

Intelligent Question answering System

V.S.Babanne¹ Dr.S.T.Patil², D.J.Joshi³

¹Research Scholar, Vishwakarma Institute of Technology, Pune, India vanita.gadekar@gmail.com

²Professor, Vishwakarma Institute of Technology, Pune, India stpatil77@gmail.com

³Asst.Professor, Vishwakarma Institute of Technology, Pune, India deepali.joshi@vit.edu

Abstract- In this paper, we have proposed a rule based Automated Question- Answering system which aims at delivering concise information that contains answers to user questions. The context would be the domain specific systems. This technique is the solution to the problem of unlimited or irrelevant data which is bombarded on the user as a result of his query on any search engine. Given a question, our system would identify those portions of the knowledge base that are relevant to the question minimizing the search space and providing exact expected answers to the user.

Keywords: KB(knowledge Base),QA (Question answer), NLP(natural language processing)

I. INTRODUCTION

Expert Systems are computer programs that are derived from a branch of computer science research called Artificial Intelligence (AI). AI's scientific goal is to understand intelligence by building computer programs that exhibit intelligent behavior. It is concerned with the concepts and methods of symbolic inference, or reasoning, by a computer, and how the knowledge used to make those inferences will be represented inside the machine. Of course, the term intelligence covers many cognitive skills, including the ability to solve problems, learn, and understand language; AI addresses all of those. But most progress to date in AI has been made in the area of problem solving -- concepts and methods for building programs that reason about problems rather than calculate a solution.

Our Question Answering system consists of two principal parts: the knowledge base; and the reasoning, or inference, engine. The knowledge base of expert systems contains both factual and heuristic knowledge. Factual knowledge is that knowledge of the task domain that is widely shared, typically found in textbooks or journals, and commonly agreed upon by those knowledgeable in the particular field. In our paper, we will be using Java as our knowledge base.

Heuristic knowledge is the less rigorous, more experiential, more judgmental knowledge of performance. In contrast to factual knowledge, heuristic knowledge is rarely discussed, and is largely individualistic. It is the knowledge of good practice, good judgment, and plausible reasoning in the field. It is the knowledge that underlies the "art of good guessing."

Knowledge representation formalizes and organizes the knowledge. One widely used representation is the production rule, or simply rule. A rule consists of an IF part and a THEN part (also called a condition and an action). The

IF part list set of conditions in some logical combination. The piece of knowledge represented by the production rule is relevant to the line of reasoning being developed if the IF part of the rule is satisfied; consequently, the THEN part can be concluded, or its problem-solving action taken. Expert systems whose knowledge is represented in rule form are called rule-based systems. We will be using first order logic in for constructing our rules.

The question answering system is very hot in natural language processing .users can propose questions in natural language and get compact and relevant answers rather than relative documents as done in the search engines. The goal of a question answering system is to retrieve answers to questions rather than full documents or best-matching passages, as most information retrieval systems currently do. There are two categories in this system viz: open and closed.

The domain specific QA system involves heavy use of natural language processing systems formalized by building a domain specific ontology QA research attempts to deal with a wide range of question types including: fact, list, definition, paragraph and cross-lingual questions. Search collections vary from small local document collections, internal organization documents to be complied with newswire reports on the World Wide Web. A QAS returns answer of a user question in concise form. In order to provide the specific answer, the system must know what precisely a user wants. The prior knowledge of the estimated answer type helps the QAS to extract correct and precise answers from the document collection.

In this paper we are proposing the expert system for Question Answering (QA) System. The user should be able to access answer of their questions from knowledge base using inference engine. This system is domain specific. This typically operated in small domains enough to be covered by a few hundred rules - knowledge bases. That converts them into English using simple techniques before presenting them to the user.

We are using the first order logic for creating the rule, then storing in the knowledge base using the function, that implies the understanding of the inference of question by the system. From the literature that for certain classes of questions, "simply" retrieving and reciting prewritten paragraphs of text can achieve high performance levels (e.g. START [3], the Loebner "Turing test" winner modern case-based help-desk applications), even though such systems can be easily tricked, and rely heavily on the user to interpret and answer the question him/herself from the recited text.

II. Related Work

The task of identifying in large collections of documents a text snippet where the answer to a natural language question lies. The answer is found either in a short or a long text. A keywords extracted from the natural language question are either within the text document, word document, pdf which forming a text paragraph. Since such paragraphs must be identified throughout voluminous collections, automatic and autonomous Q&A systems incorporate an index of the collection as well as a paragraph retrieval mechanism.

The expected answer type is determined based on the question stem, e.g. who, where or how much and eventually one of the question concepts, when the stem is ambiguous (for example what), as described in (Harabagiu et al., 2000) (Radev et al., 2000) (Srihari and Li, 2000). However finding question keywords that retrieve all candidate answers cannot be achieved only by deriving some of the words used in the question. Although simple combinations of IR and IE techniques are not practical solutions for open-domain textual Q&A because IE systems are based on domain-specific knowledge, their contribution to current open-domain Q&A methods is significant. For example, state-of-the-art Named Entity (NE) recognizers developed for IE systems were readily available to be incorporated in Q&A systems and helped recognize names of people, organizations, locations or dates.

