CONTEXT-BASED QUESTIONANSWERING SYSTEM FOR AN E-LEARNING PLATFORM WITH PRE-TRAINED TRANSFORMER MODEL

Chellatamilan T*[1], Valarmathi.B[2], Santhi.K[1]

[1]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore.
chellatamilan.t@vit.ac.in, santhikrishnan@vit.ac.in

[2]School of Information Technology and Engineering, Vellore Institute of Technology, Vellore.valarmathi.b@vit.ac.in

ABSTRACT

Many diverse day-to-day applications, such as e-commerce, marketing, supply chain, customer relationship management, and so on, have a significant impact on the Question-Answering (QA) System., It also creates a vivid emphasis among teaching learning systems, where the need for replies to inquiries is intrinsically required for e-learning application stakeholders. The authors of this paper present a deep learning approach to extract lecture notes as Texts and predicts the best acceptable or effective answer to a questionnaire.We used a pre-built BERT base model that has been fine-tuned with the Stanford Question Answering Dataset (SQuAD). The suggested transfer learning approach, we claim, aids generalizations in QA without the need of vocabulary. This unique framework focuses on developing a context-based question learning system that is customizable and can be integrated into any e-learning platform to benefit both teachers and students. It also employs a common deep learning approach known as transfer learning to enhance its ability to associate multiple practical provinces.Using a custom dataset created for the course "Object Oriented Programming," we fine-tuned the BERT transformer model. The experiment was carried out, and the results were visualised through comparisons with several models. In the testing phase, the MAP (Mean Average Precision) and MRR (Mean Reciprocal Rank) performance matrices were used to compare against the baseline of 82 percent MAP and 80.1 percent exact match score.

Keywords: Bidirectional Encoder Representations from Transformers, Deep learning, E-learning, Question-Answering System, Transfer learning.

I. Introduction:

BERT takes one or two sentences as input and utilises a specific token to distinguish between them. As the initial token of the first sentence, a special token [CLS] was added, and both of these sentences were ended with another special token [SEP]. Each sentence's word is treated as its own token, which is then added to the BERT vocabulary. BERT uses two inputs in the QA System: one is the passage and the other is the query. Based on the contextual information provided in the passage text, the BERT model predicts the answer to the question.

We built and tailored the dataset for conducting the required set of experiments in the QA System by taking a sample collection of learning material, particularly in the course Object Oriented Programming, in the form of portable document format . The materials gathered were formatted in a way that would allow the QA system to be applied in terms of passages and QA Truth.

Passage:"Object-oriented programming (OOP) is a programming paradigm based on the concepts of classes and objects." It's utilised to break down a software programme into reusable classes , which are then used to build specific objects"

Question 1:"What is OOP"?

Predicted Answer: The programming paradigm is built on the idea of classes and objects.

Question 2: What is OOP?

Predicted Answer: "Object-oriented programming "

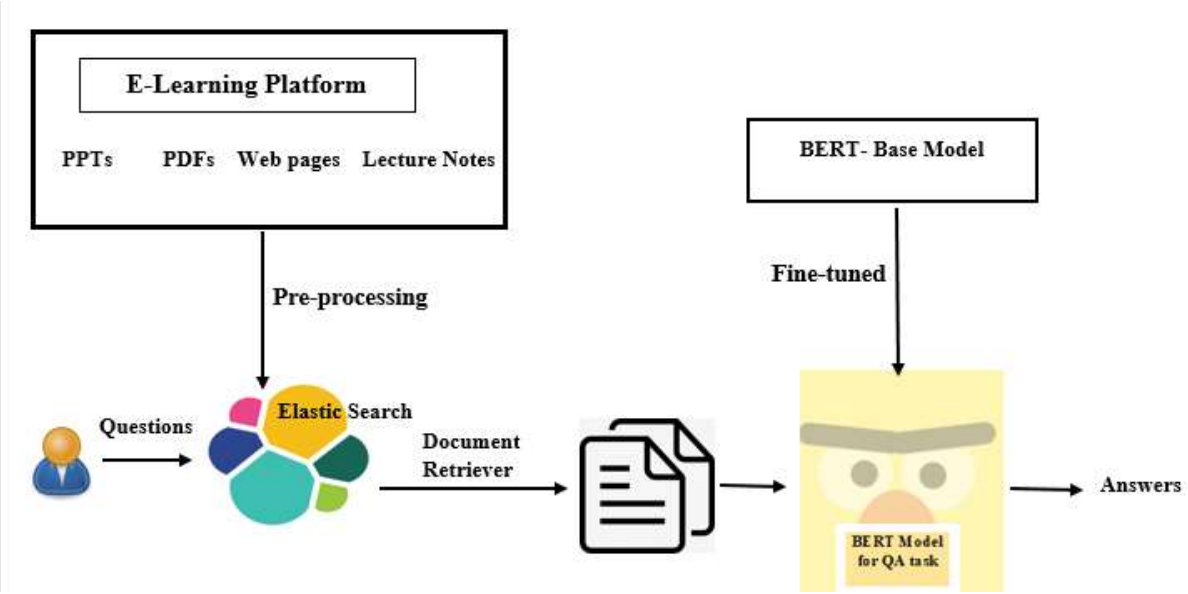


Figure 1: QA system with BERT

The simple architecture of integrating the QA model into an e-learning platform is shown in Figure 1. The platform contained course learning resources in the form of .pdf, .ppt, websites, and other formats. In order to construct the word vectors and embedding vectors of each of the para phase of the learning materials, the pre-processing techniques were applied toward the material corpus and in term that eliminates the stop words and fine tunes the text corpus suitable for applying the skip gram or bag of words model to generate the word vector. We have also used the pre-trained word embedding model such as BERT for obtaining the contextual word vectors of each word occurs in the question text. The QA is a kind of information retrieval, it retrieve answers to the given questions from the given collection of

documents. QA is playing as essential testbed for evaluating how the computing model understands the human language.

II. Literature Study

Raju Barskara et al. [1] have created a QA system for English Sentence responses. Users should be able to get responses to their questions quickly. This is accomplished by inquiring of the English paragraph System, which will then respond by searching the English dictionary repository inside the context of the paragraph for the needed response. In this article, they describe a system of questions and answers that makes use of information on categories by merging several question and response classification approaches.

Calijorne Soares, M. A., and Parreiras, F. S [2]have described the current state of the QA works, which analysed and extracted data classified by type of research and empirical research, as well as technical features such as accuracy, natural language processing methods used, application domain, applied natural language processing techniques, paradigm type, and application language.

Sriram Aryaman et. al [3] gathered 16,000,00 twitter data points from the Sentiment 140 website, which was used to detect depression. Using the BERT classifier, the technique provided in this document obtained 84.92 percent accuracy with less training and testing. That is, the BERT classifier has the greatest accuracy of 84.92 percent for the 60K (training) and 1 lakh (test) datasets.The methodology utilised was Term Frequency-Inverse Document Frequencies with n-grams with machine language processing and Term Frequency-Inverse Often Then n-grams with logistic regression classifications, with a maximum Precision of 81 percent.They plan to improve this algorithm by resolving the overfitting problem (by introducing regularization methods) in the future. They will also test variations of the BERT algorithm such as distilling BERT, A Lite BERT, Robustly Optimized BERT pre-training Approach.

