# Personalized Semantic Search based Intelligent Question Answering System using Semantic Web and Domain Ontology

S.Kalaivani
Assistant Professor / Dept Of CSE,
Institute Of Road And Transport Technology,
Erode,India

Dr.K.Duraiswamy
Dean / Academic,
K.S.Rangasamy College Of Technology,
Tiruchengode. India

## ABSTRACT
This paper proposed the architecture of question answering system for a specific domain using semantic web andontology. It defines the question answering system using semantic searching methodology and natural language processing techniques. The traversal in ontology is based upon conceptual graph matching type. The repositories are the collection of documents which hold the information related to specific domain. The main aim of this model is to give the question in natural language and the appropriate mechanism is used to retrieve the correct answer for the given question.

## Keywords
Natural Language Processing, Ontology, semantic Web, Question Answering system.

## 1. INTRODUCTION
In current situation, the bulk of knowledge is available in Internet as in the form of documents, articles, discussions, books, etc. But only the problem is when we need some relevant information from those resources, there is no mechanism currently available to find out what is relevant to our needs. In this situation, the need of question Answering system is required to find out the correct and related information which can be automatically retrieved from internet using some specific mechanism.
Already Search Engine is available to satisfy the user's need but on that also some disadvantages are there. The Search Engine not only gives related information but also the irrelevant too. The user only chose the best result from it and also the search is based upon keyword based type and it not checks the meaning of user's query. In this situation the semantic search comes into a role.
The input to the proposed question answering system is given by the user, which is converted in the form of Natural Language Processing. With the help of user feedback, the query is expanded and refined for getting relevant answer. It can be achieved in open and close domain environment.

### 1.1 Features of QA system
There are two types of Question answering System available now a day. Those are all Open Domain QA System and Closed Domain QA System. Open Domain environment is more complex compared to closed domain. The input to the question may be in any form and the answer may vary with various domains. Closed domain environment is easy because of domain specific concept and the natural language processing (NLP) can be implemented easily with the help of ontology. Closed domain environment deals with questions that depend on particular ontology.

## 2. NLP, ONTOLOGY AND SEMANTIC WEB
### 2.1 Natural Language Processing
It is one of the idea of Artificial Intelligence concept. The most related fields in Natural Language Processing are Machine Translation, Information Retrieval and spell checking. The NLP[1] allows the semantic form of text for constructing an ontology which is well suitable for semantic web concept. The named entity model like word stemming and part-of-speech tagging mechanism is implemented in our system.

### 2.2 Ontology
The concept of ontology derives from the philosophy [2].It has a close relationship with information technology, knowledge engineering and artificial intelligence. "ontology isa shared explicit specification of a conceptualization" [3]. In this definition, "shared" means that the information described by ontology is commonly accepted by users; "explicit" requires the precision of both concepts and their relationships clearly defined; "conceptualization" is referred to an abstract model of a phenomenon [4].
According to the extent of dependence on field, ontology can be subdivided into four categories, namely toplevel, domain, task and application ontology [5].
Ontology defines in the basic terms and relations comprising the vocabulary of a specific area, as well as rules for combining these terms and relations to define extensions vocabularies. Ontology's are used to represent the knowledge in the form of class/concept, relations, functions, entities and axioms. These ontology can be represented as OWL[6] , RDF languages[7] using Protégé Tool.[8].

### 2.3 Semantic Web
According to Tim Berners-Lee, semantic web in which information is given with well-defined meaning, better enabling computers and people to work in cooperation.[9].The following technologies are supported in Semantic Web such as XML (eXtensible Markup Language), RDF (Resource Description Format). The semantic web is the vision for future web in which the web is more intelligent like a human.

## 3. DESIGN OF INTELLIGENT QUESTION ANSWERING SYSTEM
In Fig 1, the general architecture of Personalized question Answering System with ontology and Knowledge Base is represented. It has the following modules.