Sethi, R. J., and Gil, Y. used an approach [1] in which a user poses a question and others give their own self-contained answers to the question. The drawback of this system is users had to make uninformed trust judgments based on the quality of an answer, also the users had to understand a complex topic by examining alternative views on it, not just by being told a single, "right" answer. As they had different beliefs about what sources are biased and which ones are trustworthy to them in specific topics the system did not give proper results. Sanda Harabagiu et al., in their paper [2] used an approach in which that uses several feedback loops to enhance the question answering performance. These feedback loops combine in a new way statistical results with syntactic, semantic or pragmatic information derived from texts and lexical databases. The Pitfall however is finding question keywords that retrieve all candidate answers cannot be achieved only by deriving some of the words used in the question.

Davide Buscaldi & Paolo Rosso[3] developed a system on the use of Wikipedia "categories" in order to determine a set of patterns that should fit with the expected answer validation consists in, given a possible answer, saying whether it is the right one or not. The disadvantage of the system is the fact that knowledge related to small-scale events or less known people usually is not included into the Wikipedia.

Another approach by Shaw-Yi Chaw et al[4].discusses about questions using knowledge bases authored by domain experts using triples. The purpose is to identify the portion of the knowledge base that is relevant for answer of question. It significantly improves run-time

performance without sacrificing coverage. The Disadvantage is an entire knowledge base is intractable.

Mohammed Akour et al.[5]in their method of query, part of speech-feature ,find-Named Entity-remove stop word-and get answer. They seek a specific and somewhat exact answer to the query. The drawback being that they excluded one or both types of questions (How and Why) from their work because of the difficulty of handling these questions sets. In the work by Teruhisa misu ,Tatsuya kawahara[6] they say that In order to make an interactive guidance system, Since users tend to use successive questions, they investigate appropriate handling of contextual information based on topic detection, The effect of using N-best information.

Bjorn Pelzer. [7] In his approach constructed an extensive knowledge base of textual sources using of deductive reasoning. The are not being information the utilization of this data is becoming an important but difficult task.

Chinese Question-Answering System by zheng-tao yu, yan-xia qiu, jin-hui deng [8]used a method in which with Question Classification which uses word, named entity, part of speech (POS) and semantics as a classic feature to classify the question. The Advantage is it like the domain question and combines lexical relationship .It needs to improve practicability of question answering system largely.

FREyA [9]. A Feedback Refinement and Extended Vocabulary. They used a Approach in which, the rules are not used here instead of that knowledge encoded in ontology for understanding the user's question. Then the syntactic parsing is used to get answer. Then the SPARQL query is generated and the answer type is identified. The problem is it gives the correct answers only if exist in the KB.

Speech-based interactive information guidance system using question-answering technique by teruhisa misu tatsuya kawahara [10] used an Approach in which QA as successful when the system made an appropriate response to the question. An answer to the question exists in the KB, QA as successful when the system presented the answer.

Another Approach by S. Kalaivani,K. Duraiswamy [11] proposes another Keyword based search is progressed and Web based query form with large collection of knowledge base. The limitation being Data heterogeneity.XIPENG QUI,LING CAO,ZHAO LIU and XUANJI NG HUANG.[12]proposed an approach Recognizing inference in text attention in NLP. Inference in texts which combines semantic analysis with all kinds of knowledge resources. Limitation being it is not implemented in English language.

III. Approach

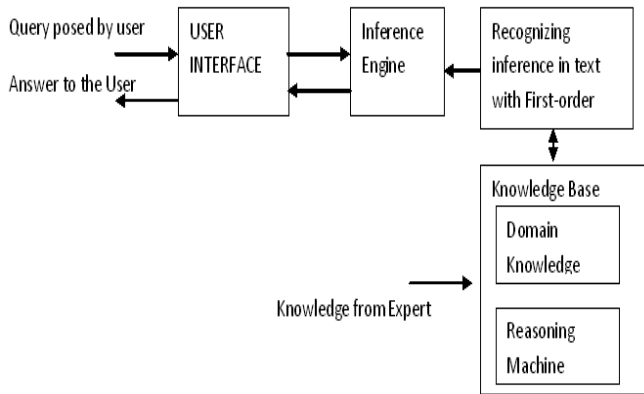


Figure:1 System Architecture

The challenge is to efficiently find relevant information in large knowledge bases. Our approach is to answer each question with a knowledge-base containing just enough information to infer an answer. Initially, the knowledge base contains only the information provided by the question. The system incrementally extends the knowledge base and includes both domain assertions and inference methods. The system succeeds if it constructs a knowledge-base that is sufficient to answer the question within time. The method consists of searching a state space in which states are knowledge-base and operators select content from the knowledge-base being searched and add it to the knowledge-base. Based on the knowledge base rules are written in drool expert system.

A. User Interface

Logging in system the user starts enters the appropriate text fields for the user to log in the username and the password.

