International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-3 Issue-4, September 2013

Semantic Web Tool: For Efficient Retrieval of Links and Required Information

Ritika Bansal, Sonal Chawla

Abstract- Web, which grows significantly, becomes a need for modern society to make transactions, search the information, and spread the information. This paper presents the prototype for the semantic web based tool for efficient retrieval of links and required information. By using this tool, users can easily access the information without technical knowledge on RDF. This paper has fourfold objective. Firstly, paper throws light on the need of the proposed semantic web based tool. Secondly, paper proposes the semantic web based tool and its adaptive view. Thirdly, paper highlights the importance of ontologies and comparison of various tools for ontology development. Lastly, paper also throws light on java based frameworks for developing semantic web based tools.

Keywords – Semantic Web, Ontology, SPARQL, RDF, Jena, Sesame, OWL API , Protégé.

I. INTRODUCTION

A huge size of information in the web and the spread of the information in many different sources make users need more time in searching for information and organizing them from many different sources manually [1]. Although web can give all information needed, it often gives irrelevant information to the user. The large and voluminous information available on the internet still requires a human eye and brain to understand and process it. Also, humans refrain from directly putting in the query in their mind due to the large number of irrelevant results fetched by regular search engines and the large differences in results caused by the most subtle difference in the query. Semantic web is a solution to solve those problems by providing knowledge based on ontology. Converting this information into a form that can be processed by the machines will open new doors to knowledge management and agent based processing[2], and enables precise in-depth searching with a considerable reduction in user time and effort.

II. OBJECTIVE

To create a Semantic Web Tool which is Intelligent agent based which will retrieve efficient links and required information easily understandable for user. It will search in RDF based libraries (ontologies) and if the solution to the query is not found then it will convert the existing web into RDF and then show in the search result. Search results should be intelligent and semantically understandable (with the help of icons etc.) Also, it will show the required information in organized manner, if user has requested for that.

Manuscript published on 30 September 2013.
*Correspondence Author(s)
Ritika Bansal, Research scholar, Department of Computer Science and Applications, Panjab University & Assistant Professor, M.C.M.D.A.V. College, Chandigarh, India
Sonal Chawla, Associate Professor, Department Computer Science and Applications, Panjab University Chandigarh, India

III. PROPOSED SYSTEM

Proposed semantic web based tool will hold domain related information in its knowledge base and a natural language querying platform that takes in user queries and fetches the relevant results from the knowledge base. Storing and extracting of domain related information from the ontology (Knowledge Base / Semantic data library) so that the user gets precise and quick results of the queries. Also, the user is provided with an interface to select options associated with the selected concepts. Interface accepts free form text queries, and fetches the relevant results found in the Ontology as shown in figure 1. The system will be time effective and helpful in making decisions with minimum queries as compared to browsing even with selected sites. The model presented can be implemented as a prototype to show the effectiveness of the semantic web in searching and extracting information in a specific domain. This will enable accurate and reliable data from one platform with the ease of Natural Language Query Processing.

