Efficient Semantic Web Data Querying And Integration Using Fuzzy Ontology

[1]M.S.Raghuram, [2]M.Thenmozhi
1 Research Scholar(M.Tech), [2]Assistant Professor,
[1]Information Technology, Dept of Information Technology, SRM University, Chennai, India,
Dept of Information Technology, SRM University, Chennai, India
[1]Raghuram.manchineni@gmail.com, [2]Mozhi_2000@yahoo.com

***Abstract* - In this technical demonstration the framework of the development of generic method concerning data integration, the team is particularly interested in the representation of heterogeneous and imprecise data and in their extended querying through the use of techniques borrowed from fuzzy logic. Moreover,**
since the OTR is at the heart of Semantic Ontology system, the team works on ontology construction and evolution .We present a system architecture relies on an Ontological and Terminological Resource (OTR) which is composed of two parts: on the one hand, a generic set of concepts dedicated to the data integration task and, on the other hand, a specific set of concepts and a terminology, dedicated to a given domain of application. The main objective of the semantic annotation method is to identify which relations of the OTR are represented in a data table those simple concepts are called in the following simple target concepts. In order to annotate a column col by a simple target concept, a score is computed for each simple target concept c of the OTR for the column, on a generic OTR expressed in OWL. our system allows XML data tables, which have been extracted from Web documents, to be annotated with fuzzy RDF descriptions and to be flexibly queried using Ontology search engine. Ontology search engine allows To retrieve not only exact answers compared with the selection criteria but also semantically close answers and to compare the selection criteria expressed as fuzzy sets representing preferences with the fuzzy annotations of data tables. finally OWL file is downloaded in client machine.

***Keywords*- Fuzzy logic, OWL(Ontological web language), XML/RDF, Ontology, Semantic annotation, OTR**

I. INTRODUCTION

The huge amount of technical and scientific documents available on the Web include many data tables. In addition to local data sources, they represent big potential external data sources for the data warehouse of a company dedicated to a given domain of application. To lighten the burden laid upon domain experts when selecting data from the data warehouse for a particular application, it is necessary to give them indicative reliability evaluations. In this paper, we present a framework to estimate the reliability of data tables collected from the Web. Due to its generic nature, this method can be reused in other data warehouses using the semantic web recommended languages. Reliability estimation is an essential part of the Semantic Web architecture, and many research work focus on issues such as source authentication, reputation, etc. For example, advocates a multi-faceted approach to trust models. They propose an OWL based ontology of trust related concepts. The idea is to provide systems using the annotation power of a user community to collect information about reliability.

System architecture relies on an Ontological and Terminological Resource (OTR) which is composed of two parts: on the one hand, a generic set of concepts dedicated to the data integration task and, on the other hand, a specific set of concepts and a terminology, dedicated to a given domain of application. System architecture is composed of two subsystems: 1) @Web subsystem designed to load an XML/ RDF data warehouse with data tables which have been extracted from Web documents and semantically annotated using concepts from the OTR. 2) Ontology search subsystem designed to query simultaneously and uniformly the local data sources and the XML/RDF data warehouse using the OTR in order to retrieve approximate answers in a homogeneous way. When a ontology search query is asked by the end user into the XML/RDF data warehouse which contains fuzzy RDF graphs generated by our annotation method to annotate XML data tables, the query processing has to deal with fuzzy values. More precisely, it has 1) to take into account the certainty score associated with the relations represented in the data tables and 2) to compare a fuzzy set expressing querying preferences to a fuzzy set, generated by our annotation method, having a semantic of similarity or imprecision.

After OTR we are using WSDL (web service description language) and SOAP(simple object access protocol) to connect to another application and get information and filtering and extraction is done this is semi-automatic process and Table Annotation is done by using SVM (support vector machine).This svm will analyze data and recognize patterns and used for classification and regression analysis and by using SOAP technology create the WSDL. Created OWL (ontology web language) file will downloaded in the client side and user will deploy that OWL file in the browser.

We have presented in this paper a complete system, called ONTOLOGY, built, using the recommendations of the W3C, on a generic OTR expressed in OWL. ONTOLOGY system allows XML data tables, which have been extracted from Web documents, to be annotated with fuzzy RDF descriptions and to be flexibly queried using SPARQL. Fuzzy RDF annotations are used to represent (1) the set of most similar symbolic concepts of the OTR which are automatically associated with the content of a cell belonging to a symbolic column, (2) imprecise values

associated with a quantity expressed in one or several numerical columns, (3) a degree of certainty associated with each n-ary relation recognized in a data table.

