

## A LITERATURE REVIEW ON QUESTION ANSWERING SYSTEMS AND TECHNIQUES

**Mr. Ch.Mahender Reddy, Mr. H.V.Ramana Rao, Mr. Kenanya Kumar Kasaraneni, Ms. S.Sravani,**  
CMR Institute of Technology Hyderabad

### **Abstract**

After the development internet technology internet users are increasing rapidly. Many times internet users posing question on search engine and they demanding answers from it, this phenomenon is called question answering. Now a days this is an very demanding issue for the researchers as users want accurate answers. For this Question answering system need to be develop by the various researchers. In this we surveyed various research papers which aims to identifying Question Answering techniques, tools and systems, as well as the metrics and indicators used to measure the approaches for Question Answering systems.

**Keywords:** Question Answering Systems; Information Retrieval; Natural Language Processing.

### **Introduction**

Question Answering systems (QAS) enable users to retrieve the exact answers for questions posed in natural language. [1] QAS is a specific type of information retrieval. Question answering is multidisciplinary, it involves information technology, artificial intelligence, natural language processing, knowledge and database management and cognitive science [2].

There are three components of Question answering system, Question classification, Information retrieval, and Answer extraction. These components play a vital role in QAS. Question classification play chief role in QA system to categorize the question based upon its type. Information retrieval method is to identify success by extracting out applicable answer post by their question answering system [3].

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, Marathi etc.

### **Literature Review**

For any research current scenario or status is a very much important part, for this we have reviewed research papers from last 02 years.

In 2003, Kumar et al.(4) developed Hindi search engine with the help of similarity heuristics. The developed search engine provide the facility to extract correlated contents from set of e-learning. Questions provided by the users were classified for selection of appropriate answers. From the query stop words are removed and relevant keywords where extracted. Query was enriched with synonyms of keywords. Finally the query were passed to retrieval engine, which on basis of locality returns top passages after ranking.

In 2005, Lopez et al.(5) developed prototype system named AquaLog which is a portable question-answering system. This system took queries expressed in natural language and ontology as input and returns answers drawn from the available semantic markup. AquaLog uses GATE NLP platform, string metrics algorithms, WordNet and novel ontology-based similarity services for relations and classes to make sense of user queries with respect to the target knowledge base.

In 2007, Wang, Chong, et al [6] designed Portable natural language interface to Ontologies, name as PANTO which accepts generic natural language queries and outputs SPARQL queries. The proposed system adopts a triple-based data model to interpret the parse trees output through parser. Stanford parser were used with multiple existing techniques. Further, these techniques were integrated with tools to interpret parse trees of natural language queries into SPARQL. To understand sense of the words in the NL queries and WordNet and string metrics algorithms are also integrated.

In 2011, Guda *et al.* [7] Investigated various Question Answering (QA) systems. the main purpose of designing QA systems are provide correct responses to the questions in a human like manner giving short and accurate answers. The investigated QA systems were categorized viz; WEBBASED QA system, Information Retrieval or Information Extraction (IR/IE) BASED QA system, RESTRICTED DOMAIN QA (RDQA) system and RULE BASED QA system. Artificial Intelligence based QA systems led to a breakthrough for new directions of research. Study reveals that each and every QA system has different aspects in answering the questions, while temporal aspect has got little attention. Further, it also gives novel idea to develop Question Answering system that can answer different types of queries.

In 2012, Chaware and Rao [8] investigated the system where Semantic matching process was performed using ontology for Hindi and Marathi languages to infer the information from knowledge base. The data and ontology were maintained in English for easy building and traversing, ontology terms were extracted to represent knowledge as an answer for the query. The approach converts local language to English using bilingual dictionary where there is more chance of translating mismatch and loosing of morphological rich words and phrases of Hindi and Marathi language, which may lead to mismatched query keywords.

Vasnik *et al.* [9] designed Hindi search engine model in which three models for query enhancement were described. These models were designed on the basis of lexical variance, user context and combination of both techniques. HWN model was applied for query expansion. This research was carried out in three different stages. In first stage synonym and hypernyms words were used for semantically related terms of query. Second stage deal with user context, in final stage first two methods was combined. The obtained results reveals that by combining the principles of search with HWN and context awareness, more relevant search results may be produced for a user and, the precision of information retrieval increases.

In the same year Sahu *et al.* [10] proposed approach to extract answers from Hindi text for a given question where the text is expressed in the form of query logic language and then relevant answer is extracted for the given question. Four kind of question were focused such as: What, Where, How many, and what time. Shallow parser was used to extract keywords. While extracting such keywords no semantic relations were consider. Traditional methods i.e. to take words as independent words during matching and just check the existence of the query keywords in the stored data were used and no relations constraints between words in a phrase or neighborhood are extracted which leads to less accuracy.