Alami Hamza et al [4]proposed a framework for organising Arabic research questions as well as a new Arabic taxonomy. They created an app to answer a series of questions derived from Text REtrieval Conference (TREC), Crosss-Language Evaluation Forum(CLEF) and Moroccan Design manuals, where they then used to construct a classifier. The proposed methodology has been aided by the quality of the question representation, which is separated via word nesting, as well as the strength of machine learning approaches.Using support vectr machine classifier and their arabic taxonomy, they were able to achieve 90% accuracy rate.This research demonstrated that their technique is an enhancement current Arab issue categorization strategies.The following are the most significant benefits: the Arab taxonomy is well-suited to the task of categorizing Arab issues; the representation of applicable words captures the syntactic and semantic relations between words; the size of the question representation is lesser than Term Frequency-Inverse Document Frequency.

Mourad Sarrouti and Said Ouatik El Alaoui [5]proposed a method for recovering biomedical Passages that uses the Stanford CoreNLP phrase limit as the passage length, Unified Medical Language System (UMLS) concepts, and stemming words as BM25 model characteristics to extract relevant Passages in systems responding to biomedical issues. They used the PubMed search engine to extract relevant records for a specific biomedical inquiry, then reclassified them based on the UMLS similarity between the ideas of the biomedical

query and each title of the obtained records.They then utilized Stanford CoreNLP to break sentences in the summaries of the top-ranked publications, resulting in a series of candidate runs. The passing time is similar to Stanford CoreNLP's Sentence. Using abbreviated words and UMLS principles as features for the BM25 prototype, they eventually created a new ranking for all candidate passes and maintained the top-ranked N's. Experiments on the 2015 biomedical semantic indexing and question answering (BioASQ) datasets showed the validity of their methods. The findings of the evaluation reveal that the proposed methodology can improve current state-of-the-art recognizing textual entailment procedures by 6.84 percent. They discovered that the length of the passage had a significant impact on performance, and they believe that achieving a uniform length of passage will improve passage recovery.

The ontology-based QAS presented by Ali Albarghothia et al. [6]analyses users' queries related to the realm of knowledge in pathology. In their method, they created an Arabic ontological model from the ground up that deals with natural issues and then translated it into 3 models for constructing protocol and RDF Query Language (SPARQL) queries and implementing them in the Jena framework. The Protégé tool was used to create the ontology, which contains over 200 instances and 1260 triples. They also show how to use the Protégé tool to build ontologies. Furthermore, how to read requests in triple models and compose the SPARQL questions that are the instrument for recovering the suitable answer from Resource Description Framework (RDF) data.NLP functions have also been utilized to cope with difficulties, namely for normalization, tokenization, deletion of stop words, stemming, and tagging. On 100 queries, including factoids and difficult investigations, examination tests were undertaken. The model exhibits encouraging results in the evaluation phase, with an accuracy of 81 percent.

Archana S.M et al. [7]proposed a Malayalam questionanswering System. They set aside the terms of the question and recognised the vibhakthi that would connect the correspondent to the response. They used a rules-based method to check vibhakthi for each word after locating the sentences containing the general answer articulations. The responses are then retrieved to see if the question and answer modules' vibhakthi are the same.Part of the speech labeling, TnT tag, composite word separator are also used for their question-response System. The proposed work has used vibhakthi characteristics for the factoid QA System.

YongGang Cao et. al. [8] introduced their online question answering System(AskHERMES), which is designed to help physicians respond quickly in a timely manner. The System is based on both supervised and non-supervised learning techniques in different components for the exploration of a variety of linguistic characteristics. Ask-HERMES is currently able to analyze and understand complex clinical issues of various types that cannot be resolved by factoids or a single phrase. Their pilot evaluation shows that AskHERMES works in a similar manner to advanced systems such as Google and UpToDate. In particular, their System establishes good ability to answer long and difficult clinical questions as other systems, demonstrating robustness among the questions of different word accounts. Overall, based on their initial findings did not indicate statistically important differences existed between AskHERMES and both other systems.

Amit Mishra and Sanjay Kumar Jain [9] They classified question answering system based on type of questions,, the types of processes it did on the question and information sources, the types ofmodels it is used, the response forms it created, and the data sources features. The

effectiveness of this is highly dependent on a respectable source corpus and explicitly documented user demands. If the corpus is well-structured and user demands and charge of deciphering the text using complicated NLP approaches. The psychology and knowledge of the customer asking the question, have an impact on the performance of the QASystem.

Jeff Rickel and Bruce Porter [10] characterized a compositional show the differences to answer prediction questions regarding a difficult framework that allows for a great deal of adaptability in developing models, the programe should address modeling issues that are undoubtedly addressed in the domain knowledge required by previous programs.It tackles these issues by defining a set of field-independent reporting constraints that define a good model. Variables and impacts play a critical role in every modeling decision due to these constraints. TRIPEL can create basic and satisfactory models for a totally mind-boggling framework, according to the evaluation. More crucially, the assessment suggests a key subject for future research: the development of more specific criteria for determining whether one element has a fundamental impact on another. **Alaa Mohasseb et. al. [11]**, for question categorization and classification, a grammar-based method was proposed. The structure is based on 3 main points: grammatical traits, features, and patterns unique to the location. In the classification of factoid questions, the performance of various ML methods such as J48 and SVM was investigated. The results demonstrate that their response resulted in a high level of classification accuracy for the questions.

G. Suresh kumar and G. Zayaraz [12]A framework for identifying its attributes and relationships from a large volume of unstructured text for domain programmed ontological modeling was developed, and exploratory results were included in this publication. Using lexico-syntaxic and lexico-semantic probabilities, the suggested iterative idea connection extraction approach based on the dependency analysis model was used to extract attributes and relationships.**Stefan Schlobach et.al [13]**Open domain questionanswering frameworks were used, which extract a huge number of respondent responses from a collection of papers and then select the most likely responses from the list. The semantic type expected by the query is a criterion for selecting answers. Web redundancy can be utilized to gain the most information, the type of likely semantic response from a candidate, when there aren't any sources of knowledge that supply typing information.**Walaa Saber Ismaila and Masun Nabhan Homsi [14]** used a dataset for the Arabic Why QAsystem (DAWQAS) with the goal of disseminating it to the research community to encourage information retrieval and language comprehension research. It contains 3205 QA pairs spanning 8 fields, making it significantly larger than comparable datasets. Following that, the responses were categorized according to their sectors. Finally, for each Sentence in the data set, the probability of rhetorical relations based on speech markers were estimated. DAWQAS is a useful tool for studying and assessing language comprehension.**Yashvardhan Sharmaa and Sahil Gupta [15]**talked about many methods from the simple natural language processing and algorithmic methods were made and the document is finally based on the recently planned Deep Learning methods. Details of implementation and many adjustments in the algorithms that created better outcomes were also discussed. The planned models were evaluated on twenty tasks in the Facebook babI dataset.

LIANG Zhenqiu [16]presented the System of automated response to questions based on case reasoning (CBR). This System allows to analyse the questions that come into the natural language, searching for candidate questions defined in the database of historical questions by

the keyword automatically. When computing Sentence similarity, similar historical responses are returned to the user. Practice demonstrates that the System already achieves practical results by responding to accuracy and an intelligent, in distance learning and other related field has some practical value.