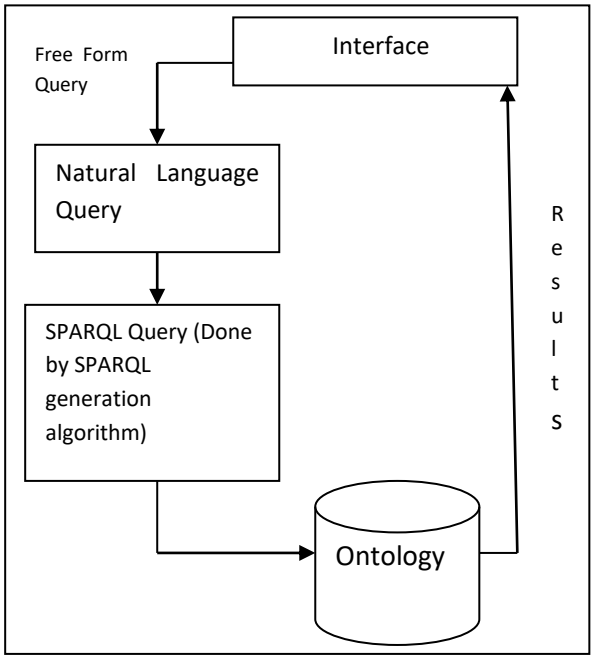


Figure 1: Proposed Semantic Web based tool

IV. COMPONENTS OF PROPOSED SYSTEM

The proposed semantic web based tool will comprise of five main components:
Component 1- Extract Real time data / Domain specific data from Web and convert it into RDF Triples.
Component 2- Ontology Development and storing RDF triples into it.
Component 3: Interface for Natural Language query.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Component 4: Convert Natural Language query into SPARQL Query.
Component 5: Output (Query Results)
Ontologies will be developed containing the domain specific data. While searching if the solution to the query is not found then the real time data available on the web will be converted into RDF triples and will be stored in the domain specific ontologies. For this conversion of data available on the web into RDF triples, web wrappers can be used. These RDF triples are stored in the ontologies and ontology is developed. Various tools are available for ontology creation like protégé tool as shown in figure 2. After the development of ontology, third component is the designing and development of interface which will accept free form queries from the user. This natural language query should be converted into the language which will help in retrieving the required solution from the ontology. SPARQL is the RDF query language, so natural language query is converted into SPARQL queries. Finally, based on the SPARQL query, results are retrieved from the domain ontology and will be displayed on the interface into the format easily understood by the user.

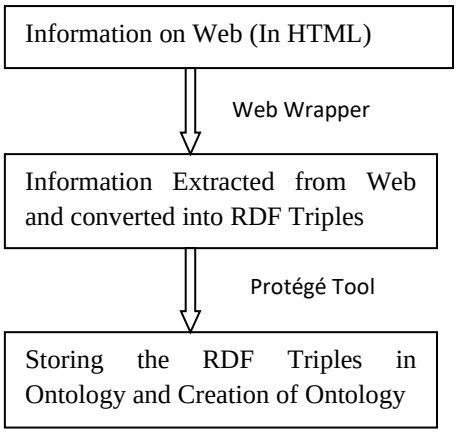
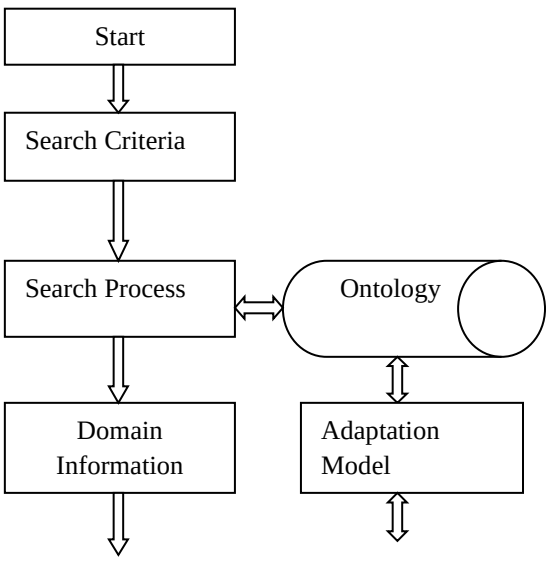


Figure 2: Steps for ontology development

4.1 ADAPTIVE VIEW OF SEMANTIC WEB BASED TOOL

To make the tool adaptive, along with all the above discussed components, user model is used to represent users and give the personalized output. The adaptation model used knowledge from the ontology to generate a personalized output based on the initialized user model which represents user's behavior to the interface as shown in figure 3 below.



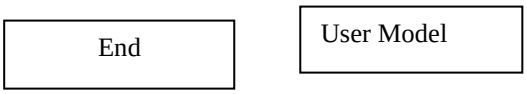


Figure 3: Searching process in adaptive semantic web based tool

V. SEMANTIC WEB

Semantic web is an extended web that provides intelligent access to heterogeneous, distributed information, enabling software products (agents) to mediate between user needs and information sources available[5]. Semantic web has provided us with a standard for the representation and retrieval of relevant data and to make such information which is machine process-able. The development of semantic web has provided with a standard for the representation and retrieval of relevant data. As, search engines used on a large scale on a daily basis shows that each word of a query being entered is matched separately, without any logical connection between the keywords. So, Semantic Web is used in order to define the logical relationships between the keywords of a query.

