PRIOR ONTOLOGY SELECTION AND QUERY TRANSLATION FOR INFORMATION SEARCH

POORNIMA N[1], SHIVAM AGRAWAL[2], SALEENA B[3]

Department of , School of Computing Science and Engineering, VIT University, Chennai, Tamil Nadu, India. Email: vijayalakshmi.av@vit.ac.in

ABSTRACT

Objective: Most of the current search engines follow informal keyword based search. Finding the user intention and improving the relevancy of results are the major issues faced by the current traditional keyword based search. Targeting to solve the problems of traditional search and to boost the retrieval process, a framework for semantic based information retrieval is planned.

Methods: Social and wine ontologies are used to find the user intention and retrieving it. User's natural language queries are translated into SPARQL (SPARQL Protocol and Resource Description Framework query language) query for finding related items from those ontologies.

Results: The proposed method makes a significant improvement over traditional search in terms of some searches required for searching a particular number of pages using performance graph.

Conclusion: Semantic based search can understand the user intention and gives better results than traditional search.

Keywords: Semantics, Data retrieval, Ontology, Resource Description Framework, SPARQL.

INTRODUCTION

Today's search engines are keyword-based. Current search engine requires significant manual effort because the relevant data sources are both syntactically and semantically heterogeneous. As a result, traditional search engines are error prone, labor intensive, and time-consuming process. Semantic technology specifications such as resource description framework (RDF) [1], RDF schema, and web ontology language (.owl) [2] gives a formal relationship between the different classes of entities. These semantic technologies can be applied to identify the relationship between the web pages. SPARQL (SPARQL Protocol and RDF query language) is a semantic query language that can be used to retrieve and manipulate data from RDF.

Semantic web
Semantic web [3] is an extension of the normal web that describes the relationship between different things. Even computers can understand, relate and combine information without human interruption. Major difference between semantic web and other technologies is that it considers only the relation and not the structure of information.

Ontology
Ontology [4] is the semantic tool used to build an application for a specific domain. Ontology captures semantic relations from different documents.

Semantic search
Semantic search engine [4] searches the web based on the meaning of the user query. This produces more relevant results than keyword based search. Semantic search engines understand user intention correctly by revealing the hidden information from the user query.

Related work
This section describes various techniques and approaches carried out for semantic web search.

Alfrjani et al. [5], aims in finding an opinion on the particular product using product domain ontologies. Domain knowledge about the product is obtained and converted into the formal ontology. Finally, procedures link formal ontology and natural language processing.

Ontology Based E-healthcare Information Retrieval System [6] focuses on retrieving health care information. Health-care information includes treatment for disease, location of hospital, travel information, and nearby hotel details. Health-care domain ontology is created using human disease treatment knowledge base. Query from user is given to WordNet through which synonyms of query are taken. These user queries are translated into SPARQL queries. SPARQL query filters the health-care domain ontology and gives the relevant information.

Vijayarajan et al., [7] proposed a method for ontology based image retrieval system. Domain ontology is built using the image description to minimize the semantic gap between the image content and user query. Natural language processing algorithm converts the query into subject-predicate-object model. From this model, images are retrieved from the auto generated SPARQL queries.

Proposed work
A methodology for ontology based data retrieval model that will boost the efficiency of data retrieval has been planned.

Retrieval process
Applets and JFrames are used to build a graphical user interface consisting of a search text field and radio buttons. The radio buttons are connected to RDF. The process of retrieval consists of a method which converts natural language query into SPARQL [8] Query. SPARQL query is created from the natural language query given in the search text field and ontology is selected through radio button. Both the inputs from search text field and radio buttons, are used to search the ontology, and relevant results are generated (Fig. 1).

Social and wine ontology
Social ontology is chosen to build a model for retrieval system. This social ontology is made up of classes, subclasses, and relationships among them. Social ontology is stored in ontology repository as shown in (Fig. 1) with file format .owl. The hierarchical structure consisting of different categories of social ontology is illustrated in Fig. 2.

Wine ontology describes all the consumable things, fruits and wine with different classes and descriptors. Structure of wine ontology is shown in Fig. 3. Wine ontology is also stored in ontology repository (as shown in Fig. 1).

Data annotations

Semantic annotation is the method of attaching more related data to various ideas represented in the ontology (e.g., people, things, places, and organizations). This annotation is not similar to classical text annotations which are used by machines for reader's reference.

When a document (or any piece of content, e.g., video) is semantically annotated, it becomes a supply of knowledge which is simple to interpret, mix and utilized by other computers. Illustrated in Fig. 1 as annotation documents.

Data properties

Data properties allow us to add/edit values of data, strings, integer, date, and time into classes of an ontology. These are similar to object properties excluding their domains which are in literals.

SPARQL query

SPARQL [8] queries for searching ontology are described as follows:

A. Based on the individual, type and ordered by class.
SELECT distinct ?ind ?class WHERE ?ind rdf:type ?class
ORDER BY ?class
Above query displays the individual, type and class which is aggregated on the basis of class.

