

An Examination System Automation Using Natural Language Processing

Indrashis Das
School of Computer Engineering
Kalinga Institute of Industrial
Technology Deemed to be University
Bhubaneswar, India
indrashisdas98@gmail.com

Bharat Sharma
School of Computer Engineering
Kalinga Institute of Industrial
Technology Deemed to be University
Bhubaneswar, India
yashsharmamgs@gmail.com

Siddharth S. Rautaray
School of Computer Engineering
Kalinga Institute of Industrial
Technology Deemed to be University
Bhubaneswar, India
siddharthfcs@kiit.ac.in

Manjusha Pandey
School of Computer Engineering
Kalinga Institute of Industrial
Technology Deemed to be University
Bhubaneswar, India
manjushafcs@kiit.ac.in

Abstract—This world has seen a lot many examination portals that are deployed over several servers which are used to conduct online examination for various purposes among which some may include conducting a test for entrance examinations, or olympiads at national and international level and while some portals are designed to conduct a test for placement purposes. But what we have seen is that mostly all the portals are designed to conduct tests that contain multiple choice questions. Here our aim is not to work on the technology that is already existing, rather some technology that is very rare. Here we talk of the descriptive online examination system. Multiple choice questions are easy to deal as they have a question, a few options and a field in the same question that stores the correct option in the database. While in the case of descriptive questions it is not so. It brings in or uses the concepts of Natural Language Processing or NLP to assign marks to answers. Answers are nothing but strings and the job of the model is to do some operations on the answer string such that it can assign the correct marks to answers written by the examinee. The data is basically collected from a descriptive online examination system. Further, it is analyzed and the designed model assigns accurate marks to the answers for the question. The back-end is written in Python where the web framework used is Django, the library used for Natural Language Processing includes NLTK and for database purpose, SQLite version 3 is used, while for the front-end HTML version-5, CSS version-3, Bootstrap and Javascript is used.

Keywords—Exam System; SQLite3; Django; Descriptive System; Natural Language Processing; NLP; Python; NLTK

I. INTRODUCTION

We come to hear news from around the globe that a particular exam was conducted for a job or for a college or examination in schools and the result was published after some time, while this is a good way to conduct an exam but it is inefficient with respect to the current world where automation is the future. The examination system relies on manual work from printing to transporting the paper to the examination hall, then invigilation and the most tedious task of checking the answer sheets which is a huge mess for any examiner which sometimes leads to resource loss. Also, we hear news about paper leaks and answer sheet being lost in the transporting process. The manual checking process will always have that human error based on certain factors like biasing, the mood of the examiner, target completion and much more such factors. Also if we take account of all the paper wastage and the stationary waste which harms our

environment leading to do more bad then good as the enormous amount of trees being chopped off across the world for the process. This helps us understand that the offline examination system is not cost effective or time efficient, resources are also wasted in the process and moreover we all know that resources are scarce in nature and we need to utilize it efficiently to get the maximum output of it.

While the offline examination system has a big disadvantage but are not getting replaced at a bigger scale because new online examination system features only multiple choice type of question's while most of the exams contain descriptive question for which multiple choice answers do not work and hence they are not that compatible and efficient to replace it at a larger level. We all know that if we have to remove a universally accepted system, the new system should not be just good, rather it should be able to make a quality difference so that the organizations accept it. While there are some examination system and they are good at evaluating the answers but they have little to no scope for the descriptive ones and the analysis is not well implemented to get meaningful results. Even most famous of them just have a simple system of storing the correct options in the database and just matching the correct option with it to calculate the result.

In the proposed model we are taking the online examination system to a new level by enabling the examinee to write descriptive answers which will get evaluated on their own i.e automating the entire offline examination system with the efficiency of computing having no human error involved, this can be done using NLP or Natural Language Processing. The evaluated answers will be stored in the database and they can be viewed anytime and a particular student profile will be maintained for better evaluation of the student.

