

Comparison of Question Answering Systems Based on Ontology and Semantic Web in Different Environment

¹S. Kalaivani and ²K. Duraiswamy

¹Department of Computer Science and Engineering,
Institute of Road and Transport Technology, Erode-638 316, Tamilnadu, India
²K.S. Rangasamy College of Technology, Tiruchengode-637 209, Tamilnadu, India

Abstract: Problem statement: Question Answering (QA) system is taking an important role in current search engine optimization concept. Natural language processing technique is mostly implemented in QA system for asking user's question and several steps are also followed for conversion of questions to query form for getting an exact answer. **Approach:** This paper surveys different types of question answering system based on ontology and semantic web model with different query format. For comparison, the types of input, query processing method, input and output format of each system and the performance metrics with its limitations are analyzed and discussed. Our question answering for automatic learning system architecture is used to overcome the difficulties raised from the different QA models. **Results:** The semantic search methodology is implemented by using RDF graph in the application of data structure domain and the performance is also analyzed. Answers are retrieved from ontology using Semantic Search approach and question-to-query algorithm is evaluated in our system for analyzing performance evaluation. **Conclusion:** Performance of question answering system of getting exact result can be improved by using semantic search methodology for retrieving answers from ontology model. Our system successfully implements this technique and the system is also used in intelligent manner for automatic learning method.

Key words: Semantic web, ontology, RDF, semantic search, question answering system

INTRODUCTION

The Question Answering system plays a major role in current era. It is needed when the user gets an in-depth knowledge in a particular domain. QA system is classified as two types namely closed domain or restricted domain and open domain model. In QA systems two types of search is available namely keywords based search and semantic search (Zhang, 2006). Normal search engines are working under keyword based searching concept. But some time, there is a problem of getting wrong answer for different meaning of same word. So, semantic search is used to solve the above problem.

Semantic search is used to improve the accuracy of search by understanding the intent of the user and the meaning of the terms in the searching sentence. Mainly there are two search are available as namely Navigation Search and Research Search.

In navigational search, the user is using the search engine as a navigation tool to navigate to a particular intended document. Semantic Search is not applicable to navigational searches. In Research Search, the user

provides the search engine with a phrase which is intended to denote an object about which the user is trying to gather/research information. Rather than Google's PageRank algorithm, Semantic Search uses semantics to produce highly relevant searching results.

This Semantic Search technique can be used to retrieve the knowledge from the data source like ontology. Ontology (Fernandez *et al.*, 2009) is a technology used to enable the domain knowledge at a high level and improve the query time used in Question Answering system.

Related work: PANTO (Wang *et al.*, 2007) model a Portable nAtural laNguage inTerface to Ontologies which accepts input as natural language form and the output is in SPARQL query. It is based on triple-based model in which parse tree is constructed for the data model using the off-the-shelf Stanford parser. Logic rules are applied for natural language queries as negation, comparative and superlative form. For mapping WordNet and String metric algorithms are used. The parse tree forms the intermediate representation as Query Triples Form. Then PANTO

Corresponding Author: Kalaivani, S., Department of Computer Science and Engineering Institute of Road and Transport Technology, Erode-638 316, Tamilnadu, India

converts Query Triples form into OntoTriples form which are represents as entities in ontology. OntoTriples are finally interpreted as SPARQL form. The performance of PANTO is analyzed by using F-Measure type. At the maximum 88.05% Precision is achieved for Geography domain with tested queries. So this system helps to bridge the gap between the real world users with the semantic web based on logic model.

FREyA (Damjanovic *et al.*, 2010) a Feedback Refinement and Extended Vocabulary Aggregation system combines syntactic parsing with knowledge in ontology for reducing customization effort. The rules are not used in this system instead of that knowledge encoded in ontology is given for understanding the user's question. Then the syntactic parsing is used to get a precise answer. In this model the ontology concepts are identified and verified initially. Then the SPARQL query is generated and the answer type is identified. Syntactic parse tree is generated using Stanford Parser. Mapping of user query with ontology concept is implemented in two ways as automatically and by the help of user. Ranking model is used by string similarity type. Answer type of this system is in graph form. JIT library is used for graph visualization. The Precision and Recall value for the tested data is reached high as 92.4% this is achieved only by the system returns the answers always as correct/partial/incorrect form. For measuring the performance of the system the Mean Reciprocal Rank (MRR) algorithm is implemented. It is a statistic for evaluation the process to a query. MRR value is achieved to 0.81. It supports high precision and recall.

