**1. Introduction**

In today’s Education System the entire entrance exam in all different fields are objective tests. Objective tests are all not that sufficient in order to test the knowledge of students. 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 situation we cannot completely judge whether the student is really intelligent or whether it is his/her luck. The student may have some knowledge about the topic but not complete, in such cases in order to actually test the knowledge of student, descriptive answers play an important role. But the evaluations of descriptive answers are mostly manual and that becomes too hectic for faculty. To solve this problem of manual checking of subjective answers we have proposed an online evaluation system of descriptive type answers.

Descriptive answers vary from student to student, so in our proposed paper to extract the meaning from the various answers the concept of semantic similarity is used. Faculty needs to give the answer alongwith some compulsory expected keywords in it. The answer will be pruned, stemmed which will reduce the size of the answer and then converted into vectors and matrix form. Depending upon the keywords used in the answer marking will be done. For this the required text will be extracted from the database by using various methods such as Term Frequency/ Inverse Document Frequency (TF/IDF). To extract the meaning from the text, techniques like LSA, SOM are used. The TF/IDF method with LSA semantic work is suitable for information retrieval, text classification , etc. The marks will be assigned using the Cosine Similarity depending upon the value of theta. In this proposed system the length of the answers will also be taken into consideration while allotting the marks. The result analysis will then be mailed to the students.

* 1. **Overview**
* To develop an online exam system for descriptive type questions.
* To implement techniques and algorithms that uses semantic similarity for evaluation of detailed type answers.
  1. **Industry profile**

iBase Electrosoft is stands on the base of intelligent of software professionals. We combine research activity with practical applications in programming, web development & integration, multimedia development, custom IT systems design, testing and consulting. Through this tight integration of basic and applied activities, we managed to be at the forefront of the IT process.

**Services :** We offer world class to class quality solution which is more efficient. We provide affordable solutions for organizations seeking web solutions, digital media solutions and design solutions that would be competitively prices and simple to maintain. We have Internet Marketing experts, far-sighted Web Developers, wellversed Content Writers and creative Web Designers. These skilled professionals have the zeal to win and strive to offer professional and competitive solutions to our clients. Our teams serve following services. 1. Software Development 2. Android App Development 3. Web Designing and Development Services 4. Web Hosting Services 5. Software consultation 6. HR Services 7. Embedded System Development

* 1. **Problem Defination**

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.

**1.4 Project Scope**

Although a lot of research work has already been done, many modifications are still required. Such as in near future we can make use of lemmatizer instead of stemmer, because it takes care of additional analysis that stemmer does not support. For instance- Lemmatizer looks at the synonym of word unlike stemmer. In future we can use Explicit Semantic Analysis (ESA) instead of LSA that uses Wikipedia as a source of knowledge and creates feature for each text where each feature vector corresponds to a Wikipedia article.

**1.5 Motivation**

* There are many online exam systems, but existing system can evaluate only objective type questions.
* There is no online system to evaluate descriptive type questions. Manual evaluation is time consuming and very lengthy process.
* Result will be declared in very short span of time because calculation and evaluations are done by the machine itself.

**2. System Analysis**

A module can encapsulates code and data to implement a particular functionality. It as an interface that lets client to access its functionality in a uniform manner and it is easily pluggable with another module that expects its interface. So satisfying all the conditions of what a module is our proposed system will have following modules:

* Exam System Admin
* Staff
* Student
* Online Exam Portal
* Result Calculation
* **EXAM SYSTEM ADMIN:** This is the module of an admin, the person or a team under whom the entire system will be controlled. The admin will monitor all the actions of the system. The subjects will be registered by the admin. He will register the staff members and create the user id and password for each member of the staff. He will supervise all the activities of the staff and will have the authority to check all the details of the students too. He will allot each member of the staff a particular subject to set papers for that subject.

