Implementation of Cognitive Approaches in Question Answering System

Mr. Naveen Chandra Gowda, Asst. Prof., School of C&IT,REVA University,Bengaluru.
Ms. Rekha K B, Asst. Prof., Dept of CSE, GITAM University, Bengaluru.

Abstract
The impact of technology is effected the human in such a way that, user will make use of internet for every simple to large purposes. To do so an effective search engine is very much required which can search and retrieve the required data from the huge chunks of data storage. The main idea of this paper is to define and implement a new question answering technique in internet using cognitive computing. The uncertainty and ambiguity exists in the traditional computing system can be solved using the cognitive computing system. The cognitive computing works as imitation of human brain.It also provides the better interaction with users. This paper is all about demonstration of searching process as an application of cognitive computing methodology. The kind of data considered here is a textual data through the analysis of natural languages query.

Keywords: cognitive computing, question answering technique, the traditional computing system, natural languages

1. INTRODUCTION

The question answering system is about to established by using cognitive computing. This concept is based on the processing the conceptual data information provided human beings memory capability.

The problem with open domain answering system is much worrying from many years [1, 2], which needs a much knowledge to establish a high coverage and request for good speed of retrieval. The data may be analyzed and understood by the answering system with increase in data volume. The information required may exists in articles, blogs, research papers and emails. It is not so easy for the traditional computing systems to analyze this kind of data and provide the answering scheme. So it requires new computing system which can think as human and do respond by considering the phrases of user query as idioms, nuances of language, text data. The data to be retrieved must be mined form the chunk of huge data collection.

The cognitive computing is an advanced approach that can be implemented as how human beings does for any query through a computerized system. The cognitive computing models can efficiently understand natural languages, process them with structured data or unstructured data finally provide the responses with evidences. The cognitive systems are capable of performing retrieval of data with automated and probabilistic techniques. A kind of processing natural language and techniques of machine learning approaches are used to provide better interaction with user for question answers. Basically the cognitive systems proposes a learning methods through experience, analyze and understands the intention of user. Here the input will be interpreted as answer for a question produced based on some notions. These notions can be item or phrases in query. Here proposed methods tried to improve the quality and validation of answers to provide the quick and correct answers to user.

2. METHODOLOGIES

Already a huge number of researchers did much work in the area of search engines. Bhatiya and Prasad [3] proposes an architecture to demonstrate e-learning scenario. It acts on huge unstructured data, makes use of map reduce techniques to figure out the answers.The question answering system was first defined by the Simmons [4] in the year of 1967 which can able to analyze many questions in English language. Then it is redefined and established at TREC-8 [5] in 1994.

There are plenty of question answer technologies exists. Few of them are database oriented, accessibility for restricted domains, some are also open for all domain. The kind of answering scheme may be of retrieving the data from structured database or unstructured data. The challenge here was a query related to natural language is to converted to database query.

Androutsopoulos et al [6] describes the template based answering method for a questions. The pattern matching methods are also implemented for natural language processing with database. The template based method of answering does not suite for text processing. The results produced are of no guarantee as answer is valid or not. The answer may be of a part of question template, retrieved through a SQL query [7].

The major part of research here is based on the semantic analysis of systems from very small domain to larger domain knowledge base. The Freebase [8] and DBpedia

[9] knowledge base, where system can get the accurate answers. But provides the least options to be solved.

The proposing methodology works on unstructured data, it do not need any knowledge base. The aim is to answer by analyzing the document containing plain text as sentences or phrases of answer keywords.

3. PROPOSED ARCHITECTURE

Internet holds bulk of peta bytes of data can be accessed by any user by using any modern search engines. But the retrieved information may be matched and useful or not for a user who expects search engine to produce some results. The general objective of this proposed work is to extract the answers to user query which matches the user expectations which he had in his mind. The answer can be retrieved from the huge amount of documents present in the local database or World Wide Web.

The general traditional search engines are of two cases [10]:

Open domain system: The main purpose of this system is to extract the answer data anything and everything from the web. The internet or web contains the huge bulk of data which is open to all, any user can query and extract the data[11]. This way of maintaining the enormous data is too complex which is again the open challenge as any search engine depend on how and huge amount of data is stored and maintained in database.

Closed domain system: The domain of the answers being extracted for any user query is limited to specific domains. The documents is restricted to specific volume and subjects. The database is defend small in size from which answer is tried to extract. So, it is limited to kind of questions accepted from the user [11].

Data sources

The major prior section of system is depends on the kind of data storage location. Here Wikipedia is considered with high priority. The data is accessed after it is cleaned, processed, indexed finally retrieved.

Psychological way of modelling a question answering system

A huge amount of research is done on how the human beings can actually answer for any questions and also analyzed the mental functionalities of operated in human brain. Arthur Greasser[12] covered a huge scenarios of psychology of answering by the human. He also noted in [12] gives an immense architecture of psychological level of answering. In order to answer for any question, human follow a procedure as,

➢ Question interpretation,
➢ Identify the kind of question,
➢ Implementation of procedures for question answering,
➢ Finalize the answers.

