Dissertation on

"Automatic Question Paper Generator System-Blooms Taxonomy"

Submitted in partial fulfilment of the requirements for the award of degree of

Bachelor of Technology
in
Computer Science & Engineering

UE17CS490A – Capstone Project Phase - 1

Submitted by:

Jyosna S	**PES1201700792**
Megha M	**PES1201700948**
Radhika Sadanand	**PES1201701702**
Roshan Bhaskar Poojary	**PES1201701628**

Under the guidance of

Dr. Uma D
Professor
PES University

August - December 2020

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

FACULTY OF ENGINEERING

CERTIFICATE

This is to certify that the dissertation entitled

'Automatic Question Paper Generator System-Blooms Taxonomy'

is a bonafide work carried out by

Jyosna S	**PES1201700792**
Megha M	**PES1201700948**
Radhika Sadanand	**PES1201701702**
Roshan Bhaskar Poojary	**PES1201701628**

in partial fulfilment for the completion of seventh semester Capstone Project Phase - 1 (UE17CS490A) in the Program of Study - Bachelor of Technology in Computer Science and Engineering under rules and regulations of PES University, Bengaluru during the period Aug. 2020 – Dec. 2020. It is certified that all corrections / suggestions indicated for internal assessment have been incorporated in the report. The dissertation has been approved as it satisfies the 7th semester academic requirements in respect of project work.

Signature	Signature	Signature
Dr. Uma D	Dr. Shylaja S S	Dr. B K Keshavan
Professor	Chairperson	Dean of Faculty

External VIVA

Name of the Examiners	**Signature with Date**
1. Dr. Suresh Jamadagni	_____
2. Dr. Prafullatha Auradhkar	_____

DECLARATION

We hereby declare that the Capstone Project Phase - 1 entitled **"Automatic Question Paper Generator System-Blooms Taxonomy"** has been carried out by us under the guidance of Dr. Uma D, Professor and submitted in partial fulfilment of the course requirements for the award of degree of **Bachelor of Technology** in **Computer Science and Engineering** of **PES University, Bengaluru** during the academic semester August – December 2020. The matter embodied in this report has not been submitted to any other university or institution for the award of any degree.

Jyosna S	**PES1201700792**
Megha M	**PES1201700948**
Radhika Sanadand	**PES1201701602**
Roshan Bhaskar Poojary	**PES1201701628**

ACKNOWLEDGEMENT

I would like to express my gratitude to Prof. Dr. Uma D, Department of Computer Science and Engineering, PES University, for her continuous guidance, assistance, and encouragement throughout the development of this UE17CS490A - Capstone Project Phase – 1.

I am grateful to the project coordinators, Prof. Rajashree Soman, for organizing, managing, and helping with the entire process.

I take this opportunity to thank Dr. Shylaja S S, Chairperson, Department of Computer Science and Engineering, PES University, for all the knowledge and support I have received from the department. I would like to thank Dr. B.K. Keshavan, Dean of Faculty, PES University for his help.

I am deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro Chancellor – PES University, Dr. Suryaprasad J, Vice-Chancellor, PES University for providing to me various opportunities and enlightenment every step of the way. Finally, this project could not have been completed without the continual support and encouragement I have received from my family and friends.

ABSTRACT

Assessments or exams generally play an essential role in education and as well as in a student's life. It is also considered as the primary indicator of students learning process. Facing exams is also considered as academic readiness and also learning progress .One of the main forms of the assessment used is exams in learning. Asking and setting proper assessment questions in order to attain the expected result from the course is a difficult work for the examiner to deal with.

So, the research on the whole focused on categorizing the queries from all the cognitive level domains that follow Bloom's Taxonomy with better model accuracy and on which technical skills are used to equally cover the questions from all the domain levels in bloom's taxonomy. Eventually, to output the standard question paper automatically few derived rule help to analyze and make it easy to make and output desired outcome. Based on the classification process, a question paper can be redesigned by examiners. Coming up with a desired good accuracy and also a better one than all the research studies has been the notion.

This study follows a rule-based approach to achieve multiclass classification. Then the method of weighting the classes accordingly based on a fuzzy model is used to correctly assign the categories to the questions. Once the questions are classified, they are grouped together according to their difficulty level in order to generate an effective and standard question paper.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER-1

INTRODUCTION

In the present world of education, the assessment process is a dominant and vital venture for educational establishments and universities for student performance evaluation. It is a demanding, time consuming and very tedious job for the tutors to come up with worthy questions.

There are many taxonomies that have been developed in order to define questions so that each question has a significance for its presence. It might be conversational, accuracy-based, social or knowledge according to one Taxonomy whereas another research says one can classify questions into easy, hard and moderate but there was a need of in-depth classification of the cognitive levels of assessment.

Then in 1956 an American educational psychologist Benjamin Samuel Bloom came up with a Rubric that systematically classified the cognitive levels .This was later named as Bloom's Taxonomy. This Taxonomy has three ranked models which are cognitive, psychomotor and affective domains. The cognitive domain is a knowledge based model which is divided into six levels of objectives. These six levels are named as knowledge/remember, analysis/analyse, application/apply, comprehension/understand, evaluation/evaluate and synthesis/create.

The affective domain is an emotion based model which is split into five levels. These five levels are responding, receiving, valuing, characterizing and organizing. The psychomotor domain is an action based model which is divided into seven levels. These seven levels are perception, guided response, set, mechanism, origination, complex overt response and adaptation. Since the main focus here is on examination for educational establishments, the first cognitive (knowledge-based) model is considered to be a better fit for this kind of research.

CHAPTER 2

PROBLEM STATEMENT

Examination is an indispensable role in assessing a student's intellectuality as it is the central and foremost approach used. Hence, there is a need for the educators to systematically frame the assessments that include various types of questions that could precisely test the student's understanding. Bloom's Taxonomy's Cognitive Model is being considered one of the best classification methods to effectively classify the learning objectives.

