ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, Vol 9(23), DOI: 10.17485/ijst/2016/v9i23/95145, June 2016 ISSN (Online) : 0974-5645

An Ontological Approach for Originating Data Services with Hazy Semantics

R. Saranya, S. Gowri, S. Monisha and S. Vigneshwari

Department of information technology, Sathyabama University, Chennai, India; saranyar.cute@gmail.com, gowriamritha2003@gmail.com, monishabtch@gmail.com, vigneshwaris@gmail.com

Abstract

Objective: The tremendous growth in volume of data is increasing the complexity of data handling. The existing content based search is not only complex and expensive, but also leads to poor processing and unacceptable latency for massive amount of data. Data accessed from large dataset using existing approach will slow down the query response rate. High resource cost is the severe performance bottleneck caused by query operations. The semantic based search can be carried out on the data, In order to reduce the complexity in data handling. The proposed method uses Resource Description Framework (RDF) with Ontology Web Language (OWL). Thus the RDF is used for non domain based search and OWL is used for domain based search. The non domain based search provides the composition of best data services from master RDF. The triplets from different RDF are combined to form master RDF. And the domain based search is one which provides recommendations for the user based on their search query. The main objective of this approach is to provide semantic search using RDF and recommendation using OWL. The search result will be the high similarity result sets as the method uses the combinational RDF to form large master RDF.Statistical **Analysis/Methods:** The user is provided with two methods of searching which is domain based search and non domain based search. The domain based search is one which provides recommendation for the user based on the search query. And the non domain based search originates the data services from master RDF. Agriculture data set is chosen as a sample data set for domain based search which can be extended further for any other domain. The web services which contain general information about the world and also related web services is chosen as a data set for non domain based search. **Findings:** The statistical analysis of ontology based semantic search with keyword based search produced the precision value of 0.8 out of 1.0 using the search results obtained from both semantic search and keyword based search. **Applications/Improvements:** The application can be further improved by adding realistic data sets and increasing the size of the database. The domain based approach is not restricted only to agriculture but it can also be extended to some other applications like Health care, pattern recognition and so on.

Keywords: Correlation, Data Services, Domain, Hazy Semantics, Ontology, Triplets

1. Introduction

Over the past few years, the growth of data is very high. The existing content based search retrieves the information based on keywords in the query statement. Thus for the massive amount of data keyword based search leads to inflexibility which causes timing and budget issues. It always provides an imperfect solution because it does not consider many relevant documents as it always looks for keyword match. For example, search for documents containing the word “bike” will only retrieve the documents having the exact keyword. However it misses the documents containing potentially relevant words such as “Honda”, and “Bajaj”. Thus by increasing the number of search terms the above problem can be compensated but still the results will likely overlook some relevant documents. The issues of existing content based search can be resolved by context based search which is semantic search. The key component of semantic search is RDF[1]. The strong Semantic Web community support, this proliferation of RDF data can also be attributed to the generality of the underlying graph-structured model that is many types of data can be expressed in this format including relational and XML data. This data representation, although flexible, has the potential for serious scalability issues. Another problem is that schema information is often unavailable or incomplete, and evolves rapidly for the kind of RDF data published

* Author for correspondence

on the web. Thus, web applications built to exploit RDF data cannot rely on a fixed and complete schema but, in general, must assume the data to be semi structured. The master RDF repository helps in composing data services. The combination of different RDF's produces master RDF[2]. The best course of action can only takes place by means of recommendations. Thus the recommendation for the user is also possible in semantic based search. It is provided by Ontology Web Language. Each time when the user is proceeding with domain based search recommendations are given to the user based on search terms. In the perspective of this rag we discussed about a unique approach for originating data service and recommendations.