C. Pechsiri and R. Piriyakul [17] the goal of the study is to design a System for automatically answering queries on community bulletin boards, particularly the questions Why and How, to aid ordinary people in preliminary diagnosis and issue solutions, such as plant diseases. ML techniques are used in the study to identify the different types of queries. To decide the visualised responses using the information retrieval technique, they employed an integrated causal diagram with procedural knowledge collected from the text. The QA method yields an accuracy of 91.1 percent and 88.9 percent for the Why and How questions, respectively, according to an experiment..

Hong Yu et al. [18]developed and tested a System for answering medical definition queries that automatically tracked a large number of electronic documents and produced brief, consistent definition answers. In 2 important efficacy criteria, such as the time spent and the number of measurements taken by a doctor to define a definition, their initial cognitive assessment indicates that it has surpassed 3 other online information systems (Google, OneLook, and PubMed). In their view, systems to answer questions that group relevant information dispersed in different documents, clinical information needs can be addressed within a time frame necessary to respond to clinician requests.

Michael Spranger and Dirk Labudde [19]have developed an integrated IT solution to support the process of evaluating forensic texts with the help of linguistic IT technologies is described. The framework in development is based on a quality assurance System and the ability to solve a specific criminal problem and depict important relationships on a case-by-case basis.For this purpose, a number of advanced methods in the areas of categorization of texts and extracting information/events is examined on the basis of their suitability for the specificities of the field in question. In addition, a number of problem-solving approaches specific to the region are announced.

Yogi Wisesa Chandra and Suyanto Suyanto [20]created a chatbot using a sequence to sequence pattern. It is made up of data from a university admissions process that has been taught. The algorithm produces a relatively high BLUE score of 41.04 based on a limited data set from Telkom University's admission to the Whatsapp instant messaging program. The model is improved by an attention mechanism based on inverted sentences, which results in a BLUE of 44.68.

Alexander A S Gunawan et al. [21] On the intelligent humanoid robot, presented their research on the Indonesian query response System to answer arithmetic problems of words using the pattern matching method. The goal of this article is to demonstrate how the Question-Response System uses natural language processing and model matching to answer arithmetical word problems.When given an Indonesian arithmetic word problem, the robot decodes the Indonesian word in the English text, solves the conjunction problem, co-references the problem, preprocesses the problems, analyses the questions, represents the knowledge, and finally answers the problem. They used the Natural Language Toolkit and English NLP in their investigation. According to the results of the trial, the QA System's

accuracy for each model ranges from 80% to 100% depending on how complex the word problem is to understand, and the response time is fairly slow, with an Average turnaround time of 1.12 minutes.

Olga Popova et al. [22] used the creation of a binary tree of the QA System to give life to a method of intelligence amplification (response to questions). They propose using the lesson activity on a remote learning System, such as Moodle, to implement the tree. They also used the Add Content page to provide their own implementation tree methodology for this activity. New reference points for future research were given in the paper. The authors demonstrated how each lesson activity was implemented using a tiny binary tree.

Olga Popova et al. [23] the description of the language of choice criterion and the language of binary relations, i.e. the 1st and 2nd language of choice, could not solve a number of difficulties. These issues were then manually handled without the use of artificial intelligence. For example, the choice of one of the 2 solutions is contingent on the presence of other options; preference has no value. The "normal," "Average," and "exceptional" selection rules, among others, have been utilized here.The authors solved one of these difficulties using their preferred 3rd language description, in which they introduced the mathematical object as an abstract data type - the binary tree of the questionresponse System. The authors used this tree to develop an amplification of the intelligent System-oriented problem "Optimal" decision-making. This methodology allowed for the selection of the "best suited" procedure from among the various options.

Junichi Fukumoto et al. [24] proposed a browsing approach for obtaining a right answer via user involvement for the QASystem. When a question is asked, the QA System narrows the search area so that the user can focus on the intended subject of the documents they wish to see. The QA System has picked a keyword to determine an appropriate topic among the recovered document topics, and it interacts with the user to determine that this might not be the case. The algorithm finds documents using these keywords and expands the search space to improve the applicants' ability to provide accurate replies.They experimented on 10 questions and demonstrated the effectiveness of their interaction methodology in answering the question.

Seena, I. T et. al. [25]Limited work has been conducted in particular for the Malayalam language due to the clustering nature of southern Indian languages. There was a requirement for the user to have a specific System that gave the proper response while they were searching for the right answer to their inquiries. Malayalam QASystem could be a solid start for future Malayalam work.

They were supposed to get factoid responses to queries in Malayalam from a set of Malayalam documents in this article. To determine the exact factual replies, the TnT tagger was used to shape the word corpus.Using a new perspective taken from model-based software development, Katia Vila et al. [26] highlighted their strategy for addressing the challenging issue of mechanically adapting the Question-Response (QA) System in limited domains in a logical, well-structured, and complete way. They demonstrated how to use several types of Knowledge Organization Systems to fit issue diagrams from an ODQA system to a new restricted field (KOS). I the relevance of a QA System to diverse regulated areas using only one document set and the existing KOS, (ii) model portability to any QA System, and (iii)

integration of different types of KOS using a single meta-model are the primary strengths of their method.

Hironori Takeuchi et al. [27] investigated an investigative System in which clients ask questions regarding a certain service and receive reliable answers. Question and answer technology was vital within the search System, and several programming modules for QA technology were given as APIs. They would compile training data with pairs of responses and hypothetical questions using a ML module such as categorization-based QA technology. They demonstrated that the survey System's quality is dependent on the training data's quality, and that training data must be relevant and accurate. To respond to relevance, they must first obtain a complete set of answers from the training data.

They proposed a technique for obtaining a set of replies by arranging knowledge information about a service necessary for a survey System to meet this problem in the creation of practical systems about training data. They organized the processes of obtaining into a customer behavior pattern and defined the different sorts of knowledge for each phase. They also established a function-service paradigm and used it to introduce steps in the collecting of service elements. They sought knowledge information for a set of replies in the training data using the service components acquired and the relationships among the sorts of knowledge information required.

In a case study, they demonstrated that using the suggested technique, they were able to collect service items more extensively than when subject matter experts collected knowledge information in their own way. They demonstrated that they can collect nearly every response that clients can request by using the suggested technique on a survey design project, avoiding situations where training information does not meet the eligibility criteria and the designer and service provider must create a new answer to a question that the System cannot adequately answer.Brijendra Singh and Anbarasi M identified IoT in the education sector in terms of campus energy management, smart classroom management, student tracking and monitoring, and intelligent learning[28].

Rajni Devi and Mohit Dua [29] The objective Hindi language questions and answers were discussed. The authors compared 9 different indicators of similarity as well as 2 different categorization algorithms for extracting relevant data. Smith Waterman surpasses other similar performance evaluation functions, according to the data. For 2 distinct test datasets, the K-Nearest Neighbor algorithm (KNN) scores 97 percent, 95.6 percent, while the NN algorithm scores 93.3 percent, 95 percent, respectively.

