bitstream/2295/303480/1/Aula6%20-%20Protege.

Authors Biography

Mr.E.Arul is working as an Assistant Professor in Coimbatore Institute of Technology. He obtained his Bachelor's degree from Anna University Chennai and the Master's degree from Anna University Chennai. His research areas include antivirus and malware detection.

Mr.C.Murale is working as an Assistant Professor in Coimbatore Institute of Technology.He obtained his Bachelor's degree from Anna University Chennai and the Master's degree from Anna University Coimbatore. His research areas include Data Mining, Service Oriented Architecture and Web Services.

Mr.M.Maranco is working as an Assistant Professor in Coimbatore Institute of Technology. He obtained his Bachelor's degree from Anna University Chennai and is doing the Master's degree from Anna University Chennai. His research areas include Internetworking and Network Management.

IJSER

IJSER

ISSN 2229-5518

IJSER