This will be a huge boost to the online examination system as this will allow it to overcome its biggest con and it will also help the online examination system to stretch its paw even in the half-yearly or annual examination conducted by schools or college for evaluating the profile of the student. This will have instant benefits like the system will relieve the burden of the teachers and professors of checking copies and in return they can be more productive with their time in teaching things, this will also eliminate biasing in answer script checking and will have leased space for any human error as copies would not be scanned and the entire marks

will be allotted according to the way answers are written by the examinee while he was on and there will be little to no space for acquisition, it will help in resource management as this will cut corners on stationery products, it will also have greater efficiency with respect to time as it will produce instantaneous results and will be more secure and reliable.

Talking about the technology used in order to build such a model for evaluating descriptive answers, NLP or Natural Language Processing is has a great role to play. NLP can do a lot of innovative jobs like predicting if a message or an email is a spam or a ham, the quality search that we can do on shopping websites like www.amazon.in and www.flipkart.com in order to search for different categories of items that include kitchen utensils, electronics gadget, apparels, food items and much more such products that are available online. The basic idea was that did anyone ever think of knowing how these search bars or how these ham-spam classifications work? The answer to this question is that rarest of the rare people have tried getting into this and tried to know what the mechanism or the back-end work in order to give such powerful search results and such predictive classification techniques. For those who are not aware of the mechanism behind this, it's all just about playing with strings of characters, numbers and special characters or what we call as string manipulations to arrive at such results.

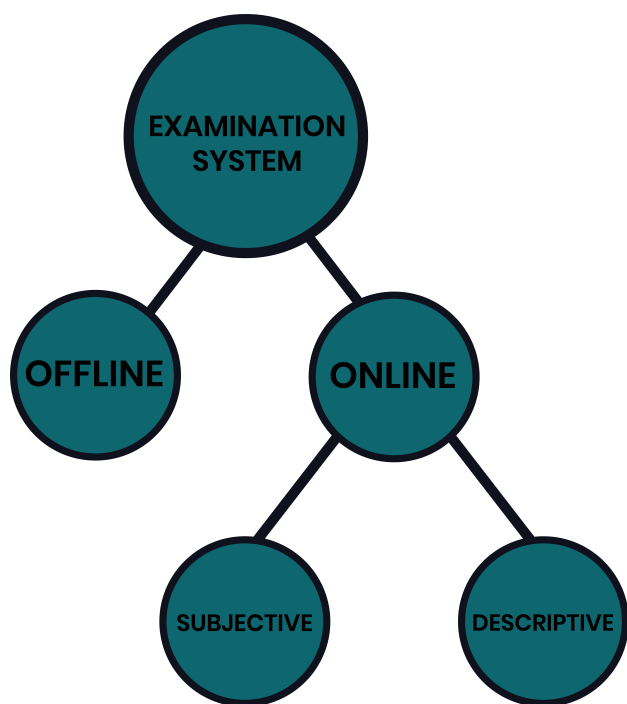


Fig. 1. Types of an examination system

For example when a product is stored in the database of the online shopping website, what happens is that additional keywords are stored for them such that they can be searched using those keywords. The string input by the customer in the search bar contains keywords using which products are searched and displayed on the customer dashboard. Even the back-end for such applications contains quality search algorithms that are out of the scope of this paper's discussion which is on the model of the descriptive online examination system. Similarly, the scenario is quite the same here for evaluating answers for questions that are descriptive in

nature. Keywords are stored for every question and on the basis their occurrence in the answer string, the examinee is allotted marks. In order to perform this in Python language, there exists a library that makes it a little easier for the algorithm developers to perform string manipulations. The name of the library used is NLTK which is specifically designed for python to work on NLP. It is discussed in detail in the section where the algorithm is discussed.

In the below section as we can see is critical analysis in the field of Natural Language Processing and online examinations system. Its speaks about the works in the same field by other researchers and what they have done in order to develop such a system. The section below covers all the fields in order to develop and derive such a system out of the demerits of the other systems that are currently in use.