Stalin *et al.* [11] designed Hindi Question Answering System which based on searching in context by using similarity heuristic and utilizes syntactic and partial semantic information. Domain-specific and question specific entities were found out after removing the stop words and also longest phrase are extracted while processing query. The obtained results reveals that Building of limited words synonyms lexicon reduces the accuracy of system due to mismatch of unavailable entities.

In 2013, Gupta [12] proposed Punjabi and English language based Question answering system. In this proposed system accept query in English or Punjabi language of which stop word is eliminated initially. Noun, adjective, verbs or adverb were extracted with the help of query string. Synonyms of Punjabi and English languages were extracted with the help of dictionary. Reformulation of query was done using the extracted keywords and its synonyms.

In 2013, Tahri, Adel, and Okba Tibermacine [13] proposed a new architecture to develop a factoid question answering system based on the DBpedia ontology and the DBpedia extraction framework. The proposed system SELNI is sentence level question answering system which integrates natural language processing, ontologies, machine learning and information retrieval techniques. While designing the proposed system three steps were followed to build this system as the comprehension of the question, detection of its answer type, Question Processing, resources and keywords extraction to build SPARQL query and execute it by interrogating the DBpedia ontology.

In 2013, Raj [14] proposed an Architecture for ontology based natural language question. In this study concept of semantics and ontology were used to facilitate better query construction and extraction of answer. The proposed architecture consists of question processing, document extraction and processing and

finally answers processing. Further in the question processing module questions were analyzed using NLP techniques like POS tagger, Parser, NER. Relevant documents were retrieved from repository based on conceptual indexing and processed to extract candidate answer set. In answer processing module candidate answers were filtered and answer were generated.

In 2014, Gupta and Gupta [15] proposed Punjabi Question Answering system based algorithm to increase the accuracy of Question Answering Systems in terms of Recall and Precision. Factoid based Punjabi questions were considered. The proposed algorithm constructed pattern finding and matching system which identifies most accurate probable answer out of multiple answers. Also, the newly designed algorithm understood the Punjabi question and expresses them in query logic language. “W” word based questions were taken for the data set consideration. The proposed system was found useful in development of other NLP applications.

Sharma et al. [16] used Named Entity based n-gram approach to answers the Hindi language question answering system. The questions were classified and analyzed to generate the query. Classification of question were used to identify the relevant type of answers. Similarity metric was used to retrieve probability of the answers. Bigram and NER relevant answers were retrieved for the given question. Bigram method produces higher accuracy.

In this study Gupta [17] applied rule based suffix stripping technique for Hindi nouns and generated 16 noun suffixes. Further, corresponding stemming rules were developed for each suffix. While performing text mining or natural language processing operation viz; text Summarization, Machine Translation, Keywords Extraction, Topic Tracking, Documents Clustering and Classification etc steamer plays a firm role. The major objective of this study was to design suffix stripping rule based approach for performing stemming of nouns. Study reveals that the accuracy of stemmer can be improved by adding more suffixes and more stemming rules.

In the same year Banerjee et al. [18] designed a factoid question answering system for Bengali language. This article also discussed the challenge that occurs during the model development. Extraction and ranking method for relevant sentences has been proposed. Further, they also proposed the sentence ranking strategy for the BFQA system. To address the low-resource languages signatures architecture was proposed. The obtained result reveals that due to low accuracies of the shallow parser the proposed system resulted poor performance. Also, the performance of the NER system depend upon the accuracy of factoid QA system and the parser.

In 2015, Dhanjal and Sharma [19] highlighted the issues associated with multiple types of QA systems architecture. The study was categorized in seven different phases. The observation of this study reveals that traditional Q&A systems have some limitations while answering the current issues of question. To tackle this issue hybrid approach can be useful in development of Q&A system. Since there is limited work has been done on Indian languages due to complex nature of Indian languages, Natural language processing based architecture can play a firm role in designing and development of Q and A system.

In 2017, Govilkar and Bakal [20] proposed question answering system for Marathi natural language by using concept of ontology as a formal representation of knowledge base for extracting answers. The major objective of this study was to design, implement and experiment a new Marathi language QA framework based on ontologies where answers to the user’s questions are provided by using predefined domain specific ontology. Domain experts help to design ontologies and queries were analyzed both syntactically and semantically. The obtained results satisfy the query raised by the user with higher accuracy.

