MODELING OF QUESTION ANSWER SYSTEM USING AI

Subodh Kant[1]
Dr. Dinesh Yadav[2]

[1]Research Scholar, Dept. of Computer Science
[2]Research Supervisor & Assistant Professor
Dept. of Computer Science
Radha Govind University, Ramgarh Jharkhand

Abstract— Automatic question-answering (QA) system is a typical problem in natural language processing task to automatically produce relevant answer to a posed question. This work provides an overview of various techniques and methods employed to solve this typical question-answering problem. The basic idea behind QA system is to support the urge for information. This paper provides a brief review of different types of QA systems and work done so far. It is observed that the lexical gap and semantics with respect to context poses new challenges in question answer system. An attempt is made to provide a review of traditional and deep learning techniques employed for solving the research problem is made in order to bring an insight to research scope in this direction. We provide a proposed framework of question answer system using deep learning approach. The paper also discusses limitation and considerations for the said system.

Keywords— *Question Answer system, knowledge base, deep learning*

INTRODUCTION

Question-answering (QA) is an active research area in natural language processing (NLP). In today's ICT enabled world, answering user questions with most appropriate content relevant to the posed question with available knowledge base is need of the hour. The solution to this problem is question -answering systems (QA systems) [1]. QA system extracts information from a knowledge base in order to answer a specific question. As defined by Wikipedia a "Question answering (QA) is a computer science discipline within the fields of information retrieval and natural language processing (NLP), which is concerned with building systems that automatically answer questions posed by humans in a natural language." The QA system is divided into two broad categories viz: open domain and closed domain. Open domain questions are not focused to any specific material whereas closed domain refers to questions having concrete limitations referring to predefined source. To answer the question, the system analyzes the problem using the context to find one or more possible answers. In case of closed domain system, the answers are extracted though knowledge base. Conversely web searches provided answer with supporting material that makes the answer understandable in an open domain system. In the evaluation of QA systems that answered the factual questions by consulting 1the documents contained in TREC collections was started in 1999. Popularly knowledge is represented using graphical model The NOK method [2] belongs to this group.The system faces many challenges like question classification, right queries formulation, ambiguity resolution, semantic relatedness, identification of temporal relationship in complex questions. Apart from these challenges identification of a perfect answer requires proper extraction and validation mechanism. This paper provides a brief summary of work done in area of question answering system. The paper provides an analysis of various results obtained from tradition machine learning algorithms. The paper also discusses the recent deep learning approaches to the question answering problem. The paper discusses thevarious problem that need to be addressed for successfulimplementation of the system.

Rest of the paper is organized as follows, Section I contains the introduction of Question Answer System, Section II contain the related work Question Answer System, Section III explains the use of deep neural network for QA system, Section IV explains the knowledge base requirements and its impact on QA system, section V contain the basic block diagram and architecture of QA system, section VI explain describes the evaluation measures. Section VII presents a discussion of QA system; Section VIII concludes research work with future directions.

RELATED WORK

The QA systems are classified on basis of:

1. Application domain
2. User Question Types
3. Source documents and analysis on users' questions
4. Other Approaches

Application Domain

The application domain for the QA system can be closed or restricted domain or open or general domain. In General domain variety of questions can be asked and system generally searches for answers within a large document collection. Kan [16] has used ontology and word knowledge for answering. Users generally ask casual questions and quality of answer is marginal.

The general domain requires large and -variety of knowledge base to answer to causal questions posed by user. Wikipedia, Newsgroup datasets are sufficient to answer the user questions. As the questions are casual in nature the quality of answer produced is low and user dependent. In Restricted domain answers are extracted from domain specific knowledge specific collection. As repository is limited in nature the accuracy increases. Use of domain ontology, knowledge base would help in selecting an appropriate answer. Various restricted or closed domain QA systems available in literature are temporal, medical, patent, geospatial, community based etc. This entire domain, if integrated together would make an open domain QA system. These systems suffice the need of expert users who require specific answers to their queries. The level of user satisfaction depends on their domain knowledge. The quality of answers is also high.