2. Resource Description Framework

RDF is the standard way of specifying the data. It is a framework for describing resources and the model for data. Thus the information is represented in the form of triplets called subject, predicate and object which is technically termed as resource, property and property value of a statement. For example, consider the statements which can be represented as triplet "The moon has the color white" as a subject denoting "the moon", a predicate denoting "has", and an object denoting "the color white". XML (eXtensible Markup Language) is used by the RDF as a general syntax which enables processing and exchanging of data. The XML syntax provides extensibility, liberty among vendors, human readability and also enables representation of complex structures. The RDF values are always atomic in nature and also have their own property[2]. The following Figure 1 is an illustration of RDF description. The two different statements can have same meaning. For example, "The moon has the color white" and "the moon is white in color" both the statement provides same meaning in human view. The machine may consider these statements as different strings. Thus problem can be solved by RDF which enables machine readable encoding. The RDF syntax which is derived from XML is one which stores instances of this model into machine-readable files and to communicate these instances among applications. RDF inflicts formal structure on XML and support the consistent representation of semantics. The effective use of information among applications requires common conventions. However there is a problem named word sense disambiguate, that is sense of the word may differ according to the place where they are used. This problem can be resolved using RDF namespace which identifies the property type in a unique manner. The RDF acts as a building block of managing information all over the web[3]. It provides the propensity for transfiguring the web into a more useful and substantial information resource.

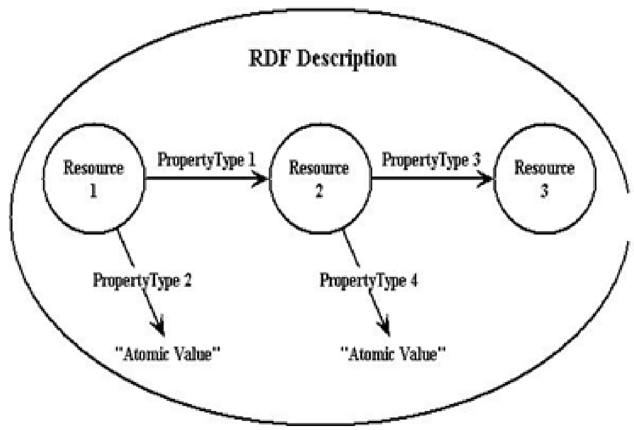

Figure 1. RDF Description.

3. Ontology Web Language

OWL is one of the W3C recommendations which are related to semantic web. It is used to represent the terms and also to find the bridging relationship between those terms. It is built on top of RDF. It was designed to provide a common way to process the content of web information. OWL differs from RDF by means of its characteristics such as stronger language, greater machine interpretability, larger vocabulary, and stronger syntax. It is relies on two properties namely data type and object[4]. Data type property is the descriptions of attributes of elements of classes. Object property is the descriptions of relations between elements of classes. OWL is written in XML by which information can easily be exchanged between different types of computers using different types of operating system and application languages[5]. OWL adds more vocabulary for describing properties and classes, relations between classes. OWL has three species namely OWL-DL, OWL Lite, and OWL Full. DL is the core dialect that includes DL primitives, but not necessarily be tractable. Lite adds restrictions to OWL-DL make it tractable. Full lifts restrictions to allow other interpretations. It is extremely general and potentially intractable. They are included only for expressiveness needs.

4. Synsets for Ontology Repository

Word Net is a lexical catalog that provides the meaning of words in human logical language, and these words are collocated in sets of synonyms, called synsets. The applications that process the natural language are much hanging on senses of words that are being processed. If a word has more than one sense then dictionary can be used to identify the correct sense, however, they were developed to serve for humans purpose[6]. The WordNet can be used to resolve this problem. It is a lexical framework that groups nouns, verbs, adverbs and adjectives as synonym sets. For example, the word "bass" has different meaning at different place based on the usage. "Bass is the nether most chunks of musical series, "bass part is the lowest in polyphonic music". "The basso is an adult male singer with the lowest voice" .The closed class lexical items are not included in the Word Net.

5. The Methodology

The user is provided with two methods of searching which is domain based search and non domain based search. The domain based search is one which provides recommendation for the user based on the search query. And the non domain based search originates the data services from master RDF. Agriculture data set is chosen as a sample data set for domain based search which can be extended further for any other domain. The web services which contain general information about the world and also related web services is chosen as a data set for non domain based search. The work flow of both the methodology is shown in Figure 2. The work flow consists of two repositories namely RDF repository and ontology repository from which information is retrieved. In background, WordNet an open source tool is used as language resource to provide ontology and lexical knowledge in Natural Language Processing (NLP) task[7]. Word Net hastily be similar to a phrasebook, in that it groups words together based on their significances. However, there are some essential distinctions. WordNet interlaces not just word forms and strings of letters but precise senses of words. As a result, words that are found in close juxtaposition to one another in the complex are semantically disambiguated. WordNett ages the semantics.