Objectives

- To retrieve not only exact answers compared with the selection criteria but also semantically close answers.
- To compare the selection criteria expressed as fuzzy sets representing preferences with the fuzzy annotations of data tables.
- We are using Ontological and Terminological based resource. So, the data must be always accurate to our query

II. MODULES

2.1. User Login & User Query Module

In this module, we are going to design web application to main originalities of our new flexible querying subsystem are: 1) to retrieve not only exact answers compared with the selection criteria but also semantically close answers; 2) to compare the selection criteria expressed as fuzzy sets representing preferences with the fuzzy annotations of data tables. Querying subsystem allows the end-user to express preferences in his/her query and to retrieve the nearest data stored in the two kinds of data sources corresponding to his/her selection criteria.

2.2. OTR Resource & Search Web

In this module, allows the complexity of the querying into different data sources to be hidden to the end user. A Ontological search query is an instantiation of a given view by the end user, by specifying, among the set of query able attributes of the view, which are the selection attributes and their corresponding searched values, and which are the projection attributes. An important feature of a Ontological search query is that searched values may be expressed as continuous or discrete fuzzy sets. A fuzzy set allows the end user to express his/her preferences which will be taken into account to retrieve not only exact answers.

2.3. Filtering & Table Extraction

Recent propositions in the Semantic Web community propose to extract, filter, annotate and query Web data tables, but they have not been designed with the same objectives as ours. Table Seer for instance allows a set of predefined metadata to be extracted from Web data tables, but it does not compare the schema of the Web data tables with preexisting schemas defined in ontology. We can also cite Web Tables which proposes a system to identify relational tables in a huge amount of tables included in HTML documents and to index them, this in order to query and rank them.

2.4. Table Annotation With OTR Based

Our method to identify relations depends on the identification of the symbolic concepts and quantities, which can be considered as a weakness. For this reason, our experimentation to automatically annotate the data tables with the relations of the considered OTR was applied without validating the intermediate steps.

2.5. Validation & Storing into RDF/XML Database

In this module, when a query is asked by the end user into the XML/RDF data warehouse which contains fuzzy RDF graphs generated by our annotation method to annotate XML data tables, the query processing has to deal with fuzzy values. More precisely, it has 1) to take into account the certainty score associated with the relations represented in the data tables and 2) to compare a fuzzy set expressing querying preferences to a fuzzy set, generated by our annotation method, having a semantic of similarity or imprecision.

2.6. User's Integrated Output

The originality of our approach in flexible querying is that we propose a complete and integrated solution which allows one 1) to annotate Web data tables with the vocabulary defined in an OTR, 2) to perform a flexible querying of the annotated tables using the same vocabulary and taking into account the fuzzy degrees generated by the annotation method according to their associated semantic.

III. SYSTEM ARCHITECTURE

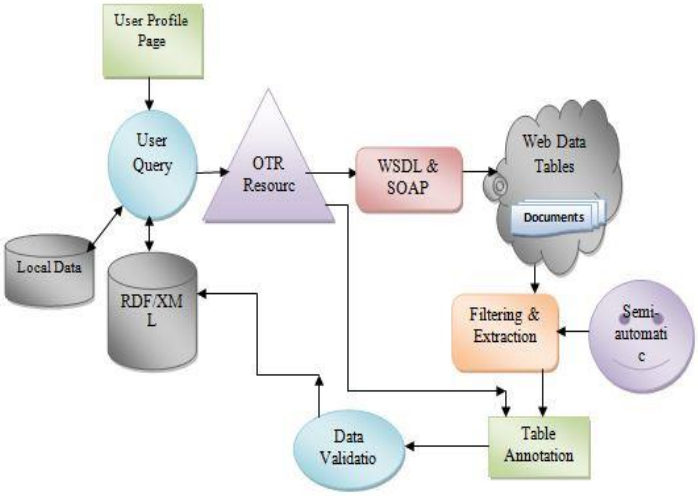
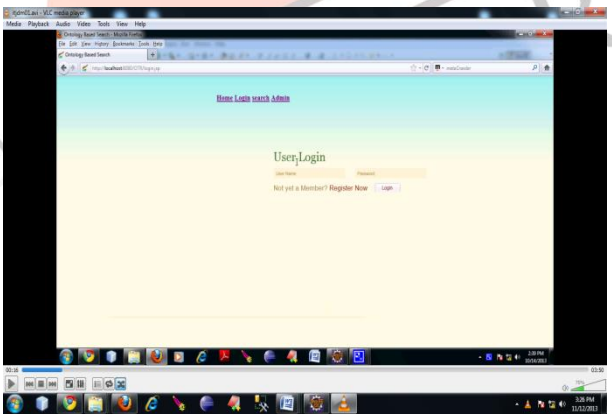


