iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** V **Month of publication:** May 2019

DOI: https://doi.org/10.22214/ijraset.2019.5228

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

Survey on Intelligent Answering System using Similarity Model based Learning

Dr. Yogish HK[1], Arun Kumar S[2], Shwetha S[3], Rachel Britto[4], Deepak N[5]
[1]Head of the Department, Professor, [2]Assistant Professor, [3,4,5]Student, Department of Computer Science and Engineering, Sapthagiri College of Engineering, India.

Abstract*: Machine Learning (ML) is an artificial intelligence (AI) area that is a set of statistical techniques to solve problems. ML techniques can be applied to a wide range of unlimited problems-vision-based research, fraud detection, price prediction, and even natural language processing (NLP). A smart, user-friendly automatic response system is developed with the ability to detect and answer questions in English. There are many response systems that use the concept of natural language processing to answer the questions, but they are not so accurate in finding the right answer. Specific predefined queries with specific format are also required. Users need to ask the questions in the given format only for such systems. Users can enter the queries as they wish in the proposed system. No specific format is required. If the query fails to match any predefined query, the user will be suggested the best matched query. The domain expert keeps the answers to such questions in a database. The best matched response searched from the database is returned to the user during the retrieval of answers. A template matching technique is used to perform this match. Thus the system is more efficient and accurate from all the perspectives.

Keywords*: Machine Learning, Artificial Intelligence, Natural Language Processing, Tokenization, Cosine Similarity Model.

I. INTRODUCTION

In the modern world, the internet technology has become so much advanced that the way of interaction has changed across the entire application domain. Natural language is being used for communication in all the fields. When a user types a question on the internet, the user is flooded with relevant and irrelevant data and its time consuming to sort it out and find the information relevant to the question which triggers a need for Automated Intelligent Question Answering System for navigating to meaningful data. This is a challenging task which can be achieved by using Natural Language Processing (NLP) and Information Retrieval (IR) techniques. NLP helps a computer to better understand the language used by user to communicate.

The skills required to build a smart response system include tokenization, parsing, speech tagging parts, question classification, query construction, sentence understanding, document retrieval, keyword ranking, classification, response extraction and validation. There are two important domains in which the QA system can be implemented: Open domain and closed domain. Open domain deals with questions in all domains. Closed domain deals with questions only in a particular domain and a database is maintained for it. NLP is a hard problem which deals with understanding the human language not only words but also how the words team up to for a meaning. The different NLP techniques are: Deep analytics, Machine translation, Named entity extraction, Co-reference resolution, Automatic summarization, Sentiment analysis. IR is used to search and retrieve knowledge based information from the database.

II. LITERATURE REVIEW

In [1] a survey is conducted on various types of system for answering questions. In the field of Information Retrieval (IR), the questioning answering system is an important research area. Different fields of research are combined in this system like Natural Language Processing (NLP), Artificial Intelligence (AI), Information Retrieval (IR), and Information Extraction (IE). It is mainly categorized into two types based on the availability of resources: open domain and closed domain. The search engines like Google comes under the open domain where the information is retrieved from the World Wide Web. This is usually done by using keywords matching and frequencies of accessing documents. It is a time consuming process for the user to search for the relevant answer from the whole bunch of documents/data obtained as a search result. The QA is a rapidly growing research field. It reduces the time for extraction of data as well provides an exact or nearest to exact answer. This time is also dependent on the domain on which the system is used. The system used on closed domain fetches more accurate result when compared to system used under open domain. The search space for the system is less in closed domain when compared to the internet. The QA system has been implemented for different languages like Korean, Japanese, and English etc. The applications of the QA system are extracting information from document, online examination system, document management, Language learning, human and computer interaction, classification of document and many more.

©IJRASET: All Rights are Reserved

In [2] an Intelligent QA system is designed for Arabic language. The main aim of this system was that it must be able to respond to the queries automatically and the answer retrieved should be accurate with respect to the query. The system answers more sophisticated questions that require some sort of temporal inference. It builds its database with different forms of question-answer pairs entered by the user and answered by the system. This will result in increased speed and accuracy. This system makes use of Information Retrieval (IR), Information Extraction (IE), and Natural Language Processing techniques. The user enters a query in spoken language and gets the answer in a word or a sentence format. Search Engines usually lack in conceptual knowledge of world and depends on probability theories and bivalent logic. It does not provide precession in the extracted answer. The retrieved snippets are not good for big data analysis and web information. But the QA system will provide only the requested answer. This system will first look up the database for the answer if it's not available then it will opt for WWW and consequently the result will be entered on to the database. If a specific question is not found in database, the question will be analysed by the module of question processing and the answer will be extracted from the web by the module of answer extraction. Usually QA system will work on the concept of matching keywords, but this does not address the issue of how to conduct extensive question analysis and understanding of natural language. This proposed QA is said to have some degree of semantic understanding and can generate answers autonomously.