The cognitive domain is a knowledge based model which is divided into six levels of objectives. These six levels are named as knowledge/remember-recalling terms which do not require any understanding, comprehension/understand-to check the understanding and the ability to organize, application/apply-problem-solving skills are tested using prior knowledge, analysis/analyze-examining relation between things, synthesis/create-building or forming things using prior knowledge and evaluation/evaluate-passing judgements based on a set of criteria.

A question falls into more than one category or level which is the common issue faced by all the researchers that worked on this specific project. When this happens it is more difficult to identify the question as to which category does it belong to. Just making sure that the dataset is filled with effective standard questions alone would help a lot while assigning the weight to each taxonomy category. Hence, this particular study presents a solution to the issue in the form of an Automatic Question Paper Generator System/ which makes use of Bloom's Taxonomy to detect the standard of the question paper generated.

CHAPTER 3

LITERATURE SURVEY AND REVIEW

PAPER – I

3.1 Automated analysis of the exam questions based on Bloom's Taxonomy [1]

In this paper the authors have presented a system which based on a set of rules categorize questions into distinct cognitive categories of Bloom's nomenclature. The rule-based approach makes use of nlp methods to recognize key verbs that aids in the classification of the given input question.

Main steps in the implementation is to design rules in a such a way that the rules should be able to identify the verbs in the input question and assign an appropriate bloom's level to the question , if the the identified verb is in multiple levels then they should refer to the expert's weights for the levels for that particular question to classify it to the correct level.

First before sending the data into this rule-based classifier the dataset of 70 training questions is first preprocessed using techniques in NLP like the removal of stopwords,lemmatization,stemming and at last POS-tagging which helps in identifying the verbs in the question that aids in classification . Then this preprocessed data is fed to the rule based classifier in which rules are designed to categorize the input question.

So if the identified verb is unique to a particular level then this level is assigned to that question else if the verb is present in many levels of Bloom's Taxonomy, then the category weighting technique is used where the experts assign weights to the levels to which the question is likely to belong based on their knowledge . In this way the questions are finally assigned a category

which has the highest experts weight.

3.1.1 STRENGTH

During the classification of the questions, the rules used to identify and determine the taxonomy of the questions are based on the verbs and keywords,So there are ambiguous cases where the keywords in a question might belong to more than one category .This raises conflicts and ambiguities like in what level the input question might fall into. For handling this type of scenario the methodology used was **"category weighting"** that helped in deciding, where weights were allocated to conflicting levels.

3.1.2 WEAKNESS

The rules used to categorize the questions into the taxonomy are very general and simple as they consider only the verbs/keywords .In the category weighting the weights are nominated to the conflicting categories by the experts. Making use of the given strategy might result in a lack of consistency because of the diverse types of cognition knowledge of the experts. There are some cases where questions cannot be put into one category which are not handled by the suggested method.

PAPER – II

3.2 Cosine Similarity and WordNet-based Classifier of Exam Questions [2]

This paper is specifically and mainly centered on classifying the test questions accordingly into its learning objectives in Bloom's taxonomy. Using nlp methods such as lemmatization, tagging ,removal of stop words and tokenization are done before producing the rules that are used for categorization. Similarity based algorithms like wordnet and cosine algorithms are utilized to produce a set of rules that recognizes the category to which the question belongs to and the weight that are given for each question based on Bloom's taxonomy.

Before implementation the test questions were extracted for the pdf documents provided by the user by using the pypdf package, then by using regular expression they got the questions which were stored in the mysql database. As a part of implementation, there are two methods that have been carried out that is the word net classifier and cosine based classifier. In wordnet classifier ,the first step involves identification of verbs for each level in the taxonomy . Now the questions in mysql database are converted into a list of tokens for each question after which enchant module was used to correct the tokens.

Then lemmatization is carried out on each word in the list instead of stemming as the lemmatization process leaves a apt word that retains the same semantic from the actual question. After this the lemmatized words in the list are tagged using POS-tagging with the help of which the verbs were extracted where the tag matched with V and W. Now for each verb in the input question the closeness is identified with the taxonomy verbs in each category which as a result gives wordnet score to each category for each input question .

Next in the cosine based classification method the first step is to identify tag patterns for each level in the taxonomy . Now using the classifier based tagger the to get the tag pattern of the input question .Then the input question's tag pattern is then matched with the taxonomy.s tag

pattern and get the cosine similarity between them .

Now the wordnet similarity and cosine similarity scores are multiple to get the final score based on which the questions were categorized into the top three highest score levels . At last rules were applied to recognise which among the three top levels must the question be put into.

3.2.1 STRENGTH

WordNet similarity algorithm accuracy was primarily dependent on key verbs which could be seen in test paper. Few verbs which were not present in Bloom's taxonomy levels, have made the categorization of the test question into different levels. By combining the question tag patterns with cosine similarity it made obvious that questions can be classified into the correct category.

3.2.2 WEAKNESS

A question can be asked in various ways which may also change the tag pattern due to this might be a case where the same question due the various tag patterns might belong to many classes. Recognition of tag pattern was not accurate as the performance of tagger used was not 100% accurate. Because of this some tag patterns failed to merge and the resulting cosine score was zero for cases. That is how this leads to misclassification of the questions.

PAPER - III

3.3 Using modified TF-IDF and word2vec, question classification based on Bloom's taxonomy cognitive domain[3]

In order to classify the questions automatically, this research suggests a way by drawing out TF-POSIDF and word2Vec as two different features. The calculation of term frequency-inverse document frequency is one feature based on parts of speech, sequence of allocation a significant weight for essential and important words in the input. The process of classification was enhanced using the prior-trained word2vec as the second feature. Then, the blend of these features was sent into three different classifiers; Logistic Regression, K-Nearest Neighbour, and Support Vector Machine for the classification process.

3.3.1 IDEA BEHIND WORK

This research followed the basic steps involved in a classification process. This process started with the collection of data which is a collection of questions from previous work done or books and several other sources. The next step consisted of various pre-processing techniques like cleaning the data by normalizing it, tokenizing and also stemming. Since the main idea here is to calculate term frequency based on parts of speech, POS-tagging is also done as a preprocessing.