B. 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 converted 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 forgetting correct answers.

C. Logic used

Once the Question posed by the user. That question goes under parsing ,segmentation, and removes the stop words from the questions .It generates the proof once it completed using the first order logic i.e. using the rule it find the answer in knowledge base and get back to the user.

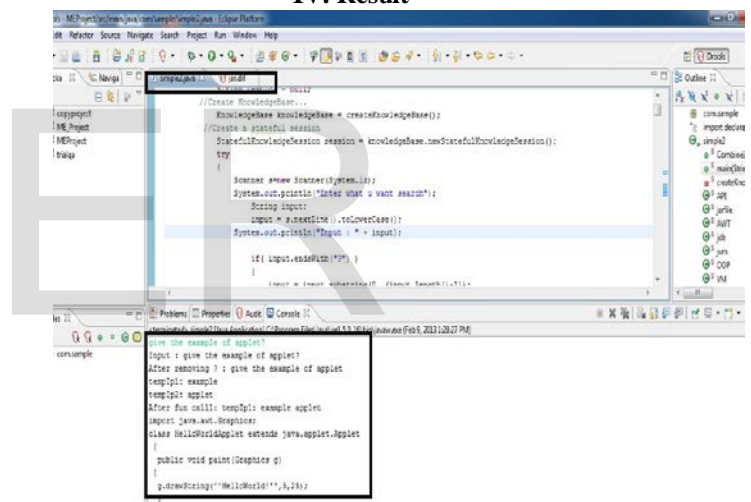
D. Knowledge Base

The Knowledge Base of this proposed system is domain specific. The storage of knowledge Base is the necessary one to retrieve the relevant and correct answer from the knowledge base using the rules. For that used the drool expert system.

E. Inference

When the queries are sent to the inference engine and the answers computed and returned. To answer these queries, the inference engine may need to apply general rules from the KB to the instance graph to infer the requested properties. The KB itself is expressed using the frame-based representation language KM [2]. Each concept in the domain is represented by a data structure, which has number slots, each of which can be filled by values. Values are themselves either frames, or expressions which evaluate to (“point to”) other frames. Formally, this is just a convenient notation for rules in typed FOL, where frame names correspond to classes (types) or instances (constants), slots correspond to relations, and frame structures correspond to a grouping of rules which share the same type restriction for their head predicate’s first argument.

IV. Result



IV. Conclusion

Performance of question answering system of getting exact result can be improved by using this methodology for retrieving answers from knowledge base. Typical information systems deal with data while knowledge-based systems automate expertise and deal with knowledge. The system gives the specific answer to your question instead of document retrieval.

REFERENCES

- [1] AAIL on Computational Models of Natural Argument, San Francisco, CA, “A Social Collaboration Argumentation System for Generating Multi-Faceted Answers in Question & Answer Communities” Sethi, R. J., and Gil, Y. 2011.
- [2] The Role of Lexico-Semantic Feedback in Open-Domain Textual Question-Answering ACL in 2010.
- [3] Mining Knowledge from Wikipedia for the Question Answering task by Davide Buscaldi, Paolo Rosso.2006.

- [4] "A Scalable Problem-Solver for Large Knowledge-Bases" By Shaw-Yi Chaw, Ken Barker, Bruce Porter, Dan Tecuci And Peter. Z. Yeh, Accenture Technology Labs. 21st IEEE International Conference on Tools with Artificial Intelligence in 2009.
- [5] "QArabPro: A Rule Based Question Answering System for Reading Comprehension To Existing Arabic QA" American Journal of Applied Sciences -2011 Science Publications system by Mohammed Akour, Sameer Abufardeh, Kenneth Magel and Qasem Al-Radaideh.
- [6] IEEE paper on "Speech-based interactive information guidance system using question-answering technique" teruhisa misu ,tatsuya kawahara 2007 .
- [7] ACM press., "An Application of Automated Reasoning in Natural Language Question Answering" by Ulrich Furbach a, Ingo Glockner b and Bjorn Pelzer. 2010.
- [8] IEEE paper Research on Chinese FAQ question answering system in restricted domain in hong kong, By zheng-tao yu, yan-xia qiu, jin-hui deng 19-22 august 2007
- [9] FREyA a Feedback Refinement and Extended Vocabulary Damljanovic et al., 2010.
- [10] "Comparison of Question Answering Systems Based on Ontology and Semantic Web in Different Environment" S. Kalaivani and K. Duraiswamy SWSE Journal of Computer Science: 1407-1413, 2012.
- [11] Recognizing inference in text with Markova Logic network "ACM transaction on Asian information processing in 2012" by XIPENG QUI LING CAO, ZHAO LIU and XUANJI NG HUANG.
- [12] A Knowledge-Based Approach to Question-Answering by Peter Clark, John Thompson Bruce Porter in 1999.
- [13] AAAI, Boris Katz. From sentence processing to information access on the world-wide web. In Proc. AAAI Spring Symposium on Natural Language Processing for the World Wide Web, CA, 1997.

IJSER