IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 08, Issue 4 (April. 2018), ||VII|| PP 18-21

A Comprehensive Review on Question Answering Systems

Anitha Mary M O Chacko[1]
[1](*Department of Computer Applications, Mar Athanasios College for Advanced Studies Tiruvalla, Kerala, India)*
Corresponding Author: Anitha Mary M O Chacko

Abstract: Question Answering (QA) is one of the most challenging tasks in the domain of Natural Language Processing and Information Retrieval. A Question answering systems (QASs) attempts to generate concise answers to user's questions asked in natural languages. Early QASs were developed for restricted domains and have limited capabilities. Current QASs focus on open domain QA which has no domain restrictions, and can almost all types of questions. This survey classifies QAs on the basis of question types and focuses on the techniques used for extracting answers from each type which can aid future researchers.

Date of Submission: 30-03-2018 Date of acceptance: 16-04-2018

I. INTRODUCTION

Information retrieval aims to obtain specific information from a collection of information resources. The present Web Search engines present a ranked list of relevant documents in response to users' queries based on various aspects such as popularity measures, keyword matching etc. However, they do not truly accomplish the task of information retrieval as users have to search each document manually for getting the desired information. An ideal search engine should return few relevant and concise sentences as answers along with their corresponding web links.

Closed domain QASs answer domain specific questions by searching domain specific document repositories. Here the qualities of answers are high as the search space is less and are beneficial to domain expert users. Open Domain QASs on the other hand, answers domain independent questions by exploiting general web ontology. These systems are beneficial to casual users but the quality of answers may be lower than Closed domain QASs.

Over the past years, works in question answering domain has been encouraged by the inclusion of a question answering task in the experimental work of the TREC conference series, which is sponsored annually by the U.S. National Institute of Standards and Technology (NIST).Competitions like the Text REtrieval Conference (TREC) (http://trec.nist.gov) or the Cross Language Evaluation Forum (CLEF) (http://www.clefcampaign.org/) have been created in order to develop and improve existing systems and to evaluate and compare their behavior. These competitions encourages active research in this domain.

This survey classifies QAs on the basis of question types and focuses on the present techniques used to answer each question type. The paper is structures as follows. Section II present the general architecture of a QAS. Section III presents the classification of QASs into different categories on the basis of question types. Finally, Section IV concludes the paper.

II. GENERAL QUESTION ANSWERING SYSTEM ARCHITECTURE

A Question answering system consists of 3 phases: Question Processing, Document Processing and Answer Processing. Fig1. presents the architecture of a general QAS.

1. Question Processing

The question processing module determines the entity answer type based on the category of questions. For eg, a question like 'Who wrote the book XX' expects person name as answer type whereas a question type like 'Where is Y' expects a location as answer. This type of question analysis can help in answering factoid/list type questions. This phase then involves identification of focus and target in the questions. The focus is the property or entity sought by the question, whereas the target is the event or object the question is about. Finally the questions are reformulated in this phase by eliminating stop words, performing query expansion and semantic reformulation. There are three main approaches for making analysis of natural language questions and source documents: Statistical based approach, Rule based pattern matching approach and hybrid approach.

2. Document Processing

This phase involves collecting relevant documents related to the question keywords from the web ontology/IR systems and extracting candidate passages from the documents which can contain answers to the questions. Some of the main passage retrieval methods include cosine similarity, tf-idf score, density based passage retrieval, N-gram approach etc.

3. Answer Processing

The answer processing module is responsible for identification, extraction and validation of answers. The candidate answers are identified from the filtered passages. This is done by deriving the answer type from the question type. Techniques such as Parts of Speech Tagging, Named Entity Recognition etc can help in identification of answer candidates. Once an answer candidate has been identified, heuristic rules can be applied in order to extract only the relevant word or phrase that answers the question. The final answers can be validated by using lexical resources and semantic relationships.

