



Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/pisc



Design and development of semantic web-based system for computer science domain-specific information retrieval[☆]

Ritika Bansal^{*}, Sonal Chawla¹

Department of Computer Science and Applications, Panjab University, Chandigarh, India

Received 5 February 2016; accepted 10 April 2016

Available online 27 April 2016

KEYWORDS

Semantic Web;
Ontology;
SPARQL;
Protégé;
QUEPY

Summary In semantic web-based system, the concept of ontology is used to search results by contextual meaning of input query instead of keyword matching. From the research literature, there seems to be a need for a tool which can provide an easy interface for complex queries in natural language that can retrieve the domain-specific information from the ontology. This research paper proposes an IRSCSD system (Information retrieval system for computer science domain) as a solution. This system offers advanced querying and browsing of structured data with search results automatically aggregated and rendered directly in a consistent user-interface, thus reducing the manual effort of users. So, the main objective of this research is design and development of semantic web-based system for integrating ontology towards domain-specific retrieval support. Methodology followed is a piecemeal research which involves the following stages. First Stage involves the designing of framework for semantic web-based system. Second stage builds the prototype for the framework using Protégé tool. Third Stage deals with the natural language query conversion into SPARQL query language using Python-based QUEPY framework. Fourth Stage involves firing of converted SPARQL queries to the ontology through Apache's Jena API to fetch the results. Lastly, evaluation of the prototype has been done in order to ensure its efficiency and usability. Thus, this research paper throws light on framework development for semantic web-based system that assists in efficient retrieval of domain-specific information, natural language query interpretation into semantic web language, creation of domain-specific ontology and its mapping with related ontology. This research paper also provides approaches and metrics for ontology evaluation on prototype ontology developed to study the performance based on accessibility of required domain-related information.

© 2016 Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

[☆] This article belongs to the special issue on Engineering and Material Sciences.

^{*} Corresponding author. Tel.: +91 09988414517.

E-mail addresses: rbansal_298@yahoo.co.in (R. Bansal), sonal_chawla@yahoo.com (S. Chawla).

¹ Tel.: +09815968085(Mobile)

Introduction

Machine processable information enables precise in-depth searching along with the reduction in user time and effort and also it will give new possibilities to knowledge management and agent-based processing (Malik, 2009). Searching in semantic web-based system is done by contextual meaning of query instead of keyword matching as done in regular search engines (Chen and Zhang, 2012). In computer science, ontology represents a common understanding of a domain in which semantics of data are machine-understandable. Ontology works on metadata and facilitates the functionality of semantic matching to the search engine along with the exchange and integration of knowledge. This paper shows how RDF-based ontology is developed specific to some particular domain and provides a model for the same. Also, retrieving the required results from the ontology through natural language query along with the evaluation of ontology with specified sample size is shown. For this, ontology-based information retrieval system for computer science domain is designed which is known as IRSCSD system. Without technical knowledge of semantic web technologies like RDF, users can easily access the information by using this system. This paper provides the complete layout of IRSCSD system by considering whole spectrum of ontology-based information retrieval system for computer science domain.

Framework for IRSCSD system

IRSCSD system provides user friendly interface that accepts queries in natural language and extracts data from ontology which is domain specific to retrieve the desired results. This system converts the natural language input query into query language of RDF database which is SPARQL, thus no need of learning SPARQL language. SPARQL query is then executed on the ontology for accessing the relevant information. High-level design of IRSCSD system comprises of three main stages. Ontology building is a first stage which is a core part of the system for which protégé tool is used. Second stage is NLQ to SPARQL conversion for which QUEPY tool is used which is a python-based framework. Third stage is the execution of converted SPARQL queries on RDF database through Apache's Jena API to get the results (Bansal and

Chawla, 2014). Lexicon library is built by extracting the entities from ontology and is expanded by adding synonyms of entities extracted with the help of Word-Net (Wang et al., 2007). Natural language query is parsed to form the parse tree after the recognition of named entities. Then query terms are matched with the concepts and properties of the ontology and finally triples are generated. SPARQL query is generated by integrating these triples. Each stage is expanded into its components and the detailed design of the framework is shown below in Fig. 1.

Prototype building for IRSCSD system

For IRSCSD system, prototype ontology is developed in computer science domain by considering two of its core topics, i.e. stack and queues. The first step towards ontology development is finding an appropriate tool. In Bansal and Chawla (2013), comparison of various ontology development tools has been done by comparing features like base language, modelling features, import/export format, graph view, consistency check and many more. Protégé is found to be the most ruling and domain independent tool. Protégé is based on java, provides GUI, is extensible and creates data in RDF format. It is a suitable base for rapid prototyping and application development as it provides a ready to use environment. In an online survey (Khondoker and Mueller, in press), it was found that protégé is used by 75% respondents. So, Protégé Tool is used for the development of prototype ontology for IRSCSD system. Global knowledge bases are used to learn ontology for web information gathering. Our ontology is based on the Word Net's database consisting of terms on data structure topic of computer science domain (Salahli et al., 2009). We also used other lexical resources: dictionaries, glossaries, etc. from the Web. These resources allowed us to extent and correct the contents of the ontology. For ontology development, UMLs are used in its initial phases as it is based on object-oriented paradigm (Gašević et al., 2005). Under object modelling group, MDA (model-driven architecture) standards are followed for ontology development. MDA has four layers named as M0, M1, M2, and M3 where various standards are defined for each separate layer. M3 layer is meta-meta model layer at the top of the MDA architecture consisting of set of concepts which

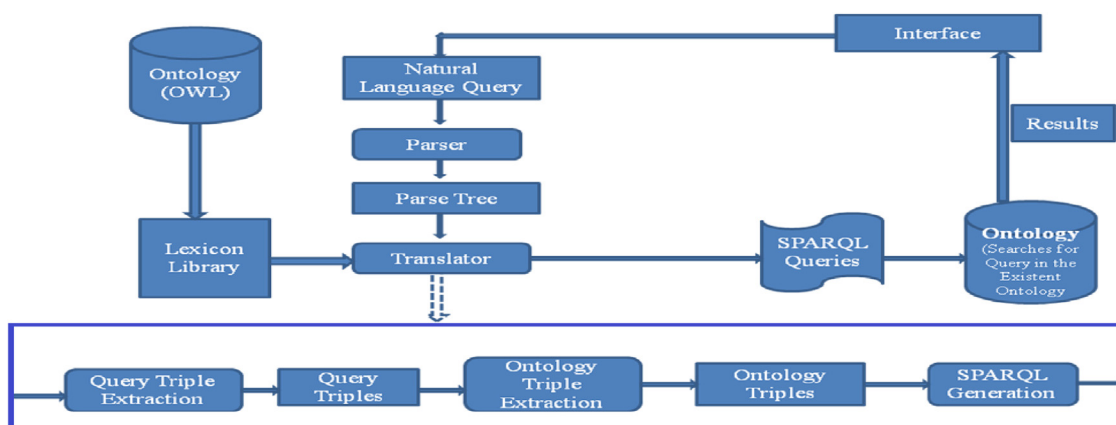


Figure 1 Framework of IRSCSD system (Bansal and Chawla, 2014).

Download English Version:

<https://daneshyari.com/en/article/5518912>

Download Persian Version:

<https://daneshyari.com/article/5518912>

[Daneshyari.com](https://daneshyari.com)