

DOMAIN SPECIFIC ONTOLOGY BASED QUESTION ANSWERING SYSTEM

^[1] Sini G.M, ^[2] Kavitha Raju , ^[3] Binu R

^{[1][2]}Dept of Computer Science and Engineering, Govt. Engineering College Sreekrishnapuram, Palakkad, Kerala, India

^[3]Asst. Prof. Dept. Computer Science and Engineering. Govt. Engineering College Sreekrishnapuram, Palakkad, Kerala, India
^[1] sinimohanan7@gmail.com , ^[2] kavitharaju18@gmail.com , ^[3] binurajappan@yahoo.com

Abstract: *Question Answering System (QAS) are basically developed to extract precise answers from a set of unstructured documents whose contents are in natural language. Information retrieval techniques are used to extract the correct answers for the natural language queries. But in recent years, when the idea of ontology came, it has been moving from the realm of artificial laboratories to the desktops of domain experts. It is viewed as a structured form of data where we can retrieve the data in the form of triplets. SPARQL, a semantic query language can be used to query from an ontology knowledge base. This paper focus on automatic query generation for natural language questions, thereby developing a domain specific natural language question answering system for ontology. Accuracy of the system is improved by using the DBpedia ontology for extracting some sort of information about the domain.*

Keywords: *Ontology; SPARQL; DBpedia*

I. INTRODUCTION

In this era of fifth generation of computers, people surfing for information on web are getting increasing day by day. Search engines are working based on three stages: contents about the query that the user searching for will be discovered in the crawling stage which is followed by indexing where the discovered data is analyzed and stored. Finally the analyzed data is retrieved in the form of list of relevant document pages. During this process user is lacking the exact precise answer for the query. Here comes the importance of question answering system which is a major application of information retrieval models. Question answering systems provides short precise answers for a specific query. In recent years with the development of ontologies, it becomes much easier to develop systems of various applications in the field of natural language processing. Ontology defines common vocabulary of basic concepts and relations among them. Such a knowledge representation is more beneficial to extract information about the domain. Ontology defines a structured form of data from where the application of question answering system can be developed in much easier way.

A. Ontology

Ontology refers to a body of knowledge which describes about the facts in a specific domain in a structured way. Ontologies are becoming increasingly popular modeling schemas for knowledge management services and applications [1]. Hierarchical structure of data (ontology) gives a common understanding among the user agents. Defining the classes is the primary stage of developing ontology. Subclass–superclass hierarchy makes it easier to pick the exact information from ontology. Reusability is the important feature for developing new systems. Protégé is the most popular tool for creating ontologies developed at Stanford University. It is a free, open source and knowledge requisition system written in Java and uses swings to create complex user interface [2]. It can create classes with several instances along with its object properties and data properties. Hozo, DOML etc. are other tools used for developing ontology. By developing such ontologies, the information retrieval becomes easier task. Thus the ontology becomes more important for applications in the field of information retrieval systems.

B. SPARQL

SPARQL (SPARQL Protocol and RDF Query Language) is a semantic query language used to retrieve and manipulate data stored in the form of RDF triplets i.e. subject-predicate-object triples. Resources are represented with URIs which can be abbreviated as PREFIX names. FOAF is standard RDF vocabulary for describing people persons, their activities and their relations to other people and objects. DBpedia is another RDF version of information which is extracted from Wikipedia. These datasets can be used for SPARQL querying.

II. PROPOSED SYSTEM

Question answering systems are the main application in the field of information retrieval tasks. Information retrieval from a knowledge base is more specific than from a set of documents. Thus the idea of ontology becomes more relevant for information retrieval task. Proposed system aims at producing the specific answers in natural language for natural language queries. SPARQL query generation for a natural language question is an important part of generating the natural language answers. SPARQL endpoints can be used to generate the answers for SPARQL queries. General architecture for QAS system is shown in Fig 1.