Then comes the most important portion of the research where they extract features required. One required feature here is to have a set of important words that are estimated using the TF-IDF method on POS-tagged words which can be used to develop a weighting method. The second feature is that it should boost the classification by mining semantic features of words based on pre-trained Word2Vec. Word2Vec model is being developed with the use of google new dataset that contains more than a billion words.

The main idea behind POS-tagging is to give the verbs that occur in the questions a higher priority. However, this does not infer that other words hold no significance in the classification

process. Because of this reason, semantic feature is also taken into consideration so that words that have similar meaning are aligned together. Combination of these two models produce a set of vectors that basically represent questions. These are fed into various classifiers discussed earlier. Classifiers used here are SVM, K-Nearest Neighbour and Logistic regression. The metrics used for evaluation here recall, precision, F1 measure and variations of the same. The results and performance of these 3 classifiers examined. LR and SVM achieved better results as compared to KNN.

3.3.2 STRENGTH

The taken out and proposed attribute Term Frequency-Inverse Document Frequency (TF-IDF) is highly dependent on Parts of Speech (POS) and is initiated as TFPOS-IDF. To find and decide the proper and suitable rank for word weight on the basis of Bloom's taxonomy, various cases have been scrutinized.

$$w_{pos}(t) = \begin{cases} w_1 & \text{if } t \text{ is verb} \\ w_2 & \text{if } t \text{ is noun or adjective} \\ w_3 & \text{otherwise} \end{cases}$$

Figure 1: Weighting algorithm

3.3.3 WEAKNESS

The dataset mentioned in this paper has only open-ended questions, which consisted of neither true or false nor multiple choices questions. Usage of weighing the words into different ranks on the basis of its type (verb, noun, adjective) can wrongly classify some questions. This is because the method would rank up insignificant words along with ranking up significant words as it carries from question to question.

PAPER -IV

3.4 Using Multi-class Text Classification, identification of Cognitive Learning Complexity of Assessment Questions [4]

The paper focuses on two descriptions of studies to identify the learning complexity of given cognitive questions automatically. The former approach used is the labeled Latent Dirichlet Allocation (LDA). This approach represents texts as arbitrary mixtures over quiescent topics, where everything could be represented by a diffusion of words in the corpus. The second approach uses the Bidirectional Encoder Representations from Transformers famously known as BERT framework that aids the multi-class text classification under deep learning. The algorithm uses prior-trained deep multi-directional presentations from the unprocessed text by jointly setting up either on right and left contexts.

3.4.1 LDA

When we are not sure of what we are looking for (unsupervised) with respect to classification, topic modelling is the method which is preferred. This is because the method discovers unseen concepts and keeps these concepts as a base for classification. LDA is one such method used for topic modelling. Implementing LDA has two diversions. One is something we know which are the words that occur in the dataset as a whole. Second is something we do not know which are words that occur in each topic (Category). In order to find these words, there are few assumptions made. These assumptions are that the number of topics are pre-decided and also the words under these topics are correct.

3.4.2 BERT

BERT is one of the most recent developments in the NLP field. This approach uses feature based learning and classification as a base. The Bi-Directional Transformer has two parts of working which is encoder and decoder. The encoder takes input and the decoder predicts the output for the task assigned. This is a very unique approach as the model processes texts both ways hence

the name Bi-directional. Also, the model masks few of the words while processing and tries predicting them by keeping in mind the words that occur to the left and right of the masked word. This special feature of BERT makes it credible to be used for small datasets and the main work done here is on the input specifically rather than the quantity of input.

3.4.3 STRENGTH

Most machine learning models train them on the text input in a sequential manner, while the entire sequence is read at a go by the Transformer encoder. This particular feature allows the model to learn the factors of a word keeping in mind all of its surrounding words.

This study uses a wide range of questions collected from different subjects and sources making the training better than other studies. Questions without Bloom's Taxonomy action verbs, too, were assigned a cognitive level correctly by the algorithms due to the usage of bidirectional approaches in understanding the context of a word.

3.4.4 WEAKNESS

The error occurred in cases where the text structure of the question stem was similar.
What does < subject > mean ? was tagged as remember / knowledge level and What does < subject > do ? was tagged as comprehension / understanding level, as the former has a more specific answer. Unless there is a lot of training data, the accuracy cannot be improved. Each training question need to be carefully annotated by the annotators so that the training data is correct.

PAPER -V

3.5 Classifying Question Papers with Bloom's Taxonomy Using Machine Learning Techniques [5]

The main goal of the researchers of this paper is to show how Bloom's Taxonomy can be employed to determine the difficulty and the precision of the given university question paper.
They have used NLP techniques , feature extraction and many ML techniques with a view to classify the questions in the question paper and compare the respective results to determine the best model to serve the purpose.

3.5.1 IDEA BEHIND WORK

The first step of implementation is feature extraction to collect the most frequently appearing words that could aid question classification. This study required proper understanding of the ML models for the sake of determining the most suitable ones. This research aided in selecting few of the models that were later used in the later stages.

Before applying these models, an ample amount of work has to be done to the data to make it suitable to be fed into these models. This work includes tokenizing of a dataset containing 1042 labeled questions. Other basic preprocessing is done like Stopword removal and Punctuation removal. Since the data is being fed into ML models, the entire data is being split into test and train. Here intervenes the first step i. e feature extraction. After this, frequencies are determined for both training and testing data. The data is then fed into the studied models.

The models selected and used for this are K-Nearest Neighbours, Neural Networks, Logistic Regression, Random Forests, Linear Discriminant Analysis (LDA), Decision Trees and Support Vector Machine (SVM). These trained models enter into the testing stage where they are being tested against the test data. This gives us an idea as to how accurately the models are trained.

3.5.2 SETUP

A web app is done so as to make the entire system usable. The user enters the questions individually or as a set. These questions are analysed by the built model and are classified accordingly. The category percentage is outputted with respect to the number of questions in the input. If no question belonged to any category, it is shown as 0.00%.

3.5.3 STRENGTH

Since more than one ML models are used for training the data, it is known to be a better approach instead of using any one of the methods individually. Also, this way one can deeply examine the behaviour of these models on feeding the different types of data.