User Question Type

QAs can be classified on basis of types of questions asked by users. The different categories are (a) factoid type questions (which, when, what, how, who) requiring answer in short form. These question answer systems have satisfactory performance. These can be answered with simple and readily available repository like Wikipedia without complex NLP processing. (b)List type questions presents answer as a list of facts or entities as answer. (c)Hypothetical type questions do not have specific answer. (d) Confirmation questions produce answer as yes or no true or false. (e) Causal questions expect answer in descriptive form. It requires explanation about the entity in question. (f) Fuzzy questions which cannot represent the need of user thereby expected answer is not fixing. (g) Dialog questions make it difficult to identify user requirement. (h) Descriptive type questions which require finding the definition or description of the term.

Source Document and Analysis of User Question These documents are classified on basis of analysis done on question. The different groups are: (1) morphological analysis type of analysis aims at separating words into individual morphemes and assigning a class to the morpheme. (2) Syntactical analysis identifies grammatical construction of words in questions and source documents (3) semantic analysis deduces the possible meaning of questions based on the words used in the questions. (4) Pragmatic and discourse analysis of the questions is done and documents are interpreted at sentence or higher level. (5) Expected answer type analysis requires answers based on the category of questions and (6) focus recognition of questions to extract correct answer.

Other Approaches

Apart from above mention approaches, the system can also be differentiated on basis of linguistic approach, statistical approach and pattern matching approach.NLP techniques and knowledge base or corpus is used for such QA system. The question answers are analyzed on basis of production rules, logic, templates, semantic relatedness and ontology. User questions are pre-processed with POS tagging and important keywords are identified which act as features for searching the answer into corpus. Around 1960 BASEBALL [5] and LUNAR [6] were NLP systems for querying a structured database. They used techniques to derive canonical forms which were converted to queries.

Statistical Approaches deal with large amount of heterogeneous data coming from online repositories and worldwide web. They employ statistical methods like SVM, EM, Bayesian methods to learn model which can also be applied to other domains. These methods produce better results than their competitors.

Major contribution in statistical QA system of system was IBM's statistical QA [7] system which employed maximum entropy method using bag of words features. This system should considerable performance improvement and reduction in error rate. Moschitti [8], Zhang et al. [9], Quarteroni et al. [10], and Suzuki et al. [11] used SVM text classifiers for question and answer categorization. Wei et al. [13], MKQA [12] used Modified Bayesian Classifier. Their experiments showed that Modified Bayesian Classifier method has better accuracy than base Bayesian method.

Pattern matching approach uses the text patterns to replace the sophisticated processing involved in other competing approaches. In recent times these QA systems are widely used as they automatically learn text patterns from text. They refrained from employing complex linguistic knowledge, parser, named-entity recognizer, ontology, WordNet, etc. to text for retrieving answers. Answer to a question is identified on the basis of similarity between their reflecting patterns viz. regular expressions having certain semantics. This system has high precision.

Zhang et al. [14] used support and confidence measures whereas Greenwood et al. [15] employed named entity tagger to identify pattern in text. Cui et al. [17] used bigram model and Profile Hidden Markov Model, and Saxena et al. [16] used pattern matching with acronym expansion to identify appropriate answers. Gunawerdena et al. [19] worked on automated FAQ systems on closed domain. Sneiders [18] and Unger et al. [20] employed template-based approach for helping answering questions.

Linguistic approach is used to handle factoid question. It requires a deep semantic understanding of text. These systems are highly reliable for extracting answer from their knowledge base but fail to handle heterogeneous data. These systems suffer from scalability issue as new rules need to be developed for every new concept and added in knowledge base. Statistical approach is generally used for both factoid and non-factoid questions. Shallow understanding of text is sufficient. As they rely on statistical methods they are scalable and follow supervised methods for forming model. This makes them suitable for heterogeneous datasets. Pattern based approach is used for all sorts of questions viz. factoid, non-factoid, definitions acronyms etc. As fewer patterns have to be learned for every concept scalability is issue. This approach is best suitable for small and medium size data corpus.

On reviewing the current scenario, more efforts are required to efficiently integrate linguistic, statistical and pattern-based techniques in order to cope with diverse users' needs. Majority of QA systems learn question answer into low dimensional space and select appropriate answer by finding feature similarity.