The main idea of proposing system of inspired by the techniques of human being answering for any question.

4. IMPLEMENTATION

The proposed work is consists of 5 major essential modules.

4.1 Indexing.

In spite of using different computational techniques will not produce the best results if there is no better understanding of type and structure of document in the database. The indexing combines different disciplinary techniques such as linguistics, mathematics, computer science, information science and cognitive sciences [13]. In Indexing search engine we can achieve the multiple document searching. Here all the documents are checked and extracted raw contents of document and analyses by using indexing process. As a part of document analysis, the content of document will be broken into tokes which will be again normalized. Finally these normalized token terms will be stored in a separate data structure named as inverted index table.

4.2 Question Processing

The answer extraction is to be done from a pool of text documents. This model is organized in two levels.

1. Parsing the query:
The initial processing is done with parsing the tokens lightly for key parts of query which saves the CPU cycles [14]. The key parts of the input query sentence will be considered such as nouns, verbs and adverbs by the user. Rest parts of the query can be ignored as excused.

2. Computing the answer type:
The system must know the kind of questions to be expected and classify them based on their type. The Table 1 in [15] shows the generalized types. They can be classified as long answers, factoid, definition, multi-lingual answers etc. After the classification, then it has to identify the type of answers. The possible answer types can be time, location, organization and person. This way of identifying the type is much important for our system. Syntactic and semantic analysis is made for the query based on the possible set of questions depends on retrieval and extraction criteria's.

4.3 Query generation.

Now we need to think about system to generate a query which helps you to extract the user passage by searching. The user passage must contain key words useful for answer type. It must make sure that one or many keywords must encounter in the passage.A new mechanism is created which can match the user query to all possible requirements. It can also be applied for the query data through speech tagging in the user passage.
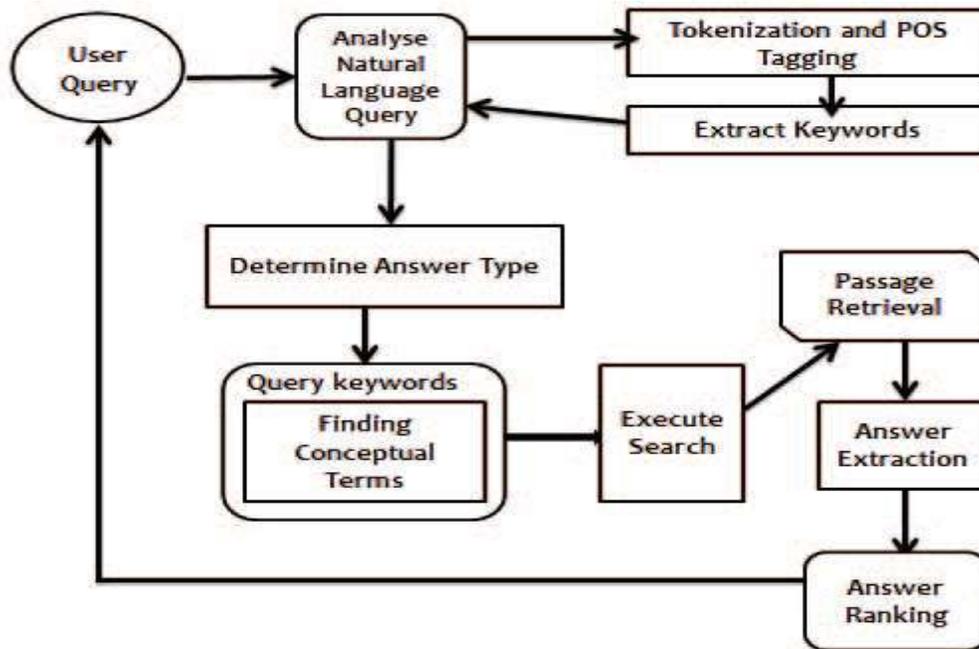
Fig: Proposed architecture.

4.4 Candidate answer extraction

Here we completely diverse in the searching process form any other information search and retrieval system. The answer in the query can be extracted by any of ways: patter based, similarity measure based, feature based analysis. In pattern based method:Much of textual patterns are learned by considering the input form websites or real text from user, will be analyzed automatically based on the kind of answers [16, 17]. For an unknown question, the patterns are used to extract and rank suitable answers from the text or web.In similarity measure based: the similarities between the candidate answers and stored question are analyzed and computed [18]. Based on which the sentence with highest similarity will be extracted as correct answer for the input query question. Where as in feature based analysis: the relevant answers will be extracted based on the SVM [21], neural network [19], regression [22]and maximum entropy [20].

