

Descriptive Answer Evaluation System using Machine Learning

Tushant A. Tayde

M Tech Student

Department of computer science and Engineering

Government college of engineering, Amravati, Maharashtra, India

Premchand B. Ambhore

Assistant Professor

Department of Information Technology

Government college of engineering, Amravati, Maharashtra, India

Abstract - Online subjective tests are very rarely proposed. All the descriptive exams set for the scholars are pen and paper based evaluated by the teachers manually. This paper presents a survey on effective way of online descriptive subjective tests. In this, the answers are unstructured data which need to be evaluated. The evaluation is predicated on the semantic similarity between the faculty's answers and students' answers. Different techniques are compared and a replacement approach is proposed to gauge subjective test assessment of text. This paper study the work done by the many researchers on online examination system or web-based exam system and many algorithm and techniques are used like, TF-IDF, NLP, LAS, SOM etc. and Also gives brief knowledge of it and its wide scope.

Keywords- Descriptive Exam, Semantic Similarity, TF-IDF, LAS.

I. INTRODUCTION

In today's Education System the whole entrance examination altogether different fields are objective tests. Objective tests are all not that sufficient to check the knowledge of scholars. Students are only judged by the answers that they have marked. In this case, there can be two situations either the answer that is marked is surely known by the student or it can be an assumed answer. So in such a situation, we cannot completely judge whether the student is really intelligent or whether it is his/her luck.

Online examination system has also brought some huge advancement to the normal assessment system. It has made the examination method automated, comprehensive, accurate and instant. Such technological evolution has not just brought ease to the teachers/educators, but has also introduced tons of advantages for the scholars and test takers. The online examination system uses online exam software which is used to create the test, conduct, and also evaluated the test.[8] This sort of examination system possesses multiple benefits few of them is that it eliminates the dependency on paper for a question and answer sheets, and eliminates any kind of manual workload which is just too much just in case of an offline test.

The student may have some knowledge about the subject but not complete, in such cases to check the knowledge of the scholar, descriptive answers play a very important role. But the evaluations of descriptive answers are mostly manual which becomes too hectic for faculty. To over-come this problem of manual checking of subjective answers I've proposed a network evaluation system of descriptive type answers.

Descriptive answers vary from student to student, so in our proposed paper to extract the meaning from the varied answers the concept of semantic similarity is employed. Faculty needs to answer along with some compulsory expected keywords in it. The answer is going to be pruned, stemmed which can reduce the dimensions of the solution then be converted into vectors and matrix form.[3]

Depending upon the keywords utilized in the solution marking is going to be done. For this, the specified text is going to be extracted from the database by using various methods like Term Frequency/ Inverse Document Frequency (TF/IDF). To extract the meaning from the given text, techniques like LSA, SOM are used. The TF/IDF method with latent semantic analysis (LSA) semantic work is suitable for information retrieval, text classification, etc. The marks are going to be assigned using the Cosine Similarity depending upon the worth of theta. In this proposed system the length of the answers also will be taken into consideration while allotting the marks.[2]

II. LITERATURE SURVEY

There are various works proposed for brief answers evaluation and objective answers. Even though there is various work proposed for short-answer evaluation, the works associated with the descriptive type answer evaluations are very limited. Some work associated with descriptive answers evaluation is mentioned below.

Menaka Sand Radha N, classified the text using keyword extraction. The keywords are extracted using term frequency-inverse document frequency (TF\IDF) and word Net [1]. TF-IDF algorithm is employed to pick the words and word Net is that the electronic database of English wont to find the similarity among the words. In this proposed work, the word which has the very best similarity is selected as keywords. Songkick Lee and Han-Joon Kim proposed a conventional TF\IDF model for keyword extraction. It involves cross-domain filtering and table term frequency (TTF) for extraction of the text from the answer [2]. Ari Aulia Hakim, Alva Erwin, Kho I Eng, Maulahi kmah Galinium, and Wahyu Muliady work on the TF-IDF algorithm creating, a classifier that may classify the net articles[3]. Stephen Robertson explains the understanding concepts of term frequency-inverse document frequency (TF\IDF) [4]. Professor Teuvo Kohonen, in conjunction with a gaggle of researchers at the Neural Networks research center in Helsinki University of Technology, developed some of the optimization techniques for SOM training.[5].