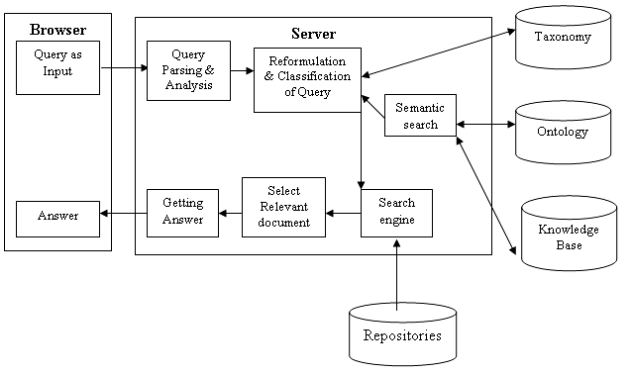


**Fig 1 Proposed Architecture of Personalized Intelligent Question answering system**

**a. Question as Input**
User enters the question from the browser. The question type, main concept of the question and searching elements are identified in this phase. The Semantic meaning of the question is given as input to the next stage.

**b. Query Parsing & Analysis**
In this phase, the analytical operation of the question is found out. This Analysis is responsible for processing Natural Language Processing (NLP). It is a technique to identify the type of a question, type of an answer, subject, verb, noun, phrases and adjectives from the question. Tokens are separated from the question and the meaning is analyzed and the reformulation of question/query is sent to the next stage.

The input is concerted into Natural Language and that is implemented using word segmentation algorithm. In word

segmentation algorithm the input query from the user is divided as keywords which is further subdivided and searched in knowledge base for getting correct answers.

The following general word segmentation algorithm is implemented in .NET environement.

```
static public string SegmentFile(string pathRed)
{
Regex r = new Regex("[\u4e00-\u9fa5]|[0-9]|[A-Za-z]|[,]|
[ , ]");
Regex r2 = new Regex("[,]|[ ]");
string soure = pathRed;
string sb = null;
string DictPath
=Path.Combine(Environment.CurrentDirectory,
"Data")+ Path.DirectorySeparatorChar;
Hashtable stopList = new Hashtable();
stopList = InitStopList(DictPath);
List<WordResult[]> result;
WordSegmentSample wss = new
WordSegmentSample(DictPath, 1);
MatchCollection mc = r.Matches(soure);
string linetach = "";
for (int i = 0; i < mc.Count; i++) linetach = linetach + mc[i];
char[] sp ={ ',', ' ', '/' };
string[] lineArr = linetach.Split(sp);
foreach (string newLine in lineArr)
{r
esult = wss.Segment(newLine.ToString());
for (int j = 0; j < result.Count; j++)
for (int k = 1; k < result[j].Length - 1; k++)
{i
f(stopList.ContainsKey(result[j][k].sWord.ToLower()))
continue;
sb = sb + result[j][k].sWord.ToLower()+ "\r\n";
}}r
eturn sb; }
```

**c. Reformulation & Classification of Query**
According to the user's choice, the reformulation of query is generated with the help of taxonomy of given topic.

**d. Semantic Search**
At final stage, the given question is taken as a word format and the relevant concept is searched in ontology and knowledge base. There are three algorithms are available for semantic search. The Search is carried out using Conceptual Graph Matching algorithm[10] which is the best technique compared to the above three algorithms. Graph patterns are important concept in semantic search. RDF model is organized and graph patterns are used to formulate and encode constraint queries for locating sub graph in RDF network.

**e. Knowledge Base**
The Knowledge Base of this proposed system is domain specific. The storage of ontology is the necessary one to retrieve the relevant and correct answer from the knowledge base[11] . In our system mysql database is used which can be easily linked in protégé. The linking step from protégé to database is given in Fig 2


**Fig 2 Protégé connected with database**

**f. Repositories**
These repositories contain all the documents related to this Data structure field. The proposed document may be structured or in unstructured format which can be retrieved by the search engine.