3.5.4 WEAKNESS

At the same time, the fact that models are trained on the dataset and so they work accordingly should not be left behind. This is because, even though the models are enhanced the main factor in better performance would be played by the training dataset.

CONCLUSION

1. In [1], a system where the questions are categorized into Bloom's levels by using a set of rules which designed based on the verbs present in the question and category weighting method is used in a case where the verb belongs to more than one level of the taxonomy so that the question is assigned to the level which more weightage.

2. In [2] ,classification of the questions are done based on combined results of WordNet and cosine classifier scores . The WordNet similarity gives the closeness score between the verbs in the question and the taxonomy verbs whereas the cosine similarity gives the closeness score between the input question tag pattern and the taxonomy tag pattern.

3. Paper [3] proposes a procedure to classify open-ended questions on the basis of Bloom's cognitive levels. It uses a modified version of TF-IDF i.e TF-POSIDF in calculating frequency of words and later uses word2Vec word embedding model to align semantically similar words as vectors together.

4. In [4], a feasible solution is provided with a computational approach that includes deep learning and ML techniques that are instructed over an exciting variety of assessment questions. Also, questions without action verbs were classified with the help of WH- words present in them.

5. In [5], performance of various ML techniques is observed by feeding them with a dataset of 1024 questions. The study concluded with the fact that ML techniques like Logistic regression and Linear discriminant analysis showed better results compared to other.

PROJECT REQUIREMENTS SPECIFICATION

PROJECT REQUIREMENTS (FUNCTIONAL):

- As an input the system must accept a number of questions and questions must be in only texts and double the number of questions expected in the paper accordingly as input.
- The system must be well-trained so as to classify each question accordingly. The questions accepted as input must be classified into any of the classes (bloom's taxonomy levels).
- The system must generate a high-quality Paper (pdf/word) as output.

PROJECT REQUIREMENTS(NON-FUNCTIONAL):

- Considering usability, the system must be easy to use with basic features while adding questions.
- When it comes to performance, the system must classify and generate paper faster i.e the response time should be as small value as possible.
- Maximized accuracy as in efficiency while classifying the questions with respect to the work that's already done using similar approaches.

Assumptions made are questions that are entered are true and meaningful and for classifying the cognitive level of a given question is that it should belong to only one particular class.

SYSTEM REQUIREMENT SPECIFICATION

The system must be well-trained so as to classify each question accordingly. The questions accepted as input must be classified into any of the classes (bloom's taxonomy levels).The system must generate a high-quality Paper (pdf/word) as output. Institutions can use our application with the help of which they can set a standard question paper for conducting admission tests.

Tutors/Teachers can also use this application to create a question paper to test their students cognitive level of thinking which actually helps them in saving their time. Our application can also be used by test generator platforms so that they can generate better test papers in such a way that they cover all cognitive levels instead of randomly choosing the questions from the question bank given by their users. The system must accept no. of questions and questions (only text and double the number of questions expected in the paper) accordingly as input.

For software requirements,

Language – Python,Scikit learn

Tools-Anaconda Navigator

Techniques – NLP, ML Algorithms and Taxonomy Rules

Data requirements – Huge Questions Data sets,SQLite

SYSTEM DESIGN

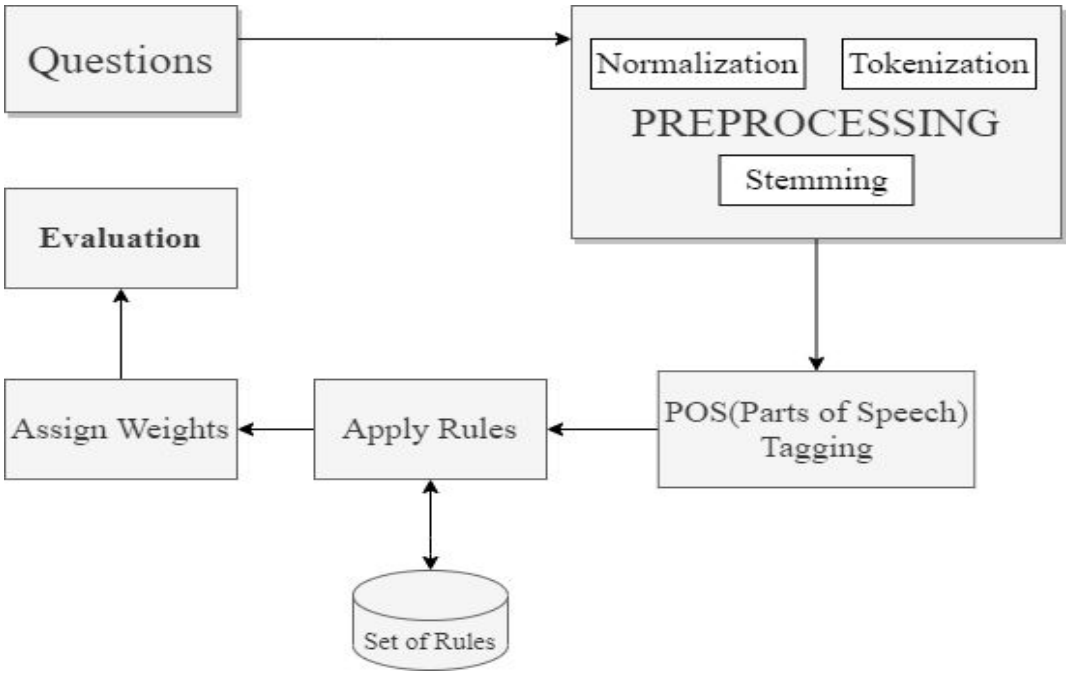

Figure 2:Design model

Designing an Automated question paper generator system that generate questions depending on the type of assessment. Given a question paper, uses Bloom's Taxonomy rules to predict the Standard of the Paper.

ASSUMPTIONS

The questions entered are true and meaningful. For classifying the cognitive level of a given question is that it should belong to only one particular class.

