ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Knowledge Baesd Question Answering System Using Ontology

R.Mervin [*1], Dr.A.Jaya[2]
[*] Computer Science and Engineering,B.S.Abdur Rahman University, India
mervinvijay@gmail.com

Abstract

In the modern era numerous information available in the World Wide Web. Question Answering systems aims to retrieve point-to-point answers rather than flooding with documents. It is needed when the user gets an in depth knowledge in a particular domain. When user needs some information, it must give the relevant answer. The basic idea of QA systems in Natural Language Processing (NLP) is to provide correct answers to the questions for the user. Here the question answering system has to be implemented as semantic web. This research will evaluate the answering system by the expert. The experts are personalized based on their domain and subject. Ontology is used to personalize the experts. Based on these, the answer has to be ranked and given back to the user. It consists of three phases such as User's zone, Interface and Expert's zone. Natural language processing techniques are used for processing the question and also for answer extraction. The domain knowledge is used for reformulating queries and identifying the relations.The highlight of this paper is providing most relative answer for the question provided by the end users in an efficient way.

Keywords: Question Answering System, User's zone, Interface, Expert's zone, Natural language processing.

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 the user need some relevant information from those resources, there is no mechanism currently available to find out what is relevant to the user 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.

There are two types of Question answering System available now a day. Those are all Open Domain QA System and Closed Domain QA System. The input to the question may be in any form and the answer may vary with various domains. Open domain QA system consists of various domain. It gives answer to the user question in any domain. So it is more complex compared to closed domain. 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.

The concept of ontology derives from the philosophy [2].It has a close relationship with information technology, knowledge engineering and artificial intelligence. " ontology is a 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 top level, 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

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

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].

The remaining part of the paper is organized as follows, **Section 2.0** deals related work ,**Section 3.0** describes Architecture diagram and **Section 4.0** gives the Experimental Results for KBQA. Finally **Section 5.0** concludes the paper by giving a brief glimpses into the future directions of research in this area.

Related work

Automatic QA systems, such as AnswerBus (Zhang et al., 2005) and MULDER (Kwok et al., 2001), extend their data resource from the local database to the web resources, which also extend the scope of the questions they can handle. In 1999, TREC set the first QA track (Voorhees, 1999). AquaLog (Lopez et al., 2007) is an ontology-based question answering system that processes input queries and classifies them into 23 categories. If the input question is classified into one of these categories, the system will process it correctly. There are a few question answering systems based on conditional knowledge structures, which were introduced by Areanu and Colhon (2009). In these systems, a conditional schema is used to generate XML-based conditional knowledge structure, which is used for question answering. Ferrnandez et al. (2009)proposed an ontology-based question answering system called QACID to answer natural language queries related to the cinema domain. This system extracts answers from a preconstructed ontology by comparing question attributes with ontology attributes. QACID was evaluated using entailment queries composed for the cinema domain. The overall official F1-accuracy reported by QACID is 93.2% with an ABI threshold of 0.5.

Athira P.M et al (2013)"Architecture of an Ontology-Based Domain-Specific Natural Language Question Answering System"[11]discussed Question Answering, the process of extracting answers to natural language questions, is profoundly different from Information Retrieval (IR) or Information Extraction (IE). IR systems present the user with a set of documents that relate to their information need, but do not exactly indicate the correct answer. In IR, the relevant documents are obtained by matching the keywords from user query with a set of index terms from the set of documents. In contrast, IE systems extract the information of interest provided the domain of extraction is well defined. In IE systems, the required information is built around in presumed templates, in the form of slot fillers.

Ontology learning is a knowledge acquisition activity that relies on automatic methods to transform unstructured data sources into conceptual structures. The first proposals for ontology learning (Maedche, 2002) built all resources from scratch, but the manner of the tackling ontology population has evolved due to the existence of complementary resources,such as top-level ontologies or semantic role repositories. Some ontology learning approaches, such as TERMINAE (Aussenac-Gilles et al., 2008), provide conceptualization guidance from natural language text integrating functions for linguistic analysis and conceptual modeling. A number of methods have already been proposed for automatically constructing an ontology from text. Graph-based approaches are very popular for representing concept relations (Hou et al., 2011). There are some approaches using mixed methodologies, such as using relational databases and semantic graphs (Ra et al., 2012). Some ontology development tools have been proposed to extract deep semantic relation between concepts using mapping functions and to generate rough schema. OntoCmaps (Zouaq et al., 2011) is an ontology development tool that extracts deep semantic relations from text in a domain-independent manner. Mining the situation context from text and constructing a situation ontology is an interesting area in information retrieval. Jung et al. (2010) have performed notable work in this area. There were a few studies that utilized lexico-syntactic patterns and lexico-semantic probabilities for automatically extracting concept relationships (Hearst, 1992, 1998) from raw text.

