International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 2, February - 2014

Semantic Search Engine for Cancer

SyamRaj B S
JayShriram Group of Institutions
Dharapuram Road, Avinashipalayam
TamilNadu, India

Sarumathi S
JayShriram Group of Institutions
Dharapuram Road, Avinashipalayam
TamilNadu, India

***Abstract*— The idea is to analyze knowledge about the real world and then create a model/standard upon stabled rules and relation types to translate the human (natural) language in a machine and human readable language for that we needed to classify and organize data such as text, pictures, videos, or database entries in a system with logical connections between data representing the knowledge shared by people. The Ontology provides a framework for the development of Semantic Web and Artificial Intelligence. Here Medical Knowledge Engineering is the Key. This paper deals with the Medical Knowledge Base to build an ontological structure. In this paper Medical Knowledge about cancer is been combined with the semantic web search engine. Based on the introduction of ontology theory, the author uses Protege 2000 of Stanford, the construction and maintenance tool of ontology, designed and completed Medical Knowledge based on Ontology. All details about cancer, cancer categories, its cause, symptoms etc. The system also learn from this details and new details from the searching process. The improvement and learning process is been by comparing the details with some knowledge organization systems. Knowledge acquisition in semantic web is done by RDF explorer. RDF scheme define relationship and those relationship make the searching in a different level**

Keywords- Protege, RDF,RDFS

I. INTRODUCTION

1.1 Semantic Web Vision:

After the invention of the World Wide Web, Tim Berners-Lee proposes the Semantic Web. The Semantic Web simply means the web of meaning. In the World Wide Web information is presented in natural human language which is not rich enough to convey formal meaning and therefore it is not machine processable. This current web contains millions and millions of resources such as HTML files, documents, images and graphics, and media files. These resources contain huge amounts of information scattered in various web pages and documents. The current web is a web of documents and understandable only to humans. This makes information retrieval processes very hard; humans alone cannot deal with this huge amount of resources on the web. Software agents or machines could help in this process but a difficulty arises from the fact that machines do not understand human language. Trying to make machines act as humans is a very complex task and needs a lot of training. The idea of the Semantic Web was introduced mainly to solve the problem that content on the current web is intended only for human consumption. The basic idea of the Semantic Web is to give information a well-defined meaning, thus better enabling agents and people to work in cooperation [1]. W3C states [2] "The Semantic Web is about two things. It is about common formats for interchange of data, where on the original Web we only had interchange of documents. Also it is about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing". We can simply say that the Semantic Web is a web of data rather than a web of documents. Semantic Web is about two things: It is about common formats for interchange of data, as opposed to documents. This data is well-defined so that agents will fully comprehend the semantics of the data. Also it is about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one knowledge base, and then move through an unending set of knowledge bases which are linked by being about the same or related domains. We can think of the Semantic Web as a mesh of information linked up in such a way as to be easily processable by machines, on a global scale. We can think of it as being an efficient way of representing data on the World Wide Web, or as a globally linked database. The challenge of the Semantic Web was to provide a language that expresses both data and rules for reasoning about the data and that allows rules from any existing knowledge-representation system to be exported onto the Web. As stated by Berners-Lee [1] "Making the language for the rules as expressive as needed to allow the Web to reason as widely as desired". The Semantic Web uses RDF (Resources Description Framework) to represent information. Each piece of information on the Semantic Web is called a resource and each resource is uniquely identified. Information about resources is represented as a Directed Graph of triples (Subject, Predicate, Object) also called RDF statements. Knowledge on the Semantic Web is stored in an ontology. The ontology holds both the data and metadata, which enables understanding the semantics. The Semantic Web structure enables not only combining Semantic Web statements to create larger pieces of information but also the ability to infer new information based on the rules defined in its ontology. Figure 1 shows the different layers of the Semantic Web. The three upper layers are still under construction and are not final yet. Logic or reasoning is one of the major important issues for Semantic Web and it is an important design issue when creating a Semantic Web agent.

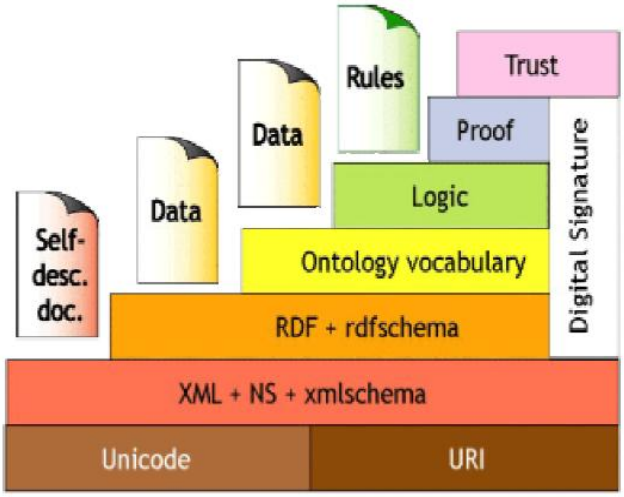

Figure 1Semantic Web layers [5]