Querix (Kaufmann *et al.*, 2006) is another ontology-based question answering system which relies on clarification dialogs in case of ambiguities. This system contains user interface, ontology manager, query analyzer, matching center, query generator, dialog component and ontology access layer. NL queries are converted into SPARQL query form and using Wordnet the synonym is identified. Standford parser is also used in this system which provides a syntax tree for NL query. Querix doesn't exploit the logic based semantic techniques.

In case of ORAKEL (Cimiano *et al.*, 2007) is used for computing intentional answers of user query. It computes wh-based questions as logical query form and knowledge is represented with F-Logic and Onto broker form. This system is used to convert question into query form and the given query is fed to bottom-up generalization model for getting intentional answer to the user. Inference engine is used to evaluate queries to knowledge base form. Customization is performed through the user interaction, using software called Frame Mapper, where the linguistic argument

structures, such as verbs or nouns with their arguments, are mapped to the relations in the ontology.

QACID (Ferrandez *et al.*, 2009) is based on collection of queries from a given domain which are analyzed and grouped as clusters and those are manually annotated using SPARQL queries. Each query is considered as bag of words, mapping between words in NL queries into KB by using string distance metrics. SPARQL generator replaces the ontology with instances mapped for original NL query. It is domain specific and the performance depends on the types of questions collected in domain.

AquaLog (Lopez *et al.*, 2007) is capable of learning the user's jargon in order to improve his experience by the time. Their learning mechanism is good in a way that it uses ontology reasoning to learn more generic patterns, which could then be reused for the questions with similar context. In this system two major models are used as Linguistic Component which is used to convert the NL questions into Query-triple format and Relation Similarity Service (RSS) which takes Query Triple form into Onto-Triple form. The data model is triple based like {Subject, Predicate, Object} type. The Performance is based on Precision, Recall and also failure types are referred separately. At average 63.5 % of successive answers are retrieved from ontology with closed domain environment.

SMART (Battista *et al.*, 2007) Semantic web information Management with Automated Reasoning Tool is an open-source system with integrated query form. Semantic bases queries are evaluated using DL queries which are mapped with SPARQL query form. The feature of this system is a semantic query with validation using DL reasoners and a graphical representation of query and mapping of DL queries to SPARQL. The retrieval of pre-computed inferences are from RDF triple format. Ontology supports URI identifiers. It uses file based system to store ontologies. Users can write syntactic, semantic and logical queries in valid form.

SWSE (Hogan *et al.*, 2011) Semantic Web Search engine supports SPARQL with RDF representation. Index structure comprises of complete index on quadruples with keyword search function based on inverted index form. Query processing components are distributed to number of machines.

Here, the Table 1 gives the comparison of different QA systems with its question type, input and processing format, the performance metrics used in each system and the limitations of each system. This is the survey conclusion described in table form.

Table 1: Comparison of Various QA systems with its Limitations

Type of QA System	Query Entry	Success %	Limitations
PANTO	NL Question	88.05	1.Work with small ontology 2.Operations stored in memory 3.No database indexing technique
FREyA	NL Question	92.4	1.Quality of annotation depends on ontology based gazetteer. 2.Suggestion is selected by user each time. 3.Lexicons are generated based on demand when the KB is loaded.
QUERIX	NL Question	86.08	1.Doesn't need for adaptation of new ontology. 2.Ask user for clarification, if ambiguity occurs.
ORAKEL	F-Logic and SPARQL	93	1.Not suitable for multi ontology open scenario.
QACID	NL query	80	1.Costly because of domain dependent 2.Can only be applied with limited coverage.
AQUALOG	NL Question	63.5	1.String based comparison method 2.Grammars are domain independent 3.Lack of appropriate reasoning services defined by ontology.
SMART	SPARQL	80.3	1.User interaction is needed. 2.Web based query form.
SWSE	Keywords	84.7	1.Keyword bases search is progressed. 2.Web based query form with large collection of knowledge base. 3.Data heterogeneity is the problem.

MATERIALS AND METHODS

Proposed architecture of QAAL system: In Fig. 1, the general architecture of QAAL System with ontology and Knowledge Base is represented. It has the following modules.

Question as Input: User enters the question from the browser. The factoid type of the question is identified and the expected answer type is also identified in this phase. The Semantic meaning of the question is given as input to the next stage.