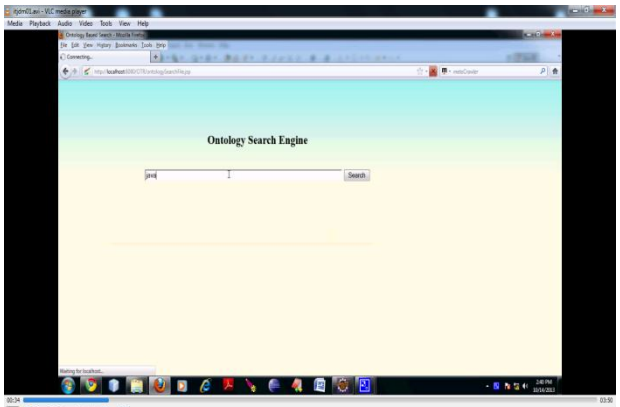
Fig 1

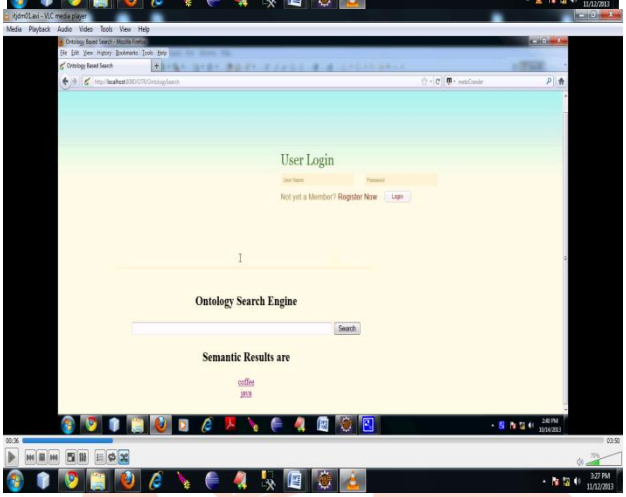
The systems architect establishes the basic structure of the system, defining the essential core design features and elements that provide the framework. The systems architect provides the architects view of the users' vision. Above diagram shows that the user profile page and user query will be converted into search based on ontology & terminology based queries. Then, the OTR based data query will give to WSDL & SOAP process to retrieve the data from web documents. Later, the data which will be available in web data tables will be filtered & extracted by using of semi-automatic process and thereafter the data will be annotated based on the OTR based phase will be done & later it will validate the data to give the integrated output.finally OWL file is downloaded in client machine and user must deploy that owl file in browser.

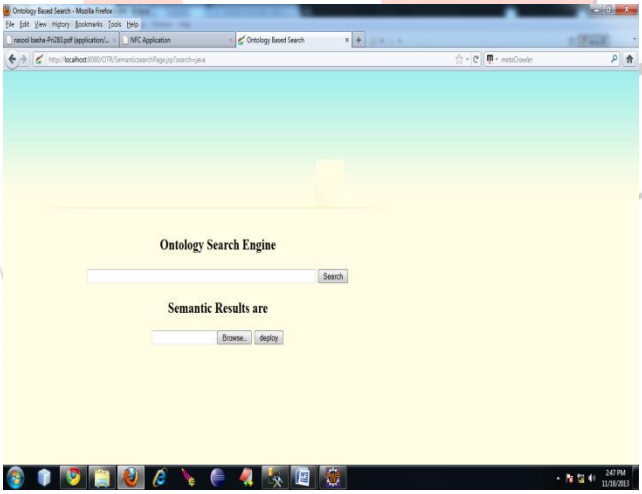
In this system architecture Semantic Web framework, Ontology search querying , OTR Resource is used and local data to store normal data and tables and RDF/XML is to store structured information/structured data.A Ontology search query is asked in a view which corresponds to a given relation of the OTR. A view is characterized by its set of queryable attributes and by its actual definition. Each queryable attribute corresponds to a simple concept of the relation represented by the view.When a Ontology search query is asked by the end user into the XML/RDF data warehouse which contains fuzzy RDF graphs generated by our annotation method to annotate XML data tables, the query processing has to deal with fuzzy values. More precisely, it has 1) to take into account the certainty score associated with the relationsrepresented in the data tables and 2) to compare a fuzzy set expressing querying preferences to a fuzzy set, generated by our annotation method, having a semantic of similarity or imprecision.