1.2 Www To Semantic Web:

Semantic Web is not intended to replace the existing web but rather it is an extension of the current Web. The current web depends on visual representation of information through HTML tags. This visual representation makes information clear for humans to understand but very difficult for machines to understand and process. For example to emphasize something it could be in a different font or colour. Some form of extraction is required to strip off the information part from the presentation part. Other techniques are used to infer meaning from this information; this leads to an increased complexity in the agents dealing with the World Wide Web. Another problem with the current web is the fact that different terms are used to represent the same meaning, for example in a shopping site could refer to the shopping cart as cart, while another would refer to it as shopping basket or basket for short, yet another site could refer to it as shopping bag. All these words refer to the same meaning or the same semantics, which is very obvious to humans while it is unknown to software agents. These agents have to be explicitly informed that the previous terms are all the same. Another example comes from the fact that the web is multi lingual; an English shopping website would use the word price to refer to an items price, while a Dutch website would use the word prijs, a French website would use the word prix, a Spanish site would use the word precio, an Italian site would use the word prezzo, and an Arabic site would use the word الثمن . An agent that is looking for a product and comparing prices to retrieve a list of the cheapest sites would have to be familiar with these terms. These are just a sample of languages that exist on the web while there are many more. The Semantic Web targets solving these problems by providing not only the data but also metadata that describes explicitly what this data means. This form of data annotation makes an agent understand the semantics behind the data and thus allows for better interpretation between data and gents and allows for better inter agent communication and collaboration. As stated by Berners-Lee [4], "this notion of being able to semantically link various resources (documents, images, people, concepts, etc) is an important one. With this we can begin to move from the current Web of simple hyperlinks to a more expressive semantically rich Web, a Web where we can incrementally add meaning and express a whole new set of relationships (hasLocation, worksFor, isAuthorOf, hasSubjectOf, dependsOn, etc) among resources, making explicit the particular contextual relationships that are implicit in the current Web. This will open new doors for effective information integration, management and automated services". The Semantic Web promises a solution in which the web becomes one big knowledge base and everyone has access to it. In order for this to happen there should be supporting technology that allows for such annotation in a formal and unified syntax, such annotation are RDF/RDFS and OWL which are standards set by the W3C. Also reasoning on the Semantic Web promises for more intelligence in services provided by the web such as personalized notifying agents, search agents, personalized search agents, e-learning and many other applications where agents would pull the information and process it having a better understanding of its meaning.

1.3 Ontology:

The term Ontology has its roots in the philosophical domain. In order to understand the basic structure of our world and the study of existence, the word ontology has been connected with a branch of metaphysics. The problem is that the philosophical definition of ontology is not easy to port to the scientific domain. Therefore Dunwoodie [2007] uses an intelligible definition of ontology:"An ontology is a detailed model/picture/schema of a slice of reality which is based on the facts that we know about that reality. This model/picture/schema is a description of some of the things and some of the relationships between the things that are known about that reality". In Helfin [2004], the term"Ontology" is defined as following:"An ontology defines the terms used to describe and represent an area of knowledge". These ontologies can be shared by different applications, people and databases within a domain.

A domain can be an area of knowledge, like medicine or a specific subject area. The definitions of ontologies are machine readable and they describe basic concepts in the domain and the relations between them. The knowledge, which is encoded in ontologies, is reusable due to the fact that the encoded knowledge can span different domains. Ontologies are able to specify the following kinds of concepts, which enable the description of almost every knowledge:

- Classes (things)
- Relationships between things
- Properties (attributes) of things

There are many motivations for developing and using ontologies:

- To share common understanding of the structure of information among people or software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from the operational knowledge