4.5 Answer ranking and scoring:

Based on the kind of question classification, a part of information retrieved from candidate answer can act as answer for user query. The answer can be retrieved by different ways as using statistical analysis, SVM or using patterns based on the kind of user question that increases the performance of system. The confidence score can be found based on different parameters. The parameters can be like number of successive terms occurred in the user query. The score can be increased when the synonyms of the keywords found the query. The confidence score can be defined individually for different user query or can be consolidated.

5. CONCLUSION

Here we presented the building an engine which work as a question to answer machine. Basically it proposes a cognitive computing that imitates the mechanism of human brain. This question answer machine depicts as human brain where it considers a question, analyzed and answered. The answers for a user question will be generated automatically. It also be applied for online learning and processing communication in different activities. This can be extended as future for recognize and analyze the speech from the end user.

6. REFERENCES

1. B. J. Grosz, D. E. Appelt, P. A. Martin, and F. C. NPereira, "TEAM: An Experiment in the Design ofTransportable Natural-Language Interfaces", ArtificialIntelligence, 32(2):173-243, 1987.
2. B. Katz, "Annotating the World Wide Web using NaturalLanguage", In RIAO, pages 136-159, 1997
3. L.Bhatia, S.S. Prasad, "COPAL — Cognitive Personalizedaid for Learning," in IEEE International Conference onCognitive Computing and Information Processing(CCIP), 2015, vol., no., pp.1-6, 3-4 March 2015.
4. Simmons, Robert F. "Answering English questions bycomputer." Automated Language Processing, edited byHarold Borko, Wiley, New York (1967): 253-289.
5. Voorhees, Ellen M., and Donna Harman, "Overview ofthe Ninth Text REtrieval Conference (TREC-9)", *TREC*.2000.
6. I. Androutsopoulos, G. D. Ritchie, P. Thanisch, "NaturalLanguage Interfaces to Databases: An

Introduction",Journal of Natural Language Engineering, 1 (1), 1995.

7. Andrea Andrenucci, EriksSneiders," Automated QuestionAnswering: Review of the Main Approaches",Proceedings of the Third International Conference onInformation Technology and Applications, 2005.

8. Q. CAI and A. Yates, "Large-scale Semantic Parsing ViaSchema Matching and Lexicon Extension", In ACL,2013.

9. C. Unger, L. Buhmann, J, Lehmann, A.-C, N. Ngomo,D.Gerber, and P. Cimiano, "Template-Based QuestionAnswering over RDF Data" in WWW,2012.

10. Hai Doan-Nguyen, Leila Kosseim, "Improving thePrecision of a Closed-Domain Question-AnsweringSystem with Semantic Information", ACL 2004Workshop on Question Answering in RestrictedDomain,2004- acl.ldc.upenn.edu.

11.Ms.Pooja, P.Walke, Mr.ShivkumarKarale,"Implementation Approaches for Various Categories ofQuestion Answering System ", Proceedings of 2013 IEEEConference on Information and CommunicationTechnologies (ICT 2013).

12. A. Graesser and T. Murachver, "Symbolic procedures ofquestion answering", in A.Graesser and J. Black,editors,The Psychology of Questions, pages 15–88, Erlbaum,1985

13. https://en.wikipedia.org/wiki/Search_engine_indexing

14. Ingersoll, Grant S., Thomas S. Morton, and Andrew L.Farris. Taming text: how to find, organize, andmanipulate it. Manning Publications Co., 2013. 251-252.

15. Mohammad Reza Kangavari, Samira Ghandchi, ManakGolpour "A New Model for Question AnsweringSystems",World Academy of Science, Engineering andTechnology 18 2008.

16. Wu Lide, Huang Xuanjing, Guo Y, et al. FDU at TREC-9: CLIR, Filtering and QA Tasks[C]. Proceedings of the9th Text Retrieval Conference, 2001: 189-203.

17. Dell Zhang, Wee Sun Lee, "Web based Pattern Miningand Matching Approach to Question Answering[C]", InProceedings of the Eleventh Text Retrieval Conference.TREC2002. http://trec.nist.gov/.

18. Cui Huan, Cai Dong-Feng, Miao Xue-Lei, "Research onWeb based Chinese question answering system andanswer extraction", Journal of Chinese InformationProcessing, 2004, 18 (3):24-31(in Chinese)

19. Zhang Yong-Kui,ZhaoZhe-Qian,Bai Li-Jun etal,"Internet-based Chinese question answeringsystem.Computer Engineering", 2003, 29(15): 84- 86

20. Abraham Ittycheriah,"Trainable question answeringsystems", USA:The State University of New Jersey,2001

21. Jun Suzuki, Yutaka Sasaki, Eisaku Maeda, "SVM answerselection for open-domain question Answering", l9thInternational Conference on Computational Linguistics(Coling2002) [C] Taipei:Howard International House,2002. 974- 980.

22. Peng Li, Yi Guan, Xiao-Iong Wang. "Answer extractionbased on system similarity model and stratified samplinglogistic regression in rare data", International Journal ofComputer Science and Network Security, 2006,6(3):189-196