

A Review on Question Answering Systems for Indian Languages



Bharat A. Shelke

Research Scholar,
Dept. of CS and IT,
Dr.B.A.M.University,
Aurangabad (MS), India

Ramesh Naik

Research Scholar,
Dept. of CS and IT,
Dr.B.A.M.University,
Aurangabad (MS), India

C. Namrata Mahender

Assistant Professor,
Dept. of CS and IT,
Dr.B.A.M.University,
Aurangabad (MS), India

Abstract

Question Answering is a rising domain of research in natural linguistic communication processing specially after the tremendous development in Internet technology. From a set of a Question Answering system (QAS) attempts to find out the correct resolution to the question pose in natural language. In this process user is assigned a particular answers in place of number of text documents or paragraphs. As compare to English and other foreign languages, very less work has done on Indian languages. This paper discusses a survey of various question answering systems based on Indian languages.

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

Introduction

Question Answering (QA) is the job of automatically answering a interrogative sentence posed in natural oral communication. QA is a specific type of information retrieval. Given a readiness of papers, an Interrogative Answering system attempts to find out the correct answer. Question answering is multidisciplinary. It involves information technology, artificial intelligence, natural language processing, knowledge and database management and cognitive science [1]. Most of the Question Answering system consist of three briny faculty s; question processing, document processing and answer processing. Question processing module plays an important part in QA systems. Answer processing module required to social status and validate candidate solvent. These proficiency aiming at discovering the short and precise answers are often based on the semantic classification. QAS gives the ability to answer questions posed in natural language by extracting, from a repository of documents, fragments of documents that contain material. [1]

Literature Survey on Question Answering Systems

Question answering systems can be categorized in many ways, such as by the application domain, answer roots, or target and source languages. But mainly it divided into two main region. i] Open Domain QA System, ii] Closed domain QA System. We are summarized the survey of various Question answering systems for Indian Languages in Table 1.

Table1. Literature Survey on Question Answering Systems.

Sr. No.	Research Work/ Language	Technique and Results	Year
1.	Sekine et al. [2] (Hindi)	A Hindi-English cross-lingual question-answering system, It uses English-Hindi bilingual dictionary for conversion. MRR= 0.25	2003
2.	P. Gupta et al [3] (Hindi)	Implemented the Hindi search engine for the retrieval of the relevant passages and accuracy is 75%	2005
3.	Reddy et al. [4] (Telugu)	A duologue based question answering system in Telugu language for railway specific domain and precision of 96.34 and dialogue success rate 83.96%	2006
4.	Pakray [5] (Bengali, Telugu)	Multilingual Restricted Domain Question Answering System with Dialogue Management (Bengali and Telugu) Dialogue Success Rate= 72.91% and Precision= 83.67% Telugu query System: Negotiation Success Rate= 89.06% 89.06% Precision= 96.49%.	2007