Vanessa Lopez et al "Question Answering on the Real Semantic Web"[13] to conclude with, Power Aqua balances the heterogeneous and large scale semantic data with giving results in real time across ontologies, to translate user terminology into distributed semantically sound terminology, so that the concepts which are shared by assertions taken from different ontology's have the same sense. The goal is to handle queries which require to be answered not only by consulting a single knowledge source but combining multiple sources, and even domains.

Vincent Barbier et al "Semantic Knowledge in Question Answering Systems" [12] explores for each question, FRASQUES question analysis module determines several kinds of information, among which three sets are more thoroughly studied in this article: i) the set of non-empty words of the question, ii) the set of their synonyms extracted from Fastr and

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

iii) the set of their synonyms extracted from EuroWordNet. Victoria Uren et al "AquaLog: An ontology-driven Question Answering system as an interface to the Semantic Web"[9], the authors approach finally, although AquaLog has primarily been designed for use with semantic web languages, it makes use of a generic plug-in mechanism, which means it can be easily interfaced to different ontology servers and knowledge representation platforms.

Architecture Diagram for KBQA

The Fig.1 shows the architecture diagram for KBQA.It is divided into three modules such as the User's zone, Interface and Expert's Zone. User's zone is responsible for collecting the questions from the user, sending the question to the question processor and also at the other end provides the answers to the viewer. Interface acts as a middle layer in KBQA system. It is used to communicate between the User's zone and the Expert's zone. It accepts the questions from the user and send it to the Expert's zone for question processing .Once the question is processed by the expert's zone, the answer will be send to the user via this interface.

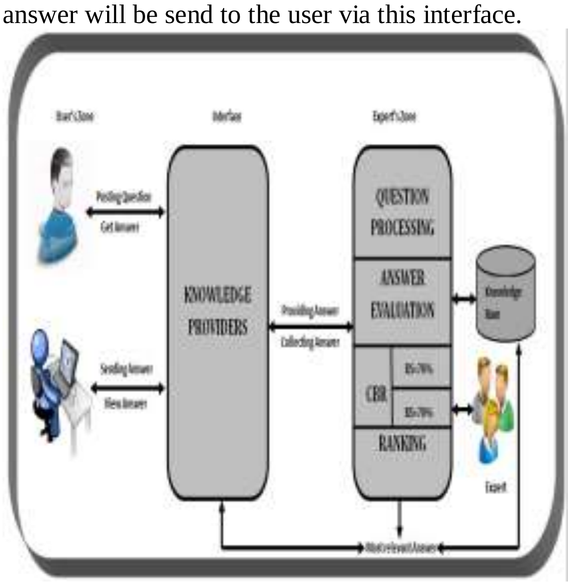

Fig.1 Architecture Diagram for KBQA

Expert's zone is divided into three parts such as Question Processing, Answer Evaluation and Ranking. The question will be posted by the user and send to the answer evaluator through the interface. The Answer evaluator will process the question and evaluate the answers posted by others in the community. Then Case Based Reasoning has to be applied for each and every answer. It will give calculate the relevance sore. If the relevance score is less than 70% it will send to Expert. Expert will process the answer with their knowledge. If the answer greater than 70% then it will go to ranking. The relevance score for each answer will be calculated. Then ranking will be done to all the answers for a particular question. Finally the most relevant answer is selected automatically and given back to the user. Knowledge Base is the repository for the KBQA system.

Domain ontology construction

The Fig.2 shows the Domain Otology construction for Data Mining.The proposed method to construct domain ontology concepts extracts the domain attributes and associations from a set of relevant documents. The Data mining domain is taken as an example. The Stanford dependency parser (Marie-Catherine de Marneffe, 2008) is used for generating a parse tree for each individual sentence in relevant documents. Then, the ontology concept schema is generated for the relevant relations. Decision rules and decision trees are used to discover the data in databases . The training sample set D is a collection of text documents that consists of the dependency parsing patterns for the corresponding sentences.