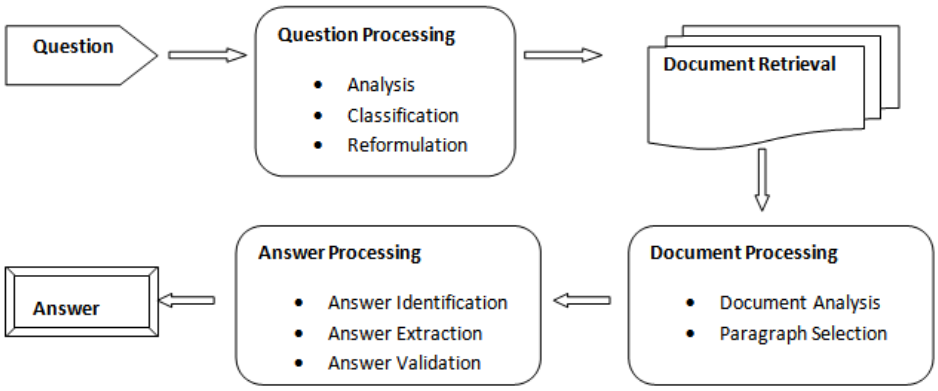

Fig1. A General Question Answering System

III. RELATED WORKS

Before the emergence of web, Question answering systems were mostly restricted to a closed domain. The first question answering system was BASEBALL (Green et al.,) [1]. This QA system answers questions about the baseball games played in American league during a particular season such as dates, location etc by using a simple database-centered system that translates a natural language question to a canonical query on the database. LUNAR (Woods et al.) [2] was another closed domain QAS that provides information to a lunar geologist about soil samples taken from Apollo lunar exploration by transforming users' questions into database queries through pattern matching rules. MASQUE (Androutsopoulos et al.) [3] makes linguistic analysis of the questions and represents these in a logic representation, and then transforms the logic query into a database query. QUARC QAS (Riloff and Thelen , 2000) [4] classifies questions into different wh-types and forms their expected answer types by lexical and semantic rules. With the emergence of World Wide Web, the focus shifted to open domain QASs.

IV. CLASSIFICATION OF QASS BASED ON QUESTION TYPES

The task of answer extraction to users' questions is dependent on the type of question asked. The different categories of questions are:

Factoid Questions

These are simple wh-questions (what, where, when, which) that expects a precise answer, usually a single word or a single sentence. For eg, 'Who is the first president of India?. These type of questions are relatively easier to handle and doesn't require complex natural language processing tasks.

A web based QAS, NIST was developed by (Radev et al.)[5] that utilizes existing web search engines to retrieve related documents on the web and processes the top ranked documents to extract a number of potential answers. They made a flat taxonomy with seventeen classes using decision rule induction and a heuristic rule-based algorithm. LASSO (Moldovan et.al,)[6] used a hierarchical taxonomy by classifying the question into nine classes and extracts answers by a combination of syntactic and semantic techniques, and lightweight empirical abductive inference.

MULDER QAS (Kwok et.al)[7] combines information retrieval ideas with statistical natural language processing, lexical analysis, query formulation, answer extraction, and voting to generate answer snippets to

user queries. MULDER parses the question using a natural language parser that constructs a tree of the question's phrasal structure which is given to a classifier to determine the answer type and a query formulator uses this parse tree to transform the questions into search engine queries. AskMSR QAS (Brill et.al) [8] integrates query-reformulation, n-gram mining, filtering, and n-gram tiling approaches to generate answers to user's questions. This system generates a number of rewrite strings, which are likely sub-strings of answers to the question.

Ravichandran et al.[9] presented a QAS system for finding answers by formulating regular expressions by exploiting surface text information using manually constructed surface patterns. QASYO (Abdullah et.al)[10] uses a semantic search technique that uses YAGO ontology and integrates NLP, ontologies and IR. AqualLog QAS (Lopez et. al.)[11] was an ontology-based QA that uses GATE processing resources for English tokenizer, sentence splitter, POS tagger, VP chunker and a regular expression based rule language for finding candidate answers.

List Questions

The list type questions require a list of entities or facts as answers e.g., – List the seven wonders of the world?. These questions allow multiple answers whose scores are above a cut-off threshold. These QAS do not need deep natural language processing tasks to retrieve answers.

FADA(Yang et. al)[12] is a web based list question answering system that uses question parsing, web page classification/clustering, and content extraction to find relevant answers. Another approach based on pattern-based method for answering list questions was proposed by Greenwood et.al[13] that uses named entity recognition and noun chunking for learning answer patterns. Wang et. al [14] uses set expansion (SE) algorithms that mine textual resources to produce an extended list including additional members of the class represented by the seeds.

Causal Questions

Causal questions(why, how) require reasons, explanations or elaborations about an entity. For eg, why the sky is blue?. The answers are not named entities as in the case of factoid type questions. These type of QASs require advance natural language processing techniques to generate answers.

Higashinaka et al. [15] proposed a corpus-based approach for answering why-questions where causal expressions are automatically collected from corpora tagged with semantic relations and features are created from the collected expressions to train an answer candidate ranker that maximizes the QA performance. Pechsiri et.al [16] developed an automatic causal Question Answering system, on community web-boards to support ordinary people in preliminary diagnosis and problem solving, such as plant disease problems. This system applies machine learning techniques for question type identification and uses an integrated causality graph with extracted procedural knowledge from text to determine the visualized answers.

Confirmation Questions

These questions require answers in the form of Yes or No. For eg. 'Is Java an Object Oriented Language?'. Inference mechanism, world knowledge and common sense reasoning are required to answer these type of questions.

Kanayama et. al [17] uses question inversion that involves generation of factoid questions by replacing some parts of a given proposition with abstract expressions. The key method used here involves conversion from yes/no questions into factoid-style questions, and aggregation of the results of factoid answers to determine yes or no. Taniguichi et.al [18] developed a yes/no question answering system for answering questions about a statute legal domain using case-role analysis, in order to find correspondences of roles and relationships between given problem sentences and knowledge source sentences.