5.	Vishal Gupta et al [6] (Punjabi)	Analysis of output of Panjabi keyword extraction has been done over 50 Punjabi documents of Punjabi news corpus. The Precision, Recall and F-Grudge of Punjabi language keywords extraction are 80.4%, 90.6% and 85.2% respectively and 14.8% fourteen .8% of errors.	2011
6.	S. Sahu et al [7] (Hindi)	An approach for finding out the answers for the questions in the Hindi language. Overall accuracy appro. 68.00 %	2012
7.	N. Vasnik et al [8] (Hindi)	Combining the principles of search with HWN and context awareness, search results may be produced for a user and the precision of information retrieval increases.	
8.	Shalini Stalin et al [9] (Hindi)	A web based application for the extraction of answers for a question from Hindi text. Inconclusive results are shown.	
9.	Banerjee et al. [10] (Bengali)	Lexical, semantic and syntactic characteristic was given for the classification of the Bengali enquiry and used the Bengali shallow parser. Lexical, semantic and syntactic features accuracy= 87.63%.	2013
10.	Poonam Gupta et al [11] (Punjabi)	Discussed an algorithm for the implementation of Punjabi Question Answering System and it helpful for developing other NLP applications	
11.	Vishal Gupta [12] (Punjabi)	Punjabi language System can only analyses starting twenty online web pages with high scores than thousands of extracted web pages. Nouns, adjectives, verbs and adverbs etc. are treated as Key terms for the system.	
12.	Pallavi Bagul et al [13] (Marathi)	A rule based POS tagger is used to assigning the tags to the ambiguous words using Marathi grammar rules.	2014
13.	B. C. Melinamath [14] (Kannada)	The proposed rule based methodology for recognition of Kannada named entities has good recognition rate and precision around 86%..	
14.	Vishal Gupta [15] (Hindi)	Stemmer for Hindi noun is applying rule based suffix stripping technique. The accuracy of stemmer is 83.65%.	
15.	Garima Joshi et al [16] (Punjabi)	Hybrid approach using synset is used for stemming in Punjabi, table lookup has benefit of giving accurate results The performance of stemmer is directly dependent on number of entries in the root word table.	
16.	R. Thaker et al [17](Urdu)	A QURdPro system for query processing of Urdu language	2015
17.	S. Govilkar et al [18] (Marathi)	A Rule-based POS tagger is used to assigns all possible tags to the Marathi word.	
19.	Prachi Dalvi et al [19] (Marathi)	Ontology model for agriculture domain in Marathi speech. According to the farmer's query, the relevant will be extracted. The average F-scoth is 76.98%	2016
18.	Sharvari S. Govilkar et al[20] (Marathi)	Ontology is used to express domain specific knowledge about semantic relations and restrictions in the given domains. Factoid questions are considered. The system is tested with Marathi documents of various domains like History, sports, festival, politics.etc and shows an overall Precision 94.44%, Recall 98.07%, Accuracy 92.72%	2017

Aim of the Study

This research work is for the information retrieval from the resources such as Internet. The main objective of Question Answering system is to get back answers of questions rather than full document. Question Answering provides perfect solution to get valid and accurate answers to user question asked in natural language instead of query.

Types of QA Systems

Different types of QA system which are divided into two major radical s based on the

methods used by them. First group of QA system belongs to simple natural language processing and information retrieval methods, while another group of QA arrangement are dependent upon the reasoning with natural language. The two QA systems are compared with characteristics of different dimension such as technique used, question that deals with and so on. The table2 provides the detail of the comparisons of these QA systems. [21]

Table 2 Characterization of QA Systems [21]

Dimensions	QA system based on NLP and IR	QA systems Reasoning with NLP
Technique	Syntax processing, Named Entity tagging and Information Retrieval	Semantic Analysis or high reasoning
Data Resource	Free text documents	Knowledge Base
Domain	Domain Independent	Domain Oriented
Responses	Extracted Snippets	Synthesized Responses
Questions Deals with	Mostly wh- type of Questions	Beyond of wh- type of questions
Evaluations	Uses existing Information Retrieval	N/ A

Types of Question Generation

Question generation from text can be broadly divided into three categories: [22]

1. Syntax-based,
2. Semantics-based,
3. Template-based

Answer Extraction Methods

1. Pattern Matching Based Methods

Answer extraction patterns may consist of several types of units, such as punctuation marks, capitalization patterns, plain words, lemmas, part-of-speech (POS) tags, tags describing syntactic function, named entities (NEs) and temporal and numeric expressions. [23]

2. Semantic parser
3. Temporal context identifier
4. Logical prover

Logical prover has access to knowledge of five different types, Extended WordNet axioms, Ontological axioms, Linguistic axioms, Semantic calculus, Temporal reasoning axioms. [24]

Question Answering Approaches

According to our survey, Question answering has three approaches are present. 1) Linguistic approach, 2) Statistical approach, 3) Pattern bases approach.