DEPENDENCIES (SOFTWARE)

Python with suitable library (Dask and FPDF) and huge data sets of respective subjects questions(Operating software ,computer networks etc) and Questions dataset which should be provided by the user.

RISKS

Poor data Quality,Low Accuracy and Computational Complexity (training huge data sets).

DESIGN CONSIDERATIONS

Design Goals

• The application's goal is to make the best possible and suitable question paper by using the questions given by the user.

• The efficiency, accuracy, speed must be better than already existing models.

• Reduce the response time.

Architecture Choices

- The architecture has a front-end web interface with an option to upload documents full of questions.
- The application sends the document to the backend service or model which has classifier to extract the keywords to classify every question on the learning level basis.
- The keywords that make questions fall under two or more than two categories will be made to undergo a fuzzy based model to be unambiguous to get classified under one category.
- As the ambiguity will be gotten rid of through fuzzy model there will a better model accuracy.

High Level System Design

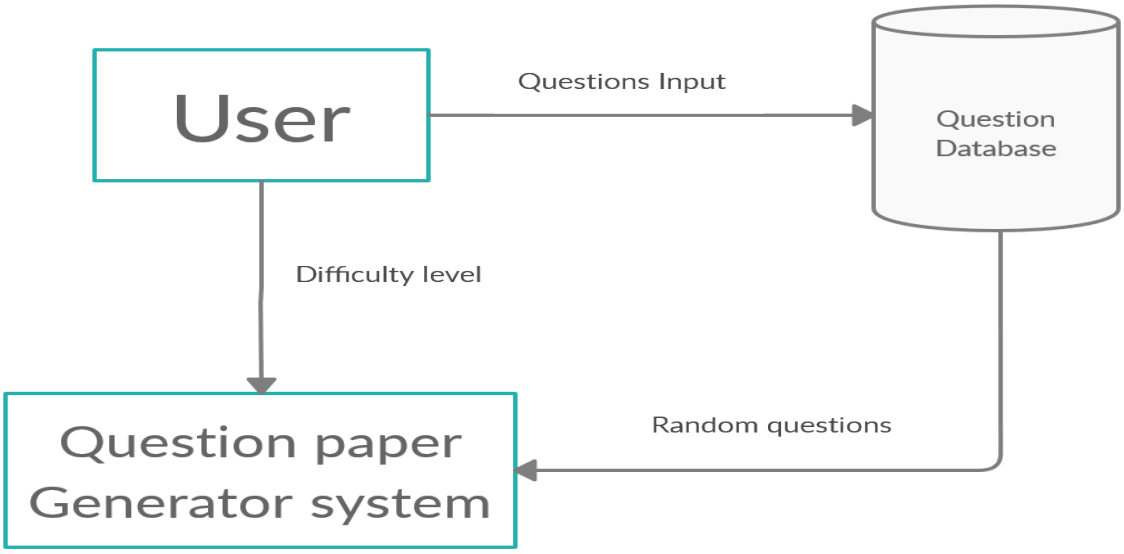

Figure 3: High level system design of the proposed model

When the user gives input as a dataset of questions, the first step of implementation is feature extraction to collect the most frequently appearing words that could aid question classification. Study needs proper understanding of the ML models in order to determine the most suitable ones. When the user gives input as a dataset of questions. In deciding the difficulty level of the paper, ML models play an important role. Questions are checked whether or not they already exist in the database. If a question already exists in the database, then the same previous already categorized classification is given.If not, it follows a procedure in which questions are picked randomly and given to the model that decides the difficulty level of the question paper.

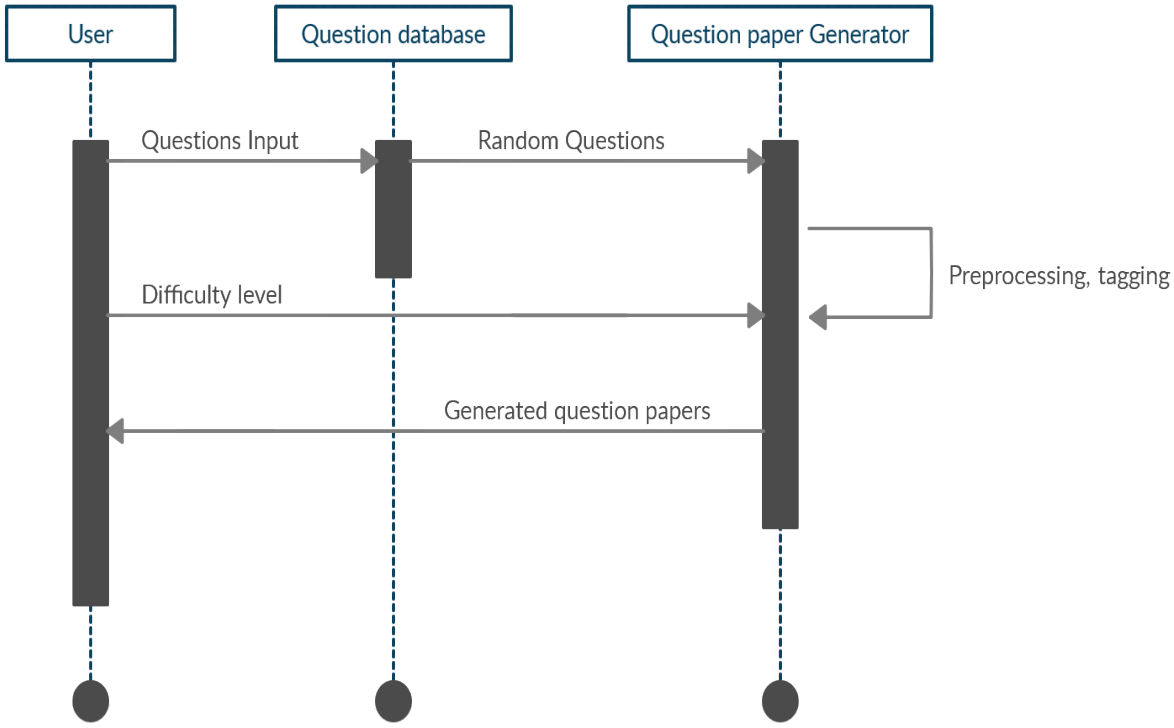

Figure 4: Running view of the system

In the above diagram as the user gives input,questions that are randomly picked are fed to the model and fuzzy based logic model. After undergoing processes like preprocessing and tagging,a set of questions are generated as output for the user's use.

