

Question Answering System, Approaches and Techniques: A Review

Ajitkumar M. Pundge
Research Student
Dept. of Computer Science &
IT
Dr. B.A.M.U, Aurangabad

Khillare S.A.
Research Student
Dept. of Computer Science &
IT
Dr. B.A.M.U, Aurangabad

C. Namrata Mahender
Assistant Professor
Dept. of Computer Science &
IT
Dr. B.A.M.U., Aurangabad

ABSTRACT

As technology developed the use of internet has tremendously increased because of the availability of huge amount of data. Question answering is a specialized area in the field of information retrieval Text Processing. Question Answering system has many application based on source of answering like extracting information from document, language learning, online examination etc.

General Terms

Natural Language Processing, Question Answering System, Information Retrieval

Keywords

Question Answering system, Classification, Information retrieval, Answer extraction.

1. INTRODUCTION

From the last decade internet users are increasing in more percentage. 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. Lot of advancement is seen in QA related to the English, Chinese, Japanese, Korean, etc. languages. Question answering is a specialized area in the field of information retrieval. There are many question answering systems having its own application area. Question Answering System (QAS) has many application based on the source of answers. Like extracting information from document, language learning, online examination system, human and computer interaction, document management, classification of document and many more[1]. Question answering system can be divided into structured data and semi-structure data. The main objective of question answering system is to retrieve answers of questions rather than full document. There are two types of question answering system i.e. open domain and closed domain. Open domain system mainly based on web based where no restriction of any era while in closed based system having limited work domain e.g. medicine or weather forecasting and etc). This paper presents the study on question answering approaches and its techniques. After the introduction framework of Question Answering system has been taken and then details of approaches has presented...

2. FRAMEWORK of QAS

Question answering framework is built using Natural language processing and information retrieval techniques. The framework is divided into four modules namely. Question processing Module, Document Processing Module, Paragraph extraction module and answer extraction module [4].

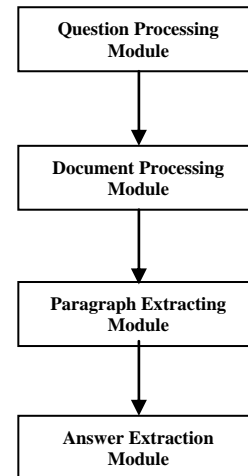


Fig 1: Framework of QA system

2.1 Question Processing Module

The question processing module converts natural language question queries for the document retriever. The process ranges simply returning the user's question as the query to employing question analysis to generate complex structured queries [4]. This module also detects the expected answer type of a question e.g. the expected answer type of "When was Shivaji born?" is date this information helps guide the answer extraction process.

2.2 Document Processing Module

This module retrieves documents from the corpus that are likely to contain answers to the user's question. It consists of a query generation algorithm and text search engine. The query generation algorithm takes an input the user's question and creates a query containing terms likely to appear in documents containing an answer. This query is passed to the text search engine, which uses it to retrieve a set of documents. [5]

2.3 Paragraph Extraction Module

Paragraph extraction algorithms take a document as a question and try to find passages from the document that contain an answer. Typical passage retrieval algorithms break the document into passages, compute a score for each passage and return the passage with the highest score. The system abstracts this mechanism so that passage tokenizing algorithms and passage scoring algorithms can be modularized as well the algorithm which cannot broken down can also support a large number of passage retrieval algorithms[4][5][6].

2.4 Answer extraction Module

The Modules takes as input a passage from the passage retrieval component and tries to retrieve an exact phrase to

return as an answer to achieve this required parsing and detailed question analysis by using of answer extraction algorithms. The identity answer extraction returns the center-point of the passage, stripping words from either end until it fits within the specified answer window.

There are many Approaches used in Question answering system based on different purpose namely linguistic-based approach, statistical-based approach and pattern matching approach. User needs precise and very specific answers. The large amount of data is continuously added in different scientific fields, in many disciplines. It's becomes challenging for researcher and many user to cope up with data. Understands the way a specific approach is supporting for full fledge development of QA system. Before start to understand the approaches used its better to first understand the major issues of QA system which makes it more difficult for progress. [4][5][6][7]

3. ISSUES OF QAS

There are various issues are present which plays vital role in question answering system such as Question Classes, Question Processing, Context and QA, Data Sources for QA, Answer Extraction, Answer Formulation, Real time Question Answering, Multilingual (or cross-lingual) question answering, Interactive QA, Advanced reasoning for QA, Information clustering for QA, User profiling for QA [8][9].

