



## Comparative Study on Question Answering Systems and Techniques

<sup>1</sup>Sunil A. Khillare\*, <sup>2</sup>Bharat A. Shelke, <sup>3</sup>C. Namrata Mahender

<sup>1,3</sup>Department of CS & IT, Dr. B. A. M. University, Aurangabad (MS), India

<sup>2</sup>Department of Computer Science, Shri Chhatrapati Shivaji College, Omerga (MS), India

---

**Abstract**— As technology highly developed, QA Answering also developing rapidly specially form the last decade. Users input the question on the net and return exact answer by various search engines. This also called as Question Answering (QA). Where various researchers work with their selected domain like closed or open domains. In this paper we reviewed several research papers on the question answering system. We also observed that more research has been done on closed domain rather than open question answering system.

**Keywords**— Question, Answer, Question Answering, Algorithm, Question Answering in Marathi

---

### I. INTRODUCTION

From the last decade internet users are increasing in more titre. In QA different queries are provided by the user in aim of getting accurate answers in Question Answering Systems. Question Answering provides perfect solution to retrieve valid and accurate answers to user question asked in natural language instead of query [25]. Various Scientists are Work on QA from last many with the English, Chinese, Japanese, Korean, etc. languages. But from some last year's scientists move their research into regional languages like Tamil, Punjabi, Hindi, Malayalam, etc. In this paper we presented the review on question answering and system of various languages. After the introduction Literature review has been taken, third chapter will present information about question answering system which creates in Devnagari script.

#### A. Question

Question, the word has never a dictionary to know, but the response needs not only a dictionary, but the whole universe sometimes to be searched for just a response. The question is a noun a sentence recorded or expressed so as the elicit information [6].

#### B. Types of Questions [3]

- Factoid Based Questions: Factoid questions are those questions which the answer is a single fact. e.g. when was Mr. Gandhi born?
- Definition Questions: The answer should be a short paragraph which briefly wishing to know features of the thing e.g. if ask about person then his birth date, height eyes color, marriage all are required
- Yes- No / True - False Questions: are those questions which answer is withering yes or no. e.g. Is Aurangabad pollution bigger than Amravati?
- Instruction Based Questions: are those questions where instructions are present e.g. How do I make Tomato Soup?
- Explanation Questions: are those questions where explanation is required e.g. Why did America enter in WWII?
- List Question: are those questions which answer is present in list format e.g. What are the ingredients are present in Tomato Soup?

#### C. Answer

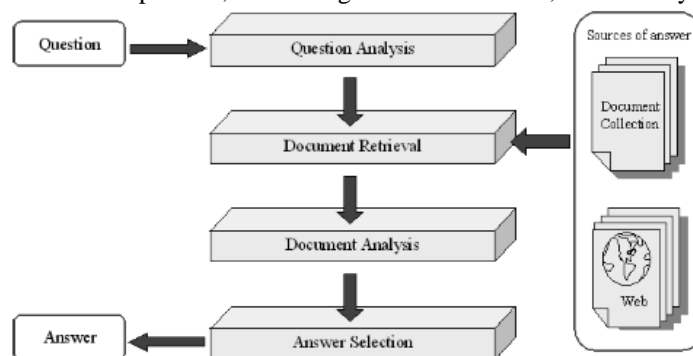
Answer is a noun, a thing that is said written or done to deal with or as a reaction, statement or situation. An answer to a particular question may use more than one way of expression. Thus tends the responses to the unique aspect relevant to each single user.

### II. QUESTION ANSWERING

Research and development of systems capable of answering questions in natural language dates back to 1959 [24], but the notion of a question answering system was born in 1950, when Turing offered a solution to the question of whether or not machines can think. He proposed a task he called an Imitation Game," which has eventually become known as the famous Turing Test, in which a human communicates with a machine via a teletype interface and asks questions of it. Mainly Question Answering (QA) is becoming an increasingly important research area in natural language processing. Since 1999, many international question answering contests have been held at conferences and workshops, such as TREC, CLEF, and NTCIR [1]. Basically QA is special task to search natural language answers from natural language questions. The main task of QA is providing a short answer to a natural language query supported by a document in an underlying

document collection [2]. Question answering aims to develop techniques that can go beyond the retrieval of relevant documents in order to return exact answers to natural language questions, such as “How tall is the Eiffel Tower?”, “Which cities have a subway system?”, and “Who is Alberto Tomba?”. Answering natural language questions requires more complex processing [3]. From last many decades various researchers are working with this domain. Question Answering (QA) systems take in a natural language question, and return the answer from the set of documents. Answer retrieval, rather than document retrieval, will be integral to the next generation of search engines. Currently, there is a web site called AskJeeves™<sup>1</sup> that attempts to retrieve documents to answer a question. Question answering systems will handle query creation, and finding the exact entity that is the answer [4].

Following figure displays the main components, of such a general architecture, and the ways in which they interact.



### 2.1 Types of Question Answering [4]

QA systems can be categorized in many ways, such as by the application domain, answer sources, or target and source languages. But mainly it divided into two main parts.