**START**

**VIEW STUDENT DETAILS**

**VIEW STAFF DETAILS**

**ALLOT SUBJECTS TO STAFF**

**CREATE STAFF LOGIN**

**REGISTER STAFF**

**REGISTER SUBJECTS**

**LOGIN ?**

**STOP**

Fig. FLOWCHART OF EXAM SYSTEM ADMIN MODULE

* **Staff:** It is the faculty engaged in the work of conducting tests for students. In the proposed system the members of the staff will be registered by the admin. They will be provided with an username and password with which they can login. After login each staff member can register the students who will be appearing for tests. The job for registration of students will be allotted to particular members of the staff. They will create the username and password for the students. All the staff members will be able to see the details of all students.

Once the staff member logins in he/she will be able to see his/her own subjects for which he/she has to set question papers. It is required to register the answers alongwith questions. The answers that the staff will enter should have list of some compulsory keywords expected in the particular answer.

After evaluation of the tests, the staff will be able to see the evaluation analysis and result of the student

**START**

**REGISTERED?**

CONTACT ADMIN

YES

**VIEW OWN SUBJECTS**

**SCHEDULE EXAM**

**REGISTER QUESTIONS**

NO

**LOGIN**

**REGISTER STUDENTS**

**CREATE STUDENT LOGIN**

**VIEW STUDENT DETAILS**

**REGISTER ANSWERS**

**VIEW STUDENTS’ ANSWERS AND EVALUATION RESULT**

**STOP**

Fig. FLOWCHART OF STAFF MODULE

* **Student:** Students will be registered by the staff. They will be provided with username and password to login in the system. If the student is not registered he/she can contact his/her staff. Once the student will login he/she will be able to see the exam scheduled by the staff. According to the time the test is available the student will give the test. In the test he/she is expected to give detailed answers of the questions. All the questions will be descriptive type. After completing the test he/she will submit the test. The test results will then be mailed to the student on their respective mail-id.

**START**

**REGISTERED?**

YES

NOO

**CONTACT STAFF**

**LOGIN**

**VIEW SCHEDULED TESTS**

**SOLVE TEST**

**SUBMIT TEST**

**RECEIVE RESULT THROUGH EMAIL**

**STOP**

Fig: FLOWCHART OF STUDENT MODULE

* **Online Descriptive Exam Portal and Evaluation**

**START**

**SCHEDULE EXAM**

**WILL**

**EVALUATION**

**MODEL ANSWER**

**WILL**

**STAFF**

**STUDENT**

**TEXT STEMMING**

**APPLY LSA**

**SOLVE AND SUBMIT TEST**

**WILL DO**

**MACHINE**

**READ ANSWERS**

**STUDENT ANSWER**

**APPLY KEYWORD EXTRACTION ALGORITHMS**

**TEXT TOKENIZATION**

**STOP WORDS REMOVAL**

**APPLY SOM CLUSTERING**

**CALCULATE COSINE SIMILARITY**

**DISTRIBUTE MARKS**

**STOP**

Fig. Online Exam Portal and Online Evaluation module

**Exam Portal:** The exam portal will be as such the staff will schedule exams. The students will solve the tests and submit it. Then the complete evaluation of the answers in the test papers will be done by the machine.

**Evaluation Process:** The evaluation process will have both the teacher’s answer i.e. model answer and students’ answers. Every student’s answer will be different. The model answer will be stored in the database. When the student submits its test those answers written by the student will be compared with that of model answer. The system will read both the model and student answer and then extract the keywords using keyword extraction algorithm such TF/IDF (text frequency/ inverse document frequency).

Keyword extraction is basically information retrieval which automatically identifies the best terms in the given document. These terms can be key phrases, key terms or just keywords. In this approach keywords are extracted. Keywords are easy to define as they are widely used within the information retrieval (IR).

Example: ”Clustering is the process of grouping the data into classes or clusters”.