II. STATE OF THE ART

The main aim of such a system is so that human effort can be reduced by introducing proficient computing technologies. Below, there is a discussion by some critical study on the existing technologies and algorithms for Natural Language Processing and examination systems that are already in use. It is by the help of these research papers that we could construct such an algorithm for conducting online descriptive examinations system.

Jayakodi et al.[1] had proposed a concept in the year 2015 to assign weights to questions for a test or basically their end semester question papers. There are multiple educational taxonomies that can be used to assess and assign weights to questions. So they use a taxonomy learning to function the above weight assignment. Here also they are using NLTK which is a library for working with Natural Language Processing along with the Wordnet package using Bloom's taxonomy.

In the year 2010 Ishikawa et al.[2], conceptualized a paper using Natural Language Processing. The idea was to detect users having multiple accounts over a community site. They had to solve this and hence they did some experimental analysis over the messages passed over Yahoo. The users who have multiple user accounts discourage other users in retrieving quality records and hence the credibility of the community site gets decreased.

Kaur et al.[3] in the year 2017 presented and introduced a paper keyword extraction from text. It uses the concepts of data mining along with Natural Language Processing in order to extract keywords. The sub-concepts used to extract keywords from the text include text mining, web mining, graph mining, temporal data mining, sequence mining, spatial data mining, multimedia mining and distributed data mining.

In the year 2009, Futrelle et al.[4] proposed a paper on NLP-NG which a new NLP system for biomedical text analysis. In this system there exist three subsystems depending on which NLP-NG works. The three components include NG-CORE which is used for language processing, NG-DB which is used for the management of the database, and finally NG-SEE that is used for active visualization. Normalization is being used to generate schema which can be used to detect domain-specific text from the main script or the text. The basic advantage of this system is that it can detect domain-specific text from a large variety of text scripts as it uses a 300 million word BioMed Central corpus.

In the year 2017, Revathy et al.[5] introduced a paper for feature extraction from text using NLP. The basic disadvantage of having an LDA system or Latent Dirichlet Allocation process was that it had a very high time complexity. In order to decrease time complexity, they used NLP or Natural Language Processing in order to do the same. LDA uses a generative statistical model that uses an unobserved group of data to explain observations for which time complexity reduces. NLP is faster as compared to LDA because it uses techniques of pattern generation and then ranks documents according to the patterns generated. What exactly happens is that documents related to the existing documents are found and further a cluster of documents are formed. Then an advanced ranking technique is applied to generate an author community so that they can communicate between them.

Nei et al.[6] in the year 2013 gave an emerging concept of simulation training which could improve the response time and could enhance the safety of the system in the real time, which would help to recreate a smart examination system. Also, the interface module works with the communication module which is connected to other modules. The analysis of data can improve the operational steps which will restore those steps and provide a base for further reports.

In the year 2009, Zhang et al.[7] introduced a SQL based online exam system which could show objective questions as well as the SQL questions but this had larger limitations that apart from the SQL questions. There is no way of showing subjective questions and also that time constraints and security could be improved to a greater extent. The system primarily focused on the processing of query statements and process of DML statements which are then compiled and the basic functions of account management and test management could be implemented in the code.

Zhai et al.[8] focussed on the online system with keeping an eye on the security and improving the performance of the system, the system basically which included the addition of the question modules storing the answers in the database, the paper also compared the traditional access methods like the MAC (Mandatory Access Control), DAC and many more such access techniques. The paper is experimenting with the basic concepts and then showing the test results on the access methods, the basic question module interacts with the test module is less reliable.

Luo et al.[9] in the year 2009 differentiated between a normal test and sequenced test and the paper focuses on order and dependency. It goes a step ahead in showing that where the executor can give executing permission or can deny it, it can also set the examinee-set which will lead to the questions module preparation and the exam system is automated as when permission is given to the executor it can start evaluating the sequence test and will generate the result, it combines the traditional model to make an automated system but the lack of security and have no scope for subjective answers is a disadvantage in the system.