Query Parsing and 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.

Reformulation and Classification of Query: According to the user's choice, the reformulation of query is generated with the help of WordNet which is implemented as semantic matching model.

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 which is the best technique compared to the above three algorithms. All the sentences in repository are framed as conceptual graph and the given question is also framed as conceptual graph. The matching of question CG with given CG are checked out using CG matching algorithms and the result us displayed at front-end of the QA system.

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.

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. 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.

Repositories: These repositories contain all the documents related to this Domain ontology field. Proposed ontology is likely related to particular domain (Robin and Uma, 2011) such as Data Structure ontology. The proposed document may be structured or in unstructured format which can be retrieved by the search engine.

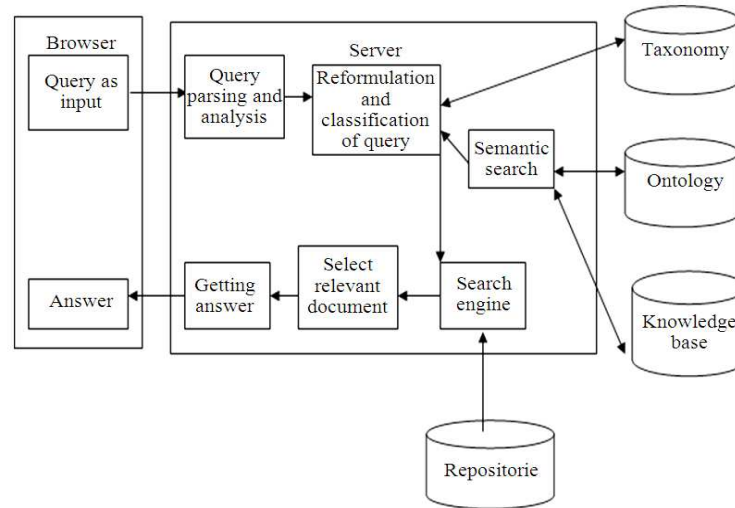


Fig. 1: QAAL System with Ontology and WordNet

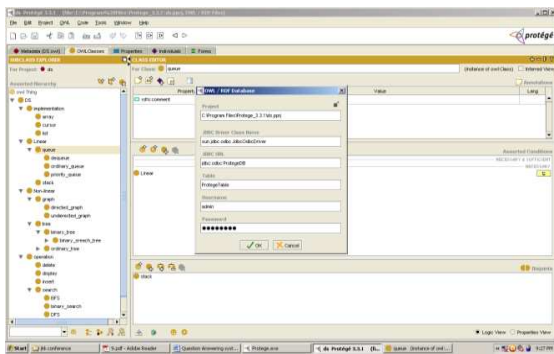


Fig. 2: Protégé connected with database

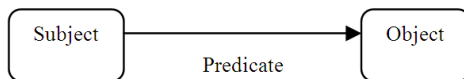


Fig. 3: Simple RDF form of a sentence.

Search engine: The user can search answers from ontology. If the concept exists in the knowledge base, the system can answer the question quickly, otherwise the user needs to apply web search as don't know method. User can call meta search engine through web search interface.

Select relevant document: Using first order logic of projection algorithm based Conceptual Graph (CG) matching with some transformation rules (Sowa, 2008) 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.

Getting answer: This is a simple pattern matching technique to choose the appropriate response in terms of accuracy and simplicity. The proposed artificial Intelligence with Fuzzy logic concept is going to be implemented to get relevant answer for the given query.

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.

Graph matching in ontology: Conceptual Graph acts as an intermediate language for mapping natural language questions and assertions to a relational database. Conceptual Graph (CG) contains concept, concept relation and argument. It is a graph which represents logic based on semantic model of artificial intelligence and existential graphs.

Resource Description Framework (RDF) is a framework to annotate information resources in a machine-understandable way. It is used for making statements rather than language itself. RDF contains triple syntax to express annotations as subject, predicate and object.

QAAL system is to be implemented as semantic web concept which can be represented by RDF. Information resources are commonly represented as uniform Resource Identifiers (URIs). URIs are described by RDF. RDF triples are visualized as directed labeled graph in which subject; objects are represented as nodes and predicates as arcs in Fig. 3.

