International Journal of Scientific Research & Engineering Trends
Volume 6, Issue 4, July-Aug-2020, ISSN (Online): 2395-566X

Research and reviews in Question Answering System

Ragini Patil, Asst. Prof. Shailendra Tiwari
Department of Computer Science & Engineering
Astral Institute of Technology & Research Indore, MP, India
raginipatil174@gmail.com, Shailendra.tiwari4@gmail.com

***Abstract*-Question Answering (QA) Systems is an automated approach to retrieve correct responses to the questions asked by human in natural language. The fundamental thought behind QA system is to assist man-machine interaction. In this paper, we propose taxonomy for characterizing Question Answer (QA) systems, briefly survey major QA systems described in literature and provide a qualitative analysis of them. Finally, a comparison between these approaches based on certain features of QA system found critical in our study has been done, in order to bring an insight to research scope in this direction.**

***Keywords*-QA sytem; linguistic approach; statistical approach; pattern-based approach; templates;**

I.INTRODUCTION

Modern information retrieval systems allow us to locate documents that might have the associated information, but the majority of them leave it to the user to extract the useful information from an ordered list. For e.g., the question "Who has won the maximum individual medals in Olympics 2012?" should get back the response "Michael Phelps" but instead the user is presented with a list of relevant documents to explore in the quest of an accurate answer. QA is recognized as a capability with great potential. QA system enables users to access the knowledge resources in a natural way (i.e., by asking questions) and to get back a relevant and proper response in concise words.

Though automatic question answering will definitely be a significant advance in the state-of-art information retrieval technology in forthcoming years but still there are many challenging issues that are yet to be resolved. One of the challenging tasks for existing QA systems is to understand the natural language questions correctly and deduce the precise meaning to retrieve exact responses. Improvement in mechanized understanding of questions faces issues like question classification, formulation of right queries, ambiguity resolution, semantic symmetry detection, identification of temporal relationship in complex questions. In the similar way identification of a perfect answer requires proper validation mechanism.

The processing of a QA system may broadly have three stages, i.e., question analysis: parsing, question classification and query reformulation; document analysis: extract candidate documents, identify answers; and answer analysis: extract candidate answers and rank the best one. Question answering task combines techniques from artificial intelligence, natural language processing, statistical analysis, pattern matching, information retrieval, and information extraction. Most of the recent works integrate some or all of these approaches to built enhanced systems that can deal with scarcity of these approaches. Moreover, the taxonomy presented here in this paper for categorizing various QA systems developed so far is, based on the overall approach followed by each system in their different processing phases, that lead us to following categorization: Linguistic approach, Statistical approach and Pattern Matching approach. This paper is organized as follows. In Section 2, we discuss classification for characterizing QA approaches. Following in next Section, we discuss pros and cons of each approach and present a comparative study of these approaches while final section presents conclusions.

II. APPROACHES

1. Linguistic approach

A question answering system requires understanding of natural language text, linguistics and common knowledge. Therefore, many of the previous researchers relied on artificial intelligence (AI) based methods that integrate natural language processing (NLP) techniques and knowledge base or corpus to build QA logics. The knowledge information is organized in the form of production rules, logics, frames, templates (represented with triple relations), ontologies and semantic networks, which are utilized during analysis of question -answer pair. Linguistic techniques such as tokenization, POS tagging and parsing were implemented to user's question for formulating it into a precise query that merely extracts the respective response from the structured database. However, deployment of a specific domain knowledge base poses portability limitation as a different application domain requires different grammar and mapping rules. Additionally, building an appropriate knowledge base is a time-consuming process, so these systems are generally applied to problems that have long-term information needs for a particular domain.

Earlier QA systems around 1960s such as BASEBALL [1] and LUNAR [3] were merely natural language front-ends for structured database query systems. The questions

presented to these systems were usually analysed using NLP techniques to produce a canonical form, which was then used to construct a standard database query. Dialogue system viz., ELIZA [2] and GUS [4] also used structured database as the knowledge source. The key limitation of these systems is that the knowledge stored in the structured database was only capable of answering questions asked within the restricted domain.

However, in recent works, this limitation of the knowledge base is accepted as the capability to provide a situation-specific answer. Clark et al. [7] presented an approach for augmenting online text (dynamic manual) with knowledge-base question answering ability. This combined approach allows users to access not only response of routine questions but also of those questions that were unforeseen at the time of system construction. This specific feature of QA system is achieved through inference engine component.