Abdelghani BOUZIANE et. al. [30], they surveyed a variety of Question-Answer Systems (QAS). They also provided statistics and analytics. This paper can pave the way for researchers to choose the right solution to their problem. They can see the mismatch; they can therefore suggest new systems for complex queries. QAS techniques can also be adapted or reused to address specific research issues.

A methodology for extracting information was proposed by Walaa A. Elnozahy et al. [31] in the context of the research project "LET'SeGA" to assist educational institutions in selecting candidates for various programs. This would also help with marketing decisions for university programs that target specific student categories. An ontological model for academic data is established as part of the suggested framework, and it is used to enhance student recruitment and retention in order to assure post-admission success.

Sanjay K. et. al. [32], a taxonomy was proposed to characterize QA systems, briefly review and provide qualitative analysis of the major QA systems described in the documentationFinally, in order to provide an overview of the spectrum of research in this area, a comparison of various methodologies based on specific elements of the QA System was selected as essential in their study.The main outcome of this paper is the development of a suggested System that incorporates the respective issues and stages that drive the user to formulate policies and measures that could efficiently solve the problems of the urban road freight transport System in its area of responsibility, according to Eustace Bouhouras and Socrates Basbas [33].

Jovita et al. [34]The purpose of this study, according to was to use the vector space model to describe knowledge and determine the answer to a given question. While certain techniques for answering questions have been created, such as the n-gram, model-oriented response, reversible transformation, and so on, attempts have been made to employ the vector space model. The inquiry will be compared to knowledge to see how similar it is. Two Indonesian ministries, the Minister of Education and Culture and the Minister of Tourism and Economic Culture, provided the sample data used to test the model. During the experiment, a total of 150 questions were asked. For each question word, a total of 25 questions were received.The research disclosed a recall of 0.662, accuracy of 0.548, and F-measure of 0.580. Unfortunately, users had to wait an Average of 29 seconds for a response.

Mounika et.al. [35] suggested an bag of words with n-grams for extraction. They used 16,000,000 tweets linked to the depression and compared with the performance of various classifiers like Logistics Regressor, Support Vector Machines, Adaboost, Bayes and MLP. Of these classifiers, logistic regression and MLP with TF-IDF + n-grams were accurate to 81%.

III. Methods for QA

For conducting our experimentation, we considered the following comprehensive methods of QASystem.

a) Feature Based Methods
b) Bidirectional Long Short-Term Memory Model (BiLSTM)
c) Pre-Training and Fine-tuning with BERT

a) Feature Based Methods

Feature is always the essential part of any based model. In QASystem. The inherent set of features of a question helps to find the précised answer from the paraphrase. The idea ofthis technique is to generate the candidate answers list suitable for the chosen or given questions with the help of the parse-tree constructed from the given collections of training samples.

The sequence of steps for building parsetree is given below.

(i) Generate a list of candidate answers F (p,q,a)
(ii) Considered only the constituents in parse trees
(iii)Define a feature vector
(iv)Word/bigram features
(v) Parse-tree matches
(vi)Dependency labels, length, part-of-speech tags

b) Bilstm-based models

One of the main drawback of the traditional RNN (Recurrent Neural Network) is vanishing gradient which made the model with lengthy sequence of neural nodes to achieve the desired result. LSTM addresses this issue by enforcing multiple number of gates over the sequence of nodes such as input nodes, output nodes and forget gates. So the information pertained at a node could be propagated to a faraway nodes down to the line.The biLSTM enables both the direction in forward and backward propagation during the training phase of building the mode. The vectors obtained in each directions are concatenated together into a single vector. It is similar to run tow LSTM on question text and context paraphrase.

The comprehensive sequence of steps shown below.

- Encode the question using word/char embedding's; pass on an biLSTM encoder
- Encode the passage similarly
- Passage-to-question and question-to-passage attention
- Modeling layer: another BiLSTM layer
- Output layer: 2 classifiers for predicting start and end points

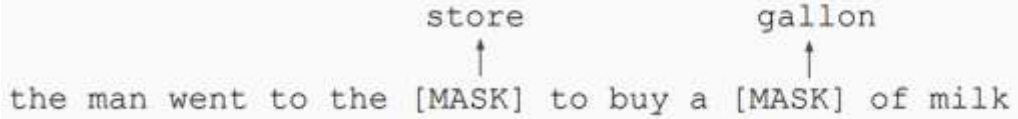
The entire model can be trained in an end-to-end way

c) BERTbased Models

The BERT is a bidirectional unsupervised pre-trained language model constructed with help of any one of the following techniques. It is similar to the concept that whenever we read any story book in English, first of all we should learn English then only we could understand the meaning of each and every sentences in the given story. Without knowing English we could not understand the story. The pre-training is similar that of learning English and the fine-tuning process is analogous that of understanding the meaning. This understanding helps for answering any question asked about the context of the story. Two familiar techniques were being used for building the word vectors from the given corpus of text.

(i) Masked Language Model (MLM)
(ii) Next Sentence Predictions(NSP)

(i). Masked Language Modeling (MLM): This model mask out 'k' percentage of the input words, then predict the masked words using sequence to sequence deep learning neural network. We always use 15% as the value of 'k' where k is the number of masked words. The sample Sentence using MLM is given below.



(ii). NSP: This model uses pairs of sentences as input and learn how to predict whether the second Sentence is the next Sentence in the original paragraph pair or not. It is given below.

$$\text{Input} = \texttt{[CLS] the man went to [MASK] store [SEP]}$$
$$\texttt{he bought a gallon [MASK] milk [SEP]}$$
$$\text{Label} = \texttt{IsNext}$$

$$\text{Input} = \texttt{[CLS] the man [MASK] to the store [SEP]}$$
$$\texttt{penguin [MASK] are flight ##less birds [SEP]}$$
$$\text{Label} = \texttt{NotNext}$$

The figure 2 shows the consolidated pre-training and refinement processes for BERT. Irrespective of outputs of the model, the same model architectures were used for preliminary training and development of streaming tasks. The standard parameters of the pre-trained model were used to initialize the models for various downstream tasks. When doing a streaming task, all parameters are fine-tuned. [CLS] is a special symbol added before each entry example, and [SEP] is a special separation token (for example split questions and answers). As shown in figure 2, the input representation of a word token is built by combining the appropriate token, segment, and position integration.