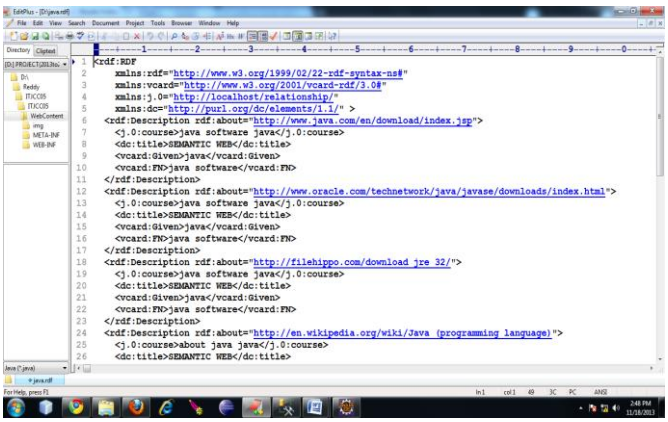
IV. RESULTS

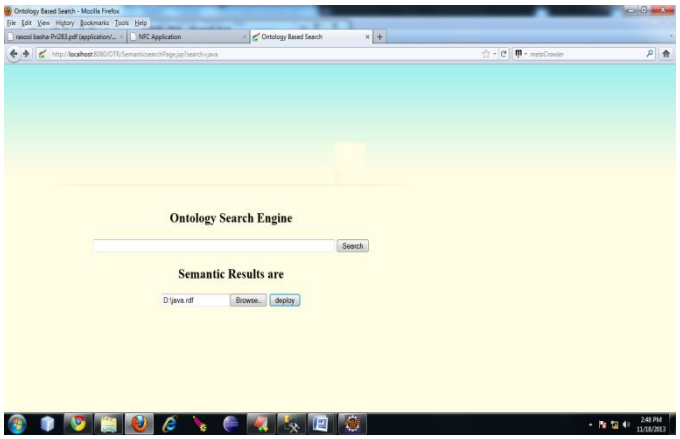



Ontology Search Engine
Search


User Login
Ontology Search Engine
Search
Semantic Results are


Ontology Search Engine
Search
Semantic Results are
Browse...
deploy


<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:vcard="http://www.w3.org/2001/vcard-rdf/3.0#"
xmlns:j.0="http://localhost/relationship/"
xmlns:dc="http://purl.org/dc/elements/1.1/" >
<rdf:Description rdf:about="http://www.java.com/en/download/index.jsp">
<j.0:course>java software java</j.0:course>
<dc:title>SEMANTIC WEB</dc:title>
<vcard:Given>java</vcard:Given>
<vcard:FN>java software</vcard:FN>
</rdf:Description>
<rdf:Description rdf:about="http://www.oracle.com/technetwork/java/javase/downloads/index.html">
<j.0:course>java software java</j.0:course>
<dc:title>SEMANTIC WEB</dc:title>
<vcard:Given>java</vcard:Given>
<vcard:FN>java software</vcard:FN>
</rdf:Description>
<rdf:Description rdf:about="http://filehippo.com/download_jre_32/">
<j.0:course>java software java</j.0:course>
<dc:title>SEMANTIC WEB</dc:title>
<vcard:Given>java</vcard:Given>
<vcard:FN>java software</vcard:FN>
</rdf:Description>
<rdf:Description rdf:about="http://en.wikipedia.org/wiki/Java_(programming_language)">
<j.0:course>about java java</j.0:course>
<dc:title>SEMANTIC WEB</dc:title>