Design Description

Master Class Diagram

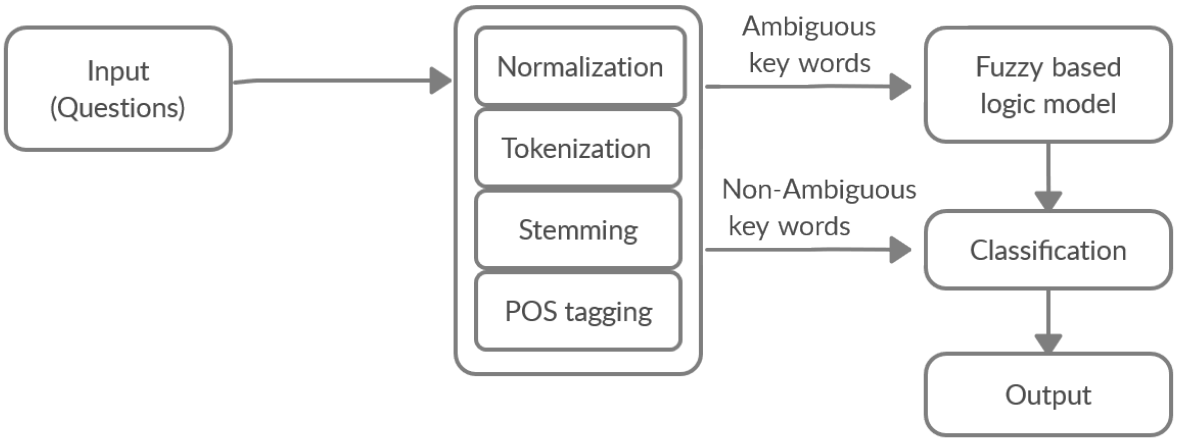

Figure 5: Master class diagram

Above diagram highlights the classification of Ambiguous keywords that are fed to fuzzy based logic model to get rid of the ambiguity while Non-Ambiguous keywords classified are merged with those keywords that are ambiguity free later for the classification.

Reusability Considerations

- Reusing the dataset of question to train the model for an accurate classification.

State Diagram

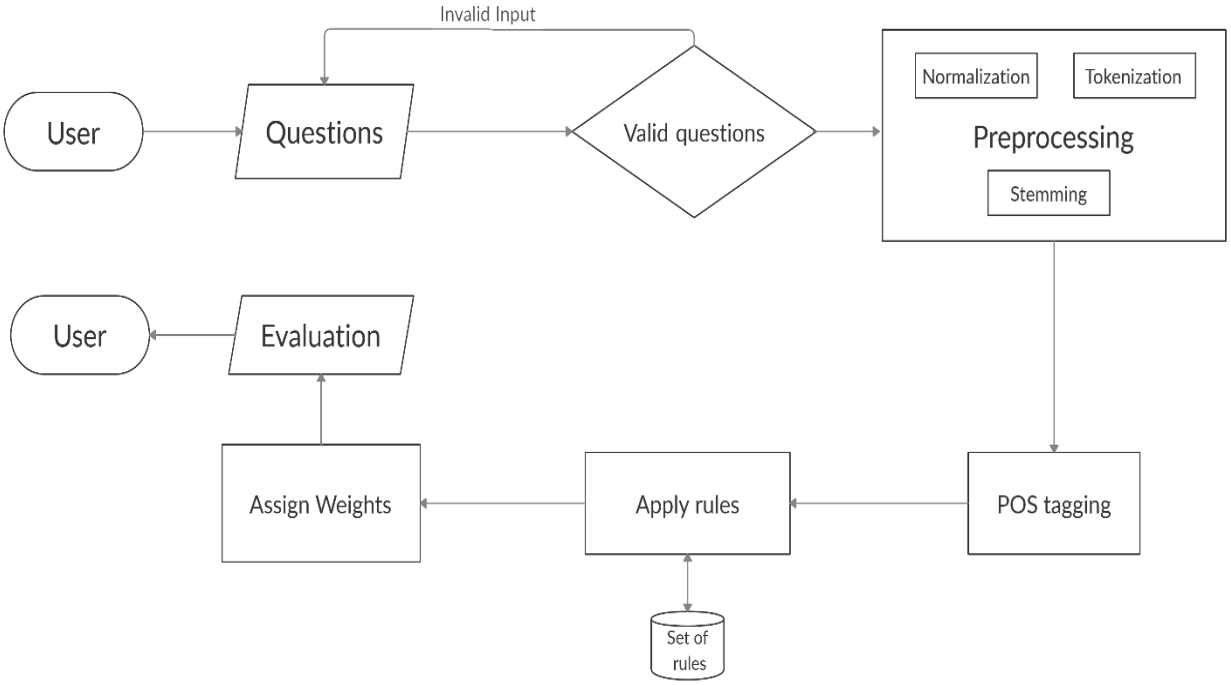

Figure 6: State Diagram

Before undergoing processes and procedures the ML model and fuzzy based logic model has to offer,only valid input is accepted. Then the questions undergo preprocessing where tasks like normalization, tokenization and stemming take place. After tagging the words set of rules which are already present in the database will be applied. Then assigning weights takes place before the final evaluation which will be presented before the user.

User Interface

The user will have an option to enter the questions dataset and an option to choose the difficulty level, after uploading the documents and selecting the application will do it's work and present the user with the final question paper.