VI. ONTOLOGY

Ontology, which is the heart of semantic web, with concept instantiations serves as a domain knowledge base i.e. semantic web technology information center. The ontology is designed for the system to incorporate the domain information in the form of instances and data type values, classes and object properties i.e. to model the knowledge domain. Ontologies allow for machine-understandable semantics of data, and facilitate the search, exchange,and integration of knowledge An ontology is always constructed with a certain task in mind, this task focus restricts the content and structure of the ontology. Many tools have been developed for implementing metadata of ontologies using these languages. Ontology tools can be applied to all stages of the ontology life cycle including the creation, population, implementation, and maintenance of ontologies (Polikoff, 2003). For a knowledge-management system, an ontology can be regarded as the classification of knowledge. Ontologies are different from traditional keyword-based search engines in that they are metadata, able to provide the search engine with the functionality of semantic matching. Ontologies are able to search more efficiently than traditional methods. Typically, an ontology consists of hierarchical descriptions of important concepts in a domain and the descriptions of the properties of each concept.

1.1 ONTOLOGY DEVELOPMENT TOOLS

Finding an appropriate tool to develop ontology is the first step towards ontology development. Ontology development is a complex and largely domain-oriented process that can be benefited from tool support. Various tools for ontology development are available like protégé, SWOOP, OilED, Apollo, RD Fedt etc. Though there are enormous ontology development tools available for free on the web. Ontology development tools are compared based on certain features such as modeling features/limitations, base language, web support and use, import/export format, graph view, consistency checks, multi-user support, merging, lexical support, and information extraction as shown in table 1 given below [6].

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Several important aspects when we analyze tools exist. Most of the tools are moving toward Java platforms and extensible architectures as well. Interoperability and storage in databases are still weak points of ontology tools. The most dominant and domain-independent tool used is protégé as it supports many features like GUI , storage through JDBC etc. which are not supported by other many tools.

Features	Protege	OilEd	Apollo	RDFedt
Import format	XML, RDF(S), XML schema	RDF(S), OIL, DAML+OIL	OCML , CLOS	RDF(S), OIL, DAML, SHOE
Export format	XML, RDF(S), XML schema, FLogic, CLIPS, Java, HTML	RDF(S), OIL, DAML+OIL, SHIQ, dotty, HTML	OCML , CLOS	RDF(S), OIL, DAML, SHOE
GUI	Via plugins like GraphViz and Jambalaya	No	No	No
Consistency check	Via plug-ins like PAL and FaCT	Via FaCT	Yes	Only checks writing mistakes
Multiuser	Limited (multiuser capability added to it in 2.0 version)	No	No	No
Web support	Via Protégé-OWL plug-in	Very limited name spaces	No	Via RSS
Merging	Via Anchor-PROMPT plug-in	No	No	N/A
Collaborative working	No	No	No	No
Ontology library	Yes	Yes	Yes	No
Inference engine	With PAL	With FaCT	No	No
Ontology storage	File & DBMS (JDBC)	File	Files	Files
Extensibility	Via plug-ins	No	Via plug-ins	No
Availability	Free	Free	Free	Free

Table 1: Comparison of ontology development tools

Moreover, it is popular as it has enough documentation on the web and is also extensible.

1.1.1 PROTÉGÉ

Protégé is a free, open source ontology editor and knowledge-base framework. The Protégé platform supports modeling ontologies via a web client or a desktop client. Protégé ontologies can be developed in a variety of formats including OWL, RDF(S), and XML Schema. Protégé is based on Java, is extensible, and provides a plug-and-play environment that makes it a flexible base for rapid prototyping and application development. Protégé is supported by a strong community of developers and academic, government and corporate users, who are using Protégé for knowledge solutions in areas as diverse as biomedicine, intelligence gathering, and corporate modeling [4].