The keywords of the above example can be clustering, process, grouping, data, classes, clusters. Keyword extraction also improves the quality of document that are mentioned in the text. Words that are occurred in the document are analyzed to represent the most appropriate words

1. **Term Frequency- Inverse Document Frequency (TF-IDF):**

TF-IDF is the weighing factor in information retrieval and text mining. It evaluates the important word in the corpus of large text. Term Frequency (TF) is the number of times the word appears in the document and Inverse Document Frequency(IDF) is the weight to measure the importance of term in text document. Weighing is generally multiplying the IDF by TF as TF\*IDF to filter out common terms where TF = C / T where C = number of times a given word appears in a document and T = total number of words in a document and IDF = D / DF where D = total number of documents in a corpus, and DF = total number of documents containing a given word

The keywords are extracted using TF-IDF and WordNet. TF-IDF algorithm is used to select the words and WordNet is the lexical database of English used to find the similarity among the words. In this proposed work, the word which will have the highest similarity will be selected as keywords.

1. **Tokenization:** Tokenization is the act of breaking up a sequence of strings into pieces such as words, keywords, phrases, symbols and other elements called tokens. Tokens can be individual words, phrases or even whole sentences. In the process of tokenization, some characters like punctuation marks are discarded. The tokens become the input for another process like parsing and text mining.

Tokenization relies mostly on simple heuristics in order to separate tokens by following a few steps:

* Tokens or words are separated by whitespace, punctuation marks or line breaks
* White space or punctuation marks may or may not be included depending on the need
* All characters within contiguous strings are part of the token. Tokens can be made up of all alpha characters, alphanumeric characters or numeric characters only.

Tokens themselves can also be separators. For example, in most programming languages, identifiers can be placed together with arithmetic operators without white spaces. Although it seems that this would appear as a single word or token, the grammar of the language actually considers the mathematical operator (a token) as a separator, so even when multiple tokens are bunched up together, they can still be separated via the mathematical operator.

1. **Stemming:** In information retrieval, stemming is the process of reducing inflected (or sometimes derived) words to their word stem, base or root form—generally a written word form. The stem need not be identical to the morphological root of the word; it is usually sufficient that related words map to the same stem, even if this stem is not in itself a valid root.

Example: A stemmer for English, for example, should identify the string "cats" (and possibly "catlike", "catty" etc.) as based on the root "cat", and "stems", "stemmer", "stemming", "stemmed" as based on "stem". A stemming algorithm reduces the words "fishing", "fished", and "fisher" to the root word, "fish". On the other hand, "argue", "argued", "argues", "arguing", and "argus" reduce to the stem "argu" (illustrating the case where the stem is not itself a word or root) but "argument" and "arguments" reduce to the stem "argument ".

1. **Stop Words Removal:** Sometimes, some extremely common words which would appear to be of little value in helping select documents matching a user need are excluded from the vocabulary entirely. These words are called *stop words* . The general strategy for determining a stop list is to sort the terms by *collection frequency* (the total number of times each term appears in the document collection), and then to take the most frequent terms, often hand-filtered for their semantic content relative to the domain of the documents being indexed, as a *stop list* , the members of which are then discarded during indexing. An example of a stop list is shown in Figure . Using a stop list significantly reduces the number of postings that a system has to store. And a lot of the time not indexing stop words does little harm: keyword searches with terms like the and by don't seem very useful.

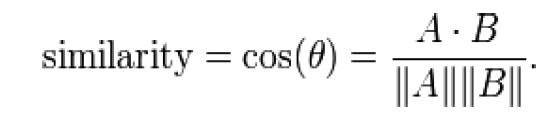
Common words: a, an, and, are, as, at, be, by, for, from, has, he, in, it, its, of, on, that, the, to, was, were, will, with, etc.

1. **Latent Semantic Analysis(LSA):** Latent Semantic Analysis is a natural language

processing technique used for analyzing the relationship between the set of responses and the terms. LSA exactly means examining the documents to discover the core meaning of those documents. It is a completely mathematical technique for mining and gathering associations of words in the documents and returns a matrix. In this matrix, rows represent the unique terms and

columns represent each paragraph. It construct a matrix(m) of size n\*d where n is the amount of terms and d is the amount of answers. This matrix(m) contains a number in each cell which specifies the exact number of appearances of every word in all answers.

