

Design Factoid Question Answering System using BERT

M. Tech. Scholar Sheetal Singh Goutam, HOD. Avinash Pal

Department of Computer Science,
Patel College of Science & Technology, Indore, India
Shitusheetalsingh@gmail.com, avinash.pal@patelcollege.com

Abstract- The field of text mining which deals with the providing of answers to the questions of the users is also one of the hot topics for researchers. In this paper Natural Language Processing (NLP) has been used which deals with the processing of the data that comes in any form like text, video, image, or audio. This NLP comes under the field of artificial intelligence (AI), which is used in the field of question answering (QA) system. Here proposed worked for designing a system that works for factoid QA which will answer the questions that are asked by the users. Lexical Chain and Keyword analysis is used in our system for the answering of questions from a given set of articles. The reasoning system is used for the validity of the answering. The experiment here is done with the SQUAD dataset. In our experiment overall average of the correct prediction of the answer the accuracy obtained for the passage retrieval using existing TFIDF is 70.30% and proposed BERT is 87.81%.

Keywords- Factoid Question Answering System, Natural Language Processing, TF-IDF, Artificial Intelligence, SQUAD Dataset.

I. INTRODUCTION

Natural language processing is the field where the question answering system can be built. It also comes under the artificial intelligence field where the answering of the questions is automatically given by the computer system [1].

The factoid QA system works with the natural language and the question is made in the natural language then the system answers for the questions asked are given through natural language processing of the available text or images [2]. In our proposed system the natural language processing is performed by which the system will be able to generate correct answers. Semantic analysis is done here for the answering and the validity is provided.

Here in this paper used the TF-IDF (Term Frequency - Inverse Document Frequency) technique for finding the relevant words in the files or documents. This measures the number of occurrences of any word in the particular data given. For this task the multiplication of two matrices is done and the occurrence of that particular word is tested. The TF-IDF is mostly done for text mining and the usage increase in the field of Natural Language processing.

In our paper a factoid QA system is proposed that will provide the answers to the user in our natural language. In the proposed factoid QA system, the first task performed is the extraction of keywords and the keyword analysis is performed for the searching for the retrieval of an answer. The removal of stop words is done and the usage of Porter Stemmer is made for the accuracy calculation.

II. NATURAL LANGUAGE PROCESSING

Natural language processing (NLP) is one of the most important techniques that change the computer into the human brain. The NLP comes under the field of artificial Intelligence where NLP is used to train the computer to understand the human language [3].

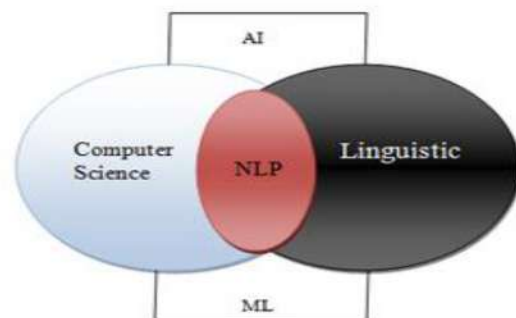


Fig 1. Natural Language Processing.

The computer is trained as a human brain for its activities of language processing. The NLP is majorly aimed for the reading of language, decipher, understanding the language and after this all making sense like the human brain and reacts as per human brain reactions [4].

So doing all these tasks makes the NLP the most important technique for building a system that can reply according to the human brain. This method of language processing is used in many applications. One of the applications of NLP is seen in the question answering system where the answering for the questions asked by the user is done [5].

This task of answering the question builds the QA system. The NLP helps in this system as it reads the question and tries to answer like a human brain.

Whereas the NLP uses machine learning for the evaluation of the proper meaning of the question and then the answering is done.

III. LITERATURE REVIEW

In this paper, authors here have developed a QA system that answers concisely for the questions through natural language processing. Their proposed system worked well than the existing QA system.

The existing system works for the Amharic QA that is based on factoid. So authors have designed an Amharic non factoid QA. The question types here solved are definition, biography, and description types. They have used lexical patterns for the extraction of answers. [6]

Here the author has stated a QA system that uses the answer triggering for the providence of answers. The answer triggering system selects an answer from the given set of the answers. For this system a cognitive approach is used for the selection of an appropriate answer from the set. They have used the WIQIKA dataset for the experiment. The precision achieved in this system is 48.89, Recall as 64.17, and F1 as 55.49. [7]

The paper here gives a survey on the QA system where the answer is provided precisely in the form of natural language. They have studied both the structured and unstructured datasets and the combination of both. State less QA is surveyed here in areas like RFD, linked data, etc. 21 system on this are studied here and 23 evaluation are analyzed. [8]. Here a factoid QA model is given by the authors for the answering of questions in the natural language. Here the authors have presented a model and named it as Temporality-Enhanced Knowledge Memory Network (TE-KMN). They have applied this model on the model of trivia named s/as quiz bowl. They have obtained 74.46% accuracy with the proposed model. [9]