VII. RDF QUERYING LANGUAGE –SPARQL

SPARQL is the standardized query language for RDF, the same way SQL is the standardized query language for relational databases. A SPARQL query consists of a set of triples where the subject, predicate and/or object can consist of variables. The idea is to match the triples in the SPARQL query with the existing RDF triples and find solutions to the variables. A SPARQL query is executed on a RDF dataset .SPARQL Protocol and RDF Querying Language is used as the channel that accomplishes the retrieval of the information.

VIII. FRAMEWORK FOR SEMANTIC WEB TOOL

There are various Java based open source frameworks for building tools based on semantic web like Jena, Sesame, and OWL API etc.

Jena is a Java framework for building Semantic Web based tools. Jena provides a collection of tools and Java libraries which helps in developing semantic web and linked-data apps, tools and servers. It provides a extensive Java libraries for helping developers develop code that handles RDF, RDFS, RDFa, OWL and SPARQL. Jena includes a rule-based inference engine to perform reasoning based on OWL and RDFS ontologies, and a variety of storage strategies to store RDF triples in memory or on disk. Jena stores information as RDF triples in directed graphs, and allows code to add, remove, manipulate, store and publish that information [7]. The Jena Framework includes: an RDF API, reading and writing RDF in RDF/XML, N3, N-Triples, Turtle, an OWL API ,in-memory and persistent storage, SPARQL and RDQL – query languages for RDF.

Sesame is a standard framework for processing RDF data which includes parsing, storing, inferencing and querying of RDF data. It also offers an easy-to-use API that can be connected to all leading RDF storage solutions. Sesame has been designed with flexibility in mind. It can be deployed on top of a variety of storage systems (relational databases, in-memory, filesystems, keyword indexers, etc.), and offers a large scala of tools to leverage the power of RDF and related standards. Sesame fully supports the SPARQL query language for expressive querying and offers transparent access to remote RDF repositories using the exact same API as for local access. Finally, Sesame also supports all main stream RDF file formats, including RDF/XML, Turtle, N-Triples, TriG and TriX[8].

OWL API is a high-level programmatic interface for accessing and manipulating OWL ontologies. The OWL API is a Java API and reference implementation for creating, manipulating and serializing OWL Ontologies. The OWL API includes various components like RDF/XML parser and writer, OWL/XML parser and writer, OWL Functional Syntax parser and writer, Turtle parser and writer etc. [9].

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

IX. CONCLUSION

A benefit of providing all the domain related and comparative information on a single interface based on user selected parameters will provide a better and more efficient approach to relevant data extraction. It is possible to reuse information that has already been acquired if the information acquisition by mechanical processing becomes possible. An enhancement of this is to work on real time data, which can be extracted at runtime after the user's query is processed and understood. Any domain which involves a lot of information sources and data exchange can be implemented with the semantic web.

REFERENCES

[1] Dwipa Search engine: When E-Tourism meets the semantic web,G.P.Kuntarto,.Gunawan, Department of Information System and Computer Science Universitas Bakrie and Universitas Multimedia Nusantara,2012.

[2] Semantic Web E-Portal for Tourism, F.F.Ahmed,S.F.Hussain,S.Hameed,S.M.Ali, Sir Syed University, Karachi, Pakistan,2012.

[3] A Semantic Web browser for novice users, Y.Kim,S.Yoo,S.Park,Computer Science Education, Korea Univeersity,2012.

[4] http://protege.stanford.edu/

[5] Wikipedia http://en.wikipedia.org/wiki/Semantic_Web

[6] RDF Primer. W3C Recommendation. Feb, 2004. http://www.w3.org/TR/rdf-primer/

[7] http://jena.apache.org/about_jena/architecture.ht

[8] http://www.openrdf.org/about.jsp

[9] http://owlapi.sourceforge.net/

[10] Ontology Development Tools for ontology-based knowledge management, S.Youn, D.McLeod, University of southern California,2006.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)