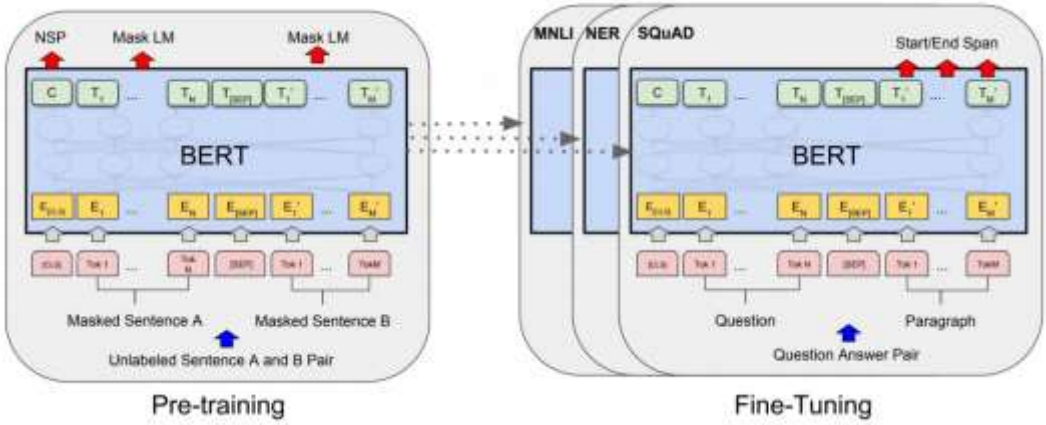


Figure 2: Standard BERT Model with pre-training and Fine-tuning for QA **[**Jacob Devlin]

Traditionally the researchers were used probabilistic language model for understanding the language and its context well. But, it creates a large dataset and need very big language model to support all the words or vocabulary of the language is concern. Later, due to the sequential structure of language, the RNN was used for training the models, but they were very slow in learning convergence. Now, people have started to use the pre-trained model constructed using bidirectional Long ShortTerm Memory (LSTM) with transformer attention mechanism to support speedy convergence and to save the training time of the model.For each down streaming task, the BERT model adds ahead over the top of the model as an additional layer. Here, to carry out the QA task, the model added a new questionanswer head

to the top of the BERT, that estimates the scope of the answers found in the given paragraph as start and end token location settings. All the words between token start and token end have been considered an answer to the question. BERT finds and highlights an extent of text segment contains the answer to the given questions just determining which token word marks the beginning of the answer which token word marks the end. This is similar to the classifier that categorizes each word token as part of the beginning or end token.

Additionally, the model generates 2 set of weights one for start token and other one for end token with same dimensionality similar to the output embedding of each word token. The dot product is performed between the output embedding's of all tokens along with the start and end token weights. Finally, the softmax activation is applied into all the tokens for the start and end token to produce a probability distribution over tokens. The final maximum probability with its respective token was identified at last.

Algorithm 1: Algorithm for QAsystem

Input : Question Text(QT), Context Paragraph(CP)
Output : Answer Text

```
1  input_ids=TOKENIZE(QT,CP)
2  tokens=conver_ids_to_token(input_ids)
3  SegA,SeqB=SplitSegment(tokens)
4  segment_ids=calculatedSegment_ids(SegA,SeqB)
5  model=Construct Model(input_ids,seqment_ids)
6  start_token=model.start_logits()
   end_token=model.end_logits()
7  Answer_Text=get_phrase_between(start_token,end_token)
```

Figure 3: Algorithm for QAsystem

The figure 3 shows the algorithm of QASystem based on the pre-trained BERT model.The collection of vocabulary present in the question text and context text are tokenized using BERT model and then the series of word vectors were generated for each of the tokens. To formulate and supply the essential inputs into the deep learning network model we generated different segments with embedding's also. The constructed deep learning model accept these input ids and segment ids and outputs 2 tokens namely start token and end token which in term decides the location of starting point and ending point of the answer.

IV. Experimental Results

The QA problem could be formulated as a text generation problem, whereas the text generated form the passage is nothing but the answer. The QA System accepts passage and questions, then try to predict the answer from the passage. The BERT, AlBERT, XLNET and ROBERT are all such common widely used retrained models for QA systems. We used the Hugging face QA pre-trained model to conduct the experiment and develop the QA model. Our own dataset, made up of questions and answers from the course "OOP," was used to fine-tune this pre-trained model.We tested our QA model with our own test dataset and the performance is measured and analyzed with various metrics as shown in the Figure....To accomplish this task we have chosen a collection of course materials prepared by the experts of the course "Oops". First, we have done the pre-processing task such as removal of stop words and conversion of all words to upper case words. The cleaned text passage had been

taken into the QA task where it find out the answers with its probabilities. The final answer has been filtered according the probability threshold and then visualize the correct answers. The experimental steps are given below.

Step 1: Install the transformation library - Higgin face Transformer for BERT

Step 2: Loading the Model BertForQuestionAnswering

Step 3: Initialize the Tokenizer using the above model

Step 4: Fine-tune the model on our own dataset constructed for the subject "Oops" (OopsQA)

Step 5: Run the OopsQA through the loaded model

Step 6: Predict the result by picking the span for the total score.

The BERT SQuAD model architecture and its parameters are shown in the table 1.

On top of the BERT pre-trained transformer model, an additional new layer of question and answering head was added to perform the QA process. This added additional layer estimate the start token and end token of the answer portion in the context paragraph(Graph shown in Figure-7). The size of beginning of token and token is same as that of the embedding vector used in the deep learning layer stack.

Model	**No. of Layers**	**Embedding Size**	**No. of Parameters**	**No. of Attention Head**
BERT-Base Uncased	12	768	110M	12
BERT Large Uncased	24	1024	340M	16

Table 1: Model Parameters of BERT

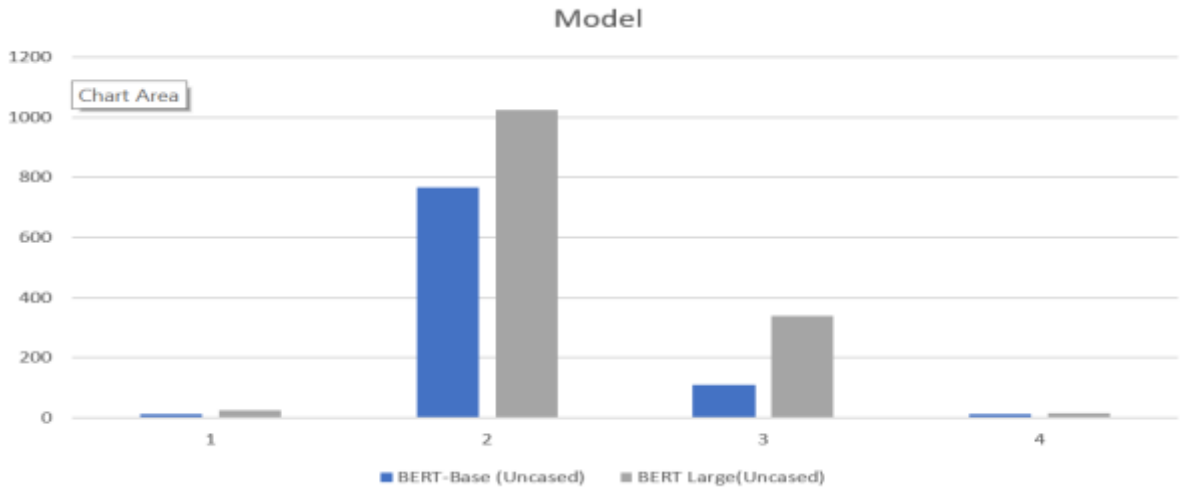


Figure 4:Model Comparision BERT-Base & BERT Large

The QASystem is a most exciting development that has been answering the questions from large unstructured data such as emails, SMS, social media posts, user's blogs, log files, financial statements, educational data and the list goes on increasing. The evaluation of the QASystem is measured by 3 parameters mostly named like Precision, recall and F-measure.

The precision is used to determine how many responses are correct for each of the questions we have in the test data and it is shown in the equation 1.