Linguistic Approach

An enquiry answering logic contains artificial intelligence based methods that integrate Natural Language processing (Human language technology) technique and knowledge base. The knowledge entropy is organized in the form of production principle, logic frames, template, ontology and semantic study. It is used during the analytic thinking of QA pair. Parsing, Tokenization, and POS tagging are linguistic proficiency, it implemented to substance abuser question for formulating it into a precise query that specified extract the respective answer from structural database. Nowadays Marathi Shallow parser mostly carry to perform on Linguistic approach. Existing question answering [25], QA arrangement have acquired web as their knowledge resourcefulness. [26][27]

Statistical Approach

Availability of huge amount of data on internet increased the importance of statistical coming. A statistical learning method gives the better results than other approaches. One disadvantage of statistical approach is it treats each terminal figure independently and fails to identify linguistic characteristic for a combination words or phrase. Statistical techniques successfully applied to the different stages of the Question Answering system. The technique used for classification purpose is Maximum entropy models, manakin, support vector machine (SVM) classifiers, Bayesian classifiers. The important work based on the statistical method was IBM's statistical QA [28] system. It used maximum entropy model for question/answer based various N-gram features

Pattern Matching Approach

The pattern matching approach uses the expressive power of textbook convention. It replaces the sophisticated processing involved in other competing approaches. There are two approaches: Surface Pattern based and Template based. Most of the traffic pattern matching QA systems use the surface text patterns while some of them also rely on templates for response generation. [29]

Surface Pattern Based

Surface Pattern-based approach is automatically learned patterns through examples. Pattern learned by in semi-automatic and the most compatible application program area is small and spiritualist size website. [30]

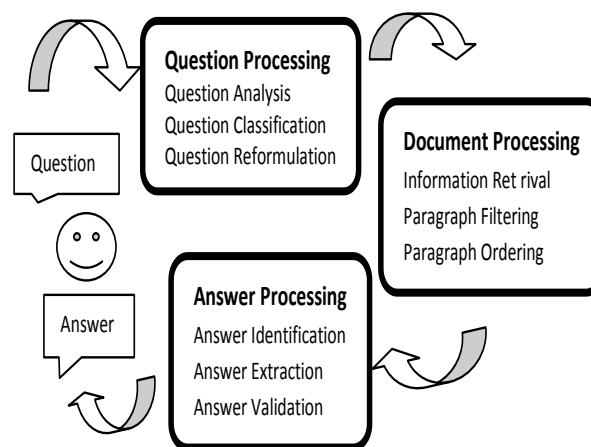
Template Based

This glide slope makes use of preformatted patterns for questions. The main stress of this approach is more on demonstration rather than explanation of questions and result. The template set is body-build in order to contain the optimum number of templates protect that it sufficiently cover the space of trouble , and each of its appendage represents a wide stove of questions of their own type. The entity one-armed bandit of Templates, which are missing elements bound to the concept of the question that has to be filled to generate the interrogation template to retrieve the corresponding reception from the database. The response returned by enquiry will be a raw data; it is going binding to the user. [31][32]

Architecture of Question Answering System

As shown in Figure 1, a typical Question Answering System (QAS) consists of three modules. Interrogation Processing Faculty , Document Processing Module and the Answer Processing Module and Question Categorization , Information Recovery , Answer Descent as their core ingredient s respectively.

Fig. 1.: Question Answering System Architecture [33]



Question processing module identifies the prime focus of the question, focusing of the interrogative sentence, and classifies the type of query. In English six eccentric of query are

present such as Who, what, why, when, Which, How much/How many. It also breakthrough the answer type expected, and then devises multiple semantically equivalent tokens from the question. [34]

Reformulation of a question into similar signification questions boost up the recall of the information retrieval (IR) system, organization, which is very important for question answering, because if no correct reply are present in a document, no further processing can be carry out to find the answer. [35]

The final examination component in a Question Answering system is the Answer Processing module. It is the distinguishing feature between Question Answering systems and the usual sense of text retrieval systems because it's engineering science is an influential and decisive on question answering system for the final termination. [33][34]

Question Answering for Marathi Language

In Republic of India after the development of TREC, various question answering system researcher moving their research from international language like English, Japanese, and Chinese to Indian regional languages like Telgu, Bengali, Urdu, Kannada, Marathi, and Hindi etc. But research in Marathi Linguistic process has not as equivalence to other languages. Some other techniques/ algorithm are also developed in another Indian languages [36] but on the Marathi Language very few researchers has worked with root, removing the stop tidings, rule based POS tagger to assigns all possible tags to the Marathi discussion. [13][18].