QA SYSTEM AND DEEP NEURAL NETWORKS

In recent years deep, neural network has shown promising solution to many of the AI problems. QA has also been used to develop dialog systems [1] and chat bots [2] designed to simulate human conversation. However, with recent developments in deep learning, neural network models have shown promise for QA. IBM Watson employed deepQA, massively parallel software architecture to examine natural language content in both clue and its database to extract relevant answers for a given clue. It then ranks the answer on basis of relevance with evidence for each answer. Sophia, the humanoid is programmed to give the pre-responses for the set of specific questions or the phrases. The inputs and the responses are processed with block chain technology and the information is shared in a cloud network employing deep learning.

lthough these systems generally involve a smaller learning pipeline, they require a significant amount of training. GRU and LSTM units allow recurrent neural networks (RNNs) to handle the longer texts required for QA. Iyyer [21] employed recursive neural network for question answering system. Weston [22], Hochreiter [24] and Tan [23] proposed a memory network model using LSTM for QA system. Kalchbrenner [25] and Feng [26] used Convolution neural network (CNN) based model for modelling questions and answer selection. Authors [27],[28],[29],[30] have used various methods of feature based learning and neural networks.

Yu et al[31] used distributed representation learning using logistic regression to match question answer pair through semantic encoding. Severyn and Moschitti [32] proposed a novel deep CNN method to find answers to factoid QA system. They used multiple convolution layers to form a metric of question and answers. Using frequency statistics, the two metrics are compared for identifying similarities. Wang and Nyberg [33] used a three-layer, stacked bidirectional long short-term memory (LSTM) network model with matching question to answer. Attention-based self-matching networks was used by Wenhui et al. [34] to locate the answers from the text. Wang et al [35] has used deep belief network to extract answers from cQA and forum datasets. Ying et al [36] has proposed KABLSTM, a Knowledge-aware Attentive Bidirectional Long Short-Term Memory to leverage external knowledge from knowledge graphs (KG) to enrich their presentational learning of QA sentences. The system was evaluated on WikiQA and TREC QA datasets. Atsushi [38] has tried to bridge the lexical gap between queries and document by using encoder decoder deep neural model. The model learns keywords from answer for questions for a multi domain FAQ system.

Author [39] has investigated on usefulness of modularity to insert background knowledge into deep networks. They proved that this improves the learning performance of system and also is suitable for extracting knowledge from trained deep networks. They proposed an algorithm for inserting hierarchical knowledge into deep networks during training using confidence rules

Current models use complex linguistic tools in obtaining linguistic features and external dictionaries like WordNet to achieve notable results. The system needs to handle task such as synonymy, polysemy, word order, question length, and data sparsity.

I. REQUIREMENTS OF KNOWLEDGE BASE

The question answer system uses 3 types of Knowledge base viz text corpus, knowledge base corpus and hybrid corpus.

1. **Text Corpus:** CLEF and TREC are factual questions. TREC questions are constructed from logs of search engines and answers have to be extracted from texts. Datasets of question-answer pairs are freely available. These datasets do not provide summaries to support the answer given. News dataset contains news from newspapers and time relevant. Encyclopaedia information is time irrelevant. Now a day's corpus like Stanford Question Answering Dataset (SQuAD) [37] is widely used for deep learning. It is a corpus of more than 100 000 question-answer pairs concerning more than 500 articles of Wikipedia, collected by crowd sourcing. Amazon Product datasets contain Customer Questions & Answers section with 20,120 question–answer pairs. Microsoft Community Questions & Answers consists of 31,000 question–answer pairs. Yahoo! Answers is another dataset of research QA.
2. **Knowledge base corpora:** The QALD datasets are made of about200 textual questions used in combination with DBpedia. The answers retrieved are DBpedia URIs. Low number of questions makes it difficult to use in learningmethods. WebQuestions [38] is a question-answer dataset on Freebase build using Google Suggest API queries and Amazon Mechanical Turk is used to filter them. Answers arethe Freebase URIs. It is constituted of 3778 examples for training and 2032for test. The AAAS Project contains MCQ with 775 available questions containing 26 pictorial or numerical answers and 749 questions covering topics in life sciences, physical science, earth science, and the nature of science.
3. **Hybrid corpus:** Since 2014 QALD is engaged in creating a hybrid QA since 2014. The objective is to answer questions by using both triples from DBpedia and Wikipedia article. Till data it has around 150 question answer pairs to evaluate a question answering hybrid system.