Precision (Q) = (Total no. of correct answers for the Question Q) / (number of System answers for question Q) Eq. (1)

The Precision is defined as the ratio of the number of responses to the total number of projected answers.The Recall parameter indicates how many of the answers that are predicted are in the ground truth gold standard / total number of ground truth gold standard answers for Q.

Recall(Q) = Total no. of correct answer predicted by the System for the question Q Eq. (2)

Recall formula is shown in Eq.(2)

In the golden standard of worldly truth, recall is the ratio of the total number of shared responses to the total number of answers.

The third parameter F-measure or F-Score is the weighted mean between Precision and recall and has been estimated as follows:

F-measure = (2x Precision x Recall) / (Precision + Recall) Eq. (3)

F-measure formula is shown in the equation 3.The next one is the exact match.For each QA pair, it computes the score by matching all the characters in the true answer with the characters of the predicted answers. It is always either 0 or 1.The intent of the Bilingual Evaluation Understudy (BLEU) score is to count the n-gram overlaps in the responses; it takes the highest count for each n-gram and fixes the n-gram count in the candidate response to the highest count in the gold standard reference. The other popular metrics are given below:

(i) Mean Average Precision (MAP) and
(ii) Mean Average Precision (MAP)

Mean Average Precision (MAP) is one of the pupular performance metric to measure the score of any information retireval System compared with ground truth to this.

If all of the relevant answers returned by the QA System matches exacltly with the ground truth, then the MAP score of the System is 100%.

Consider the rank position of each of the relevant answer returned by the System as AA1, AA2, AA3...AAN. The precision value at each such answer have been evaluated. The average of these Precision is nothing but the MAP.If we receive 5 top answers for the questions and 3 of them are relevant but 2 of them are irrelevant in the order shown as

Answer	**Sequence**	**relevant/not**
AA1	1	Relevant

AA2	2	Irrelevant
AA3	3	Relevant
AA4	4	Irrelevant
AA5	5	Relevant.

Then the MAP of the answer = 1 / Total number of relevant answer (Precision@each sequence)

=1/3(1/1+0/2+2/3+0/4+3/5) =0.76;

MRR(Mean Reciprocal Rank)is a metric used to evaluate the QA System hich returns a list of ranked answers to a particular question while the rank indicates the position of the highest ranked answer. It returns 0 if there is no answer found by the QA System. If there are multiple questions, the MRR is the Average of the Q reciprocal rank is shown in the equation 4.

$$\mathrm{MRR} = \frac{1}{Q}\sum_{i=1}^{Q}\frac{1}{\mathrm{rank}_i}$$

Eq. (4)

The scores of the MAP and MRR with BERTBase model and OopsQA dataset were shown in the table 2. The plot of the same has been visualized using XY line chart is shown in the Figure. As we observe the scores and plots the performance of the QASystem has been increased upon applying the course specific dataset over the BERT-base model. The gap between the MAP and MRR of the oopQA fine-tuning dataset is small as we compare the gap between the MAP (Shown in Figure 5) and MRR of the BERT-base model (Shown in Figure 6).

	MAP Fine-tuning		**MRR Fine-tuning**	
Epochs	**BERT**	**oopQA**	**BERT**	**oopQA**
1	0.78	0.89	0.81	0.91
2	0.79	0.891	0.82	0.912
3	0.72	0.92	0.74	0.942
4	0.7	0.932	0.74	0.956
5	0.823	0.94	0.84	0.956
6	0.832	0.96	0.85	0.975
7	0.65	0.968	0.68	0.975
8	0.814	0.971	0.68	0.98
9	0.821	0.974	0.801	0.98