Yang et al.[10] in the year 2012 gave an examination system which had a lot of features including the paper generation, to a better mix-up of library system to keep the papers and then arrange them accordingly and another automatic exam generating system which was there for the generation for the exam, the system being good for a hybrid platform for exam system has a little scope for subjective questions, even adding the feature will lead to system being check by an

examiner as opposite to the automated system works, though it has new features of difficulty coefficient and also analysis of the paper but this does not use the full potential of automated online exam system.

In the year 2013, Treenantharath et al.[11] also talked about the online examination system with the use of the thin client i.e less connection need in compare to the fat client where we need a centralized server in respect to controlling the online activity also the fat client server typically exist with the client server architecture which is quite heavy to load, using thin client will provide solution to the described problem but it has little to no scope for the descriptive answers to be taken into consideration and thus the scope is only limited to the online multiple choice questions.

As discussed by Atoum et al.[12], the approach taken by the authors was a bit different in this paper as it was supposed to be a monitored online examination system as opposite to the offline center where the invigilator will have a close eye on the examinee there was an online system created to handle that problem by using the webcam and the microphone which will search for any kind of the anomaly and according to that actions will be taken if any kind of malpractice is detected while this system is better but it has loopholes where the above condition will fail, also the system will only have the MCQ based examination which leave us with little to no way for the descriptive answers to be evaluated

III. ALGORITHM OF ONLINE DESCRIPTIVE ANSWER MARKING SYSTEM

As discussed earlier it is seen that there exist two types of examination portal or systems that include an offline examination system and an online examination system. Further, the online examination system can be grouped into two examination systems wherein one subjective questions exist while in the other descriptive questions exist. Subjective questions are those in which there exists a question for which a set of multiple choice options are provided such that the examinee has to select one of them. But then it may also happen that there may be multiple answers to a question, then, in that case, multiple check-boxes can be selected. While in the case of the descriptive examination system, a question exists for which the examinee writes an answer that is stored as a string and further marks are assigned to the answer according to the quality and correctness of the answer. Basically, keywords are stored in for the particular question by the teacher who uploads the test and its questions on the portal. The answer's correctness is judged by matching the answer with the keywords i.e basically the words of the answer with the keywords. Further, a sparse matrix is formed and hence it shows the count of each and every keyword in the answer string. Then there occurs some mathematical operation on it so as to count the actual accurate marks for the answer written.

As Python is used in order to code out the algorithm, a module or as we say a library named NLTK is used. This is the library in python that helps out to perform Natural Language Processing. It makes it easier for the algorithm developer to work with NLP. It consists of a list of common words such as prepositions and articles like a, an and the which have no use in the final answer evaluation, so they are removed. These common words are known as "stop-words".

After they are been removed, the keyword existence in the answer is calculated using a sparse matrix. For keyword that exists as a word in the answer string, we put 1 in the cell and go to the next keyword. If the keyword is not present as a word in the list of words in the answer string, then 0 is placed in the cell. This way the sparse matrix is formed. Further with the keyword existence count in the string, marks to be allotted to the answer is calculated.

There are some terms that are important while working on NLP and they are stated below, consider two strings as "Red Pen" and "Blue Pen". Here the keywords that can be obtained are "Red", "Blue", and "Pen". So, the keyword vector can be calculated in the format (Red, Blue, Pen) as,

$$A = (1, 0, 1) \quad (1)$$

$$B = (0, 1, 1) \quad (2)$$

Using this information we can calculate the similarity index between two strings by applying Cosine Similarity on Vectors.

$$\text{sim}(A, B) = \cos \theta = A \cdot B / \|A\| \|B\| \quad (3)$$

The next term is Term Frequency which is nothing but frequency of a term 't' in a document 'd'.

$$TF(d, t) = \# \text{ of occurrence of term 't' in a document 'd'} \quad (4)$$

where,

t = The term

d = The document

Another term to be introduced is Inverse Document Frequency which is nothing but the importance of the term in the corpus.