Packaging and Deployment Diagram

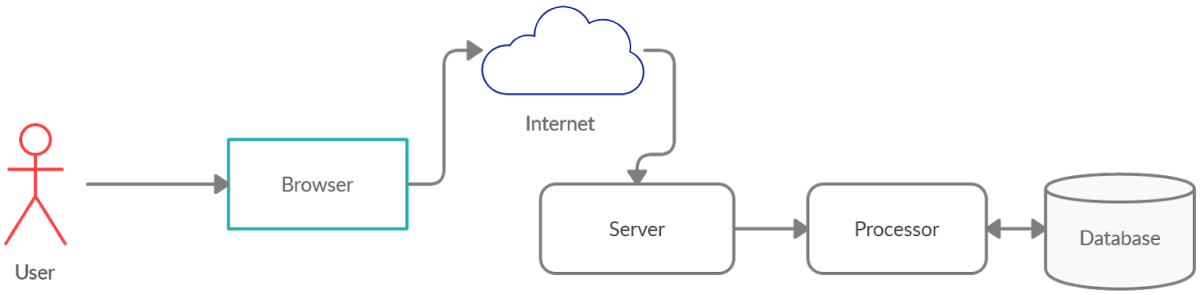

Figure 7: Packaging and Deployment diagram

Deployment diagram is used to visualize the execution architecture of the system. The user interacts with the browser to access the application, which uses the internet to fetch it from the server it is based in. The application runs with the help of a processor which interacts with the database to get the required data.

Help

- Will provide detailed instruction in the help section of the website.
- Will provide contact details in the help section (email ids).

CHAPTER 7

IMPLEMENTATION

This study needs a detailed understanding of the Bloom's Taxonomy levels. Below shown figure depicts the hierarchical model of the cognitive model of Bloom's Taxonomy starting from remember (being the least) to create (being the highest). The figure also shows what each of these levels signifies.

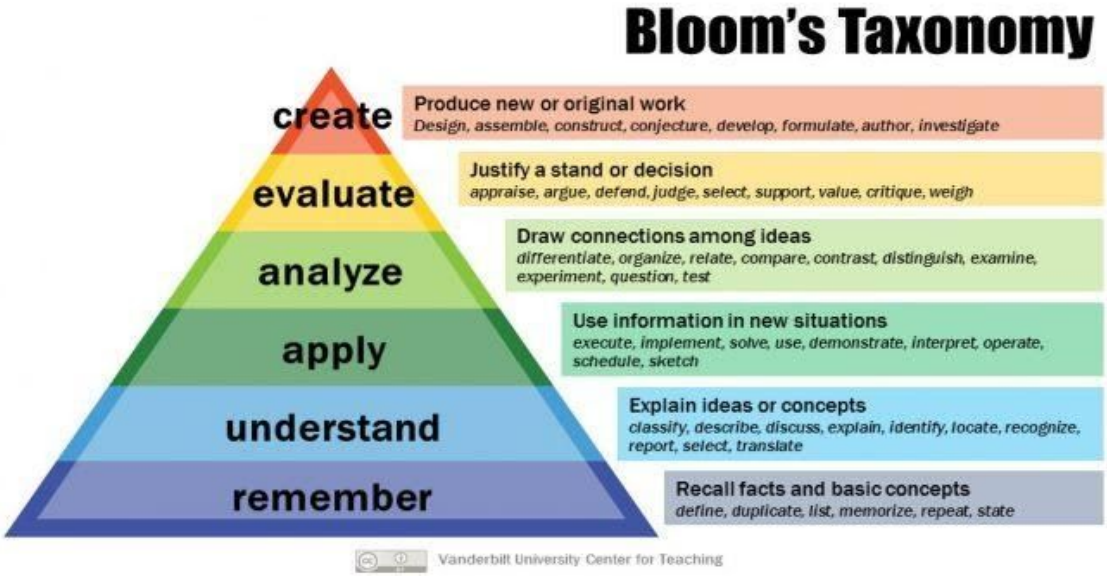

Figure 8: Blooms Taxonomy levels and their significance

7.1 DATA COLLECTION:

As seen above, the main requirement with respect to data is to have a proper dataset which consists of a set of questions that are labelled according to the above mentioned Blooms taxonomy levels.

The questions are collected from University Question Papers, Online sources and Textbooks. These questions are related to various subjects under Computer Science like Operating Systems, Databases, Machine Learning, Networking etc.,

Here the data set has two columns; one is questions columns and another one is its respective classified classification column. So when we look at the very first question ,the question starts with or has a keyword which is 'what', which has been classified under knowledge domain level, just like this remaining keywords like recite, define, name etc., are classified into knowledge domain level. Just like this the classification goes on with different key words in different questions.

The dataset consists of 1100 labelled Questions.

Name the factors that affect the performance of the network?	Knowledge
Define the terms Unicasting, Multiccasting and Broadcasting	Knowledge
Describe at one disadvantage of a peer to peer network.	Comprehension
How can you manage a network using a router?	Application
One way of securing a network is through the use of passwords. What can be considered as good passwords?	Analysis
What happens when you use cables longer than the prescribed length?	Knowledge
You need to connect two computers for file sharing. Is it possible to do this without using a hub or router?	Evaluation
When you move the NIC cards from one PC to another PC, does the MAC address gets transferred as well?	Analysis
How are IP addresses arranged and displayed?	Knowledge
When troubleshooting computer network problems, what common hardware-related problems can occur?	Analysis
How does dynamic host configuration protocol aid in network administration?	Application
Explain profile in terms of networking concept?	Comprehension
How do bridges pass spanning tree information between themselves?	Application
Describe a recent short term stressful situation and how you managed it.	Comprehension

Figure 9: Glance of the dataset being used for the research

7.2 PREPROCESSING

STEPS:

- The entire data is converted into lower case to make the further steps of data processing easier

- Removal of Punctuations: Questions usually consist of Question marks, Hyphens, colons and semicolons. These are unwanted symbols which are of no use for classification processes.
- Removal of stopwords: This is an important step in Pre-processing as Questions consists of many stopwords that occur commonly among all the Questions and do not add any unique information for the classification process. The nltk corpus of stopwords includes wh- words and similar words which actually are important for the classification in this case as we deal with questions. So a list of these words are excluded from the process of removal.