5.1 Domain Based Search

The module consists of different local ontology repositories like plants ontology, pest ontology, and soil ontology and so on. The farmer has to register details. The server in turn stores the farmer information in its database. Farmer enters soil test value N, P, and K to the server, now soil test value checks in Soil ontology. The three numbers on fertilizer represents the value of three macro-nutrients. They are Nitrogen (N), phosphorus (P) and potassium (K) or NPK for short. Each N, P, K value checks in each data type attribute in Ontology. Plants will be recommended to the farmer by the strategies like exact or above N, P, K values are checked with data type attribute in Soil Ontology, if both matches, relevant plants will be recommended first. Next one or two N, P, K values, if matches found then relevant plants will be recommended. Farmer information saved as an individual in farmer Ontology. Example 4.1 illustrates the storage structure of data which is OWL format. Farmer uploads a pest image to the server. The server checks in Pest Ontology to retrieve the solution of an uploaded pest image. Everyday system monitors weather report of the farmer district and stores weather information in database. Whenever the farmer login, past 3 days weather status informed to the farmer. If farmer request weather report and server provides solution from Weather Ontology. Farmer request fertilizer mix for plants to Server by giving the input as acre, plants, reason. Server checks in Soil Ontology and retrieve N, P, K values of the plant. Then check in Farmer Ontology and retrieve farmer plants N, P, K values. By comparing both values and also checks the reason Growth, pests, server provides fertilizer mixing for plants. Admin maintains price details of vegetables and fruits. Farmer request price and marketing information of plants, server provides solution from Marketing Ontology. The Figure 2 is the general architecture for both domain and non-domain based search.

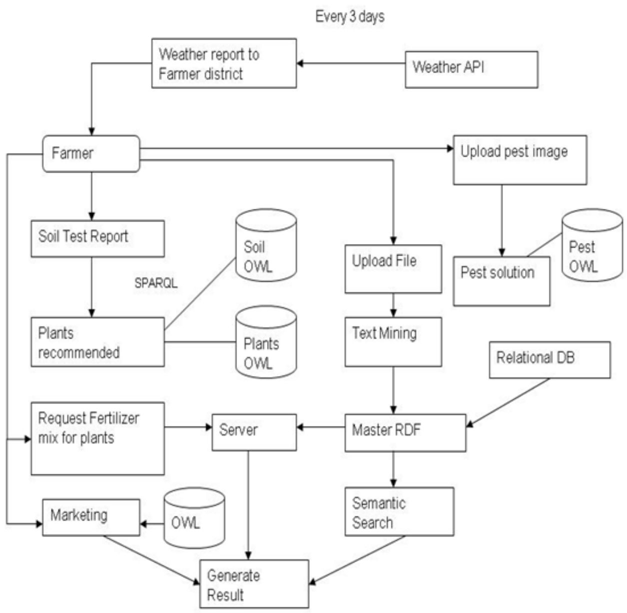

Figure 2. Semantic Search Work Flow Architecture.

5.2 Non Domain Based Search

A Schema is identified for the relational database and a RDF representing the schema of the database is constructed through model provided by the JENA-API[8]. JENA is a java API which supports the development of semantic web applications. It is an open source which supports RDF, RDF schema, and OWL, and also includes inference mechanism. The JENA transforms ontology into object-oriented abstract data model, and allows its primitives such as classes and relationships to be treated as objects. It allows manipulation of ontology elements using object oriented programming languages.

```
<rdf:RDFxmlns:rdf="http://www.w3.org/1999/02/22-
rdf-syntax-ns#" xmlns:owl=
http://www.w3.org/2002/07/owl#
xmlns:dc="http://purl.org/dc/elements/1.1/">
<owl:Class
rdf:about="http://www.agriculture.com/ontology#Soil"/>
</rdfs:subClassOf>
</owl:Class>
<owl:Class
rdf:about="http://www.agriculture.com/ontology#Tomato"
>
<rdfs:subClassOf
rdf:resource="http://www.agriculture.com/ontology#Soil"/
></owl:Class>
<owl:Class
rdf:about="http://www.agriculture.com/ontology#Rose">
<rdfs:subClassOf
rdf:resource="http://www.agriculture.com/ontology#Soil"/
></owl:Class>
<j.0:Chilli>
<j.0:Potassium>80</j.0:Potassium>
<j.0:Name>Chilli</j.0:Name>
<j.0:Phosphorus>80</j.0:Phosphorus>
<j.0:Nitrogen>120</j.0:Nitrogen>
</j.0:Chilli>
<j.0:Marigold>
<j.0:Name>Marigold</j.0:Name>
<j.0:Phosphorus>90</j.0:Phosphorus>
<j.0:Nitrogen>90</j.0:Nitrogen>
<j.0:Potassium>75</j.0:Potassium>
</j.0:Marigold>
</rdf:RDF>
```