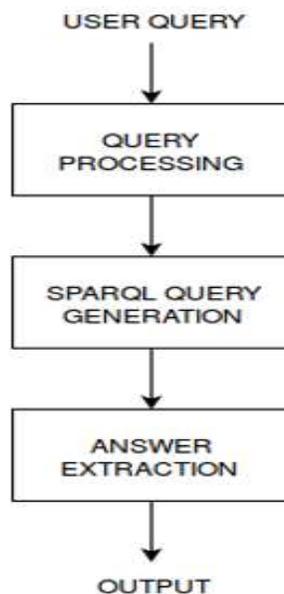


Fig. 1. Basic Architecture Of Ontology Based Qas System

A. QAS System

Proposed Question answering system focused in the domain of news events of new appointments in various organizations. A manually created ontology which have information about news events mainly focused on the organizational changes. This manually created ontology can give information only about the events that are added. In order to improve the accuracy of the system, DBpedia ontology is used to extract some basic information about the organization, personalities etc. DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia and Wikidata and make this information available on the Web.

The main stages of developing QAS system include:

1. Ontology creation & population
2. Query processing
3. SPARQL query Generation
4. Answer Retrieval

1. Ontology creation & population

Primary stage of developing the QAS system is the Ontology creation. Corpus of news articles related to the new appointments in organizations are collected (domain is fixed as the news articles related to the new appointments in

organizations). Protégé is the well-known tool for creating ontologies. From the corpus of news articles, each news events is added as events in the base ontology. Base ontology contains the classes for an event, instances of persons and organizations as shown in Fig 2. Classes can be connected by using the object properties. A specific news contains the information about the person who is appointed in an organization, designation, date of appointment, location etc. Base ontology contains 'Person' class to add the personalities, 'Organization' class to add the instances of organizations etc. A person will

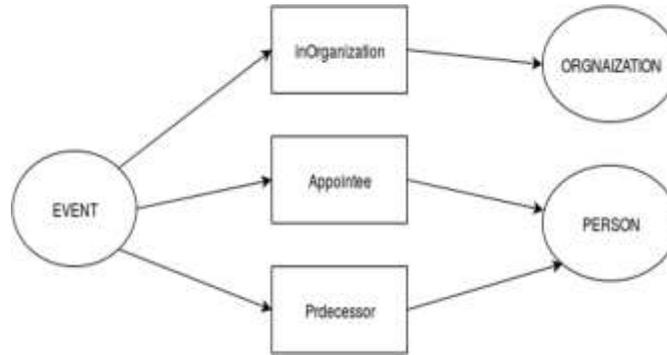


Fig. 2. Base Ontology for the domain of news articles.

have data property “designation” which contain the position in a particular organization to which a person belongs to. News article may mention some designation as CEO or Chief executive or Chief executive officer or Head etc. A unique value has to be given to designation field in order to avoid the problem while querying in ontology. In the same way the organization names are also added as acronyms eg: “The Board of Control for Cricket” as BCCI.

2. Query processing

Processing the natural language question is an important part in developing the QAS System. Answer depends on how we are analyzing the question. Question processing is difficult because a natural language query can be in any forms. Same question may be asked in different forms. E.g. “Who is appointed as the new CEO in BCCI?” The same query can be asked as “Present chief executive in BCCI?”, “Newly appointed CEO in BCCI?” etc. For the retrieval of correct answer we need to collect some information from the question. i.e. BCCI - Organization, Position - CEO . Thus it is important to extract the relationship between the entities in the question. Proposed work concentrate mainly on 5 types of questions. i) In which organization the event happened? ii) Who is appointed in the event? iii) What is the designation of a person? iv) Who is the previous person in position? v) When the event happened? Organization name, person name, designation, date are the respective answers for the above type of questions.