Out[107]:

	Questions	Category
0	Can NS-3 simulator be used to simulate Fat tre...	Analysis
1	What are the simulators available for quantum ...	Knowledge
2	Explain the Different layers of the OSI model.	Comprehension
3	How do you implement routing protocol for vein...	Analysis
4	What is the best tool/way for fast parsing of ...	Knowledge
5	Difference between bit rate and baud rate.	Comprehension
6	What is usage of Sequence Number in Reliable T...	Knowledge
7	Define Network?	Knowledge
8	Name the factors that affect the performance o...	Knowledge

Out[110]:

```
0       can ns 3 simulator used simulate fat tree bcub...
1       what simulators available quantum key distribu...
2                    explain different layers osi model
3                implement routing protocol veins omnet
4       what is best tool way fast parsing 80 gb pcap ...
5                        difference bit rate baud rate
6       what is usage sequence number reliable transmi...
7                                          define network
8                name factors affect performance network
```

Figure 10: Data before and after stopword removal

- Questions are grouped together based on their labelled classes and among those most common words are extracted for each of the categories as a part of training.
- POS tagging: POS(Parts-of-speech) tagging is done to these words.
- Sets of verbs are extracted from these collected pos-tagged words as verbs are of more importance in the Blooms classification process.
- Tokenization: Before feeding the questions to the classifier, tokenization is done as a part of pre-processing.

7.3 CLASSIFIER:

- Rule Based Classifier: Rules are written to classify a given question into one or more than one category using the list of verbs collected for each category.
- Since the above classifier classifies questions into more than one category, weighting the questions is used as a technique to determine the more suitable category among them.

CONCLUSION OF CAPSTONE PROJECT PHASE-1

Expected Capstone-I deliverables were collection, cleaning and labelling of the data, study of algorithms and their feasibility considering the requirement (random, unbiased, unique) and eventually, coming up with a classifier that classifies the questions into expected classes.

Required amount of work has been done in achieving the expected deliverables.

PLAN OF WORK FOR CAPSTONE PROJECT PHASE-2

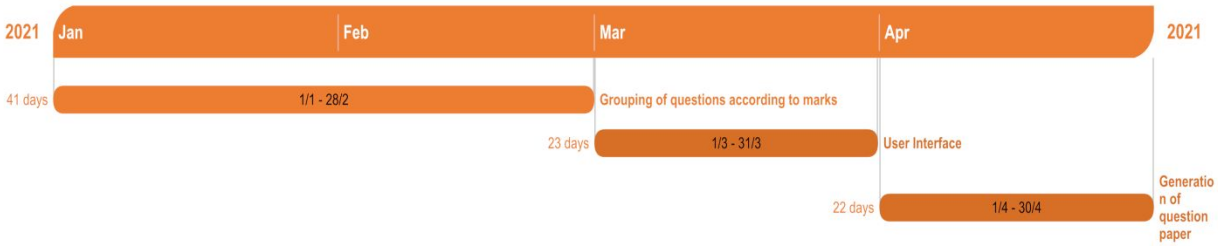

Figure 11: Timeline for CAPSTONE PROJECT PHASE – 2

Capstone Project-2 will focus on developing the User Interface for the entire system. In addition to it, more focus will be on enhancing the accuracy of weighting the categories for each question taken as input.

The questions taken as input would be randomly grouped together keeping in mind the percentage allowance of each category in order to generate a standard question paper.

Further, these questions would be allocated marks according to their difficulty level so as to match the total marks specified as input.

Given a question paper, the system would be trained to predict the difficulty level of ti considering the Bloom's Taxonomy cognitive levels.

REFERENCE/ BIBLIOGRAPHY

1. [1] Nazlia Omar, Syahidah Sufi Haris, Rosilah Hassan, Haslina Arshad, Masura Rahmat, Noor Faridatul Ainun Zainal, and Rozli Zulkifli. "Automated Analysis of Exam Questions According to Bloom's Taxonomy" Procedia - Social and Behavioral Sciences, vol. 59, 2012. doi:10.1016/j.sbspro.2012.09.278
2. [2] Jayakodi, Kithsiri & Bandara, Madhushi & Perera, Indika & Meedeniya, Dulani. (2016). WordNet and Cosine Similarity based Classifier of Exam Questions using Bloom's Taxonomy. International Journal of Emerging Technologies in Learning (iJET). 11. 142. 10.3991/ijet.v11i04.5654.
3. [3] Mohammed M, Omar N (2020) Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and word2vec. PLoS ONE 15(3): e0230442. https://doi.org/10.1371/journal.pone.0230442
4. [4] Das, Syaamantak & Mandal, Das & Basu, Anupam. (2020). Identification of Cognitive Learning Complexity of Assessment Questions Using Multi-class Text Classification. Contemporary Educational Technology. 12. ep275. 10.30935/cedtech/8341.
5. [5] Jain M., Beniwal R., Ghosh A., Grover T., Tyagi U. (2019) Classifying Question Papers with Bloom's Taxonomy Using Machine Learning Techniques. In: Singh M., Gupta P., Tyagi V., Flusser J., Ören T., Kashyap R. (eds) Advances in Computing and Data Sciences. ICACDS 2019. Communications in Computer and Information Science, vol 1046. Springer, Singapore. https://doi.org/10.1007/978-981-13-9942-8_38
6. https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/

APPENDIX A DEFINITIONS, ACRONYMS AND ABBREVIATIONS

Bloom's taxonomy: Bloom's Taxonomy is a classification of the different educational learning levels with objectives and skills that instructors/tutors/examiners set for their students. It has six levels which are Knowledge (remembering), Comprehension (understanding), Application (applying), Analysis, Synthesis (creating), and Evaluation.

NLP: Natural Language Processing

BERT: Bidirectional **E**ncoder **R**epresentations from **T**ransformers

LDA: Latent Dirichlet Allocation

Normalization: Normalization is a procedure that transforms a list of words to a more consistent or uniform sequence

Tokenization: Tokenization is the process of splitting given larger character sequence and a defined document unit up into pieces, called tokens.

Stemming: Stemming is the task of lessening a word to its word stem that affixes or suffixes and prefixes or to the roots of words called lemma.