- To analyze domain knowledge

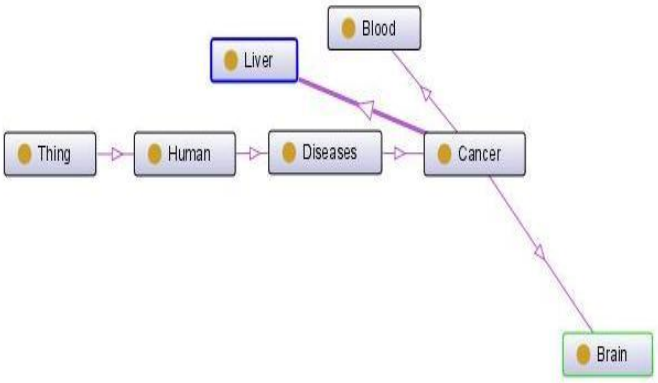

Figure 2 Cancer Detection Ontology

1.4 Rdf Schema (Rdfs):

RDF is a metadata language that does not provide special vocabulary for describing the resources. It is often essential to be able to describe more of a subject than saying it is a resource. Some form of classification for these resources is often required to be able to be able to provide a more precise and correct mapping of the world. The basic idea behind Semantic Web is to provide meaning of resources, as defined in the Knowledge Representation domain, "knowledge is descriptive and can be expressed in a declarative form" [16]. The formalization of knowledge in declarative form begins with a conceptualization. This formalization includes the objects presumed or hypothesized to exist in the world. This is why RDF schema (RDFS) was introduced as a language that provides formal conceptualization of the world. RDF Schema semantically extends RDF to enable us to talk about classes of resources, and the properties that will be used with them. The RDF schema defines the terms that will be used in RDF statements and gives specific meanings to them. It provides mechanisms for describing groups of related resources and the relationships between these resources. Meaning in RDF is expressed through reference to the schema. RDFS consists of a collection of RDF resources that can be used to describe properties of other RDF resources this makes it a simple ontology language which allows more capture of semantics than just pure RDF. The most important resources described in RDFS are:

● *Classes:* RDFS deals with classes through the term rdfs:Class which defines a class. RDFS allows for creating hierarchies of classes in which a class is defined as the subclass of another class using the rdfs:subClassOf property. RDFS also allows for creating instances of a class; that is data that are of the type of that class. The rdf:type property may be used to state that a resource is an instance of a class. RDFS defines a special class rdfs:Literal which is the class of literal values such as strings and integers, rdfs:Literal is an instance of rdfs:Class [11].

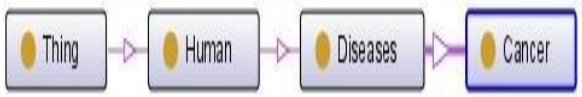

Figure 3 Example RDF representation

1.5 Semantic Content In Owl:

OWL (Web Ontology Language) was introduced to provide richer vocabulary than RDFS. OWL semantically extends RDF/RDFS which means that an OWL ontology is an RDF graph. A formal semantics describes the meaning of knowledge precisely and rich semantics describes fine grained knowledge. OWL was introduced by W3C to provide a richer ontology language that allows for: a well-defined syntax, efficient reasoning support, a formal semantics and sufficient expressive power [7]. The main content of OWL ontology is carried in its axioms and facts, which provide information about classes, properties, and individuals in the ontology. There are a several major capabilities that OWL adds to RDF and RDFS. The first is the ability to create local range restrictions. In RDFS, a property is allowed to have only one class as its range while in many cases there is a need defining more restrictions on the property range.

2. CONCLUSIONS

Right now the semantic web techniques cannot replace a human. He still must validate all the results that a computer generates. Still the human is the one to formally define concepts, things, and events, real live and presented in a machine-understandable form. Even if the vision about the Web of trust can be still far way, we pointed out the important steps already achieved. Our present work Cancer Ontology covered almost all the life cycle for a semantic application.

This includes:

- Cancer and its Property Definition.
- Cancer Ontology implementation through
- URI, XML, RDF, RDFs, OWL.
- Discovery of new semantic communities in cancer ontology
- Browse the Machine Processable OWL data through Ontology Browsers (Data Link Explorer, Manchester Ontology Browser) and DL query execution (SPARQL, RIFs).

We hope that all the above steps contributed to an extent in which the inferred knowledge about Cancer is presented in Machine Understandable as well as human understandable semantic form.

10. REFERENCES

[1] L. Miller, A. Seaborne, and A. Reggiori, "Three Implementations ofSquishQL, a Simple RDF Query Language," Proc. Int'l Semantic WebConf. (ISWC 02), LNCS 2342, Springer, 2002, pp. 423–435.

[2] T. R. Gruber, A Translation Approach to Portable Ontology Specifications[J], Knowledge Acquisition, vol. 5, no.2, 1993, pp.199~220.

[3] J. Broekstra, A. Kampman, and F. van Harmelen, "Sesame: A GenericArchitecture for Storing and Querying RDF and RDF Schema," Proc.Int'l Semantic Web Conf. (ISWC 02), Springer, 2002, pp. 54–68.

[4] Taboada, D.Martinez, J.Mira; Experiences in reusing knowledge sources using Protege and PROMPT; Int.J.Human-Computer Studies 62(2005) 597-618

[5] P. Martin and P. Eklund, "Knowledge Retrieval and the World Wide Web," IEEE Intelligent Systems, vol. 15, no. 3, 2000, pp. 18–25.