In this paper authors have worked on a Chinese QA system. They have worked on the application of both the techniques like Named Entity Recognition (NER) and the Metric Cluster (MC). The model proposed by them worked well for factoid QA and achieved an MRR value of .6883. [10]

The authors have presented a semantic-based QA for answering in this paper. The questions here asked and answered are for the tourism area. So this model can also be said as tourism QA. Firstly the detection of question is done then the SPARQL query is optimized. The accuracy found here is 80%. [11]

The paper works on the QA system for factoid QA. They have worked on the development of a web-scale factoid QA system and the open domain is presented here for the factoid questions. They have divided the questions into major 5 categories and worked on the answering of these questions. They have achieved accuracy with the Wikipedia data upto 62.11%. [12]

In this paper a survey on the accuracy evaluation for the QA system based on the web is done. The paper states the study done with the three categories of QA system that are, the extraction of an answer, answer scoring, and the aggregation of the answers. They stated that the survey will help in the selection of appropriate QA systems for usage. [13]

Here the authors have developed a QA system for the biomedical field and the complexity in this field for the answering is solved with their approach. Here they have used a multi-label classification method for the classification of questions and types that are listed for QA of this field. They have enhanced the F1 by 2% and the MRR by 3%. [14]

1 Research GAP:

- There is a need for a system that answers the question in natural language.
- The processing of the passage retrieval by making use of the proper technique is needed to be improved.
- The natural language processing for the factoid QA system need more accuracy.
- As per the previous study the factoid QA system is difficult for implementation, so a proper system should be developed.
- The variety of QA should be increased to cover a vast area of applications.

IV. PROPOSED IMPLEMENTATION PHASE

In our implementation phase the different operations are performed using the AI and NLP. Firstly, we did the Natural Language Processing, this is done for a better understanding of the question. After this the AI is applied for the answering of questions in a proper sense.

1. Extraction of Keyword:

Here the keyword extraction is done in the very first stage, the keywords are the words that specially specify the main objective of the sentence. The keywords help to define the content asked in the question. These keywords maybe the words that relate to the question or maybe available in the question. The keyword will help for the retrieval of the questions and also the answers from the passage or the available content from which the answer to be provided. Detection of improper keywords will result in providing less accurate or wrong answers.

Also, here we remove the stop words from the question, the stop words are those which are not related or not worthy. We here improved we do the list of stop words that we extracted from internet addition and removal of some words. We used the steaming process for the extraction of keywords from the question asked by the user.

2. Passage Retriever:

In the QA system the Corpus is used for the searching of the answer for the questions asked. The characteristics of corpus is seen as it should define only one topic and should be available in the natural language. The Corpus specifically works in two ways for the accessing that are: First the Offline way, in this the electronic media is used for the accessing the articles. Here text files are stored in the database. Second is the Online way, in this the text files or data is obtained from the web. So, for our system the knowledge is required from where we need to access the articles. Here passage retrieval is done by using the TF-IDF and also used the articles over 442 that is accessed from the SQUAD. Some changes are made to the existing system for the proper answering of the question and improving the accuracy of the system. The factoid questions are answered well with this approach.

3. Searching Phase:

the Searching phase in which the searching of an answer for the question is done. The proposed steps have used here is the Key Word Analysis for the extraction of the proper answer. Here in this phase the breaking of question is done and the keywords are selected. The next step involves the creation of the lexical chain for the found keywords. Now the system analyzes the obtained keywords and then the creation of the answer list is done. If the asked question by the user relates with the FAQs, then it is added to the FAQ list and if it is different then it is answered. If it does not belong to the FAQ then the keywords are processed for answer retrieval. The lexical chain is used for the retrieval of answer and also the AI and NLP is used for the retrieval of answers for the asked question.

4. Proposed Architecture:

The proposed architecture is shown in below figure2. The architecture gives the details of every single section that is used for the retrieval of answers for the factoid QA system. In proposed architecture have used here the IR system, the Fuzzy team, the logic used in the server, the extraction of the answer, the query, and the database used.

The implementation phase works with the asking of questions to the system and then all the sections will work for the asked questions. The analysis of the questions and the processing along with the retrieval of the answer is done in our proposed architecture. And finally, the user is answered for the asked question.