In [3] a system is developed which takes fact database as input and has the capability to answer questions of wide range of complexities. It implements a machine which will convert the input into a computable format so as to retrieve an answer for the question asked. The fact database consists of input facts and questions. The implemented system is evaluated based on whether it could answer simple deduction questions which produces yes/no as the answer and also on its capability to relate facts and answer a complex question. The tasks used for this evaluation are Dynamic Memory Networks (DMN), which passes the bAbi tasks of Facebook and uses this model to evaluate the data set released for the ARISTO challenge of Allen Institute, which contains increasingly difficult science questions. The knowledge base for this system is the Facebook bAbi tasks. Developing a system which gives consistent result for both simple and complex questions is a challenging task. A bAbi task consists of 5 cases. Case 1: require only single facts to answer the question. Case 2: requires multiple facts to be considered in order to conclude on an answer. Case 3: decides the answer on the basis of keywords and also takes into account the position of the keyword in the question. Case 4: deals with the entire yes/no questions. Case 5: deals with counting cases like questions with "how many?" criteria. The system takes into account all the relevant facts necessary to conclude on the final answer. To implement the above methods the system should have the capability to understand the facts given Neural networks has shown reliable average performance in this field but the performance is lowered as complex tasks such as bAbi are considered. Neural networks have made a good progress in the field of image and text classification. Recently, performance has increased for complex tasks due to the addition of memory and attention components to neural networks. In [4] the major role of the Questioning Answering System is to simulate answer required by the user. People usually use browsers to access information of the web. Blogs and Forums are a source of dynamic information sharing. Due to the widespread use of smart phones and apps for a huge number of purposes and need of information in short span of time, a mobile based questioning answering system is developed which provides personal assistance in learning and also for providing information for the users stored on the computers. It is used as a surface to fulfill the requirement of the users with related content. it uses natural language for the communication purpose and displays the optimized result. The mobile QA system will be taking input in natural language and analyses it and match it with the information stored in knowledge base and display result. Amazon book reviews, 20newsgroup and Yahoo datasets are used to build the knowledge base. The answers are stored as content - specific clusters in the knowledge base and display the output as snippets. Use Sentiment Analysis to decrease the vocabulary gap between user query and the retrieved response. The results of the proposed interface are evaluated using standard metrics such as Precision, Recall, F1-Score, Inverse Precision and Inverse Recall to return the relevant response. The open domain system is capable of answering dynamic data irrespective of domain nature, but the closed domain uses datasets for retrieving the answer. The learner who is in a need of static data will use the closed domain QA system like tourism, medical health, historical data etc. Recently, many efforts have focused on question answering systems based on social media networks for getting the precise information.

In [5] a research is made on the Questioning Answering Systems and the techniques used for development. As the information and communication technology has increased very much, the necessity and importance for the question answering system has increased rapidly. Complex assessment techniques are necessary for its development. A system must first analyse the question to answer a question and search for one or more possible answers and present it in a user - friendly form. The systems usually use graphical methods for representing knowledge. The knowledge gained is converted to a model and then the model is searched for answers. The Nodes Of Knowledge (NOK) method is one such example. The QA system is capable of answering factual questions by referring to the collection of documents by combining Information Retrieval and Natural Language Processing Techniques. There are two prominent fields in research of the QA system: resource and evaluation. Kupiec-a system used trivial method for collecting

©IJRASET: All Rights are Reserved

questions and Internet encyclopaedia as source for answer. A set of multiple choice questions makes the evaluation easy. The evaluation of questions with no answers and questions with many answers is a complex process which is still under research. The six criteria for assessment of a QA system are: relevance, correctness, conciseness, completeness, coherence and justification.