KristaLagus received the M.Sc. degree in computer engineering science from Helsinki University of Technology, Espoo, Finland, in 1996. She has been a Search Associate at the Neural Networks Research Centre, Helsinki University of Technology, since 1995. Her main research interests are associated with neural networks, especially self-organizing maps, and their application to tongue processing and data processing. Jarkko Salojärvi received an M.Sc. degree in technical physics from Helsinki University of Technology, Espoo, Finland, in 1998. His main research interests are associated with neural networks, the stress being on self-organizing maps and their application to data mining [6] Stop words By Vangie Beal Words that are filtered out by Web search engines and other searching and indexing platforms. Stop words are natural language words that have very little meaning, such as "and", "the", "a", "an", and similar words.[8].

In the University of Adelaide, a study has been performed to compare LSA vector with word and n-gram feature vectors. When compared with word and n-gram, LSA gives better performance because according to an entropy measure LSA vectors are weighted [7]. P.Y.Hui, and H.Y.Meng, used LSA for a semantic explanation of a multimodal language with speech and gestures [9]. V.Balakrishnan and E.Lloyd-Yemoh compared the information retrieval performance using stemming and lemmatization techniques [7]. Stemming and lemmatization improve the language model [1]. The process is used in removing derivational suffixes as well as infections(i.e. suffixes that change the form of words and their grammatical functions)[7]. There are many stemming algorithms available. Stemming techniques are many, including the Paice/Husk stemmer, Porter's stemmer, and Levin's stemmer [6].In the Plaice/Husk stemmer, a file is created which holds a set of rules, and these rules are read by an array that implements it until a final stem is achieved [7]. The Lovins stemmer is a single pass, a context-sensitive algorithm that only removes one suffix from a word by utilizing a list of 250 suffixes and removing the longest suffix that it finds attached to the given word [1].

Porter's stemmer is one of the widely used stemmers in information retrieval [6]. Firstly it will remove all the stop words, this is the words that frequently occur in our answers like 'and', 'the', etc. The next step is going to be to get rid of endings that make the keyword plural (e.g. -s, -es), past tense (-ed), and continuous tenses (-ing)[7]. The stemmer then moves on to examine and convert double suffices to single suffice. Other suffices such are -is, -full, -ness,-ant, -once[1]. Lemmatization also helps to match synonyms by the utilization of a thesaurus in order that when one searches for "hot" the word "warm" is matched as well[2].In the feature extraction phase, several methods were discussed to find the semantic similarity[4]. For this proposed fast SOM clustering technology for text information [3]. Y.C.Liu, C.Wu, and M.Liu proposed a rapid Self Organizing Map(SOM) clustering technique for passage information[2]. SOM has been used to projects the documents.

K.Appiah, A.Hunter, Aloft, C Waltham, and P.Dickinson [2] used SOM to mechanically categorize the hidden location of a moving object in the covered surroundings [4]. T.Kohonen used SOM for data investigation in linguistics, finance, and industry and clustering problems [4]. Cosine similarity is used to identify the similarity between two vectors. It will generate a worth that tells how two answers are related by watching the angle. This survey discussed the methods which are suitable for the assessment of descriptive type answers. In this paper, detailed forms of answers are assessed with Latent Semantic Analysis and Self-Organizing Map[2].

III. PROPOSED METHODOLOGY

The proposed work is an educational based system. In this system the web exam are going to be descriptive unlike other objective online exams. The system is going to be administered by the exam system admin. The teaching staff will conduct exams and students will be involved in solving the tests. Now days the online examination is only get conducted on objective types of question and answers. There is not any system available to get the answering for the descriptive type of examination so that it is necessary to stop the traditional ways of taking exam because it is time consuming so it is best to get the solution which will automatically generate results by efficient pattern matching technique . Traditional exams taking long time for results evaluation and will required large human resource to manipulate.