4. APPROACHES

4.1 Linguistic Approach

Linguistic approach understands natural language text, linguistic & common knowledge Linguistic techniques such as tokenization, POS tagging and parsing. These were implemented to user's question for formulating it into a precise query that merely extracts the respective response from the structured database [10].

Table 1 Linguistic based QA systems

QA system	Domain	Description
BASEBALL by Green et al[11]	Closed domain	Answering Question about Baseball game Front ends to databases
LUNAR Woods[12]	Closed Domain	Compare and evaluate the chemical analysis data on lunar rock and soil.
ELIZA Joseph Weizenbaum[13]	Closed Domain	Attempt to mimic basic human interaction Question and answer exchanges
GUS Bobrow et al.[14]	Closed Domain	A frame-driven dialog system Genial Under stander system also used structured database as the knowledge source
Clark et al[15]	Closed Domain	knowledge-base question answering ability through inference engine component

STARTQA System[16] Boris Katz.	Closed Domain	Web-Base QA system the system can answer millions of English questions about places (e.g., cities, countries, Etc)
Mishra et al[17]	Closed Domain	Web documents in the local knowledge database
Quarc Rilloff et al[18]	Closed Domain	Rule-base QA system for reading comprehension tests
Cquarc Hao et al [19]	Closed Domain	Rule-based Chinese QA system for reading Comprehension test

4.2 Statistical Approach

Availability of huge amount of data on internet increased the importance of statistical approaches. A statistical learning method gives the better results than other approaches. Online text repositories and statistical approaches are independent of structured query languages and can formulate queries in natural language form. Mostly all Statistical Based QA system applied a statistical technique in QA system such as Support vector machine classifier, Bayesian Classifiers, maximum entropy models [20] [21].

Table 2 Statistical approach based various QA systems and their technique

QA system	Domain	Technique	Description
Ittycheriah et al [22]	Open	Maximum Entropy Model	Maximum entropy model for question/ answer classification based on various N-gram or bag of words features.
Cai et al.[23]	Open	Sentence Similarity Model	Web-base Chinese QA system with answer validation
Soricut et al.[24]	Open	N-gram mining	used a statistical chunker questions into chunks/phrases asked to the search engine
Suzuki et al[25]	Open	Support Vector Machine	SVM answer selection for open domain QA system.
Rocchio Moschitti [26]	Open	Support Vector Machine text classifier	Question and answer categorization and tested his approach on Reuters-21578.

Zhang et al[27]	Open	Support Vector Machine based on the features of words	Chinese QA system with question classification and answer clustering
Quarteroni et al[28]	Open	SVM classifier for question classification	Designing an Interactive QA system for QA classification
Berger et al[29]	Open	N-gram Mining	Statistical approach to answer finding task in QA

4.3 Pattern Matching Approach

Pattern matching approach deals with expressive power of text pattern, it replace the sophisticated processing involved in other computing approaches. Most of the pattern matching QA systems uses the surface text pattern, while some of them also rely on templates for response generator[20][21].

Table 3 Pattern Matching Approach

Pattern Matching approach	Surface Pattern Based [31]	Finding answers to factual question answer are limited to one or two sentences
	Template based [32]	Used for Closed domain focuses on interpretation

4.3.1 Surface Pattern based

Surface Pattern-based approach is automatically learning-based pattern or it is human crafted. It replaces the sophisticated processing involved in other competing approaches [30] [33].

Table 4 Surface Pattern Based QA system

QA system	Descriptions
Hovy et al.[34]	Learning surface text patterns for QA system Implemented an automatic learning environment
Soubbtin et al[35]	Pattern of potential answer expression as clues to the right answers.
Zhang et al[36]	Web-base pattern mining and matching approach to question answering
Greenwood et al [37]	Using name entity tagger to generalize surface matching text pattern for question answering
Cui at al[38]	Soft pattern matching model for definitional QA system. Bigram model and PHMM
Saxena et al[39]	Using pattern matching semantic type and semantic category. For difficult question

4.3.2 Template Based Approach

This approach makes use of preformatted patters for questions. The Main focus of this approach is more on demonstration rather than explanation of question and answer [40][41].