In [6] an intelligent answering system is proposed based on the research in the field of Artificial Intelligence (AI). It aims at delivering concise information which may contain the answer for the question asked. This system is a solution for reducing the time consumed in searching the relevant answers from the data bombarded on the users. AI's goal is to develop a system that has intelligent behaviour i.e. it has the ability to solve different problems, learn and understand languages. This system consists of two primary parts: knowledge base and inference engine. Usually the knowledge base is made up of factual and experimental data and the inference engine is used to determine the answers. The system designed is for closed domain and aims at retrieving the exact answer for the query rather than bombarding the user with a set of documents as done by search engine. The closed domain QA system requires extensive use of natural language processing and a lot of research to be done on factual data, definition lists, and paragraphs and cross lingual questions. When a question is given only the relevant part of the knowledge base is searched for thus reducing the search space. Knowledge base is represented as a rule i.e. it contains an IF and a THEN condition. The systems with this kind of representations are called as rule based systems. In [10] a questioning answering system is proposed which concentrates more on the abbreviations appearing in a question asked by the user. The proposed system aims at extracting a set of documents based on the abbreviations present in the query. It includes the methods for identifying abbreviations in the question, retrieving the documents which are related to the abbreviation terms and filtering all the noun phrases from the result as the full forms of the abbreviation. This iQA system is an information retrieval system rather than a document retrieval system. The different techniques used for implementation are query rewrites and formulation, question classification, information retrieval, passage retrieval, answer extraction, and answer ranking and justification. The performance of the system will entirely depend on all these techniques. The semantic gap between the abbreviations written in the question and potential answers has been reduced. WorldNet is used in the system for identifying abbreviations. The query is sent to the WorldNet and if the words in the query are not found then they are considered as abbreviations. These abbreviations are searched on search engines and the documents retrieved are then processed to find the possible long forms.

III.CONCLUSION

This paper emphasizes on various question answering systems and question answering techniques. There are many question answering systems which return answers to the questions entered by the user but fail to return appropriate and accurate answers. It is a challenging task to develop a system which would understand natural language questions correctly and provide exact answers to the user. The quality of answer returned by closed domain question answering system is high because the system is restricted to a particular domain, whereas the quality of answer returned by open domain system is low. The performance of a system can be improved to return exact results by retrieving answers from knowledge base. The knowledge based system returns the specific answer to the user instead of a document.QA system will help all users to retrieve exact information easily.

REFERENCES

[1] A. Chandra Obula Reddy, Dr. K. Madhavi, "A survey on types of Question Answering system", IOSR Journal of Computer Engineering (IOSR-JCE) E-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 19, Issue 6, PP 19-23, Ver. IV (Nov.- Dec. 2017).

[2] Waheeb Ahmeda1, Ajusha Dasana, Babu Anto, "Developing an Intelligent Question Answering System", India Received: 03 January 2017; Accepted: 28 March 2017; Published: 08 November 2017.

[3] Vinay Chandragiri, "Intelligent Question Answering System", ISSN: 0975-9646. International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 7 (5), 2016.

[4] J. Tomljanović, M. Pavlići M, "Intelligent Question – Answering Systems: Review of research", AšenbrenerKatić Polytechnic of Rijeka Vukovarska 58, Rijeka, Croatia. I.J. Information Engineering and Electronic Business, 2018, 1, 16-23 Published Online January 2018.

[5] Karpagam K, "A Mobile based Intelligent Question Answering System for Education Domain", Received: 21 July 2017; Accepted: 22 September 2017; Published: 08 January 2018.

[6] V.S. Babanne1 Dr. S.T. Patil2, D.J. Joshi3, "Intellegent Question Answering System", International Journal of Scientific & Engineering Research, Volume 4, Issue 5, ISSN 2229-5518 May-2016.

[7] Deepa Yogish, Manjunath T N, Ravindra S Hegadi, "Survey on Trends and Methods of an Intelligent Answering System", 7 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 2017.

[8] Sameerchand Pudaruth, Kajal Boodhoo and Lushika Goolbudun, "An Intelligent Question Answering System for ICT", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), November 2016.

[9] Ahlam Ansari, Moonish Maknojia, Altamash Shaikh, "Intelligent Question Answering System based on Artificial Neural Network", 2nd IEEE International Conference on Engineering and Technology (ICETECH), March 2016.