Hypothetical Questions

Hypothetical questions ask for answers related to any hypothetical event and generally begin with 'what would happen if'. There are no specific correct answers of these questions. Obeid et. al [19] describes a computational model which provides answers to modal and hypothetical questions in knowledge base systems (KBSs) using semantic theories on modality and question answering (which have been mainly addressed by linguists and semanticists), intensionality, partiality and belief revision.

V. CONCLUSION

Question answering systems (QASs) attempts to generate concise answers to user's questions asked in natural languages. Early QASs were developed for restricted domains and have limited capabilities. Current QASs focus on open domain QA which has no domain restrictions, and can almost all types of questions. This paper presented a comprehensive review on the current status of Question Answering systems. Further, QASs

were classified on the basis of question types into 5 categories: Factoid questions, List questions, Causal questions, Confirmation questions and Hypothetical questions. This paper also analyzed and presented the key works in each of these categories which can aid future researchers.

REFERENCES

[1]. B.F. Green, A.K. Wolf, C. Chomsky, and K. Laughery ,Baseball: An automatic question answerer, In Proceedings Western Computing Conference, 1961.

[2]. Woods W. , Progress in natural language understanding an application to lunar geology, American Federation of Information Processing Societies (AFIPS) Conference Proceedings,1973.

[3]. Androutsopoulos, I., Ritchie, G.D., Thanisch, P., 1993. MASQUE/SQL – an efficient and portable natural language query interface for relational databases. In: Proc. of the 6th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems. Gordon and Breach Publisher Inc., Edinburgh, pp. 327–330.

[4]. Riloff, E., Thelen, M., 2000. 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, pp. 13–19.

[5]. Dragomir Radev, Weiguo Fan, Hong Qi, Harris Wu, Amardeep Grewal, " Probabilistic Question Answering on the Web",Proceedings of the 11th international conference on World Wide Web, ACM 2002

[6]. Moldovan D., "Lasso: A Tool for Sur_ng the Answer Net" In Proceedings of the Eighth Text Retrieval Conference (TREC-8).

[7]. Kwok C., Etzioni O., and Weld D. S., " Scaling question answering to the web ", In the Proceedings of the 10th World Wide Web Conference (WWW 2001), Hong Kong.

[8]. Michele Banko, Eric Brill, Susan Dumais, Jimmy Lin , "AskMSR: Question Answering Using the Worldwide Web",In Proceedings of AAAI Spring Symposium on Mining Answers from Texts and Knowledge Bases, March 2002, Palo Alto, California

[9]. Deepak Ravichandran, Eduard Hovy, " Learning Surface Text Patterns for a Question Answering System ", In Proceedings of the ACL Conference, 2002.

[10]. Abdullah M. Moussa and Rehab F. Abdel-Kader, "QASYO: A Question Answering System for YAGO Ontology", in International Journal of Database Theory and Application , Vol. 4, No. 2, June, 2011

[11]. Lopez V., Uren V., Motta E. and Pasin M., "AquaLog: An Ontology-driven Question Answering System for Organizational Semantic Intranets", An International Journal of Web Semantics: Science, Services and Agents on the World Wide Web 5, pp. 72-105, 2007

[12]. Hui Yang, Tat-Seng Chua, Web-Based List Question Answering, In Proceedings of the Twelth Text Retrieval Conference (TREC-12)

[13]. Mark A. Greenwood and Horacio Saggion, A Pattern Based Approach to Answering Factoid, List and Definition Questions

[14]. Richard C. Wang, Nico Schlaefer, William W. Cohen, and Eric Nyberg, Automatic Set Expansion for List Question Answering, In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, pages 947–954, Honolulu, October 2008.

[15]. Higashinaka, R., Isozaki, H., 2008. Corpus-based question answering for why-questions. In: Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), pp. 418–425.

[16]. C.Pechsiri" R.Piriyaku, Developing a Why–How Question Answering system on community web boards with a causality graph including procedural knowledge, Information Processing in Agriculture Volume 3, Issue 1, March 2016, Pages 36-53

[17]. Hiroshi Kanayama, Yusuke Miyao, John P rage, Answering Yes/No Questions via Question Inversion, Proceedings of COLING 2012: Technical Papers, pages 1377–1392, COLING 2012, Mumbai, December 2012.

[18]. Ryosuke Taniguchi, Yoshinobu Kano, Legal Yes/No Question Answering System Using Case-Role Analysis, JSAI International Symposium on Artificial Intelligence, Lecture Notes in Computer Science, Springer

[19]. Nadim Obeid, Answers to modal and hypothetical questions in knowledge base systems, In Applied Intelligence, December 1992, Volume 2, Issue 4, pp 353–367.

Anitha Mary M O Chacko "A Comprehensive Review on Question Answering Systems." IOSR Journal of Engineering (IOSRJEN), vol. 08, no.4, 2018, pp. 18-21.