Using different type of corpus different type of QA systems are developed till date. It is believed that as size of knowledge corpus increases, accuracy of system increases. Experiments with traditional IR systems with corpora in the range 10-100GB [40] indicates that system performance is directly related to corpus size. Authors [41] have proved that performance of QA system does improve up to 400-500GB but reaches to an asymptote and actually declines slightly after that. A research in this direction is required to determine if the observed asymptote represents a real effect, or is a drawback of system used with given evaluation methodology. The answers returned by the system suggest a possible weakness in the evaluation methodology of QA system. Use of deep learning techniques in QA systems would solve the said drawback and improve accuracy of the QA system.

II. METHODOLOGY AND PROPOSED SYSTEM

The QA system consists of three core components: question analysis which includes question parsing, question

classification and query formulation, document analysis viz extract candidate documents and answer extraction by identifying appropriate candidate answers and finding themost appropriate one through ranking [1]. Question classification plays an essential role in QA systems by classifying the submitted question according to its type. Most of the recent works integrate artificial intelligence, natural language processing, statistical analysis, pattern matching and information retrieval. Figure 1 shows a typical block diagram of QA system.

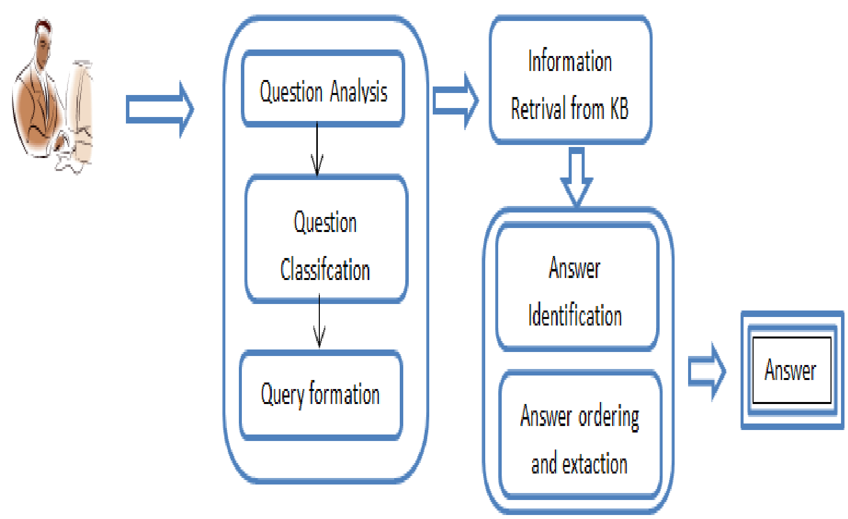

Figure 1: Basic QA system

The system executes in three phases: Query processing phase, Document Processing phase and Answer Processing phase. The question given by user is taken as input to Query processing phase. The module identifies the intent of question and classifies it into predefined categories of question type. This helps to identify the expected answer pattern. The question is added with classification information to reformulate the query. The formulated queryis given input to information retrieval system. The system extracts answers to the query fired to the information base. This knowledge base can be internet for open domain or predefined knowledge base in case of closed domain. The system extracts the probable candidates for the query referring to input question on basis of keywords in the query. The retrieved relevant documents are filtered and shortened into summaries that are expected to contain the answer. These summaries are ordered on basis of relevance. A set of heuristics is defined in order to extract only the relevant word or phrase that answers the question. Question processing is the module which identifies the focus of the question, classifies the question type, derives the expected answer type, and reformulates the question into semantically equivalent multiple questions. Reformulation of a question into similar meaning questions is also known as query expansion and it boosts up the recall of the information retrieval system. Information retrieval (IR) system recall is very important for question answering, because if no correct answers are present in a document, no further processing could be carried out to find an answer [3].Answer extraction is the final component in question answering system, which is a distinguishing feature between question answering systems and the usual sense of text retrieval systems. Answer extraction technology becomes an influential and decisivefactor on question answering system for the final results.