Example 4.1. Data Stored in OWL Repository

The data level procedures are exposed to assembly level dispensation by indexing the semantic data elements. Multiple RDF's are grouped and structured together to form a master RDF data that holds all the semantic information's of a server that support reasoning in any

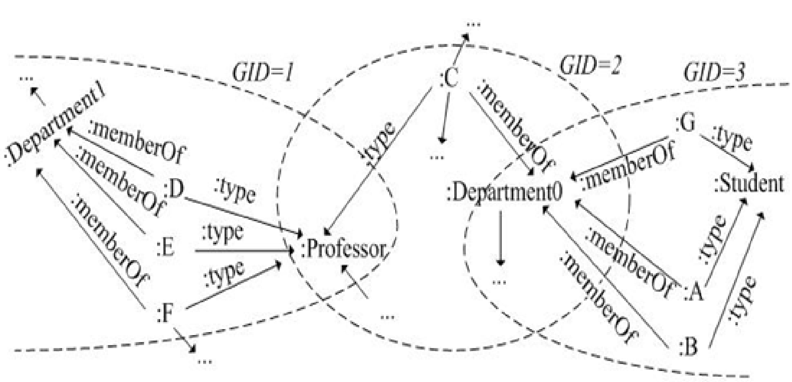

Figure 3. Partitioned RDF Graph.

formats of query processing. The Different resources are interlinked with high degree of relational factors by the predicates in the triples. The Query processing is handled directly in the RDF file by iterating the triples forming a discrete relation with the Service query and the URI representing the location of the resource is returned.

6. Querying the Master RDF Repository

The multiple RDF repositories are queried using SPARKQL. Thus the data in RDF structure is translated into graph model. From each RDF repository graph is obtained and ID called GID is assigned[10]. Then the patronized triplet table is obtained from graph as illustrated in Figure 3. The Table 1 illustrates the triplet table. The master RDF is a combination of different RDF which makes the highest similarity measure to form highest information gain.

The data from RDF one is assigned group ID as GID=1, the data from RDF two is assigned group ID as GID=2 and so on. The example considered are A,B,G are the members of Deparment0 who are all students and D,E,F are the member of department 1 who are all Professor. Sometimes there may not be any meaning in triples obtained[9].

Table 1. Partitioned Triplet Table

SUBJECT	PREDICATE	OBJECT	GID
:D	:Type	:Professor	1
:E	:Type	:Professor	1
:F	:Type	:Professor	1
:D	:memberOf	:Department1	1
:E	:memberOf	:Department1	1
:F	:memberOf	:Department1	1
:C	:Type	:Professor	2
:C	:memberOf	:Department0	2
:A	:Type	:Student	3
:B	:Type	:Student	3
:G	:Type	:Student	3
:A	:memberOf	:Department0	3
:B	:memberOf	:Department0	3
:G	:memberOf	:Department0	3

From the Table 1 the dependency between group 1, 2 and 3 can be easily found. For example consider A, where person A belongs to department0 and person A is a student. Thus only by grouping the data elements additional information of the same object can be obtained.

7. Related Work

7.1 Probabilistic Approach for Devising Web Services

The query and data services with doubtful semantics use Probabilistic Service Registry (PSR) to represent the services along with their possible semantic views. The schema for the registry is service, view and probability[10]. It follows two types of search mechanisms like name based matching and content based matching. Block Independent Disjoint (BID) model is used for finding correlation between tuples. It involves Probabilistic Service Registry transformation and query rewriting. The doubtful compositions are interpreted through the possible words. This approach cannot handle the uncertainty at the data level, that is, when both the semantics of a service and its returned data are uncertain it may produce poor results.

7.2 Thematic Similarity Based Search

The thematic similarity approach has been used for information retrieval to capture the context of concepts. It resolves the issues of semantic heterogeneity in retrieving information. It uses distance based-approaches with the help of domain ontology. It involves obtaining triple frequency for each document. Initially based on the term similarity scores the documents are ranked[11]. Then based on the user query and the search results viewed by the user the rank of the document is upgraded. The current system does not consider partially or incomplete matched RDF triples that may contain important information.