B. Based on type and winery SELECT ?x WHERE
?x wine:has Flavor wine:Strong.
?x wine:locatedIn wine:NewZealand Region.
?x wine:has Sugar wine:Dry
Above query displays the wine with a specific flavor, which is located or made in New Zealand Region, which has sugar and is dry by nature.

C. Based on wine, property and value.
SELECT DISTINCT ?wine ?property ?value WHERE
?o1 a ?class
?wine a ?o1
?wine ?property ?value.
Above query displays names of wine with properties and values aggregated on the basis of class.

D. Based on product and issue.
SELECT ?product ?issue
FROM name graph WHERE
?product issues:has Issues ?issue.
Above query displays product name and their issues.

Implementation details

All the information about various foods, hotel, and restaurant are collected. After collecting all the information, relationship forms are created such as one food to another, one tourist place to hotel, and one hotel to shopping mall. These relationships are created through ontology using protégé tool. Protégé is an OWL file editor that not only creates the relationship form but also collects all the irrelevant information by activating responder. All these relations are considered according to user needs and context dependencies.

After storing all the relationship information regarding food and various food related entity, now the relevant information needs to be filtered from a large database. To do this, an efficient ontology query language
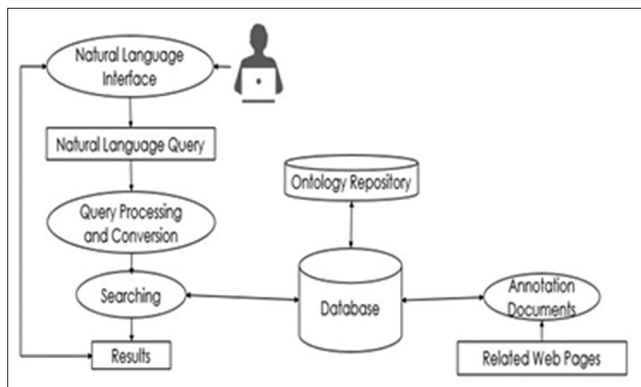
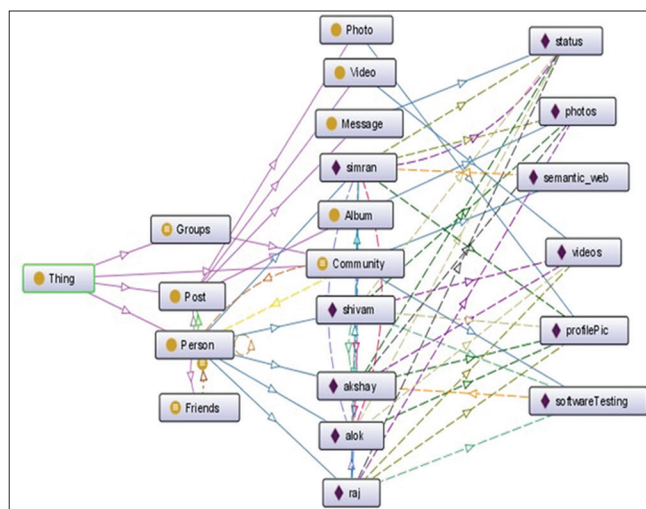