Conclusion

As compare to English and foreign languages very less research has been done in Indian regional languages. We found only question answering systems in Hindi, Punjabi, Telugu, Malayalam and Bengali. We have found less research on question answering system for Marathi language.

Acknowledgment

We are thankful to CSRI DST Major Project sanctioned No.SR/CSRI/71/2015(G), 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.

Endnotes

1. Poonam Gupta, Vishal Gupta, (2012), "A Survey of Text Question Answering Techniques", *International Journal of Computer Applications* (0975- 8887), Vol. 53- No.4.
2. S. Sekine and R. Grishman, (2003) "Hindi-English Cross-Lingual Question Answering System," *ACM Transaction on Asian Language Information Processing*, vol. 2, pp.181-192.
3. P. Gupta, S. Kashyap, A. Mittal and S. Gupta, (2005), "A Hindi Question Answering System for E-learning documents," *In Proceedings of IEEE International Conference on Intelligent sensing*

and Information processing, Bangalore, India, pp. 80-85

4. R. Reddy, N. Reddy and S. Bandyopadhyay, (2006), "Dialogue based Question Answering System in Teulgu," *In Proceedings of EACL Workshop on Multilingual Question Answering*, pp. 53-60.
5. P. Parkray, (2007), "Multilingual Restricted Domain QA System with Dialogue Management (Bengali and Telugu)," *Master's Thesis Report*, Jadavpur University, Kolkata.
6. S. Sahu, N. Vashnik and D. Roy, (2012), "Prashnottar: A Hindi Question Answering System", *International Journal of Computer Science and Information Technology (IJCSIT)*, vol.4, no.2, pp.149-158.
7. Vishal Gupta, Gurpreet Singh Lehal, (2011), "Automatic Keywords Extraction for Punjabi Language", *IJCSI International Journal of Computer Science Issues*, ISSN (Online): 1694-0814, Vol. 8, Issue 5, No 3, September.
8. Nandkishor Vasnik, Shriya Sahu, Devshri Roy, (2012), "TALASH: A Semantic and Context based optimized Hindi Search Engine", *International Journal of Computer Science, Engineering and Information Technology (IJCEIT)*, Vol.2, No.3.
9. S. Stalin, R. Pandey and R. Barskar, (2012), "Web based Application for Hindi Question Answering System," *International Journal Computer Science Engineering*, vol. 2, pp. 72-78.
10. S. Banerjee and S. Bandyopadhyay, (2012), "Bengali Question Classification: Towards Developing QA System," *In Proceedings of the 3rd Workshop on South and Southeast Asian Language Processing (SANLP), COLING*, pp. 25-40.
11. Poonam Gupta, Vishal Gupta, (2013), "Algorithm for Punjabi Question Answering System", *International Journal of Advanced Research in Computer Science and Software Engineering*, ISSN: 2277 128X, Volume 3, Issue 7.
12. Vishal Gupta, "Hindi Rule Based Stemmer for Nouns", (2014), *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 4, Issue 1, ISSN: 2277 128X
13. Pallavi Bagul, Archana Mishra, Prachi Mahajan, Medinee Kulkarni, Gauri Dhovavkar, (2014), "Rule Based POS Tagger for Marathi Text", *(IJCSIT) International Journal of Computer Science and Information Technologies*, Vol. 5 (2), 1322-1326.
14. B. C. Melinamath, (2014), "Rule based Methodology for Recognition of Kannada Named Entities", *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, ISSN: 2278-621X, Vol. 3 Issue 4.
15. Vishal Gupta, (2013), "A Proposed Online Approach of English and Punjabi Question Answering", *International Journal of Engineering Trends and Technology (IJETT) – Volume 6, Issue 5.*