## **Conclusion**

In this study we have gone through various researchers work on QA with various languages and regional languages too. It is as observed that very few researchers work with Marathi language. Also, Marathi language based question answering not developed by the researchers yet. We can conclude that although very less number of language resources are existing for Indian languages but still lot of research and development is going on for them. On the other hand much of research is going on for developing question

answering systems for other languages in the world. In this study we explored various techniques and different methodology used by the researchers. We also observed that till now there is very less work done on why type of questions where answers are descriptive while development of factoid based question answering system are more. Also, the reviewed papers describes the different question answering approaches and different types of Q&A technologies like basic pattern matching, statistical analysis, artificial intelligence etc. For future studies automation based tools can be developed to minimize the manual intervention in QA process.

### Acknowledgment

We are thankful to Computational and Psycholinguistic Research Lab and Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (Maharashtra) for providing facility for carrying out the research.

### References

1. Marco Antonio Calijorne Soares, Fernando Silva Parreiras, "A literature review on question answering techniques, paradigms and system", Journal of King Saud University, Computer and Information Sciences journal, 12 August 2018
2. Poonam Gupta, Vishal Gupta, (2012), "A Survey of Text Question Answering Techniques", International Journal of Computer Applications (0975– 8887), Vol. 53– No.4
3. Muthukrishnan Ramprasath and Shanmugaundram Hariharan, "A survey on question Answering System," International J. Research and Reviews in Information Sciences, vol.2 no.1, March 2012.
4. Kumar, Praveen, et al. "A query answering system for E-learning Hindi documents." South Asian Language Review 13.1&2 (2003).
5. Lopez, Vanessa, Michele Pasin, and Enrico Motta. "Aqualog: An ontology-portable question answering system for the semantic web." The Semantic Web: Research and Applications. Springer, Berlin Heidelberg, 2005.
6. Wang, Chong, et al. "Panto: A portable natural language interface to ontologies." The Semantic Web: Research and Applications. Springer Berlin Heidelberg, 2007.
7. Guda, V., Sanampudi, S. K., & Manikyamba, I. L. (2011). Approaches for question answering systems. International Journal of Engineering science and technology (IJEST), 3(2), 990-995.
8. Chaware, S., and S. Rao. "Ontology supported inference system for Hindi and Marathi." Technology Enhanced Education (ICTEE), 2012 IEEE International Conference on IEEE. 2012.
9. Vasnik, N., Sahu, S. & Roy, D. (2012). Talash: A Semantic And Context Based Timized Hindi Search Engine. International Journal of Computer Science, Engineering and Information Technology (IJCEIT), 02, 71-79.
10. Sahu, Shriya, N. Vashnik, and Devshri Roy. "Prashnottar: A Hindi Question Answering System." International Journal of Computer Science and Information Technology (IJCSIT) 4.2(2012): 149 - 158.
11. Stalin, Shalini, Rajeev Pandey, and Raju Barskar. "Web based Application for Hindi Question Answering System." International Journal of Electronics and Computer Science Engineering (2012): 72-78.
12. V. Gupta, "A Proposed Online Approach of English and Punjabi Question Answering," International Journal of Engineering Trends and Technology (IJETT), vol. 6, 2013, pp. 292-295.
13. Tahri, Adel, and Okba Tibermacine. "DBpedia Based Factoid Question Answering System." International Journal of Web & Semantic Technology 4.3 2013.
14. Raj, P. C. "Architecture of an Ontology-Based Domain-Specific Natural Language Question Answering System." arXiv preprint arXiv: 1311.3175 (2013)
15. Gupta P., Gupta V. (2014) Hybrid Approach for Punjabi Question Answering System. In: Thampi S., Gelbukh A., Mukhopadhyay J. (eds) Advances in Signal Processing and Intelligent Recognition Systems. Advances in Intelligent Systems and Computing, vol 264. Springer, Cham. [https://doi.org/10.1007/978-3-319-04960-1\\_12](https://doi.org/10.1007/978-3-319-04960-1_12)
16. Sharma, Lokesh Kumar, and Namita Mittal. "Named Entity Based Answer Extraction form Hindi Text Corpus Using n-grams." (2014)
17. Gupta, V. (2014). Hindi rule based stemmer for nouns. International Journal of Advanced Research in Computer Science and Software Engineering, 4(1), 62-65.
18. Banerjee S., Naskar S.K., Bandyopadhyay S. (2014) BFQA: A Bengali Factoid Question Answering System. In: Sojka P., Horák A., Kopeček I., Pala K. (eds) Text, Speech and Dialogue. TSD 2014. Lecture Notes in Computer Science, vol 8655. Springer, Cham.
19. Dhanjal, Gursharan & Sharma, Sukhwinder. (2015). Advancements in Question Answering Systems Towards Indic Languages. International Journal of Research in Computer Science. 5. 15-26.
20. Govilkar, Sharvari & W, Bakal. (2017). Question Answering System Using Ontology in Marathi Language. International Journal of Artificial Intelligence & Applications. 8. 53-64. 10.5121/ijaiia.2017.8405.