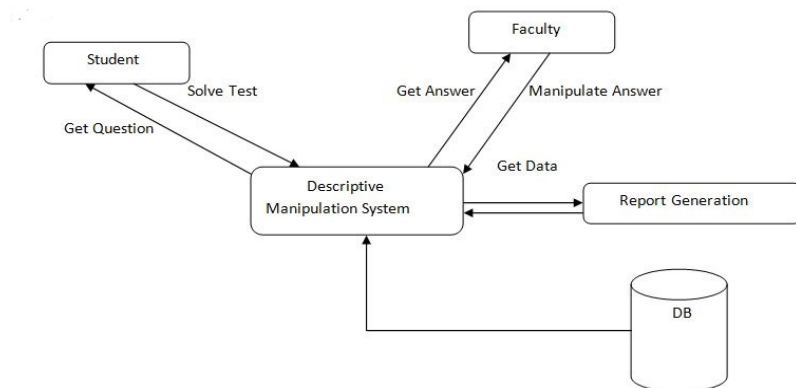


Figure 1: Data flow Diagram of Propose Methodology

The DFD diagram explains that the faculty will provide questions along with the answers and keywords that may be stored within the database. The questions are sent to the web evaluation system. The system will provide the questions to the scholar which the staffs have set. Student's answers are going to be given to the evaluation system where it'll be compared with the information within the database and also the system will check the answers and it'll be passed to the report generation where the marks are going to be calculated and also the result are going to be sent back to the system which is able to then will mailed to the scholars.

IV. MACHINE LEARNING APPROACH

A. Naive Bayes algorithm

Naive Bayes classifier is a family of probabilistic algorithm that takes a bonus of probability theory and Bayes theory to predict the category of sample. A Naive Bayesian model is straightforward to create, with no complicated iterative limitation approximation which makes it particularly useful for very large datasets. The probabilities are considered by the word frequencies. It is calculated using some basic properties of probabilities and therefore the Bayes theorem. The conditional probabilities just like the one discussed here will suits for the Bayes theory. To analyze the probability the subsequent formula is used:

$$P(A|B) = (P(B|A) * P(A)) / P(B)$$

B. Term Frequency- Inverse Document Frequency (TF-IDF)

TF-IDF is that the weighing factor in information retrieval and text mining. It evaluates the important word within the corpus of huge text. Term Frequency (TF) is that the number of times the word appears within the document and Inverse Document Frequency (IDF) is that the weight to measure the importance of term in text document. Weighing is mostly multiplying the IDF by TF as TF*IDF to separate common terms where $TF = C / T$ where C = number of times a given word appears in an exceedingly document and T = total number of words in an exceedingly document and $IDF = D / DF$ where D = total number of documents in an every corpus, and DF = total number of documents containing a given word. The keywords are extracted using TF-IDF and Word Net. TF-IDF algorithm is employed to pick out the words and word Net is that the on-line database of English accustomed find the similarity among the words. During this proposed work, the word which is able to have the best similarity are going to be selected as keywords.

C. Latent Semantic Analysis

Latent Semantic Analysis is a natural language processing technique used for analysing the connection between the set of responses and also the terms. LSA exactly means examining the documents to get the core meaning of these documents. it's a totally mathematical technique for mining and gathering associations of words within the documents and returns a matrix. during this matrix, rows represent the unique terms and columns represent each paragraph. It construct a matrix(m) of size n*d where n is that the amount of terms and d is that the amount of

answers. This matrix(m) contains variety in each cell which specifies the accurate number of appearances of each word in all answers.