- i] Open Domain QA System
- ii] Close domain QA System

Title of Paper	Technique/ Methodology	Conclusion	Year
ASK Jeeves System[7]	ASK Jeeves open domain system (open Domain)	Now days it is one of the topmost web search engine. Now day’s system includes in top most search engines.	1995
A Simple Question Answering System[2]	Language used: Perl CMU Link Parser used (Open Domain)	Link parser analysis was only used for the proportion of questions that deal with ambiguous answer types. When, Where, Why, Describe and define are easiest question words to process because they map directly to the answer types. What, Which, and name question words are extremely ambiguous when it comes to determining the question’s answer concept.	2000
SHAPAQA: Shallow Parsing for Question Answering on the World Wide Web[15]	Memory based shallow parser (Open domain on web)	In combination with two back-off methods m mean reciprocal rank of .45 is achieved	2001
Algorithms To Improve Performance of Natural Language Interface[16]	Present algorithms to handle problems caused by semantic symmetry and ambiguous modification shallow Parser.	System gives good result while it handles semantic symmetry and Ambiguous Modification and in the time manner also.	2003
Perspectives on Chinese Question Answering Systems [1]	Machine Learning Approach Deep and Shallow Parsing (Closed Domain)	Reviewed 3 QA contents: TREC QA Track, QA@CLEF & NTCIR CLQA. Only TREC QA is dedicated to monolingual QA. The other two contests deal with both monolingual and multilingual QA	2005
Context-Aware Restricted Geographical Domain Question Answering System[14]	Geographical Domain Question Answering System (Closed Domain)	The overall system performance of the system is 80% on questions starting with Find.	2010

A Generative Model for Parsing Natural Language to Meaning Representations[17]	Presented a New Generative Model	Model achieves a 6.7% and 9.8% absolute improvement over two other state-of-the-art models WASP and KRISP respectively.	2008
Natural Language Interface Using Shallow Parsing[18]	Semantic symmetry; ambiguous modification; Shallow parsing based algorithms. (Closed Domain)	Algorithms developed for ENLIGHT system are useful to improve precision of search Engines. This system does not compromise with response time when dealing problems which with Problems caused by linguistic phenomena	2010
PRASHNOTTAR: A Hindi Question Answering System[19]	Question Preprocessing Question classification used for question analysis (Open domain Web System)	Working with When, Where what time, how many question in Hindi language. Overall accuracy of the system is approximately 68.00 %.	2012
TALASH: A Semantic and Context Based Optimized Hindi Search Engine[20]	Method-I: Query enhancement using lexical resource Method-II: Query Enhancement using user Context Open Domain Query Enhancement using hybrid technique Hindi word Net used. (Closed Domain)	Method III has given better results other than two. Works with Hindi question Answering System	2012
Web Based Application for Hindi Question Answering System[21]	Query Formulation, Query interface, Query classification has been used. The implementation is done in java and it also integrates CYK parser.	NLP building modules, Query formulation and Concept of entities makes this system to work in any domain. If the Query is based on Hindi text then the total answers based on query are displayed and if the answer is not present in the Hindi text then the system ask the answer from Google.	2012

### III. QUESTION ANSWERING DEVNAGARI SCRIPT

In India after the development of TREC and various questions answering system researchers moving their research from international language like English, Japanese, and Chinese to Indian regional languages like Tamil, Punjabi, Bengali, Urdu, and Oriya, Hindi etc. But research in Devnagari script has not as compare to other languages. Sahu et. al. [19] developed PRASHNOTTAR Question Answering system in Hindi language. It is based on understanding the meaning of the given question and expressing them in query logic language. The experiment is performed on stored Hindi text data. The Hindi text data is collected from web. Vasnik et. al. [20] presents a Hindi search engine TALASH. To developed model in which used three models for the query enhancement which are based in lexical variance, user context and combination of both techniques. Author also proved that combination of both techniques giving good results than individuals. Shalini et. al. [21] developed the Web application system in Hindi language which summarizes a web application to extract answers from Hindi text for a given question. It is based on understanding the meaning of the question and performs parsing on it to extract the answers.

The same or Some other techniques/ algorithm are also developed in another Indian regional languages [22][23][24]but on the Marathi Devnagari script very few researchers has worked with stemming specifically or on removing the stop word[25].

### IV. CONCLUSIONS

As the technology growing rapidly use of internet also grown. On the base of internet question answering era tremendously developed online as well as offline. In this paper we observed that various researchers work on QA with various languages and regional languages too. But very few researchers work with Hindi language. But until now Marathi language based question answering not developed by the researchers. We also observed that rather than work with open domain question answering various scientists work with closed domain question answering system.

### ACKNOWLEDGMENT

I am very much thankful to Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University Aurangabad to give me essential equipments.

### REFERENCES

- [1] Cheng-Wei Lee, Cheng-Wei Shih, , Min-Yuh Day, Tzong-Han Tsai, Tian-Jian Jiang, Chia-Wei Wu, Cheng-Lung Sung, Yu-Ren Chen, Shih-Hung Wu, Wen-Lian Hsu, "Perspectives on Chinese Question Answering Systems".
- [2] Richard J Cooper and Stefan M R., "A Simple Question Answering System", Proceedings of the 9th Text Retrieval Conference (TREC9), NIST, 479-488. News article from TIPSTER and TREC CD's, (Pg 1-7).