Table 2: Epoch versus Performance MAP and MRR Scores of BERT-base and oopQA

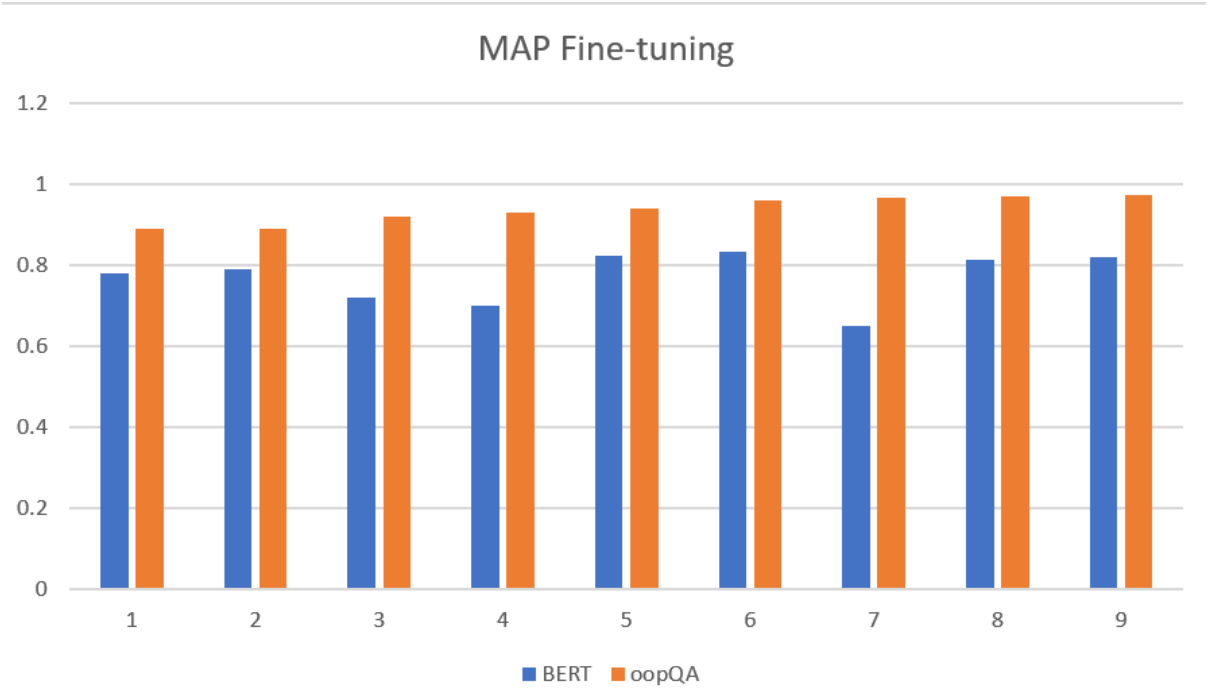


Figure 5: MAP Fine-Tuning Graph

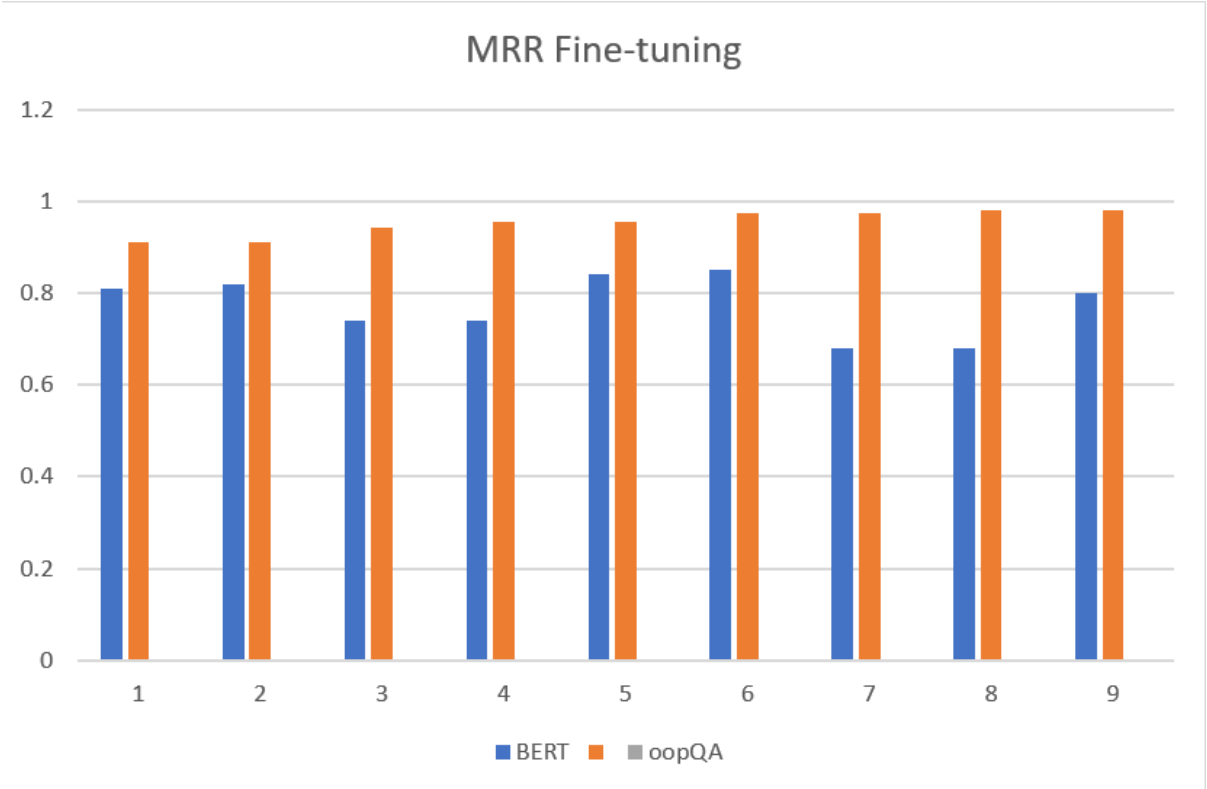


Figure 6: MRR Fine-tuning Graph

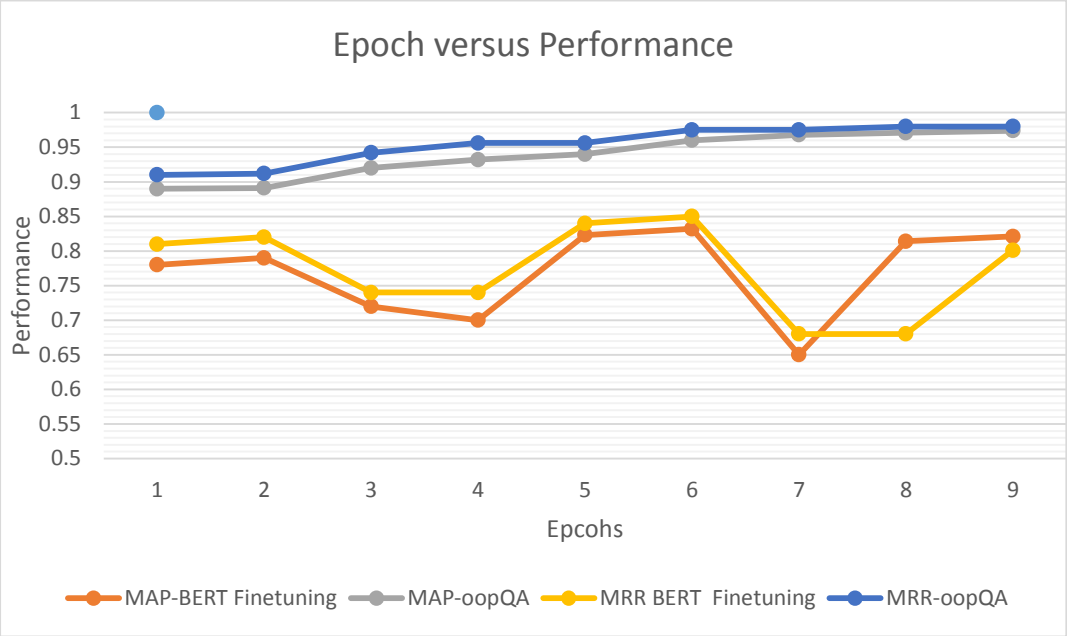


Figure 7: MAP and MRR Plot of BERT based model with oopQA

V. Conclusion

In this paper, we explored how BERT architecture was used on a questionanswering task trained a version of the BERT-base model on SQuAD version 2 data. The performance of the trained model has been visualised in terms of 2 measures namely MAP and MRR. We encapsulated the e-learning learning resources as source of paragraph or context and then prepared the various questions and its answers training set. We have used a transformer based on QA systems along with pre-trained BERT transformer fine-tuned with customized dataset named oopsQA. To obtain the global maximum with acceptable performance, we require labeled dataset which could be constructed with the help of meta learning techniques such as few shot learning as an opportunity for the future enhancement over the transformer model which we have implemented in this paper. The empirical statistical results about the BERT and its fine-tuning were discussed and analyzed well.

References

1. Raju Barskara, Gulfishan Firdose Ahmed, and Nepal Barskar.: An Approach for Extracting Exact Answers to Question Answering (QA) System for English Sentences. International Conference on Communication Technology and System Design 2011, vol. 30, pp. 1187-1194. (2012)
2. Calijorne Soares, M. A., and Parreiras, F. S. (2020, July 1). A literature review on question answering techniques, paradigms and systems. Journal of King Saud University - Computer and Information Sciences. King Saud bin Abdulaziz University, vol. 32, pp. 635-646. (2020)
3. Sriram Aryaman, Sinha Diptanshu and B, Valarmathi and N, Srinivasa Gupta, Performance Comparison of Deep Learning Algorithms for Sentiment Analysis (January 18, 2021). ICICNIS 2020, Available at SSRN: https://ssrn.com/abstract=3768228 or http://dx.doi.org/10.2139/ssrn.376822, (2021)
4. Alami Hamza, Noureddine En-Nahnahi, Khalid Alaoui Zidani, Said El Alaoui Ouatik.: An arabic question classification method based on new taxonomy and continuous distributed