The results are analyzed on following factors:

1 relevance: the answer should answer a given question
2 accuracy: the answer should be factually correct
3 conciseness: the answer should not contain extraneous orirrelevant information
4 completeness: the answer should be complete, i.e. partialanswer should not get full credit
5 Rationality: the answer should be coherent, so that thequestioner can read it easily.
6 validation: the answer should be supplied with sufficientcontext to allow a reader to determine why this was chosen as an answer to the question

The proposed system will use **S**tanford **Qu**estion **A**nswering **D**ataset (SQuAD1.1) dataset consisting of questions posed by crowd workers on a set of Wikipedia articles, where the answer to every question is a segment of text, or *span*, from the corresponding reading passage. The proposed framework has 3 major layers Embedding layer, encoder layer and attention layer as shown in figure 2.

Embedding layer: The training dataset for the model consists of context and corresponding questions. Both of these can be broken into individual words and then these words converted into Word Embedding using pre-trained vector like GloVectors. Word Embeddings capture the context around the words than using a one hot vector for every word.

Encoder layer: A bi-directional GRU/LSTM is used with RNN. This encoder layer is used to store context of words before and after it. It is also used to create question vectors. A series of hidden vectors are generated and concatenated them.

Attention layer: It helps to decide question and its context. A dot product of question and answer vector is calculated as attention. A softmax over the product ensures that sum of all product is 1. Calculate the product of attention distribution and question vector.

Identify candidate answer vectors and rank according to relevance. A vector with highest confidence value is selected as answer.

In case an appropriate answer is not identified, the knowledge base needs to be upgraded. The query is fired to a search engine and documents extracted need to be given as a training data to the model developed in previous steps. This would help in incremental learning.

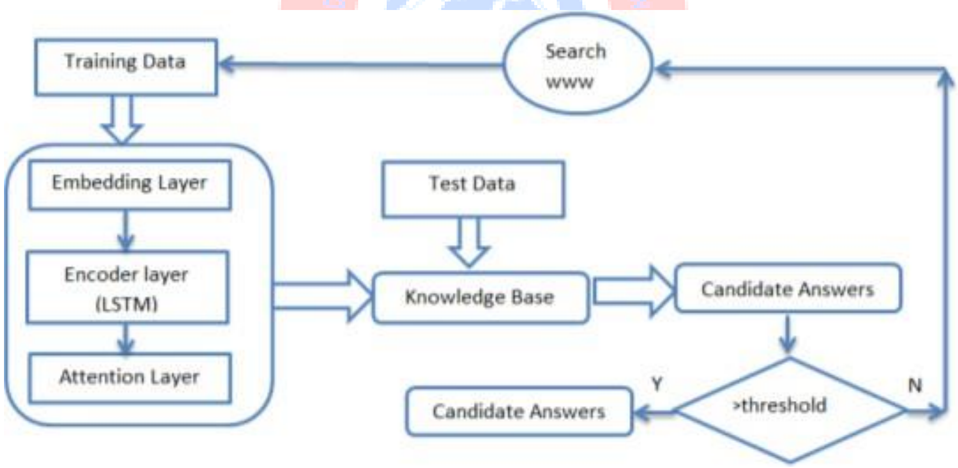

Figure2: Block diagram of proposed system

III. EVALUATION MEASURES

Precision, recall, and F-measure are performance parameters for evaluation of QA system. In a QA system, precision (P) also known as positive predictive value, demonstrates the closeness of answer with the given question. Accuracy refers to the closeness of a mapped answer to actual correct answers. Recall on the other hand indicates fraction of relevant answers that have been retrieved over the total amount of relevant answers. Let P be precision, RA be relevant answer and RT be the retrieved answers so, precision P is defined as:

$$P = \frac{|\{RA\} \cap \{RT\}|}{\{RT\}}$$

Recall for a QA system is defined as

$$R = \frac{|\{RA\} \cap \{RT\}|}{\{RA\}}$$

F1 score of test accuracy is the given by

$$F1score = 2 * \frac{(P * R)}{(P + R)}$$

IV. DISCUSSION

Though many attempts are made to design a good question answer system still some points need to be discussed. One of the main challenges in querying a Knowledge Base with a natural language question is the lexical gap between the question and its meaning. The main concerns are in the alignment of a question with the knowledge base triple and its parsing technique to determine the named entities and its relationship. A semantic relevance occurs between the question and their expected candidate answers. Due to scarcity of word embedding this relevance need to be modelled and quantified appropriately. The appropriateness and completeness of answers depends solely on the nature and power of knowledge base. The knowledge base requires to be reliable, complete for at least a given domain, updated with latest information. This corpus needs to be constantly upgraded and incorporated in learning model to improve accuracy of system.

Incremental learning should be a feature of knowledge base and model thus built. To explore features of a question and relevant answers in feature space and sample space i.e. knowledge base, ensemble learning methods can be used. Classification of questions with respective answers incurs high labelling cost and problem of concept drift also exists. Active learning should be incorporated in knowledge base which requires no fixed labelling and on demand request labelling. The problem of concept drift can be handled by dynamic adjusting the threshold for relevant answers when old concept is suddenly changes to new. Generating short descriptive answers to question and answer generation according to the type of question and its assessment still remain a challenge.

V. CONCLUSION AND FUTURE SCOPE

The paper reviews various question answer systems using traditional and deep learning approaches. Relevance, accuracy, conciseness, completeness, rationality and validation are important factors which defines the quality of QA system. Traditional machine learning SVM, EM and Bayesian methods are commonly used statistical methods for QA systems. Open and closed application domain systems are commonly employed in temporal, medical, patent, geospatial, community-based systems using ontology and word knowledge for answering. Use of LSTM and GRU with RNN is common choice for deep learning-based question answering system. Few authors have also used CNN for the QA system. The proposed system uses attention based RNN with LSTM. Knowledge Base is continuously upgraded for newer information making system more learnable. The system tries to overcome the problem of limited knowledge base. Active learning can be considered as a direction for future work.

REFERENCES

[1]. L. Hirschman and R. Gaizauskas, "Natural language question answering: the view from here," *Natural Language Engineering*, vol. 7, no. 4, pp. 275-300, 2001

[2]. 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.

[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]. 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.

[5]. 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.

[6]. 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.

[7]. 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.

[8]. 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.

[9]. 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.

[10]. 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.

[11]. Zhang D and Lee WS. "Web based pattern mining and matching approach to question answering", Proceedings of the 11th Text Retrieval Conference, 2002.

[12]. 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.

[13]. Saxena AK, Sambhu GV, Kaushik S, and Subramaniam LV, "Iitd- ibmirl system for question answering using pattern matching, semantic type and semantic category recognition", Proceedings of

the TREC, Vol. 2007, 2007.

[14]. 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.

[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]. 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.

[17]. Unger C, Bühmann L, Lehmann J, NgongaNgomo AC, Gerber D and Cimiano P, “Template-based question answering over RDF data”, Proceedings of the ACM 21st international conference on World Wide Web, 2012, pp. 639-648.

[18]. M Iyyer, JL Boyd-Graber, LMB Claudino, R Socher, IH Daume, “A Neural Network for Factoid Question Answering over Paragraphs”, In Conference on Empirical Methods on Natural Language Processing, 2014.

[19]. J Weston, S Chopra, A Bordes, “Memory networks”. arXiv preprint arXiv:1410.3916, 2014.

[20]. M Tan, B Xiang, B Zhou, “LSTM-based Deep Learning Models for non-factoid answer selection”, arXiv preprint arXiv:1511.04108, 2015.1

[21]. S Hochreiter, J Schmidhuber, “Long short-term memory”, Neural computation 9.8: 1735-1780, 1997

[22]. N Kalchbrenner, E Grefenstette, P Blunsom, “A convolutional neural network for modelling sentences”, In Proceedingsof the 52nd Annual Meeting of the Association for Computational Linguistics. June, 2014.