Table 5 Template based QA systems

QA system	Technique	Description
Sneiders [42]	Frequently answer question (FAQ)	Automated QA using question template that cover the conceptual model of the database
Gunawerdena et al[43]	Pre-processed text to identify best matched template-answer	A closed domain system to understand SMS language
Unger et al[44]	Resource description framework (RDF)	Template based QA system over RDF data
SPARQL [45]	Resource description framework (RDF)	SPARQL query language for RDF

Table 6 Comparison of Approaches

Approaches	Questions types	Domain
Linguistics approach	Factoid	Closed
Statistical approach	Complex non-factoid	Open
Pattern Approach	Factoid, definition, acronym, date of birth	Closed

Table 7 QA System based on Hybrid Approach

QA system	Domain	Technique	Descriptions
Kwork et al MULDER [47]	Open	Based on integration of linguistic and statistical approach	Fully Automated General purpose QAS
Chakrabarti et al[48]	Open	Linguistic and pattern based	WorldNet structure to determine the answer type
Xia et al [49]	Closed	Rule-based & SVM classifier	An integrated approach for question classification in Chinese cuisine QA system
Lee YH et al	Open	Surface pattern & entropy	Complex question answering with

ASQA [50]		method	ASQA. Deal with Definition and relation questions
Ferrucci D et al IBM's WATSON [51] [2010]	Open	Surface pattern & entropy	A overview of DeepQA Project

5. CONCLUSION

The QA system can be Learning Companion it can help in our education system as in this global era need QA system which can solve closed domain problem and can give relevant answer to the question. It can be able to construct systems that could evaluate or grade answers with results consistent with human performance. New possibilities have come up in the education system which allows students to do self paced learning. Knowledge representation , precise representation for proper understanding , paraphrasing , conceptual learning , online accessing of descriptive Questions, Evaluation of answers, inclusion of figures , tables , mathematical equations are few major challenges which are in front of researcher. As the aspects involved are complex their solutions are even difficult but the application needs are high. Thus there is great potential for exploring the challenges in QA domain.

6. ACKNOWLEDGMENTS

Our thanks to my research guide Dr. C. Namrata Mahender for her support for this paper. We are thankful to the computational and psycho-linguistic Research Lab, Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Auranagabad (MS), for providing facility for carrying out the research.

7. REFERENCES

- [1] Chun-Chia Wang; Hung, J.C.; Che-Yu Yang; Shih, T.K., "An Application of Question Answering System for Collaborative Learning," in Distributed Computing Systems Workshops, 2006. ICDCS Workshops 2006. 26th IEEE International Conference on , vol., no., pp.49-49, 04-07 July 2006
- [2] Hirschman L., Gaizauskas R. Natural Language Engineering Cambridge University Press. (2001)
- [3] Ajitkumar Pundge, et al, "online learning different approaches and necessity for assessment", proceeding of the international conference on recent trends and challenges in science and Technology (RTCST2014) on 20-22 August 2014 at padmashri vikhe patil college of Arts, science and commerce, pravaranagar Ahmednagar (MS), India.
- [4] Stefanie Tellex, "Pauchok: A Modular Framework for question Answering", Master Thesis Submitted to the Department of Electrical Engineering and computer science, Maccachusetts institute of Technology, June 2003.
- [5] Haiqing Hu, "A study on Question Answering System Using Integrated retrieval method", Phd Thesis Submitted to Graduate school of engineering at the University of Tokushima, February, 2006
- [6] Yun Niu, "Analysis of Semantic classes: Toward Non-Factoid question answering", Phd Thesis submitted Graduate Department of Computer science, University of Toronto, 2007
- [7] Hakan Sundblad, "Question Classification in Question Answering systems", Phd Thesis Submitted to Department of Computer and information Science at Linkoping University, 2007
- [8] Biplab Ch. Das, "A Survey on Question Answering System", Department of Computer Science & Engineering, Indian Institute of Technology, Bombay.
- [9] John Burger et al, issues, task and program structure to roadmap research in question answering (Q&A). 1st January 2001
- [10] Unmesh sasikumar, Sindhu L, "A survey of Natural Language question answering system", international journal of computer applications (0975-8887), volume 108 -No 15. December 2014
- [11] Green BF, Wolf AK, Chomsky C, and Laughery K. Baseball: An automatic question answerer. In Proceedings of Western Computing Conference, Vol. 19, 1961, pp. 219–224.
- [12] Woods W. Progress in Natural Language Understanding - An Application to Lunar Geology. In Proceedings of AFIPS Conference, Vol. 42, 1973, pp. 441–450.
- [13] Weizenbaum J. ELIZA - a computer program for the study of natural language communication between man and machine. In Communications of the ACM, Vol. 9(1), 1966, pp. 36-45.
- [14] Bobrow DG, Kaplan RM, Kay M, Norman DA, Thompson H, and Winograd T. Gus, a frame-driven dialog system. Artificial Intelligence, Vol. 8(2), 1977, pp. 155-173.
- [15] Clark P, Thompson J, and Porter B. A knowledge-based approach to question answering. In Proceedings of AAAI'99 Fall Symposium on Question-Answering Systems, 1999, pp. 43-51.
- [16] <http://start.csail.mit.edu/index.php>
- [17] 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.
- [18] Riloff E and Thelen M. A Rule-based Question Answering System for Reading Comprehension Tests. In ANLP /NAACL Workshop on Reading Comprehension Tests as Evaluation for Computer-Based Language Understanding Systems, Vol. 6, 2000, pp. 13-19.
- [19] Hao X, Chang X, Liu K. A Rule-based Chinese question Answering System for reading Comprehension Tests. In 3rd International Conference on International Information hiding and Multimedia Signal Processing (IIH-MSP 2007), Vol. 2, 2007, pp.325-329.
- [20] Unmesh sasikumar, Sindhu L, "A survey of Natural Language question answering system", international journal of computer applications(0975-8887), volume 108 -No 15. December 2014