7.3 X-ENS: Semantic Enrichment of Web Search

This model combines the pros of Semantic Web standards and common Web Searching. X-ENS discovers entities of absorption in the snippets of the best search results which can be later utilized in a faceted search-like interaction scheme, and as a result it can help the user to limit the very large search space to those hits that contain a specific piece of information[12]. Moreover, X-ENS allows the exploration of the discovered entities by utilizing semantic repositories. The entities are used as the "glue" for automatically connecting the documents with data and knowledge. The quality of the results, that highly depends on the quality of the snippets. Entity disambiguation is a problem that affects the quality of the presented entities.

8. Conclusion

The semantic search over comes the drawback of keyword based search. The Resource description Framework and Ontology Web Language plays a significant role in semantic web. Thus the ontology and RDF based semantic search resolves the issues of thematic similarity based semantic search and probabilistic approach for devising web services. In Addition to that Recommendations for the search query is provided using OWL. The constitution of data services in non domain based search provides the best result for the user as itis a combination of several RDF's that forms master RDF. The recommendations in domain based search provides advocacy for the user who does not have any prior knowledge about the domain. The future work can be enhanced by applying this novel methodology for various real time scenarios. Thus the domain based and non domain based search can be extended further and applied for large dataset.

9. REFERENCES

1. Mozhdeh Nazari Soleimandarabi, Seyed Abolghasem Mirroshandel. A Novel Approach for Computing Semantic Relatedness of Geographic Terms. Indian Journal of Science and Technology. 2015 October; 8(27). DOI: 10.17485/ijst/2015/v8i27/60811
2. Benjelloun O, Sarma AD, Halevy AY. Databases with uncertainty and lineage. VLDB J. 2008; 17(2):243–64.
3. Abdelamid Malki, Mahmoud Barhamgi, Sidi-Mohamed Benslimane, Djamal Benslimane, Mimoun Malki. Composing Data Services with Uncertain Semantics. IEEE Transactions on Knowledge and Data Engineering. 2015 April; 27(4).
4. Pavlos Fafalios, Yannis Tzitzikas. X-ENS: Semantic Enrichment of Web Search Results at Real-Time. Institute of Computer Science, FORTH-ICS, GREECE, and Computer Science Department, University of Crete, GREECE. 2013 July; 1-8
5. Hogan A, Harth A, Umbrich J, Kinsella S, Polleres A, Decker S. Searching and browsing linked data with WSE: the semantic web search engine. Web Semantics: Science, Services and Agents on the World Wide Web, 2011; 9(4).
6. Benouaret K, Benslimane D, et.al. Fudocs: A web service composition system based on fuzzy dominance for reference query answering. PVLDB. 2011; 4(12):1430–33.
7. Barhamgi M, Benslimane D, Medjahed B. A query rewriting approach for web service composition. IEEE Trans. Serv. Comput. 2010 Sep.; 3(3):206–22.
8. Martin DL, Paolucci M, Burstein MH, McDermott DV, Parsia B, Sycara KP. Bringing semantics to web services: The owl-s approach. In Proc. Semantic Web Services Web Process Composition. 2004; 26–42.
9. Ying Yan, Chen Wang, Aoying Zhou, Weining Qian, Li Ma, Yue Pan. Efficiently Querying RDF Data in Triple Stores, Beijing, China, 2008 April 21-25; 1053-54. DOI: 2008 978-1-60558-085-2/08/04
10. Tumer D. Shah MA, Bitirim Y. An empirical evaluation on semantic search performance of keyword-based and semantic search engines: google, yahoo, msn and hakia. Dept. of Comput. Eng., Eastern Mediterranean Univ., Famagusta; Shah, M.A. Bitirim, Y. IEEE. 2009; 51-55. ISBN: 978-1-4244-3839-6. DOI: 10.1109/ICIMP.2009.16.
11. Sharifullah, Jibran Mustafa. Effective semantic search using thematic similarity. Journal of King Saud University - Computer and Information Sciences. 2014 July; 26(2): 161–69.
12. Shabana Asmi P, Justin Samuel S. An analysis and accuracy prediction of heart disease with association rule and other data mining techniques. Journal of Theoretical and Applied Information Technology. 2015 20th Sep.; 79(2):254-60.