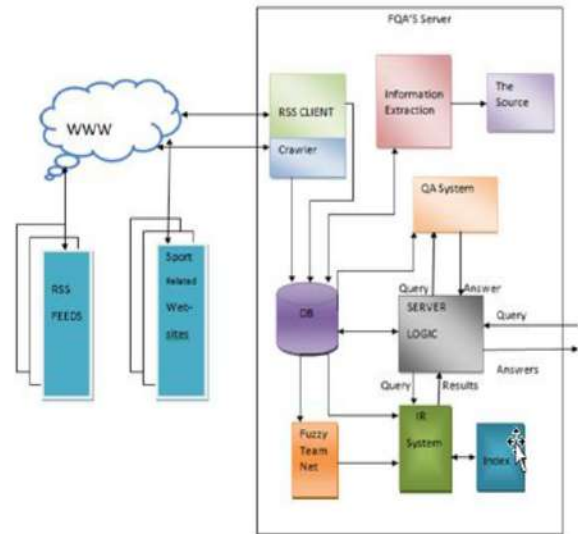


Fig 2. Proposed Architecture for the Factoid QA system.

The architecture of this bot closely follows the architecture described in the book. Main modules of the QA System are:

4.1 Question Processing: In this step, the bot identifies the type of question and type of answer it expects

4.2 Passage Retrieval: It generates question vector and vectors of the passage using TF-IDF as a feature, it computes cosine similarity between question vector and passage vector returning top 3 closely resembling passage. Further improvement to this step has been done by removing Stop Words and using Porter Stemmer

4.3 Sentence Retrieval: After retrieving passage, it tokenizes sentences and computes n-gram similarity between question and sentence. Thus, identifying the most relevant sentences.

4.4 Answer Processing: Based on the expected answer type, then it processes the answer sentence to identify a particular entity using a named-entity recognition technique and part of a speech tagging technique

4.5 Text Summarization: If the type of question is definition or bot is unable to identify named-entity from a question, it summarizes the text using n-gram tilting technique

V. RESULT

The obtained results are stated in this section and the various details related to the experiment is stated here. The results obtained are satisfying and the techniques used here worked well for our system.

1. Hardware used for Implementation:

The following configuration is used for the implementation of the algorithms. For the design Python Programming Language, 15.6 in HD WLED touchscreen (1366 x 768), 10-finger multi-touch support, 10th

Generation Intel Core i7-1065G7 1.3 GHz up to 3.9 GHz. 8GB DDR4 SDRAM 2666MHz, 512GB SSD, No Optical Drive. Intel Iris Plus Graphics, HD Audio with stereo speakers. HP True Vision HD camera.

Realtek RTL8821CE 802.11b/g/n/ac, Bluetooth 4.2, 1 HDMI 1.4, 1 USB 3.1 Gen 1 Type-C, 2 USB 3.1 Gen 1 Type-A. The Python Programming was run on Windows 10 64-bit Operating System platform. The python library was used during implementation like NumPy, Pandas, Matplotlib, SciPy, Scikit-Learn, PyTorch, Seaborn, Plotly, Tensor Flow, Keras, Seaborn.

2. Dataset Used:

We use the SQUAD dataset and our own developed dataset for the experiment. This dataset is used for the experiment of the factoid QA. The dataset provides various articles on different topics. Also the dataset is having a variety of topics so that the analysis is done properly and be tested on a variety of questions.

The dataset having the following properties:

- Over 442 articles.
- Approx 87599 cross-domain questions.
- Articles like Buddhism, alloy, USB, etc are found in the dataset.
- Anthropology (dataset/Anthropology.txt)

Sample Questions:

- What is anthropology a study of?
- How many educational institutions had some curriculum in anthropology by 1898?
- Where did early anthropology originate?
- Why type of anthropology is the study of social organization a central focus of?
- When was anthropology used as a term for comparative anatomy?

3. Result of Passage Retrieval:

- Accuracy of passage retrieval using TF-IDF was 87.81% when tested over 422 articles of SQUAD dataset and approx 87599 questions cross-domain.
- After removing stop words and using Porter Stemmer, this accuracy parameter improved to 94.44%.
- Further analysis, 100% of passage retrieval contains the equivalent paragraph in the top 3 returned paragraphs.

4. Result of Question Answer System:

Overall Average over the entire dataset was 83% correct answer prediction. The dataset included in this repository has the following accuracy:

The table shown above the results obtained from the experiment. The accuracy of every article and question is shown in the table. Over 4000 questions were done for the analysis and the correct answers obtained were over 3000. So, the proposed system worked well for the answer retrieval for factoid questions.

Table 1. Results for the Factoid QA on various topics for Accuracy.