$$IDF(t) = \log (D / t) \quad (5)$$

where,

t = # of documents with the term

D = # of total documents

The final term to be introduced is TF-IDF. It gives the word count along with a notation of how important a word is in a document or a list of documents.

$$W_{x,y} = tf_{x,y} \log (N / df_x) \quad (6)$$

where,

$tf_{x,y}$ = frequency of term x in document y

df_x = # of documents containing term x

N = # of total documents

Without these terms, NLP is not possible. Before working on NLP, one should know the meaning of these terms, what they are and how they can be used.

Below we discuss the algorithm used to work the descriptive answer marking system in detail. The merits and demerits of this system are also discussed in detail in the further sections. As we move on through the further sections, we can detect them.

Algorithm for Online Descriptive Answer Marking System

In the below algorithm, variables named KEYWORD and ANSWER are used to store the keywords from the database for the question and answer string input by examinee respectively. While variables NK is used to store number of keywords for the question with which answer string is to be checked, NY is the number of keyword columns for an answer string for which 1 is stored in the cell, MM is the maximum marks assigned to the question and MA is the marks assigned to the student for his answer. Below is the equation used to calculate the marks assigned to the student,

$$MA = (NY / NK) * MM \quad (7)$$

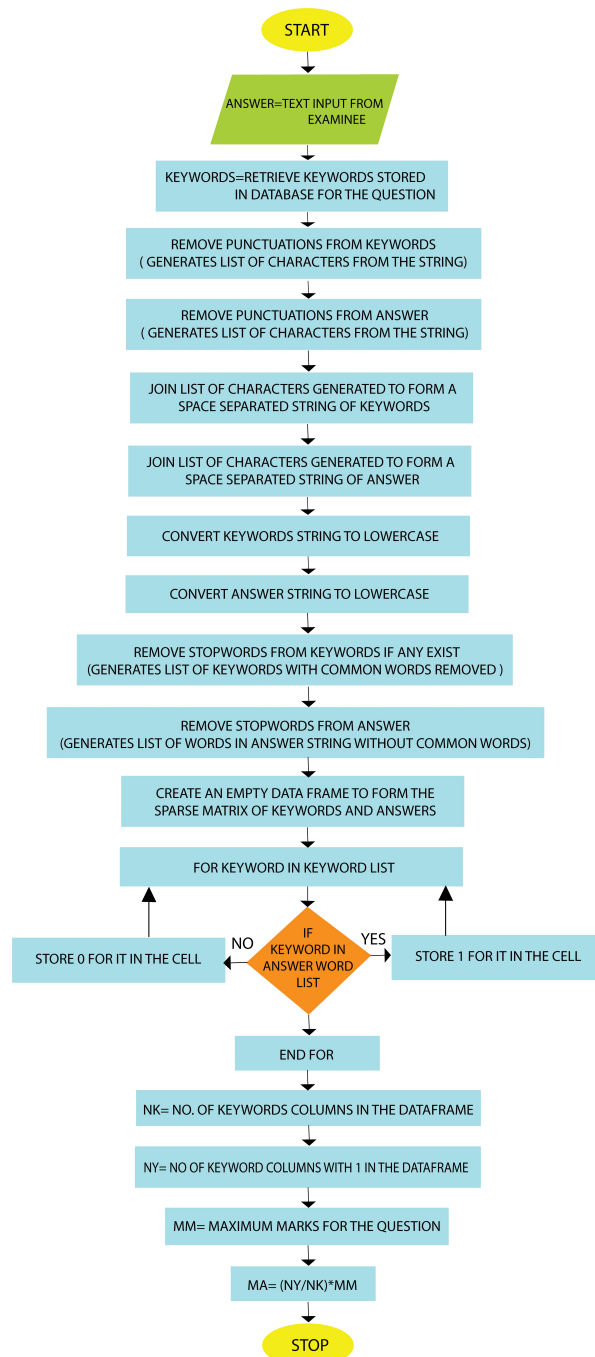


Fig. 3. Algorithm for Online Descriptive Answer Marking System

Now we would take an example of how the algorithm is working. Consider a question as "What is Data Preprocessing ?" for this question the keywords could be