- [3] A Ph.D. Thesis Entitled as “Open-Domain Question Answering”, Submitted by Mark Andrew Greenwood, to Department of Computer Science University of Sheffield, UK September 2005
- [4] A M.Sc.Thesis “Entitled Question Answering Using Document Tagging and Question Classification”, Stephen Dubien, Department of Computer Science University of Lethbridge, Lethbridge, Alberta, Canada.
- [5] [5] 2007 North American Computational Linguistics Olympiad <http://namclo.linguistics.org/cool/cfm>
- [6] A M. Phil. Thesis Entitled as “Development of Inference Engine to Automate the Descriptive Examination System”, Submitted by H. R. Gite, to Department of Computer Science, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad.2014
- [7] "Ask Jeeves," <http://www.ask.com/>
- [8] Nitin Indurkha, Fred J. Damerau, “HandBook of Natural Language Processing”, Second Edition Chapman & Hall/CRC Machine Learning & Pattern Recognition Series. SBN13: 978-1-4200-8593-8, 2010.
- [9] William A. Woods, R. M. Kaplan, and B. L. Nash-Webber. “The Lunar sciences natural language information system”. Final Report 2378, BBN, Cambridge, MA, 1972.
- [10] A M. Tech. Thesis Entitled as “Query Expansion Techniques for Question Answering”, submitted by Matthew W. Bilotti, for the degree of Master of Engineering in Electrical Engineering and Computer Science to the Department of Electrical Engineering and Computer Science , Assachusetstts Institute Of Technology, May 2004.
- [11] Sanjay K Dwivedi, Vaishali Singh, “Research and reviews in question answering system”, International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013. Proceeding Published in Elsevier.
- [12] Biplab Ch. Das, Seminar on “A Survey on Question Answering System”, 2008
- [13] Chung H, Song YI, Han KS, Yoon DS, Lee JY, and Rim HC. A Practical QA System in Restricted Domains. In Workshop on Question Answering in Restricted Domains. 42nd Annual Meeting of the Association for Computational Linguistics (ACL), 2004, pp. 39-45.  
Mishra A, Mishra N, Agrawal A. Context-aware restricted geographical domain question answering system. In Proceedings of IEEE International Conference on Computational Intelligence and Communication Networks (CICN), 2010, pp. 548-553.
- [14] Sabine Buchholz, Walter Daelemans, 2001, "SHAPAQA: Shallow Parsing for Question Answering on the World Wide Web" , Proceedings Euroconference Recent Advances in Natural Language Processing, 2001
- [15] Rajendra Akerkar and Manish Joshi,2003, " Algorithms To Improve Performance Of Natural Lan-guage Interface ", IJCSA Vol. 5, No. 2, pp 52 - 68.
- [16] Wei Lu, Hwee Tou Ng, Wee Sun Lee, Luke S. Zettlemoyer, 2008," A Generative Model for Parsing Natural Language to Meaning Representations", EMNLP 2008 - 2008 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference: A Meeting of SIGDAT, a Special Interest Group of the ACL.
- [17] Rajendra Akerkar and Manish Joshi, 2010, Natural Language Interface Using Shallow Parsing, International Journal of Computer Science and Applications, Technomathematics Research Foundation Vol. 5, No. 3, pp 70 – 90.
- [18] Shriya Sahu, , Nandkishor Vasnik and Devshri Roy, April 2012, PRASHNOTTAR: A Hindi Question Answering System, International Journal of Computer Science & Information Technology (IJCSIT) Vol 4, No 2.
- [19] Nandkishor Vasnik, Shriya Sahu, Devshri Roy, TALASH: A Semantic and Context Based Optimized Hindi Search Engine, June 2012, International Journal of Computer Science, Engineering and Information Technology (IJCEIT), Vol.2, No.3.
- [20] Shalini Stalin, Rajeev Pandey, Raju Barskar, “Web Based Application for Hindi Question Answering System”, 2012 , International Journal of Electronics and Computer Science Engineering ISSN- 2277-1956
- [21] Vishal Gupta, “Hindi Rule Based Stemmer for Nouns”, January 2014, International Journal of Advanced Research in Computer Science and Software Engineering , Volume 4, Issue 1, ISSN: 2277 128X
- [22] Vishal Gupta, “Suffix Stripping Based Verb Stemming for Hindi”, January 2014 , International Journal of Advanced Research in Computer Science and Software Engineering , Volume 4, Issue 1, ISSN: 2277 128X
- [23] Han L, Yu ZT, Qiu YX, Meng XY, Guo JY and Si ST. Research on passage retrieval using domain knowledge in Chinese question answering system. In Proceedings of IEEE International Conference on Machine Learning and Cybernetics, Vol. 5, 2008, pp. 2603-2606.
- [24] Jaspreet Kaur, Vishal Gupta, July 2013, “Comparative Analysis of Question Answering System in Indian Languages”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7, ISSN: 2277 128X.