V. RESULT

In information retrieval frameworks, precision and recall are defined in terms of a cluster of retrieved documents. Within the sector of data retrieval, precision is that the portion of retrieved documents that are relevant to the query:

$$\text{precision} = \frac{|{\text{relevant document}} \cap {\text{retrieved documents}}|}{|{\text{retrieved documents}}|}$$

For example, for a text search on a bunch of documents, precision is that the amount of correct results divided by the amount of all returned results. Precision takings all retrieved documents into consideration, but it also can be evaluated at a given cut-off rank and considered only the topmost results returned by the system. This measure is termed precision at n . Precision is utilized with recall, the percent of all relevant documents that's returned by the search. And the recall is that the fraction of the relevant documents that are successfully retrieved. Recall is a metric that quantifies the amount of correct positive predictions made out of all positive predictions that might are made. As an example, for a text search on a collection of documents, recall is that the number of correct results divided by the amount of results that ought to be returned.

$$\text{recall} = \frac{|{\text{relevant document}} \cap {\text{retrieved documents}}|}{|{\text{retrieved documents}}|}$$

Results Calculations By Probabilistic Keyword Comparison:

| Test ID | Uid | Total Marks T(M) | Obtain Marks By Keyword Comparison T(M) | Actual Manual Score AMS | Difference D(M) |
|---------|-----|------------------|---|-------------------------|-----------------|
| 1 | 56 | 20 | 10.0 | 11.0 | 1.0 |
| 2 | 57 | 25 | 12.0 | 13.9 | 1.9 |
| 3 | 58 | 20 | 11.0 | 16.0 | 5.0 |
| 4 | 59 | 30 | 17.2 | 24.7 | 7.5 |
| 5 | 54 | 30 | 20.9 | 22.3 | 1.4 |
| 6 | 57 | 30 | 16.0 | 16.5 | 0.5 |
| 7 | 58 | 35 | 21.5 | 23.8 | 2.3 |
| 8 | 59 | 20 | 18.1 | 20.1 | 2.0 |
| 9 | 56 | 25 | 15.7 | 17.0 | 1.3 |
| 10 | 55 | 20 | 11.4 | 12.9 | 1.4 |

T(h)=Threshold Constant Consider =2.0

Total Positive Results T(P)=7 ($D(M) \leq T(h)$)

Total Negative Results T(N)=3

Total Numbers Of Retrieved Results T(R)=10

Recall= T (P)/T(R) =7/10=0.7

Precision =T (N)/T(R) =3/10=0.3

Here, Test-id 1 we are taken which is the user id is 56. Total mark is 20, And marks obtain by the keyword comparison is 10. Actual Manual score is given 11 then there will be difference of 1. Suppose there are the 10 number of test which is given by the different user having id from 55 to 59 randomly. Then we get the result of actual manual marks and the marks obtain by the keyword comparison in above table. Here we have to calculate the precision and recall values which are used for classification accuracy, and it is a single measure used to calculate model performance. Precision is Suitable when minimizing false positives is the focus. Recall is Suitable when minimizing false negatives is the focus.

VI. SCREEN SHOTS

1. Admin-login page

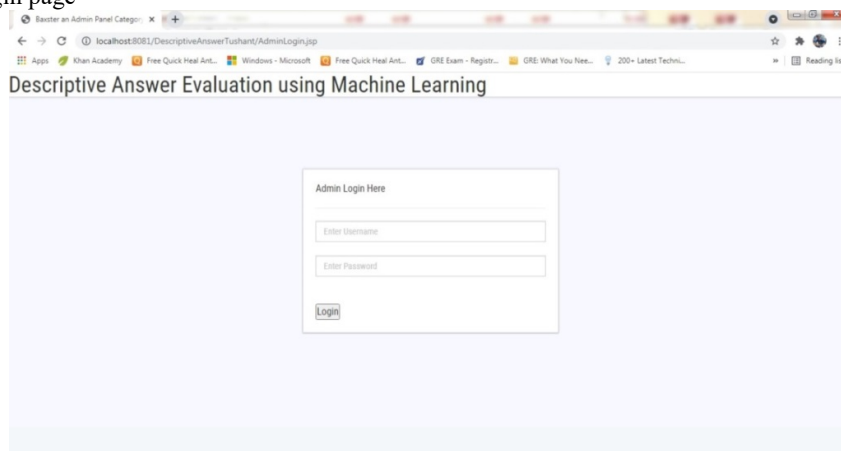


Figure 2: Admin-login page

This is the admin loginpage, where admin or staff of the college can login through given username andpassword. And they also can add student and create test for students.