KEYWORD = "data, preprocessing, mining, machine, learning, gathering, range, values, combinations, missing". Now consider the input answer to be ANSWER = "Data preprocessing is an important step in the data mining process. Data gathering methods are often loosely controlled, resulting in out of range values, impossible data combinations, missing values, etc.".

After applying all the string manipulations that include removing punctuation from KEYWORD and ANSWER, then converting the list of characters in KEYWORD and ANSWER into a string and then converting KEYWORD and ANSWER string into lowercase and then on removing stop-words or the common words from KEYWORD and ANSWER and finally after calculating keyword presence in ANSWER, it can be seen that the sparse matrix returned in the form of data-frame looks something like the figure below.

	Answer	data	preprocessing	mining	machine	learning	gathering	range	values	combinations	missing	Marks
0	Data preprocessing is an important step in the...	1	1	1	0	0	1	1	1	1	1	4.0

Fig. 2. Sparse Matrix formed on checking the presence of keywords in the answer string

Further, from here it can be seen that NY = 8, NK = 10, and let MM = 5 marks assigned to the question. Then, marks assigned would be MA = 4.0 using equation (7). Figure (2) shows the algorithm for Online Descriptive Answer Marking System.

IV. RESULTS

The data that we received had the algorithm for online descriptive answer marking system applied to it. Below in figure (4) the sparse matrix for five students who answered the question " What is Data Pre-Processing? " can be seen.

Name	What is Data Pre-processing?	Answer Words List	data	preprocessing	mining	machine	learning	gathering	range	values	combinations	missing	Marks
0	Lipika	[data, preprocessing, gathering, raw, matter...]	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.5
1	Adarsh	[manipulating, data, way, fulfills, need, expect...]	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5
2	Bharat	[removing, irrelevant, redundant, value, also...]	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	2.0
3	Kashyap	[data, preprocessing, preparing, data, record...]	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
4	Krishnanshu	[transforming, data, usable, format, removing...]	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5

Fig. 4. Applied algorithm of the online descriptive answer marking system on answer generated by five students

In the above figure, it can be clearly seen that there are five rows with students naming Lipika, Adarsh, Bharat, Kashyap and Krishnanshu. Further, there are 14 columns among which the first column contains the Name of the student, answer to the question " What is Data Pre-Processing? ", Answer Word List which contains the keywords generated from the answer generated by the students and then come to the columns which are mandatory keywords that need to exist in the answer and the last column is Marks that consists of the marks obtained by the student.

It can be clearly seen from the algorithm in figure (3) that how the marks are being calculated from the keyword existence in the answer.

Here we have a small data for explaining the instance, while in practical situations when the number of students is huge i.e approximately in the range of 500-1000 or above than that, then quality analytics can be applied to know about the trend of the marks by plotting out some graphs by Python's visualization libraries that include Matplotlib and Seaborn.

Below is a histogram showing the frequency of the marks obtained by the number of students.

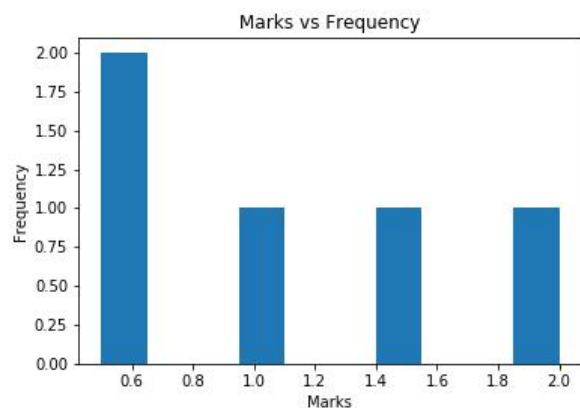


Fig. 5. Histogram for determining frequency of students marks

In the histogram, it can be clearly seen that two students have got the same marks who are Adarsh and Krishnanshu which is 0.5 marks. Further Lipika has got 1.5 marks as three of the keywords match with the answer string input by her. Similarly marks obtained by Bharat is 2.0 and by Kashyap is 1.5. Such histograms are beneficial to know the trend in the marks obtained by students.