Some of the existing QA systems such as START [5], QA system by Chung et al. [18] and Mishra et al. [30] have acquired web as their knowledge resource. These systems apply their own heuristics to store information from web documents in the local knowledge database that has to be later on accessed and rely on linguistic techniques for answer generation.

Table 1 Comparison between NLP based and rule-based approaches.

	NLP-based	Rule based
Source organization	Structured Knowledgebase	Structured or semi-structured documents but generally developed for Comprehension (text) based systems.
Question class interpretation	Not required	Annotated heuristic rules to identify classes.
Linguistic analysis	Deep NLP Techniques	Shallow NLP techniques
Knowledge or domain	spanOften small but sometimes broader when web is used as knowledge resource.	Limited to pre-stored comprehension or documents
Learning data	Not required	Require to learn rules from training data.

In addition, with NLP techniques, some of the knowledge base QA systems rely on the rule-based mechanism. After applying general purpose NLP techniques, rules are further built to identify question classification features. Quarc [8] developed by Rilloff et al., and Cqarc [23] developed by Hao et al., used heuristic rules that look for lexical and semantic clues in question to identify the question class. However, question class taxonomy may vary from one system to another. Some systems exploit general taxonomy for semantic classes like who, when, what, where and why type questions while some others utilize domain specific taxonomy. Table 1.1 shows the difference between NLP based and rule-based QA systems.

2. Statistical approach

In the current research scenario, rapid growth in available online text repositories and web data has increased the importance of statistical approaches. These approaches put forward such techniques, which cannot only deal with the very large amount of data but their heterogeneity as well. Additionally, statistical approaches are also independent of structured query languages and can formulate queries in natural language form. These approaches basically require an adequate amount of data for precise statistical learning but once properly learned, produce better results than other competing approaches. Furthermore, the learned statistical program or method can be easily customized to a new domain being independent of any language form. However, one of the major drawbacks of statistical approaches is that they treat each term independently and fail to identify linguistic features for combination of words or phrases.

In general, statistical techniques have been so far successfully applied to the different stages of a QA system. Support vector machine (SVM) classifiers, Bayesian classifiers, Maximum entropy models are some techniques that have been used for question classification purpose. These statistical measures analyze questions for making prediction about users' expected answer type. These models are trained on a corpus of questions or documents that has been annotated with the particular mentioned categories in the system.

One of the pioneer works based on the statistical model was IBM's statistical QA [9] system. This system utilized maximum entropy model for question/ answer classification based on various N-gram or bag of words features. Moschitti [17] had used Rocchio and SVM text classifiers for question and answer categorization and tested his approach on Reuters-21578. A Chinese question answering system developed by Zhang et al. [32] has also used SVM classifier based on the features of words, part of speech (POS), named entity and semantics. Quarteroni et al. [28] proposed an interactive QA system which implements SVM classifiers for question classification.

Table 2 Statistical models applied for question classification.

Statistical techniques/models	Major QA system(s)	Performance
Maximum entropy model	IBM's statistical QA [9]	Question classification error rate significantly reduces.
Support Vector Machine	VectorSystem by Moschitti [17],SVM has shown quite good performance and accuracy for question classification and preferred in QA community.	Quarteroni et al. [29], Zhang et al.[32]
Modified Bayesian Classifier	Wei et al.[36], MKQA[35]	Method has better accuracy than base Bayesian method.

Berger et al. [10] has investigated the prospects of applying statistical methods to answer finding task in QA and discovered that these techniques performed quite well depending on the characteristics of the underlying data set– vocabulary size, the overlap between question and answers, and between multiple answers, etc. Statistical techniques such as N-gram mining, sentence similarity models and Okapi similarity measurement are applied to answer finding tasks in a QA system. These techniques analyze question and document based on various similarity features in order to determine the closeness of candidate documents or answers with respect to question. The notion of answer validation could also be implemented incorporating statistical approaches via relevance feedback mechanism.