Fig. 1: Information retrieval process
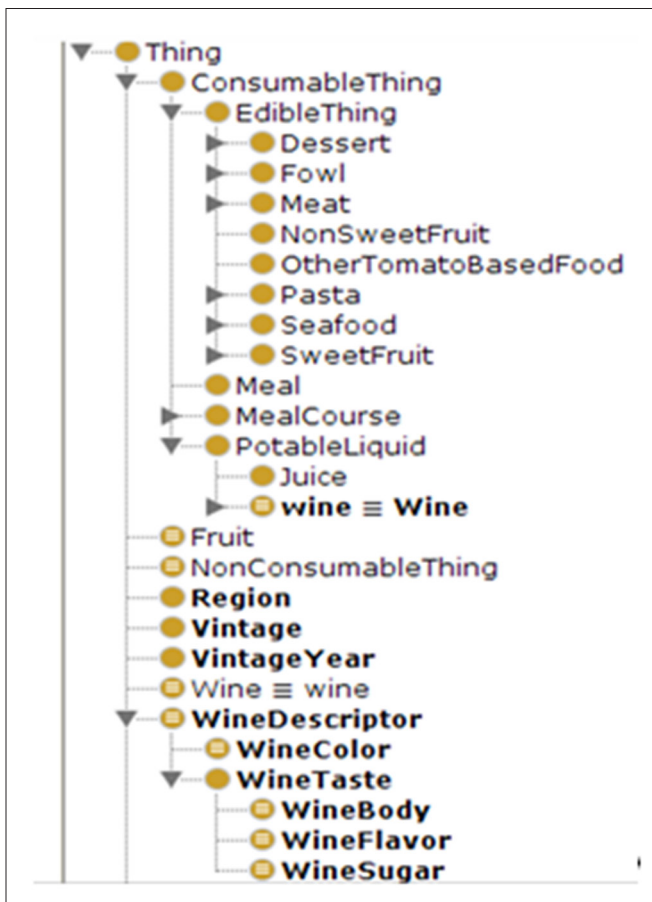

Fig. 2: Structure of social ontology

Fig. 3: Structure of wine ontology
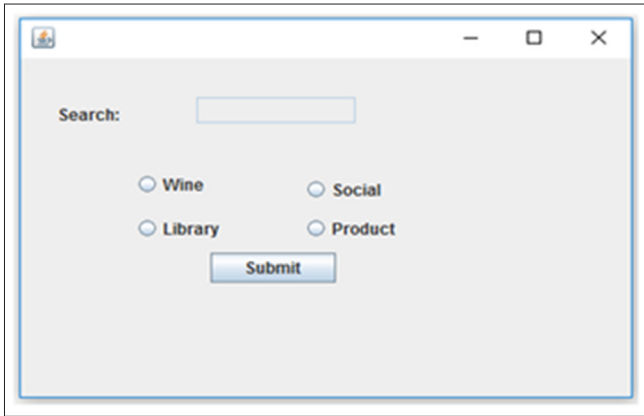
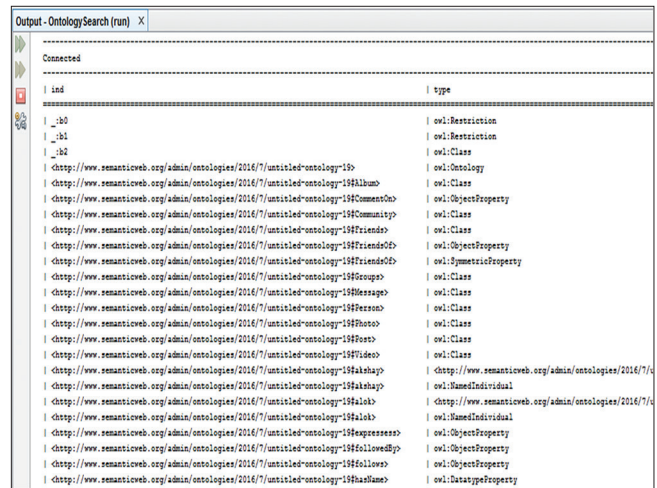Fig. 4: Dialog box for user request and choosing ontology

Fig. 5: Java code showing creation of SPARQL query and getting results from ontology

Fig. 6: Results based on search query given and ontology chosen by user

called SPARQL [8] is used. SPARQL is an intermediate that collects the data from ontology; then it will be matched with values stored in the database. SPARQL is a query language that stores, manipulate and retrieve the information from RDF [1].

After collecting all the relevant information from the database, Java Jena API is used to show all the outputs to the user. Java Jena API supports SPARQL query so that it can collect information from the ontology.

User search requests are taken through dialog box shown in Fig. 4. This dialog box consists of the text field and four different radio buttons. Search query is given through text field and ontology in which text search is to be done can be selected through radio buttons.

Fig. 5 shows how SPARQL queries are created from user natural language query using Java Jena API code. Code also shows how results are produced.

Performance analysis

Semantic web search works quicker than any everyday search engine since all the information in database are divided into many domains of comparable ontologies. For a general internet search engine, the task remains solely looking domain-wise, and thus resultant page are

given.

Performance analysis is carried out by comparing the number of searches need to get the relevant resultant page from the total number of web pages. Since semantic search filters the information into domains and subdomains, a number of searches needed to spot a relevant information from the database is lesser comparing to the general search engine.

CONCLUSION

Semantic web is an extension of current day web. Any web page can be described by domain, subdomain, classes, subclasses, etc., using semantic web technologies. Semantic web technologies find application in finding relationships, sharing and reusing resources, and so on. This paper presents a new methodology of searching relevant information from the large database using different ontologies. This methodology also focuses on translating natural language query (keyword based query) into SPARQL query for searching inside ontologies. Semantic web search reduces the amount of time spent on searching the web and also reduces the amount of energy spent on analysis the web pages. Future work aims in improving the semantic web search and also to remove the selection of particular ontology from the given ontology list.

REFERENCES

1. Available from: https://www.w3.org/RDF.
2. Available from: https://www.w3.org/OWL.
3. Rajasurya S, Muralidharan T, Devi S, Swamynathan S. Semantic information retrieval using ontology in university domain. Int J Web Seman Technol 2012;3(4):55-68.
4. Remi S, Varghese SC. Domain ontology driven fuzzy semantic information retrieval. In: International Conference on Information and Communication Technologies ICICT. Vol. 46. ICICT; 2014. p. 676-81.
5. Alfrjani R, Osman T, Cosma G. A new approach to ontology-based semantic modelling for opinion mining. In: 18th International Conference on Computer Modelling and Simulation; 2016. p. 267-72.
6. Raj TF, Govindarajan P, Ravichandran KS, Gayathri M, Uma R. Ontology based E-healthcare information retrieval system: A semantic approach. Int J Recent Innov Trends Comput Commun 2016;4(4):365-9.
7. Vijayarajan V, Dinakaran M, Tejaswin P, Lohani M. A generic framework for ontology based information retrieval and image retrieval in web data. Hum Centric Comput Inf Sci 2016;6(18):1-30.
8. Available from: https://www.w3.org/TR/rdf-sparql-query.