Standard Dataset	Own Dataset	No. of Question	Correct Retrieval	Proposed Accuracy	Existing Accuracy
Alloy	AI	48	43	89.58	67.71
Anthropology	Big data	35	32	91.42	88.18
North western University	Cloud computing	42	38	90.47	69.25
Buddhism	Data science	22	18	81.81	70.16
Mammal	Edge computing	31	26	83.87	67.05
Marvel Comics	ML	28	19	67.85	62.6
Modern history	Object detection	17	16	94.11	72.54
New York City	OS	23	20	86.95	62.79
Queen Victoria	Raspberry	27	25	92.59	55.59
Rajasthan	Testing Data	36	34	94.44	68.91
USB	Virtual reality	14	13	92.85	88.51
Overall Accuracy	Overall Accuracy			87.81	70.30

The highest accuracy is obtained for the Testing Data QA as 94.44%. and the average accuracy for the system is found as 87.81%. The graph below states the results obtained in terms of accuracy.



Fig 3. Graph for accuracy obtained for various articles for our factoid QA.

Here the graph states the highest accuracy is obtained by the TestingData QA as 94.44% and the lowest is seen for theMLArticle as 67.85%.

VI. CONCLUSION

Factoid QA is one of the important systems for answering the user's questions. The factoid QA system deals with a variety of articles and a variety of QA is done in this system. The presented system also works for the variety of QA and the methods used are lexical chain, NLP and BERT, AI and keyword analysis.

The SQUAD dataset and own dataset are used here, which contains various articles. Our experiment is tested on a variety of articles and the system tests a large number of questions. Some of the articles analyzed are AI, Bigdata, Cloud computing, Data science, Edge computing, ML, Object detection, OS, Resberrypi, Testing Data, Virtual reality. The accuracy obtained for the system is of average 87.81%. The presented system can be used for the factoid QA system.

REFERENCE

- [1] G. S. Pirtoaca, T. Rebedea, and S. Ruseti, "Improving Retrieval-Based Question Answering with Deep Inference Models," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2019-July, no. July, pp. 1–8, 2019, doi: 10.1109/IJCNN.2019.8851826.
- [2] M. Khvalchik and A. Kulkarni, "Open-domain Non-factoid Question Answering," pp. 1–8.
- [3] H. H. Hsu and N. F. Huang, "Xiao-Shih: The Educational Intelligent Question Answering Bot on Chinese-Based MOOCs," *Proc. - 17th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2018*, pp. 1316–1321, 2019, doi: 10.1109/ICMLA.2018.00213.
- [4] X. Zhang, M. H. Chen, and Y. Qin, "NLP-QA Framework Based on LSTM-RNN," *Proc. - 2nd Int. Conf. Data Sci. Bus. Anal. ICDSBA 2018*, pp. 307–311, 2018, doi: 10.1109/ICDSBA.2018.00065.
- [5] K. Saengthongpattana, T. Supnithi, and N. Soonthornphisaj, "Quality Classification of ASEAN Wikipedia Articles using Statistical Features," *2018 Int. Jt. Symp. Artif. Intell. Nat. Lang. Process. iSAI-NLP 2018 - Proc.*, pp. 1–6, 2018, doi: 10.1109/iSAI-NLP.2018.8692954.
- [6] T. Abedissa and M. Libsie, *Amharic Question Answering for Biography, Definition, and Description Questions*, vol. 2. Springer International Publishing, 2019.
- [7] K. N. Acheampong, Z. Pan, E. Zhou, and X. Li, "ANSWER TRIGGERING OF FACTOID QUESTIONS: A COGNITIVE APPROACH," pp. 33–37, 2016.
- [8] E. Dimitrakakis, K. Sgontzos, and Y. Tzitzikas, "A survey on question answering systems over linked data and documents," 2019.
- [9] X. Duan et al., "Temporality-enhanced knowledge memory network," vol. 19, no. 1, pp. 104–115, 2018.
- [10] Y. Hong, F. Li, and L. Dong, "Application of NER and MC in Answers Extraction of Factoid Questions," *2019 Int. Conf. Comput. Network, Electron. Autom.*, pp. 36–41, 2019, doi: 10.1109/ICCNEA.2019.00017.
- [11] M. Rahim, Z. Turabee, Q. Rajput, and S. A. Khoja, "Semantic Based Question Answering System on Travel Ontology," *2019 Sixth Int. Conf. Soc. Networks Anal. Manag. Secur.*, pp. 67–74, 2019.
- [12] P. Ranjan, "QUESTION ANSWERING SYSTEM FOR," pp. 221–224, 2016.
- [13] A. Ali, S. Sri, D. Ravana, S. Hamid, and M. A. Ismail, "Accuracy evaluation of methods and techniques in Web-based question answering systems: a survey," *Knowl. Inf. Syst.*, 2018, doi: 10.1007/s10115-018-1203-0.
- [14] M. Wasim, W. Mahmood, M. N. Asim, and M. U. Ghani, "Multi-Label Question Classification for Factoid and List Type Questions in Biomedical Question Answering," vol. 7, 2019, doi: 10.1109/ACCESS.2018.2887165.