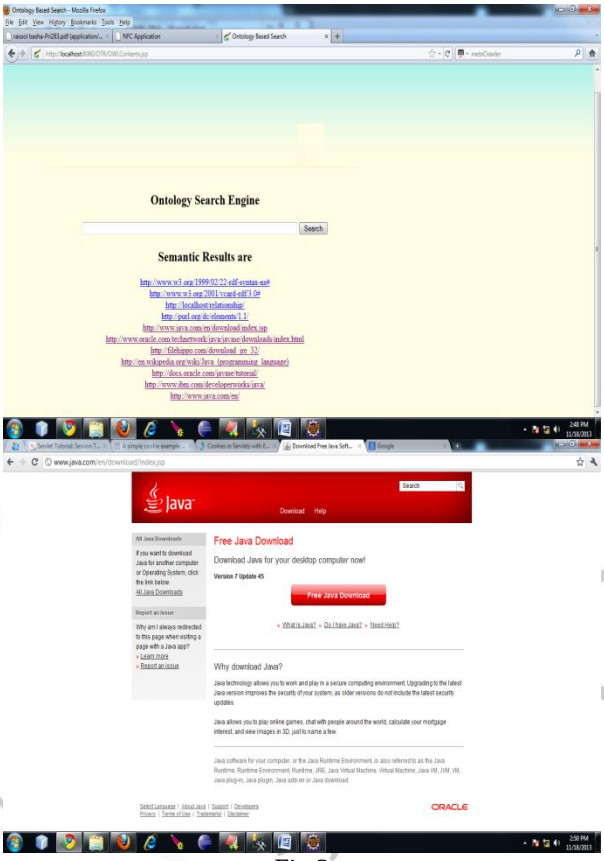


Fig 2

V. CONCLUSION

We have presented in this paper a complete system, called Ontology, built, using the recommendations of the W3C, on a generic OTR expressed in OWL. Ontology system allows XML data tables, which have been extracted from Web documents, to be annotated with fuzzyRDF descriptions and to be flexibly queried using ontology search. Fuzzy RDF annotations are used to represent (1) the set of most similar symbolic concepts of the OTR which are automatically associated with the content of a cell belonging to a symbolic column, (2) imprecise values associated with a quantity expressed in one or several numerical columns, (3) a degree of certainty associated with each n-array relation recognized in a data table.

VI. FUTURE ENHANCEMENTS

The other perspectives concern the improvement of Ontology system by

1) Completing the cosine similarity measure used to compare terms with other syntactical and semantic techniques
2) Completing the semantic annotation of data tables in Web documents with the annotation of the text using the OTR.
3) Managing OTR evolution by taking into account annotation results and other ontology's.

7. ACKNOWLEDGEMENT

The authors are grateful to express sincere thanks and
gratitude to our Department of Information Technology, SRM University for encouragement and the facilities that were offered to us for carrying out this project.

REFERENCES

[1] Patrice Buche, Juliette Dibie-Barthelemy, Liliana Ibanescu, and Lydie Soler" Fuzzy Web Data Tables Integration Guided by an Ontological and Terminological Resource" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 4, APRIL 2013.

[2] P. Buche and O. Haemmerle´, "Towards a Unified Querying System of Both Structured and Semi-Structured Imprecise Data Using Fuzzy Views," Proc. Linguistic on Conceptual Structures: Logical Linguistic, and Computational Issues (ICCS), pp. 207-220, 2000.

[3] P. Buche, C. Dervin, O. Haemmerle´, and R. Thomopoulos, "Fuzzy Querying of Incomplete, Imprecise, and Heterogeneously Structured Data in the Relational Model Using Ontologies and Rules," IEEE Trans. Fuzzy Systems, vol. 13, no. 3, pp. 373-383, June 2005.

[4] G. Hignette, P. Buche, J. Dibie-Barthe´lemy, and O. Haemmerle´, "An Ontology-Driven Annotation of Data Tables," Proc. WISEWorkshops Web Data Integration and Management for Life Sciences, pp. 29-40, 2007.

[5] G. Hignette, P. Buche, J. Dibie-Barthe´lemy, and O. Haemmerle´, "Fuzzy Annotation of Web Data Tables Driven by a Domain Ontology," Proc. Sixth European Semantic Web Conf. The Semantic Web: Research and Applications (ESWC), pp. 638-653, 2009.

[6] P. Buche, J. Dibie-Barthe´lemy, and H. Chebil, "Flexible Sparql Querying of Web Data Tables Driven by an Ontology," Proc. Eight Int'l Conf. Flexible Query Answering Systems (FQAS), pp. 345-357, 2009.

[7] P. Cimiano, P. Buitelaar, J. McCrae, and M. Sintek, "Lexinfo: A Declarative Model for the LexiconOntology Interface," J. Web Semantics, vol. 9, no. 1, pp. 29-51, 2011.

[8] J. McCrae, D. Spohr, and P. Cimiano, "Linking Lexical Resources and Ontologies on the Semantic Web with Lemon," Proc. Eight Extended Semantic Web Conf. The Semantic Web: Research and Applications (ESWC), pp. 245-259, 2011.

[9] T. Declerck and P. Lendvai, "Towards a Standardized Linguistic Annotation of the Textual Content of Labels in Knowledge Representation Systems," Proc. Seventh Int'l Conf. Language Resources and Evaluation (LREC '10), 2010.

[10] A. Reymonet, J. Thomas, and N. Aussenac-Gilles, "Modelling Ontological and Terminological Resources in OWL DL," Proc. OntoLex 2007 - Workshop associated with ISWC '07, Sixth Int'l Semantic Web Conf. (ISWC '07), 2007.

[11] S´ebastien Destercke and Patrice Buche and Brigitte Charnomordic" Data reliability assessment in a data warehouse opened on the Web" Flexible Query Answering Systems, Ghent : Belgium (2011).