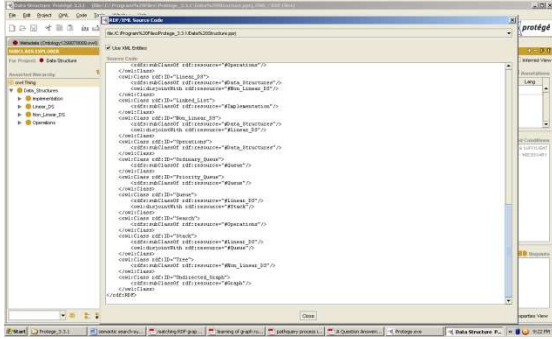


Fig. 4: RDF format for Data Structure Ontology

Graph consists of concept which can be understandable by User in an easy manner. Graph theories can be easily plotted in RDF and RDF is also used for implementing semantic web applications. For these reasons only, we propose RDF representation for our QAAL system.

We develop Data Structure ontology represented in Fig. 4 the RDF form like below.

There are three main algorithms are available for implementing semantic search in QA system. Those are finding the most specific answer, concept matching and identifying related nodes in RDF/OWL documents. In our QAAL system, Graph matching in ontology is used for implementing semantic search model. In a Graph, two steps are followed for finding the matching. We propose the Graph Matching Similarity (GMS) Algorithm for implementing semantic search technique. Similarity is to be implemented for concept, relation and function.

Spread activation: Spread activation (Suchal, 2007) is a method for searching the nodes in ontology as in semantic manner. It exploits relations between nodes in ontology. Nodes may be terms, class, object etc. Relations are labeled directed or weighted manner. SA algorithm creates initial nodes that are related to the content of the user’s query and assign weights to them. After that, nodes will activate with different nodes on ontology by some rules.

Semantic Search Algorithm is based on Conceptual Graph form of user query and domain ontology.

```

Get user query and generate query tree
For (the entire resource tree in the domain ontology)
{
    Find the match for root node of the query tree in the resource tree
    and create resource sub trees.

    For (all the resource sub trees)
    {
        Find the similarity between the query tree and
        resource sub tree
    }
}
    
```

```

}
Find the best match from the above combinations of resource sub
trees with the query tree and select it as the similarity value between
the query tree and the resource tree.
}
    
```

Semantic similarity (Seco *et al.*, 2004) calculation with WordNet is described below.

Let T1 and T2 be two nodes. Let L1 and L2 be the lists of their RDF sentences, with respective length n1 and n2. Let s_{ij} be the ith element of the list L_j. The semantic similarity between T1 and T2 is denoted by Eq. 1:

$$\left[\text{Sim}(T1, T2) = \frac{1}{n1 + n2} \left(\sum_{i=1}^{n1} \text{Sim}(s_{i1}, L2) + \sum_{j=1}^{n2} \text{Sim}(s_{j2}, L1) \right) \right] \quad (1)$$

where, the similarity between a sentence and a list of sentences is the maximum similarity between this sentence and all the sentences of the list.

Spread Activation Algorithm for best mapping of User query and domain ontology:

- The algorithm considers each of these nodes as having the same weight (set to 1.0)
- All nodes which are not in the initial set have their initial activations set to zero
- Place initial nodes in priority queue
- Process the nodes based on the activation values
- Start with the node having highest activation value
- Propagate the activation from the current node to its neighbors
- Add the activated neighbors to the priority queue
- Place the processed nodes are placed in results list
- Attenuate the activation with every propagation
- Repeat steps 2 and 3 based on the distance constraint or until there are no more nodes in the priority queue

Question classification methods: There are basically three types of question classification methods are available. Those are machine learning approaches, knowledge based approach and template based approach.

In our QAAL system we use template based approach for fast retrieval of answer. If the question is already asked in that system, the retrieval get from question template table form, otherwise matching is performed using matching algorithm.

Here are the sample Question template model which hold the question model and also its corresponding knowledge representation. QAAL is used for implementing factoid based question types. Basic terms in factoid model is who, whom, why, what,

where, when, what, which, whose type wh questions. Here sample questions in our ontology with SPARQL queries are represented:

Question 1: How to implement stack?

[implement]
(Subject) [types *] ?
(object) [stack]

SPARQL Query used in our system:

```
SELECT ?object
WHERE {?subject :is_implemented_by
?object .
?subject :has_value "stack"
}
```

Question 2: What are the types of Graph Searching Technique?