**g. Search Engine**
The user can search answers in natural language. If the concept exists in the knowledge base, the system can answer the question quickly, otherwise the user needs to apply web search. User call meta search engine through web search interface.

**h. Select Relevant document**
Using some transformation rules, the possibility of answer will be identified from various documents, from that the punctuation marks are removed. Those documents are recovered and ordered in a specific manner.

**i. Getting Answer**

This is a simple pattern matching technique to choose the appropriate response in terms of accuracy and simplicity.

**j. Answer**

Finally, after ranking, the answer will be displayed in the text field of a Browser. The user can accept the answer or if he needs more information regarding it, the query will be given to server once again. According to user satisfaction, the correct answer can be selected.

## 4. IMPLEMENTATION

This system is implemented in .NET environment for Front end design with MySQL connectivity. The ontology with 1000 nodes is constructed using Protégé tool. The relations, associativity for each node is properly set and tested. The ontology is in RDF/OWL model so the ontology representation can be used as Jena Code with Eclipse environment. Because,

Jena holds the built-in namespaces and it supports SPARQL query execution methods in it.

## 5. EXPERIMENTS RESULTS

We have implemented our test in the knowledge-base which has 1000 nodes in ontology. The parameters are taken is a=0.5, b=0.3, and c=0.6. In the process of conceptual graph matching, the question which has large similarity( threshold level is >= 0.67) is chosen. After several iterations, the result is given in the table format. If the maximal similarity is less than 0.65, it is taken as that no solution in knowledge base. The accuracy is taken as more than 80%. Two users with 75 trained questions are tested respectively. The result is shown in table format as Table 1.

| User | No.of Questions given to the system | Average interaction time | Accuracy (%) |
|---|---|---|---|
| 1 | 75 | 1.86 | 81 |
| 2 | 75 | 2.12 | 87 |

**Table 1 : Experimental training Result**

## 6. ACKNOWLEDGEMENT

The use of Ontology and semantic web in the question answering system is the new field for improving the search timing and also getting the relevant result. For this, the NLP technique is the base one. There is a need of certain standard for retrieving documents from repositories. The knowledge base which helps to get an answer using some searching algorithms that can classify the question and allow to locate the question in the concept. In a preliminary work, we will create domain specific ontology and the result will be getting from the documents also. After successfully implementing this project, it'll be extended to search from real search engine using web.

## 7. REFERENCES

[1] "Applying natural language processing techniques to domain knowledge capture" Alain Auger, Defence Research and Development Canada-Valcartier.

[2] Z.Y.Feng, W.J.Li, and X.H.Li, Ontology engineering and application, Tsinghua University press, China, 2007

[3] T. R. Gruber, "A Translation Approach to Portable ontology Specifications", Knowledge Acquisition, Academic Press Ltd, London UK, 1993, pp.199-220.

[4] R.R.Studer, R.Benjamins, and D.Fensel, "Knowledgeengineering: princles and methods", Data and knowledge engineering, 1998, pp.161-197.

[5] N.Guarino, "Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration", Lecture Notes in Computer Science, Springer-Verlag, London UK, 1997, pp.139-170.

[6]. "Web Ontology Language: OWL " Grigoris Antoniou ,Frank van Harmelen.

[7] "*A Comparison of RDF Query Languages*" Peter Haase1, Jeen Broekstra2, Andreas Eberhart1, Raphael Volz1

[8] "*The Prot´eg´e OWL Plugin: An Open Development Environment for Semantic Web Applications*" Holger Knublauch, Ray W. Fergerson, Natalya F. Noy and Mark A. Musen.

[9] *"The Semantic Web"* Berners-Lee, Tim, Fischetti, Mark and Lassila, Ora, San Francisco: s.n 2001.

[10] J.Zhong, H.Zhu, J.Li, Y.Yu "*Conceptual Graph Matching for Semantic Search* "

[11] W.Bao, G.Y.Li, "Ontology Storage Technology Research", Computer Technology and Development, China,2008, pp.146 150.