V. CONCLUSION

It can be seen by conducting tests using such an algorithm at regular intervals that one can determine the trend in the marks obtained by different students and we can give them an analyzed report on the different subjects they need to focus on for which they are weak. With the existing data, we can also implement a predictive machine learning model on the data so that it can predict marks that the students will score in the future. It is observed that students mainly study those subjects that are placement oriented or which are required for placement purpose only. While students neglect the subjects of their core domain. Deep knowledge in the domain is required as it is of no use to study if you do not have a core domain knowledge. So it can help students get quality knowledge as everything will be digital and there will be no cumbersome process of conducting a pen-paper test. Also, answers are evaluated at that moment itself and the student can see the solutions and can correct the mistakes or errors committed while appearing for the exam.

The algorithm is also efficient as the faculty conducting the test can himself set the question along with the desired keywords he is expecting in one's answer. With this, there is a generalized model using which such answer marking systems can be developed and all the tests can be conducted over an easy to use dashboard oriented user login portal.

Also, talking about the future works on it, it can be tracked using the system that which student has cheated from any other student. This will again use the concepts of Machine Learning and Data Science to work upon these.

REFERENCES

- [1] K. Jayakodi, M. Bhandara and I. Perera "An automatic classifier for exam questions in Engineering: A process for Bloom's taxonomy", IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), (2015)

- [2] N. Ishikawa, K. Umemoto, Y. Watanabe, Y. Okada, R. Nishimura and M. Murata "Detection of users suspected of using multiple user accounts and manipulating evaluations in a community site", IEEE Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering, (2010)
- [3] B. Kaur, and S. Jain "Keyword extraction using machine learning approaches", IEEE 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA) (Fall), (2017)
- [4] R. P. Futrelle, J. Satterley, and T. McCormack "NLP-NG — A new NLP system for biomedical text analysis", IEEE International Conference on Bioinformatics and Biomedicine Workshop, (2009)
- [5] M. Revathy, and M. L. Madhavu "Efficient author community generation on Nlp based relevance feature detection", IEEE International Conference on Circuit ,Power and Computing Technologies (ICCPCT), (2017)
- [6] W. Nei, Y. Wu, D. Hu, L. Wang, and Y. Li "Data Management and Analysis of Intelligent Examination Scoring System of Simulation Training System", IEEE 5th International Conference on Intelligent Human-Machine Systems and Cybernetics, (2013)
- [7] G. Zhang, and H. Ke "Design of Paperless Examination System for Principles of Database Systems", IEEE International Conference on Research Challenges in Computer Science, (2009)
- [8] L. Zhai, and T. Gong "The research of examination management system based on network flat", IEEE 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), (2011)
- [9] S. Luo, J. Hu and Z. Chen "Task Based Automatic Examination System for Sequenced Test", IEEE International Conference on Electronic Computer Technology, (2009)
- [10] N. Yang, X. Chenguang G. Weiwei and M. Xianmin "The design of exam system on the basis of .net technology", IEEE Symposium on Robotics and Applications (ISRA), (2012)
- [11] T. Treenantharath and P. Sutheebanjard "Secure Online Exams on Thin Client" , IEEE 11th International Conference on ICT and Knowledge Engineering, (2013)
- [12] Y. Atoum, L. Chen, A. X. Liu, S. D. H. Hsu, and X. Liu "Automated Online Exam Proctoring" , IEEE Transactions on Multimedia, (2017)