[graph-searching]
(Subject) [types *]?
(object) [Graph]

SPARQL Query used in our system:

```
SELECT ?s ?o
WHERE { ?s :has_type_search ?o }
```

In this way, question templates are generated in our particular domain. Depending upon the user's question first the matching is searched with the question template and the answer is retrieved from it only if the matching is success. Otherwise, semantic searching is processed by implementing query reformulation strategy.

RESULTS AND DISCUSSION

RDF Graph matching is achieved and practically implemented by using SPARQL language in ontology environment. Jena framework is successfully used to implement RDF model with SPARQL query. Eclipse environment is used for jena model in it.

In the following Fig. 5, the performance of our QAAL system is represented with different Mean Average Precision value with the frequency.

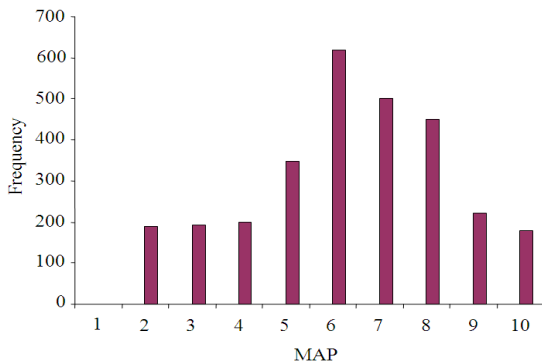


Fig. 5: Mean Average Precision (MAP) with its frequency of question template

CONCLUSION

We suggest different classification of searching techniques in Question Answering System and also we propose the Graph Matching Algorithm for query matching with the ontology using Spread Activation Algorithm.

Our QAAL system supports this criterion and finally we conclude that the system performance can be improved with the use of semantic search model instead of using normal keyword search model.

REFERENCES

Battista, A.D.L., N. Villanueva-Rosales, M. Palenychka and M. Dumontier, 2007. SMART: A web-based, ontology-driven, semantic web query answering application.

Cimiano, P., P. Haase, J. Heizmann, M. Mantel and R. Studer, 2007. Towards portable natural language interfaces to knowledge bases-the case of the ORAKEL system. *Data Know. Eng.*, 65: 325-354. DOI: 10.1016/j.datak.2007.10.007

Kaufmann, E., A. Bernstein and R. Zumstein, 2006. Querix: A natural language interface to query ontologies based on clarification dialogs. *Proceedings of the 5th International Semantic Web Conference, (ISWC' 2006)*, Citeulike, pp: 980-981.

Damljanovic, D., M. Agatonovic and H. Cunningham, 2010. Natural language interfaces to ontologies: Combining syntactic analysis and ontology-based lookup through the user interaction. *Semantic Web: Res. Appl.*, 6088: 106-120. DOI: 10.1007/978-3-642-13486-9_8

Fernandez, O., R. Izquierdo, S. Ferrandez and J.L. Vicedo, 2009. Addressing ontology-based question answering with collections of user queries. *Inform. Proces. Manage.*, 45: 175-188. DOI: 10.1016/j.ipm.2008.09.001

Hogan, A., A. Harth, J. Umbrich, S. Kinsella and A. Polleres *et al.*, 2011. Searching and browsing linked data with SWSE: The semantic web search engine. *J. Web Semantics*, 9: 365-401. DOI: 10.1016/j.websem.2011.06.004

Sowa, J.F., 2008. *Conceptual Graphs and Book of Knowledge Representation*. Elsevier.

Lopez, V., V. Uren, E. Motta and M. Pasin, 2007. AquaLog: An ontology-driven question answering system for organizational semantic intranets. *J. Web Semantics Sci. Service Agents World Wide Web*, 5: 72-105. DOI: 10.1016/j.websem.2007.03.003

- Robin, C.R.R. and G.V. Uma, 2011. Design and development of ontology suite for software risk planning software risk tracking and software risk control. *J. Comput. Sci.*, pp: 320-327. DOI: 10.3844/jcssp.2011.320.327
- Seco, N., T. Veale and J. Hayes, 2004. An intrinsic information content metric for semantic similarity in wordnet. Department of Computer Science, University College Dublin.
- Suchal, J., 2007. Caching spreading activation search. Slovak University of Technology.
- Wang, C., M. Xiong, Q. Zhou and Y. Yu, 2007. PANTO: A portable natural language interface to ontologies. Proceedings of the 4th European Semantic Web Conference, (ESWC' 07), Publication post of DBLP, pp: 473-487.
- Zhang, A., 2006. Research and Implementation of ontology-based intelligent question answer system. Comput. Appli. Software, China.