Adapt Intent Parser; open source software library can be used for converting natural language into machine readable data structure which is used for obtaining the information from natural language query. E.g.: “Who is appointed as the chief executive officer in The Board of Control for Cricket?” Output for this stage:

```

{
  "intent_type": "CLASS2",
    "ORGANIZATION": "BCCI",
    "KEYWORD": "appointed"
    "POSITION": "CEO"
}
  
```

Here the designation “chief executive officer” is mapped to CEO and “The Board of Control for Cricket” to BCCI. With the help of this information an intermediate query will be generated to match the class to form the SPARQL query. Intermediate representation for the above question would be “CLASS2 CEO BCCI”.

3. SPARQL query generation

Creating this SPARQL query manually is a tedious task, which is time consuming also. Thus an automatic SPARQL generation is needed. This can be achieved by using quepy(Python framework).. Quepy is a tool for converting natural language input queries into SPARQL query [3]. Quepy is a python framework to transform natural language questions to queries in a database query language. It provides a basic structure of SPARQL query. It can be easily customized to different kinds of questions in natural language and database queries. Quepy mainly specifies three types of classes: FixedRelation (It states that "A is related to B" through the specific relation), FixedType (To define something with a special type), FixedDataRelation ("A is related to Data" through a specific relation). Object properties and data properties are to be defined in classes in order to generate the queries. From the intermediate query generated during the Query processing stage is used here. E.g. "CLASS2 CEO BCCI" . Query generated as:

```
PREFIX ns:<http://www.semanticweb.org/user/ontologies/2016/2/untitled-ontology-74#>
PREFIX xsd : <http://www.w3.org/2001/XMLSchema#>
```

```
SELECT DISTINCT ?x3 WHERE {
?x2 ns:name ?x3.
?x2 ns:designation "CEO"^^xsd:string.
?x1 ns:Appointee ?x2.
?x1 ns:inOrganization ?x0.
?x0 ns:name "BCCI"^^xsd:string.
}
```

4. Answer Retrieval

SPARQL query is used to extract answers from the ontology. During the last stage, answer for a natural language query can be retrieved which is in the form of triplet. RDFLib is a Python library for working with RDF, a simple yet powerful language for representing information. Using RDFlib python package the answer for the particular query can be obtained from the graph structure (ontology).

III. CONCLUSION AND FUTURE WORK

Ontology is fast growing field in this fifth generation of Natural language processing and artificial intelligence. One of the main applications using ontology is the question answering systems. Ontology based question answering system gives a platform for people surfing for precise answers on the web. Querying from an ontology using SPARQL is an important part which can be achieved using python quepy framework. Thus ontology based question answering system becomes much easier task than before.

Proposed work limits within a specific domain. The DBpedia community project extracts structured, multilingual knowledge from Wikipedia and makes it freely available on the Web using Semantic Web and Linked Data technologies [4]. Proposed work can be extended to extract basic information from Dbpedia ontology, so that the system can handle more questions about the domain.

ACKNOWLEDGMENT

This work is done as a part of academic thesis at Govt. Engineering College Sreekrishnapuam, Palakkad, Kerala. I express my sincere gratitude to my guide and all others who helped me to do this work successfully.

REFERENCES

- [1] R. Sivakumar, P. Arivoli, "Ontology visualization Protege tools – a review," in proc: International Journal of Advanced Information Technology (IJAIT) vol. 1, No.4, August 2011.
- [2] Vishal Jain, dr. Mayank Singh, " Ontology Development and Query Retrieval using Protege Tool" , in proc: I.J. Intelligent Systems and Applications, DOI:10.5815/ijisa.201.2013.

[3] Ms. Pratibha S. Sonakneware and Prof. S. J. Karale , “Ontology Based Approach for Domain Specific Semantic Information Retrieval System” in proc: International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, April 2014.

[4] Jens Lehmann et al. “Dbpedia a large scale ,Multilingual knowledge Base Extracted from Wikipedia” , in proc: Semantic Web (2012) 1-5 1570-0844/ IOSpress, 2012.