representation of words. Journal of King Saud University –Computer and Information Sciences, pp. 1-7. (2019)
5. Mourad Sarrouti and Said Ouatik El Alaoui.: A passage retrieval method based on probabilistic information retrieval model and UMLS concepts in biomedical question answering. Journal of Biomedical Informatics, vol 68, pp. 96-103. (2017)
6. Ali Albarghothia, Feras Khatera, Khaled Shaalana.: Arabic Question Answering Using Ontology. 3rd International Conference on Arabic Computational Linguistics, ACLing 2017, vol. 117, pp. 183-191 (2017)
7. Archana S.M., Naima Vahab, Rekha Thankappana, C. Raseek.: A Rule Based Question Answering System in Malayalam corpus Using Vibhakthi and POS Tag Analysis. International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015), vol. 24, pp. 1534 – 1541.(2016)
8. YongGang Cao, Feifan Liu, Pippa Simpson, Lamont Antieau, Andrew Bennett, James J. Cimino, John Ely, Hong Yu.: AskHERMES: An online question answering system for complex clinical questions. Journal of Biomedical Informatics, vol.44, pp. 277-288, (2011)
9 Amit Mishra, Sanjay Kumar Jain: A survey on question answering systems with classification. Journal of King Saud University – Computer and Information Sciences, vol. 28, pp. 345–361, (2016)
10. Jeff Rickel, Bruce Porter.: Automated modeling of complex systems
to answer prediction questions. Artificial Intelligence, vol.93, pp. 201-260, (1997)
11. Alaa Mohasseb, Mohamed Bader-El-Den, Mihaela Cocea.: Classification of factoid questions intent using grammatical features. ICT Express, vol.4, pp. 239–242, (2018)
12. G. Suresh kumar, G. Zayaraz.: Concept relation extraction using Naı¨ve Bayes classifier for ontology-based question answering systems. Journal of King Saud University – Computer and Information Sciences, vol. 27, pp. 13–24, (2015)
13. Stefan Schlobach, David Ahn, Maarten de Rijke, Valentin Jijkoun.: Data-driven type checking in open domain question answering. Journal of Applied Logic, vol. 5, pp. 121–143, (2007).
14. Walaa Saber Ismaila, Masun Nabhan Homsi.: DAWQAS: A Dataset for Arabic Why Question Answering System. The 4th International Conference on Arabic Computational Linguistics (ACLing 2018), Procedia Computer Science vol.142, pp.123–131, (2018)
15. Yashvardhan Sharmaa, Sahil Gupta.: Deep Learning Approaches for Question Answering System. International Conference on Computational Intelligence and Data Science (ICCIDS 2018), Procedia Computer Science, vol. 132, pp. 785-794, (2018).
16. LIANG Zhenqiu.: Design of Automatic Question Answering System Base on CBR. 2012 International Workshop on Information and Electronics Engineering (IWIEE), Procedia Engineering vol. 29, pp. 981-985, (2012)
17. C. Pechsiri, R. Piriyakul: Developing a Why–How Question Answering system on community web boards with a causality graph including procedural knowledge. Information Processing in Agriculture, vol. 3, pp. 36-53, (2016)
18. Hong Yu, Minsuk Lee, David Kaufman, John Ely, Jerome A. Osheroff, George Hripcsak, James Cimino.: Development, implementation, and a cognitive evaluation of a definitional question answering system for physicians. Journal of Biomedical Informatics, vol. 40 pp. 236–251, (2007)
19. Michael Spranger, Dirk Labudde.: Establishing a Question Answering System for Forensic Texts. Procedia - Social and Behavioral Sciences, vol. 147, pp. 197 – 205, (2014)
20. Yogi Wisesa Chandra, Suyanto Suyanto.: Indonesian Chatbot of University Admission Using a Question Answering System Based on Sequence-to-Sequence Model. 4th

International Conference on Computer Science and Computational Intelligence 2019 (ICCSCI), Procedia Computer Science, vol. 157, pp. 367–374, (2019)

21. Alexander A S Gunawan, Pribadi R Mulyono, Widodo Budiharto.: Indonesian Question Answering System for Solving Arithmetic Word Problems on Intelligent Humanoid Robot. 3rd International Conference on Computer Science and Computational Intelligence 2018, Procedia Computer Science, vol. 135, pp. 719–726, (2018)
22. Olga Popova, Boris Popov, Vladimir Karandey.: Intelligence Amplification in Distance Learning through the Binary Tree of Question-answer System. Worldwide trends in the development of education and academic research, Procedia - Social and Behavioral Sciences, vol. 214, pp. 75-85, (2015)
23. Olga Popova, Boris Popov, Vladimir Karandey, Marina Evseeva.: Intelligence Amplification via Language of Choice Description as a Mathematical Object (Binary Tree of Questionanswer system). Worldwide trends in the development of education and academic research, Procedia - Social and Behavioral Sciences, vol. 214, pp. 897 – 905, (2015)
24. Junichi Fukumoto, Noriaki Aburai, Ryosuke Yamanishi.: Interactive Document Expansion for Answer Extraction of Question Answering System. 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013, Procedia Computer Science, vol. 22, pp. 991 – 1000, (2013)
25. Seena, I. T., Sini, G. M., Binu, R.: Malayalam question answering system. International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015), Procedia Technology, vol. 24, pp. 1388 – 1392, (2016)
26. Katia Vila, Jose-Norberto Maz´on, Antonio Ferr´andez.: Model-driven adaptation of question answering systems for ambient intelligence by integrating restricted-domain knowledge. Procedia Computer Science, vol. 4, pp. 1650–1659, (2011)
27. Hironori Takeuchi, Satoshi Masuda, Kohtaroh Miyamoto, Shiki Akihara.: Obtaining Exhaustive Answer Set for Q&A-based Inquiry System using Customer Behavior and Service Function Modeling. International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2018, Procedia Computer Science, vol. 126, pp. 986–995, (2018)
28. Singh, B. and Masilamani, A., 2021. IoT in the Education Sector: Applications and Challenges. Applications of Artificial Intelligence for Smart Technology, pp.192-205.
29. Rajni Devi, Mohit Dua.: Performance Evaluation of Different Similarity Functions and Classification Methods using Web Based Hindi Language QA System. 2nd International Conference on Intelligent Computing, Communication & Convergence (ICCC-2016), Procedia Computer Science, vol. 92, pp. 520 – 525, (2016)
30. Abdelghani BOUZIANE, Djelloul BOUCHIHA, Noureddine DOUMI and Mimoun MALKI.: Question Answering Systems: Survey and Trends. The International Conference on Advanced Wireless, Information, and Communication Technologies (AWICT 2015), Procedia Computer Science, vol. 73, pp. 366 – 375, (2015)
31. Walaa A. Elnozahy, Ghada A. El Khayat, Lilia Cheniti-Belcadhi and Bilal Said.: Question Answering System to Support University Students' Orientation, Recruitment and Retention. Procedia Computer Science, vol. 164, pp. 56–63, (2019)
32. Sanjay K. Dwivedi, Vaishali Singh.: Research and reviews in question answering system. International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013, Procedia Technology, vol. 10, pp. 417 – 424, (2013)
33. Eustace Bouhouras, Socrates Basbas.: Urban Road Freight Transport Systems: Questions and Answers. Procedia - Social and Behavioral Sciences, vol. 48, pp. 2501 – 2512, (2012)

34. Jovita, Linda, Andrei Hartawan, Derwin Suhartono.: Using Vector Space Model in Question Answering System. International Conference on Computer Science and Computational Intelligence (ICCSCI 2015), Procedia Computer Science, vol. 59, pp. 305 – 311, (2015)
35. Mounika, M., Srinivasa Gupta, N., Valarmathi, B.: Detection of Depression Related Posts in Tweets Using Classification Methods – A Comparative Analysis. In: Pandian A., Palanisamy R., Ntalianis K. (eds) Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI - 2019). ICCBI 2019. Lecture Notes on Data Engineering and Communications Techn