Information retrieval for this (IBM's QA) system used a two pass approach based on Okapi formula and expansion of queries based on TREC-9 QA corpus. Answer selection phase of this system relied on various heuristic distance metrics to search for an answer. Moschitti [17] implemented similarity measurement model for calculating the similarity score between query and documents or sentences from corresponding collections. The similarity model presented by Cai et al. [20], relied on a sentence similarity model to calculate the similarity between question and answer. This model accounted on different features such as keyword similarity, length similarity, order similarity and distance similarity of the keywords used in question and answer. A system developed by Soricut et al. used a statistical chunker that implements a dynamic programming algorithm to chunk the natural language questions into chunks/phrases asked to the search engine and N-gram co-occurrence statistics for answer extraction. This system could answer complex questions and non-factoid questions too.

Table 3. Statistical models applied for answer finding

Statistical techniques/models	Major QA system(s)	
Okapi Similarity Measurement	IBM's statistical QA [9]	This metric has so far achieved an average level of mean reciprocal ratio for factoid questions in restricted domain.
Sentence Similarity Model	System by Cai et al. [20]	Answer selection accuracy is significantly increased and above average precision is achieved.
N-gram mining	System by Soricut et al. [22]	Satisfactory performance even for the non-factoid questions.
SVM	System by Suzuki et al.[37]	SVM outperforms than other models in answer selection phase too.

3. Pattern matching approach

This approach uses the expressive power of text patterns to replace the sophisticated processing involved in other competing approaches. For example, the question **"Where was Cricket World Cup 2012 held?"** follows the pattern **"Where was <Event Name> held?" and its answer pattern will be alike "<Event Name> was held at <Location>"**. Currently, many of the QA systems automatically learn such text patterns from text passages rather than employing complicated linguistic knowledge or tools viz., parser, named-entity recognizer, ontology, WordNet, etc. to text for retrieving answers. Simplicity of such systems makes it quite favourable for small and medium -size websites, which cannot afford complex solutions that require much time and rare human skills to install and maintain the system. Most of the patterns matching QA systems use the surface text patterns while some of them also rely on templates for response generation.

3.1. Surface Pattern based

This approach extracts answers from the surface structure of the retrieved documents by relying on an extensive list of patterns. Answer to a question is identified on the basis of similarity between their reflecting patterns having certain semantics. These patterns are like regular expressions. Though designing such set of patterns requires a lot of human skill and time but the approach has shown high precision too.

Initially, the surface pattern-based method aims at finding answers to factual questions, as their answer is limited to one or two sentences. In order to design an optimal set of patterns, most of the recent surface pattern-based system used method described by Hovy et al. [13]. They implemented an automatic learning method which used bootstrapping to built a large set of patterns starting only with a few examples of QA pair from the web. Motivation behind their work is surprising strength of such patterns proposed by Soubbtin and Soubbtin [11] in the TREC- 10 Question answering evaluation track.

Another related concept based on surface patterns is proposed by Zhang et al. [14] who augmented surface patterns with 'support' and 'confidence' measures from data mining community. This system showed very high precision but low recall. Greenwood et al. [16] integrated surface patterns with named entity tagger to generalize these patterns induced from free text. System developed by Cui et al. [25] used soft pattern matching based on bigram model and Profile Hidden Markov Model (PHMM) instead of regular expression-based hard matching patterns to identify answer sentences. Some other QA systems have also followed this approach to improve their question answering mechanism. Saxena et al. [24] used pattern matching as an alternative approach for difficult questions like acronym expansion questions, date of birth questions and location questions.

Table 4 Comparison between two pattern matching approaches.

	Surface pattern based	Template based
Basic mechanism	Either human crafted or automatically learned patterns through examples.	A template is preformatted framework for questions which have entity slots to be dynamically filled by parameters.
Answer Extraction	Answer sentence is extracted using statistical techniques or data mining measures.	Uses structured query to extract answer from database.
Answer Representation	Not necessarily formatted answers.	Focus on generating formatted answers.
Pattern Learning	Semi-automatic.	Manually but automatic for semantic web.
Most compatible application area	Small and medium size websites.	Semantic web.

3.2. Template based

A template-based approach makes use of preformatted patterns for questions. The focus of this approach is more on illustration rather than interpretation of questions and answers. The set for templates is built in order to contain the optimum number of templates ensuring that it adequately cover the space of problem, and each of its members represents a wide range of questions of their own type. Templates have entity slots, which are missing elements bound to the concept of the question that has to be filled to generate the query template to retrieve the corresponding response from the database. The response returned by query would be raw data, which is returned to the user. System developed by Sneiders [15] also utilizes answer templates to pose the answer in a formatted manner.