2. Dashboard page .

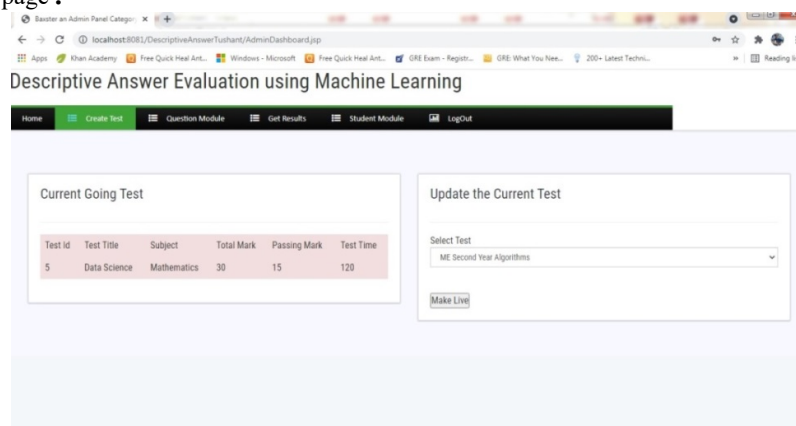


Figure 3: Dashboard page

Here, the test or exam created by the staff is displayed and also thecurrent ongoing test is also displayed stafff can update the current test fromselecting the test from deop-down list.

3.Create Test

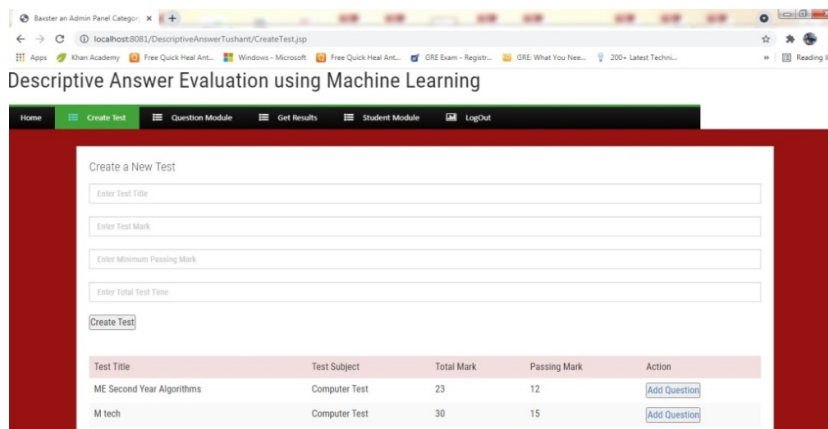


Figure 4 : Test Create page

On this page, the admin can create a test by giving test title, mark for the test and passing marks for pass the test and duration for the test.

4 . Adding Question

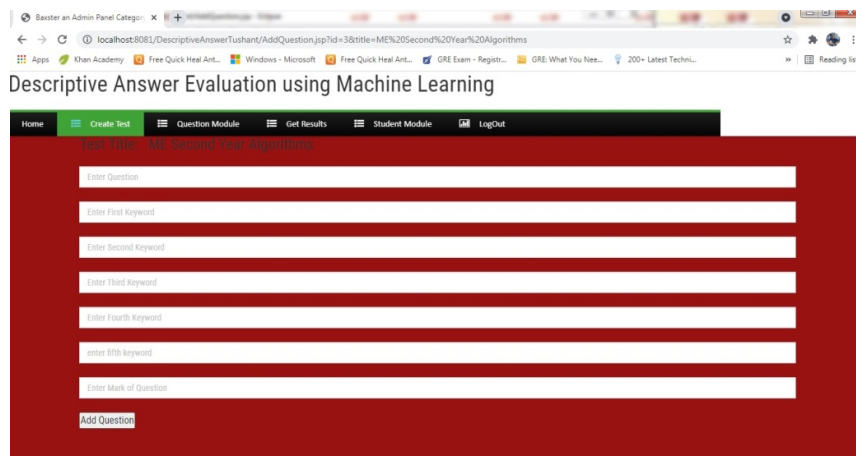


Figure 5: Question adding page

This page, Admin can add question for the examination and also they have to provide some keywords which are related with the question and answer.