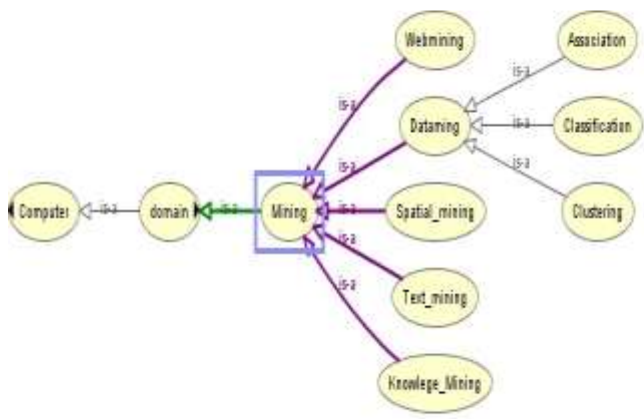

Fig.2 Domain Ontology construction for Data Mining

Question Processing

The user will post the question to the expert's one via the interface. Here the question has to be processed. The same information request can be expressed in various ways, some interrogative ("What is Data Mining?") and some assertive ("Tell me about Data Mining."). A semantic model of question understanding and processing has to be build. This model would enable the translation of a complex question into a series of simpler questions, would

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

identify ambiguities and treat them in context or by interactive clarification.

3.3 Answer Evaluation

The Fig.3 Shows the Answer Classification. Every answer has to be classified either as Relevant or Irrelevant based on the Evaluation Score. Once the answer is collected ,it has to be classified either relevant or irrelevant. The Fig.4 is used to find out the Evaluation Score. First all the answers have to be collected. Then it will be processed and send to the key mapping to find out the evaluation score.It will be done by calculating Evaluation score for every answer.

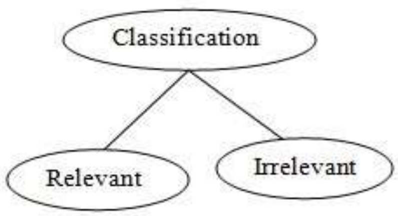

Fig.3.Answer Classification

Answer evaluation that uses the Web relevance score and the translation probability:

EvalScore(Q,A)=P(Q,A)1−γ·Web_relevance(Q,A)γ (1)

P(Q,A)=P(Q|A)/P(A) (2)

where P(Q,A) represents the probability of occurrence , *Web_relevance*(*Q*,*A*) denotes the score using Web relevance score and γ represents the weighting parameter. The equation (1) is equivalent to the translation probability when γ=0 whereas it is the same as the Web relevance score when γ=1.

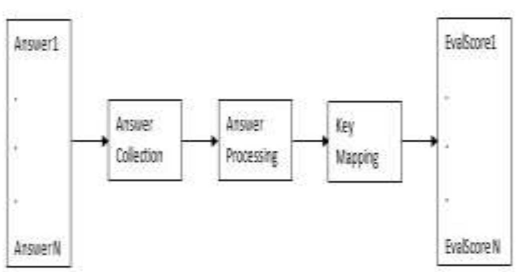

Fig.4.Answer Evaluation

Ranking

Once the relevance score has been completed the ranking has to be done with the answer. This will give the most relevant answer to the user. Here the Normalized Discounted Cumulative Gain(NDCG) is used for ranking. Every answers will get a weight between 0.0 to 1.0 based on the EvalScore which is computed in (1).The gain is accumulated from the top of the result list to the bottom with the gain of each result discounted at lower ranks. Search result lists vary in length depending on the query. This is done by sorting answers of a result list by relevance, producing the maximum possible DCG till position p, also called Ideal DCG (IDCG) till that position. Finally it will give the best answer by calculating nDCG. The *normalized discounted cumulative gain*, or nDCG, is computed as:

$$\mathrm{nDCG_p} = \frac{DCG_p}{IDCG_p} \tag{3}$$

Experimental results

The proposed relation extraction algorithm was implemented using Java. It is implemented with knowledge base which has 1000 nodes in ontology. If the maximal similarity is less than 0.70, 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. In total, 500 new input queries were generated. These new queries were used to adjust the entailment decision threshold and to evaluate the final system performance. The accuracy was calculated using the ratio between the number of questions correctly answered by the system and the total number of questions submitted to the system. Fig.5 shows the sample output screen for Search Result.

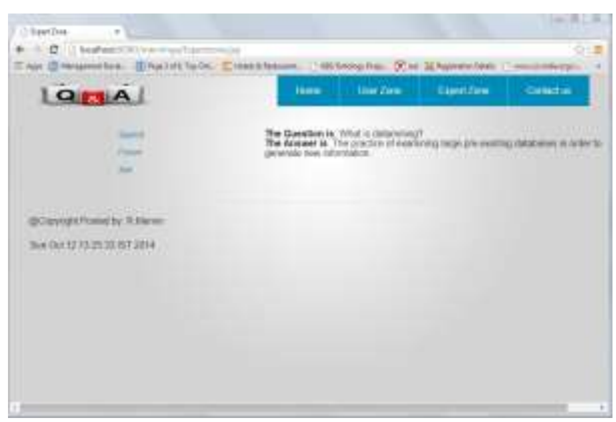

Fig.5 Screen shot of Search Result

Conclusion and future enhancements

A system for automatically extracting attributes and associations from a large volume of unstructured text for automatic domain ontology modelling was successfully developed. The empirical results were encouraging, and it has been proven that our proposed method outperforms similar well-performing knowledge extraction methods. The suitability of the constructed concept relational ontology for use with ontology portable question answering systems was experimentally evaluated using our KBQA ontology based question answering framework. 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.