16. Garima Joshi, Kamal Deep Garg, (2014), "Enhanced Version of Punjabi Stemmer Using Synset", *International Journal of Advanced Research in Computer Science and Software Engineering*, ISSN: 2277 128X, Volume 4, Issue 5.
17. Rukhsana Thaker, Dr.Ajay Goel, (2015), "QURDPro: Query processing system for Urdu Language", *International Journal of Engineering Research and Applications*, ISSN : 2248-9622, Vol. 5, Issue 6, (Part -3), pp.106-108.
18. Sharvari Govilkar, Bakal J. W, Shubhangi Rathod, (2015), "Part of Speech Tagger for Marathi Language", *International Journal of Computer Applications (0975 – 8887)*, Volume 119 – No.18.
19. Prachi Dalvi, Varsha Mandave, Madhu Gothkhindi, Ankita Patil, S. Kadam and Soudamini Pawar, (2016), "Ontology Extraction for Agriculture domain in Marathi language using NLP Techniques", *ICTACT Journal on Soft Computing*, ISSN: 2229-6959 (ONLINE), Volume: 07, Issue: 01
20. Sharvari S. Govilkar, J. W. Bakal, (2017), "Question Answering System using Ontology in Marathi Language", *International Journal of Artificial Intelligence and Applications (IJAIA)*, Vol.8, No.4.
21. Poonam Gupta, Vishal Gupta, (2012), "Survey of Text Question Answering Techniques", *International Journal of Computer Applications (0975 – 8887)*, Volume 53– No.4.
22. David Lindberg, Fred Popowich, John Nesbit, Phil Winne, (2013), "Generating Natural Language Questions to Support Learning On-Line", *Proceedings of the 14th European Workshop on Natural Language Generation*, pages 105–114, Sofia, Bulgaria.
23. Deepak Ravichandran and Edouard Hovy, (2002), "Learning surface text patterns for a question answering system", *In the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002)*, pages 41–47, Philadelphia, PA.
24. Ellen M. Voorhees and Hoa Trang Dang, (2005), "Overview of the TREC 2005 question answering track", *Proceedings of the Fourteenth Text REtrieval Conference (TREC 2005)*, Maryland, United States.
25. Katz B., (1997), *Annotating the World Wide Web using natural language*. In *Proceedings of the 5th RIAO conference on Computer Assisted Information Searching on the Internet*, pp. 136-159.
26. Chung H, Song YI, Han KS, Yoon DS, Lee JY, and Rim HC,(2004), *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), pp. 39-45.
27. Cai D, Dong Y, Lv D, Zhang G, Miao X, (2005), *A Web-based Chinese question answering with answer validation*. In *Proceedings of IEEE International Conference on Natural Language Processing and Knowledge Engineering*, pp. 499-502.
28. Ittycheriah A, Franz M, Zhu WJ, Ratnaparkhi A and Mammone RJ, (2000), *IBM's statistical question answering system*. In *Proceedings of the Text Retrieval Conference TREC-9*.
29. Saxena AK, Sambhu GV, Kaushik S, and Subramaniam LV, (2007), *litd-ibmirl system for question answering using pattern matching, semantic type and semantic category recognition*. In *Proceedings of the TREC 2007*.
30. Ravichandran D and Hovy E., (2002), *Learning surface text patterns for a question answering system*. In *proceeding of 40th Annual Meeting on Association of Computational Linguistics*, pp. 41-47.
31. Ajitkumar M. Pundge, Khillare S.A., C. Namrata Mahender, (2016), "Question Answering System, Approaches and Techniques: A Review", *International Journal of Computer Applications (0975 – 8887)*, Volume 141 – No.3.
32. Christina Unger, Lorenz Buhmann, Jens Lehmann, (2012), "Template-based question answering over RDF data", *WWW2012-session: Ontology Representation and Querying:RDF and SPARQL*.
33. Ali Mohamed Nabil Allam and Mohamed Hassan Haggag, (2012), "The Question Answering Systems: A Survey," *International Journal of Research and Reviews in Information Sciences (IJRRIS)*, vol.2, no.3.
34. Gursharan Singh Dhanja, Prof. Sukhwinder Sharma, (2015), "Advancements in Question Answering Systems towards Indic Languages", *International Journal of Research in Computer Science eISSN 2249-8265 Volume 5 Issue 1*, pp. 15-26
35. Svetlana Stoyanchev, Young Chol Song, and William Lahti, (2008), "Exact Phrases in Information Retrieval for Question Answering," in *2nd workshop on Information Retrieval for Question Answering (IR4QA)*, Manchester, UK, pp. 9-16.
36. Jaspreet Kaur, Vishal Gupta, (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.