- [21] Sanjay K. Dwivedi and vaishali singh “Research and reviews in question answering system”, International Conference on Computational Intelligence: Modeling Techniques and Applications CIMTA) 2013 Procedia Technology 10 (2013) 417 – 424
- [22] Ittycheriah A, Franz M, Zhu WJ, Ratnaparkhi A and Mammone RJ. IBM’s statistical question answering system. In Proceedings of the Text Retrieval Conference TREC-9, 2000.
- [23] Cai D, Dong Y, Lv D, Zhang G, Miao X. 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, 2005.
- [24] Soricut R and Brill E. Automatic question answering using the web: Beyond the factoid. In Journal of Information Retrieval-Special Issue on Web Information Retrieval, Vol. 9(2), 2006, pp. 191-206.
- [25] Suzuki J, Sasaki Y, Maeda E. SVM answer selection for open domain question answering. In Proceedings of 19th International Conference on Computational linguistics, COLING’02, Vol. 1, 2002, pp. 1-7.
- [26] Moschitti A. Answer filtering via text categorization in question answering systems. In Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence, 2003, pp. 241-248.
- [27] Zhang K, Zhao J. A Chinese question answering system with question classification and answer clustering. In Proceedings of IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Vol.6, 2010, pp. 2692-2696.
- [28] Quarteroni S, and Manandhar S. Designing an interactive open-domain question answering system. Natural Language Engineering, Vol.15(1), 2009, pp. 73-95.
- [29] Berger A, Caruana R, Cohn D, Freitag D, and Mittal V. Bridging the lexical chasm: statistical approaches to answer-finding. In Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, 2000, pp. 192-199.
- [30] Ravichandran D and Hovy E. Learning surface text patterns for a question answering system. In proceeding of 40th Annual Meeting on Association of Computational Linguistics, 2002, pp. 41-47.
- [31] Deepak Ravichandran, Abraham Ittycheriah and salim roukos, “Automatic Derivation of surface text pattern for a maximum Entropy Based question answering system”, Work done while the author was an intern at IBM TJ Watson research center during summer 2002.
- [32] Vanitha guda et el, "Approaches for question answering systems", international Journal of Engineering science and technology (IJEST), ISSN:0975-5462, Vol.3 No.2 Feb 2011.
- [33] Chung H, Song YI, Han KS, Yoon DS, Lee JY, and Rim HC. A Practical QA System in Restricted Domains. In Workshop on Question
- [34] Answering in Restricted Domains. 42nd Annual Meeting of the Association for Computational Linguistics (ACL), 2004, pp. 39-45.
- [35] Ravichandran D and Hovy E. Learning surface text patterns for a question answering system. In proceeding of 40th Annual Meeting on Association of Computational Linguistics, 2002, pp. 41-47.
- [36] Soubotin MM and Soubotin SM. Patterns of Potential Answer Expressions as Clues to the Right Answer. In Proceeding of the TREC-10, NIST, 2001, pp. 175-182.
- [37] Zhang D and Lee WS. Web based pattern mining and matching approach to question answering. In Proceedings of the 11th Text Retrieval Conference, 2002.
- [38] Greenwood M. and Gaizauskas R. Using a Named Entity Tagger to Generalise Surface Matching Text Patterns for Question Answering. In Proceedings of the Workshop on Natural Language Processing for Question Answering (EACL03), 2003, pp. 29-34.
- [39] Cui H, Kan MY and Chua TS. Soft pattern matching models for definitional question answering. In ACM Transactions on Information Systems (TOIS), Vol. 25(2): 8, 2007.
- [40] Saxena AK, Sambhu GV, Kaushik S, and Subramaniam LV. Iitd-ibmrl system for question answering using pattern matching, semantic type and semantic category recognition. In Proceedings of the TREC, Vol. 2007, 2007.
- [41] Christina Unger, Lorenz Buhmann, Jens Lehmann, “Template-based question answering over RDF data”, WWW2012-session: Ontology Representation and Querying:RDF and SPARQL, April 16-20, 2012, lyon, France.
- [42] 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.
- [43] Sneider E. Automated question answering using question templates that cover the conceptual model of the database. In Natural Language Processing and Information Systems, Springer Berlin Heidelberg, 2002, pp. 235-239.
- [44] Gunawardena T, Lokuhetti M, Pathirana N, Ragel R, Deegalla S. An automatic answering system with template matching for natural language questions. In Proceedings of 5th IEEE International Conference on Information and Automation for Sustainability (ICIAFs), 2010, pp. 353-358.
- [45] Unger C, Böhmann L, Lehmann J, Ngonga Ngomo AC, Gerber D and Cimiano P. Template-based question answering over RDF data. In Proceedings of the ACM 21st international conference on World Wide Web, 2012, pp. 639-648.
- [46] Prud’hommeaux E, Seaborne A(eds.). SPARQL Query Language for RDF. <http://www.w3.org/TR/rdf-sparql-query/>, 2007.
- [47] Hao X, Chang X, Liu K. A Rule-based Chinese question Answering System for reading Comprehension Tests. In 3rd International Conference on International Information hiding and and Multimedia Signal Processing (IIH-MSP 2007), Vol. 2, 2007, pp.325-329.