The basic principle followed by template-based question answering system is much similar to the automated FAQ (Frequently Asked Questions) answering system that responds with pre-stored answers to user question but unlike static FAQs, the question templates are filled dynamically with parameters. One of such systems by Gunawerdena et al. [31] has been built to meet the needs of the close domain system to understand SMS language from a mobile phone in addition to natural language questions in English. This system used pre-processed text to identify best matched template-answer pair stored in database. Each of such templates is defined to match many different variants of the same question but not the question of same type making it too constrained.

While for Question Assistant [15], Sneiders designed templates so that single template could cover a wide range of data instances relevant to the entity slot and almost all questions of its type. Entity slots within a question template represent concepts or entities contained in database and relationship between these concepts is represented by templates themselves. However, if fresh relationship has to be added, a new template is required.

Another QA system making assistance of template-based approach is proposed by Unger et al. [34]. He used this technique over RDF (Resource Description Framework) data utilizing SPARQL [38] template. Thus, given technique is quite adaptable to semantic web. A SPARQL template directly reflects the internal structure of the questions and maps natural language question to domain vocabulary. This system, however, also applies deep linguistic analysis to generate the SPARQL template as this template not only focuses on syntactical pattern but also on semantic understandings.

III. DISCUSSION

Different approaches that have been so far discussed in above section perform fairly well for their domain of scope. QA systems relying on linguistic approach were basically built upon a knowledge base for specific domain, which provides an efficient and reliable response for short answers. Answer extraction mechanism from the knowledge base is supported by deep linguistic analysis to identify the relevant answer. In addition, building of an appropriate knowledge base with hand crafted rules requires a lot of human expertise and time and poses portability limitation. In some of the recent linguistic approach-based systems, web is used as the knowledge resource for local knowledge base, which not only led to enhanced knowledge within the domain but paved the way for thought of question answering with the integration of the local knowledge base in an open domain too.

Some of these earlier systems also relied on heuristic rules to identify question class and applied shallow NLP techniques, but success of this rule-based mechanism is so far restricted to the systems having only text documents as their knowledge resource. In addition to it, construction of proper rules required sufficient amount of training data and time along with skillful human effort.

Statistical approach is most likely to be useful for large quantity of data having enough word for statistical comparisons to be considered significant. The obvious choice of large data set for this approach is made to provide the sufficient amount of learning data while training statistical models. However, once statistical models have been properly trained, these systems could successfully provide the response of even complex questions.

Pattern based approach utilize eloquence of text patterns instead of critical linguistic analysis. The shallowness of the pattern-matching would often lead to some failures but, it has also been a surprisingly efficient technique for exploiting the Web as a data source. The basic idea behind pattern matching not only reduces linguistic computations but also aid in automatic wrapper generation for handling heterogeneous web data. However, this technique lacks in semantic understandings and reasoning.

All the elementary approaches discussed so far in this paper, perform fairly well for their respective application areas but suffer from certain limitations when implemented

beyond that. This fact leads to the development of QA system with hybrid the approach that would not only overcome the constraints but also exploit the potential, resulting from an individual approach. In recent times, therefore, many systems were developed combining the abilities of individual elementary approaches.

Table 5 Overall comparison between three approaches.

	Linguistic	Statistical	Pattern
Question type handled	Factoid questions	Complex non-factoid along with factoids	Factoids, definition, acronym, birth date.
Semantic understanding	Deep	Shallow	Less than all other competing approaches.
Heterogeneous data handling.	Quite difficult as knowledge base are generally designed only to handle their prestored data type.	Statistical similarity measurements are used to integrate data.	Easily possible as pattern aids in wrapper generation.
Reliability	Most reliable as answers are extracted from self- maintained knowledge base.	Reliable as most of these systems use supervised approach.	Depends on the validity of knowledge resource.
Scalability	Quite complex as new rules have to be introduced in the knowledge base for every new concept.	Most suitable for handling large data once properly trained.	Less as new patterns have to be learned for each new concept.
Evaluation Technique/Test Collections	Domain specific manually developed test collections.	TREC, CLEF, NTIRC test collections.	Domain specific manually developed test collections.
Application area	Systems that have long term information needs for specific domains	Quite suitable in handling large volume data e.g. web	Best suits to small and medium size websites, Semantic web.