5 . Test Result

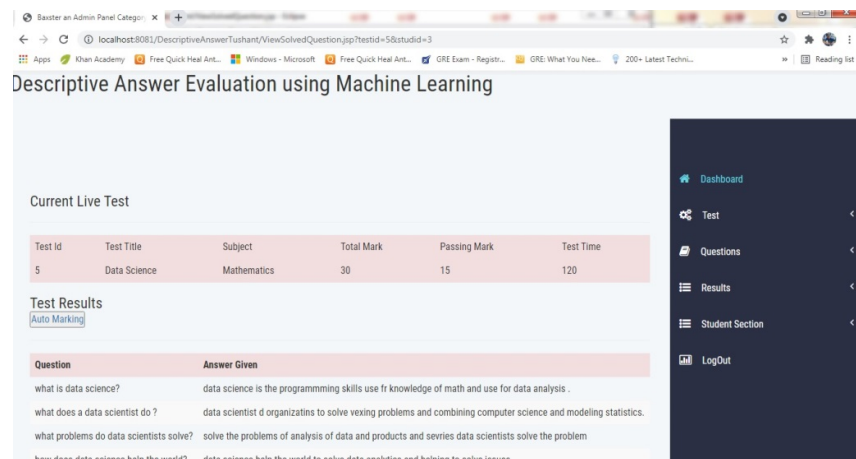


Figure 6: Test Result

On this page, the answer given by the student is display where staff can evaluate those answer by clicking on auto-marking button. The marks are automatically evaluated by machine using various techniques.

6. Answer evaluation

| Test Id | Test Title | Subject | Total Mark | Passing Mark | Test Time |
|---------|--------------|-------------|------------|--------------|-----------|
| 5 | Data Science | Mathematics | 30 | 15 | 120 |

| Question | Answer Given | Option1 | Option2 | Option3 | Option4 | Answer Score |
|---|---|---------|---------|---------|---------|----------------|
| what is data science? | data science is the programming skills use fr knowledge of math and use for data analysis . | ✓ | ✓ | ✓ | ✗ | 5 3.5294118 |
| what does a data scientist do ? | data scientist d organizations to solve vexing problems and combining computer science and modeling statistics. | ✗ | ✓ | ✓ | ✓ | 5 4.6 |
| what problems do data scientists solve? | solve the problems of analysis of data and products and services data scientists solve the problem solve? | ✗ | ✓ | ✓ | ✓ | 5 3.5625 |
| how does data science help the world? | data science help the world to solve data analytics and helping to solve issues | ✓ | ✓ | ✓ | ✓ | 5 3.5714286 |
| how data science can help us? | imdestand gigantic data multiple source vauable | ✗ | ✗ | ✗ | ✓ | 10 4 |

Figure 7: Answer Evaluation page

This is answer evaluation page, show the question and answer given by the student and how many keywords are match with the answer. If three keywords are matched then this shows in green tick otherwise in red tick, as shown for the first answer there are three keywords are matched and one is not. So, on basis of keywords matching and the length of the answer machine gives a score to the answer. But staff can also have an authority to edit the given marks given by the machine and submit their manual marks to the given answer.

VII. CONCLUSION

The existing online exam systems are mostly objective exams because online evaluation for multiple-choice questions could be a very simple task. The proposed system aims on evaluating descriptive answers. From the proposed system it's clear that descriptive answers can also be evaluated automatically. This will reduce the work of manual evaluation of a number of answer sheets. Various algorithms and techniques like TF-IDF, tokenization, stop words removal, stemming, LSA, cosine similarity has effectively contributed to evaluating the students' answers though each students' answer is different. These algorithms will give appropriate marks to the answer. The proposed system will surely help the academic system in getting the accuracy for marks allocation.