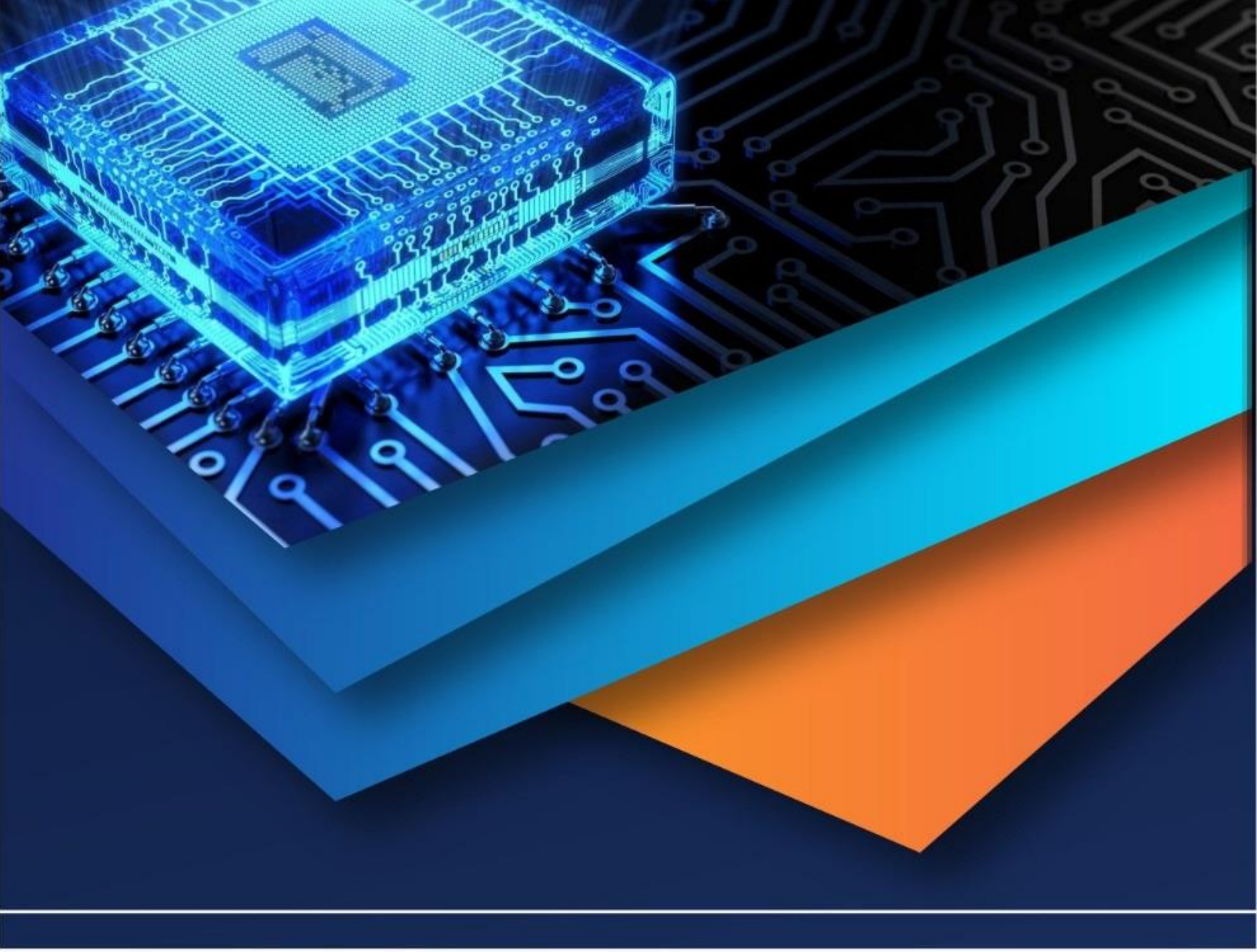
[10] ShubhangiRathod, SharvariGovilkar, "Survey of various POS tagging techniques for Indian regional languages",(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2525-2529.

©IJRASET: All Rights are Reserved

ABOUT THE AUTHORS

	**Dr. Yogish HK** Head of the Department, Professor Dept. of Computer Science & Engineering Sapthagiri College of Engineering Bengaluru - 560057
	**Arun Kumar S** Assistant Professor Dept. of Computer Science & Engineering Sapthagiri College of Engineering Bengaluru - 560057
	**Shwetha S** Student (B. E.) Dept. of Computer Science & Engineering Sapthagiri College of Engineering Bengaluru - 560057
	**Rachel Britto** Student (B. E.) Dept. of Computer Science & Engineering Sapthagiri College of Engineering Bengaluru - 560057
	**Deepak N** Student (B. E.) Dept. of Computer Science & Engineering Sapthagiri College of Engineering Bengaluru - 560057

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)