MULDER [12] developed by Kwok et al. was first fully automated, general purpose QAS to generate the response from web with less user effort. The system was based on integration of linguistic and statistical approach and has a high recall of 64.5. QA system by Chakrabarti et al. [19] has followed linguistic and pattern-based approach and specifically relied on surface patterns from WordNet structure to determine the answer type. Similarly, Xia et al. [26] has integrated rule-base classifier and SVM classifier and showed precision and recall of 96.22%. Some other systems based on hybrid approach are ASQA [27] and IBM's WATSON QA System [33]. The ASQA has used surface patterns for biography questions and entropy methods to deal with definition and relation questions while WATSON competed against human grand champions in real time on the American TV quiz show Jeopardy! and is giving precision of 70%.

Most of QA systems relying on hybrid approach show different level of efficiency for different type (acronym, definition, list, etc.) of questions. To summarize, in the current state of art on question answering, more promising effort is still required to efficiently integrate linguistic, statistical and pattern-based techniques in order to cope with diverse users' needs.

IV. CONCLUSION

Our effort has been to take a comprehensive overview of the question answering research to meet the challenges due to information explosion in this information and communication technology era. We observed that the choice of a technique is highly problem specific. Often a hybrid approach, judiciously blending apparently different techniques, provides improved results in the form of faster speed, increased relevancy, and higher precision and recall measures. It is, however, realized that question answering techniques, based on linguistic approach, statistical approach and pattern-based approach will continue to remain in sharp focus, receiving attention of a large number of Question Answering System researchers.

REFERENCES

[1] 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.

[2] 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.

[3] Woods W. Progress in Natural Language Understanding - An Application to Lunar Geology. In Proceedings of AFIPS Conference, Vol. 42, 1973, pp. 441–450.

[4] 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.

[5] Katz B. Annotating the World Wide Web using natural language. In Proceedings of the 5th RIAO conference on Computer Assisted Information Searching on the Internet,1997, pp. 136-159.

[6] Voorhees EM. The TREC-8 question answering track report. In Proceedings of TREC-8, 1999, pp. 77-82.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] Soubbotin MM and Soubbotin SM. Patterns of Potential Answer Expressions as Clues to the Right Answer. In Proceeding of the TREC-10, NIST, 2001, pp. 175-182.

[12] 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.
[13] 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.
[14] Zhang D and Lee WS. Web based pattern mining and matching approach to question answering. In Proceedings of the 11th Text REtrieval Conference, 2002.
[15] Sneiders 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.
[16] 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.
[17] 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.
[18] 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.
[19] Ramakrishnan G, Chakrabarti S, ParanjpeDand Bhattacharya P. Is question answering an acquired skill? In Proceedings of the 13th ACM international conference on World Wide Web, 2004, pp. 111-120.
[20] 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.
[21] Whittaker E, Furui S and Klakow D. A Statistical Classification Approach to Question Answering using Web Data. In IEEE International Conference on Cyberworlds, 2005, pp. 8-pp.
[22] 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.
[23] 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.
[24] Saxena AK, Sambhu GV, Kaushik S, and Subramaniam LV. Iitd-ibmirl system for question answering using pattern matching, semantic type and semantic category recognition. In Proceedings of the TREC, Vol. 2007, 2007.
[25] 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.
[26] 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.
[27] 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.
[28] 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.
[29] Quarteroni S, and Manandhar S. Designing an interactive open-domain question answering system. Natural Language Engineering, Vol.15(1), 2009, pp. 73-95.
[30] 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.
[31] 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.
[32] 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.
[33] 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.
[34] Unger C, Bühmann L, Lehmann J, NgongaNgomo 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.
[35] Fu J, Xu J, and Jia K. Domain ontology based automatic question answering. In IEEE International Conference on Computer Engineering and Technology, Vol. 2, 2009, pp. 346-349.
[36] Ying-wei L, Zheng-tao Y, Xiang-yan M, Wen-gang C, Cun-li M. Question Classification Based on Incremental Modified Bayes. In Proceedings of IEEE Second International Conference on Future

Generation Communication and Networking, Vol. 2, 2008, pp. 149-152.

[37] 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.

[38] Prud'hommeaux E, Seaborne A(eds.). SPARQL Query Language for RDF. http:/ www.w3.org/TR /rdf-sparql query/, 2007.