REFERENCES

- [1] Menaka S and Radha N, —Text Classification using Keyword Extraction Technique, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 12, December 2013.
- [2] Sungjick Lee and Han-joon Kim, —News Keyword Extraction for Topic Tracking, Fourth International Conference on Networked Computing and Advanced Information Management, 2008.
- [3] Ari Aulia Hakim, Alva Erwin, Kho I Eng, Maulahikmah Galinium, Wahyu Muliady, —Automated Document Classification for News Article in Bahasa Indonesia based on Term Frequency Inverse Document Frequency (TF-IDF) Approach, 6th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia, 2014.
- [4] Stephen Robertson, —Understanding inverse document frequency: on theoretical arguments for IDF, Journal of Documentation, Vol. 60, No.5, pp 503-520, 2004
- [5] Teuvo Kohonen, Fellow, IEEE, Samuel Kaski, Member, "Self Organization of a Massive Document Collection" IEEE, Krista Lagus, Jarkko Salojärvi, Jukka Honkela, Vesa Paatero, and Antti Saarela IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 11, NO. 3, MAY 2000
- [6] H.Ahonen, O.Heinonen, M.Klemettinen and A.I.Veramo, "Applying data mining techniques for descriptive phrase extraction in digital document collections"; Proc. IEEE International Forum, Digital Libraries, 1998
- [7] P.Y.Hui and H.Y.Meng, "Latent Semantic Analysis for Multimodal User Input With Speech and Gestures. Audio, Speech, And Language Processing IEEE/ACM Transactions", Volume 22, Issue 2, 2014, pp.417-429.
- [8] L.D.Lee, H.Chuang and K.Seamons, "Document ranking and the vector-space model. Software", IEEE, Volume 14, Issue 2, 1997
- [9] Y.C.Liu, C.Wu and M.Liu, "Research of fast SOM clustering for text information", Expert Systems with Applications, Volume 38, Issue 8, 2011
- [10] D.S.McNamara, "Computational methods to extract meaning from text and advance theories of human cognition", Topics in Cognitive Science, Volume 3, Issue 1, 2011
- [11] K.Meena and R.Lawrance, "Evaluation of the Descriptive type answers using Hyperspace Analog to Language and Self-organizing Map", Proc. IEEE International Conference on Computational Intelligence and Computing Research, 2014

- [12] J.Vesanto and E.Alhoniemi, "Clustering of the Self- Organizing map", Neural Networks, IEEE Transactions, Volume 11, Issue 3.
- [13] K.P.N.V.Satya sree , Dr.J V R Murthy,"CLUSTERING BASED ON COSINE SIMILARITY MEASURE" [IJESAT]INTERNATIONAL JOURNAL OF ENGINEERING SCIENCE & ADVANCED TECHNOLOGY Volume-2, Issue-3, 508 – 512
- [14] Ankita Patil and Prof. Achamma Thomas," A Survey of Effective Techniques for Subjective Test Assessment" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 4 Issue: 3 400 – 403
- [15] Meena,K and Lawrance,R ,” Semantic Similarity Based Assessment of Descriptive Type Answerstment “2016 IEEE International Conference
- [16] Sonal N. Deshmukh and Ratnadeep R. Deshmukh,” Cosine Similarity for Substituted Text Detection ”Volume 4, Issue 1, January 2014 ISSN: 2277 128X International Journal of Advanced Research in Computer ScienceandSoftwareEngineering.
- [17] Birpal Singh j. kapoor, Shubham M. Naggure” An Analysis of Automated Answer Evaluation Systems based on Machine Learning “International conference on Inventive Computation Technologies(ICICT-2020)
- [18] Piyush Patil, Sachin Ptail” Subjective Answer Evaluation using Machine Learning” International Journal of pure and Applied Mathematics” volume 118 no. 24 2018
- [19] <https://www.techopedia.com/definition/13698/tokenization>
- [20] P.Willett, “The Porter stemming algorithm: then and now”, Program, Volume 40, Issue 3, 2006, pp.219-223