- [48] Kwok C, Etzioni O and Weld DS. Scaling question answering to the Web. *ACM Transactions on Information Systems (TOIS)*, Vol.19 (3), 2001, pp. 242-262.
- [49] Ramakrishnan G, Chakrabarti S, Paranjpe D and Bhattacharya P. Is question answering an acquired skill?. In *Proceedings of the 13th ACM international conference on World Wide Web*, 2004, pp. 111-120.
- [50] Xia L, Teng Z, and Ren F. An Integrated Approach for Question Classification in Chinese Cuisine Question Answering System. In *IEEE second International Symposium on Universal Communication*, 2008, pp. 317-321.
- [51] Lee YH, Lee CW, Sung CL, Tzou MT, Wang CC, Liu SH, Shih CW, Yang PY and Hsu WL. Complex question answering with ASQA at NTCIR-7 ACLIA. In *Proceedings of NTCIR-7 Workshop Meetings, Entropy*, 1, 10, 2008.
- [52] Ferrucci D, Brown E, Chu-Carroll J, Fan J, Gondek D, Kalyanpur AA, Lally A et al. Building Watson: An overview of the DeepQA project. *AI magazine* 31, no. 3, 2010, pp. 59-79.
- [53] Khillare S. A., Dhokrat A. V. and Mahender C.N., Analysis Of Question Answering And Parsing Techniques in Natural Language Processing : A Review, *Advances in Computational Research*, BIOINFO Publications, ISSN: 0975-3273 & E-ISSN: 0975-9085, Volume 7, Issue 1, 2015.