The future enhancements that can be done in system are as follows other ranking algorithm can be applied to provide ranking in exact manner.

References

1. *Agichtein, E., Gravano, L., 2000. Snowball: extracting relations from large plain-text collections. In: Fifth ACM Conference on Digital libraries, 2–7 June 2000, San Antonio, TX, USA, pp. 85–94.*
2. *Areanu, N.T., Colhon, M., 2009. Conditional graphs generated by conditional schemas, Annals of the University of Craiova. Math.Comput. Sci. Ser. 36, 1–11.*
3. *Aussenac-Gilles, N., Despres, S., Szulman, S., 2008. The TERMINAE method and platform for ontology engineering from texts. In:Proceeding of the 2008 Conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge, pp.199–223.*
4. *Banko, Michele, Michael, J., Cafarella, Stephen Soderland, Matt Broadhead, Oren Etzioni, 2007. Open information extraction from the web. In: Proceedings of the Twentieth International Joint Conference on Artificial Intelligence, pp. 2670–2676.*
5. *Berant, Jonathan, Ido Dagan, Jacob Goldberger, 2011. Global learning of typed entailment rules. In: Forty Ninth Annual Meeting of the Association of Computational Linguistics: Human Language Technologies, 19–24 June 2011, Portland, Oregon, USA, pp. 610–619.*
6. *Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.H., 1984.Classification and Regression Trees. Chapman & Hall, Wadsworth,Belmont, CA.*
7. *Brin, S., 1998. Extracting patterns and relations from the world wideweb. In: WebDB Workshop at EDBT, 27–28 March 1998,Valencia, Spain, UK, pp. 172–183.*
8. *Carlson, Andrew, et al., 2010. Toward an architecture for never ending language learning. In: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, pp. 1306–1313.*
9. *QinglinGuo and Ming Zhang "Question Answering System Based on Ontology and Semantic Web" G. Wang et al. (Eds.): RSKT 2008, LNAI 5009, pp. 652–659, 2008.*
10. *Mohammad Reza Kangavari, Samira Ghandchi, ManakGolpour "A New Model for Question Answering Systems" World Academy of Science, Engineering and Technology 18 2008.*
11. *PengfeiZhang,Chunhua Wu ; Cong Wang ; Xiaohong Huang "Personalized Question Answering System Based on Ontology and Semantic Web"Industrial Informatics, 2006 IEEE International Conference on16-18 Aug. 2006.*
12. *Athira P. M., Sreeja M. and P. C. Reghuraj "Architecture of an Ontology-Based DomainSpecific Natural Language Question Answering System" International Journal of Web & Semantic Technology (IJWesT) Vol.4, No.4, October 2013.*
13. *Vincent Barbier, Brigitte Grau, Anne-Laure Ligozat, Isabelle Robba and Anne Vilnat "Semantic Knowledge in Question Answering Systems".*
14. *Vanessa Lopez, Enrico Motta, Victoria Uren, Michele Pasin" An ontology-driven Question Answering system to interface the Semantic Web".*

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Author Biblography

	R.Mervin R.Mervin pursuing her Ph.D. in Computer Science and Engineerin from B.S.Abdur Rahman University , India and she is working as an Associate Professor, with the department of Computer Science and Engineering Engineering in Saveetha Engineering College. Email: mervinvijay@gmail.com
	**Dr.A.Jaya** Dr. A. Jaya, Professor in Department of Computer Applications at B.S.Abdur Rahman University.She works on the techniques of Case based Reasoning and Natural Language Processing for Automatic construction of stories. Also, her research interest includes Ontology, Automatic Story generation, Web mining, Knowledge mining etc. She published 45 research articles out of which 15 in International journals and presented 30 papers in both International and National conferences including three research articles in IEEE conferences, one in ACM Digital Library, and two in Springer conferences. Her research articles were recognized with Best paper awards, in IEEE International Conference on Intelligent Agent and Multi-Agent Systems 2009 organized by Vinayaka Mission University. Email: jayavenkat2007@gmail.com