1. **Self Organizing Map Clustering (SOM):** Self -Organizing Map is an unsupervised learning process which is normally used in pattern recognition, image processing, natural language processing and data mining. SOMs map multidimensional data onto lower dimensional subspaces where geometric relationships between points indicate their similarity. The main advantage of using a SOM is that the data is easily interpreted and understood. The reduction of dimensionality and grid clustering makes it easy to observe similarities in the data. SOMs factor in all the data in the input to generate these clusters and can be altered such that certain pieces of data have more/less of an effect on where an input is placed. SOMs are capable of handling several types of classification problems while providing a useful, interactive, and intelligible summary of the data.
2. **Cosine similarity**: Cosine similarity is used to measure the similarity between two vectors. It will generate a value that tells how two answers are related by looking at the angle. Cosine similarity measurement is used to compare the key vectors with the answer vectors of the students. Cosine similarity function is given by

****

where ‘A’ will be the answer by a student and ‘B’ will be the corresponding model answer . This function is used to calculate the angle between an answer and the corresponding keyword. If the value returned by cosine function is 1, both answers are similar. Based on the value returned by the similarity measure, marks will be awarded.

Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them. The cosine of 0° is 1, and it is less than 1 for any other angle. It is thus a judgment of orientation and not magnitude: two vectors with the same orientation have a cosine similarity of 1, two vectors at 90° have a similarity of 0, and two vectors diametrically opposed have a similarity of -1, independent of their magnitude. Cosine similarity is particularly used in positive space, where the outcome is neatly bounded in [0, 1]. Cosine similarity is most commonly used in high-dimensional positive spaces. For example, in information retrieval and text mining, each term is notionally assigned a different dimension and a document is characterized by a vector where the value of each dimension corresponds to the number of times that term appears in the document. Cosine similarity then gives a useful measure of how similar two documents are likely to be in terms of their subject matter.

The technique is also used to measure cohesion within clusters in the field of data mining.

For text matching, the attribute vectors *A* and *B* are usually the term frequency vectors of the documents. The cosine similarity can be seen as a method of normalizing document length during comparison.

In the case of information retrieval, the cosine similarity of two documents will range from 0 to 1, since the term frequencies (tf-idf weights) cannot be negative. The angle between two term frequency vectors cannot be greater than 90°.

**3. System Requirement Specification**

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

Menaka Sand Radha N, have classified the text using keyword extraction. The keywords are extracted using TF-IDF and WordNet[1]. TF-IDF algorithm is used to select the words and WordNet is the lexical database of English used to find the similarity among the words. In this proposed work, the word which have the highest similarity are selected as keywords. Sungjick Lee and Han-joonKim proposed conventional TFIDF model for keyword extraction. It involves cross domain filtering and table term frequency(TTF) for extraction[2].

Ari Aulia Hakim, Alva Erwin, Kho I Eng, Maulahikmah Galinium, and Wahyu Muliady works on the TF-IDF algorithm which create a classifier that can classify the online articles[3]. Stephen Robertson, explains the understanding concepts of IDF[4]. Professor Teuvo Kohonen, along with a group of researchers at the Neural Networks Research Center in Helsinki University of Technology, developed a few optimization techniques for SOM training.[5]. KristaLagus received the M.Sc. degreein Computer Science from Helsinki University of Technology, Espoo, Finland, in 1996. She has been a Research Associate at the Neural Networks Research Centre, Helsinki University of Technology, since 1995.

Her main research interests are related to neural networks, especially self-organizing maps, and their application to natural language processing and data mining. Jarkko Salojärvi received the M.Sc. degree in technical physics from Helsinki University of Technology, Espoo, Finland, in 1998. His main research interests are related to neural networks, the emphasis being on self-organizing maps and their application to data mining[6] Stop words *By* [VangieBeal](http://www.webopedia.com/author/Vangie-Beal) Words that are filtered out by Web [search engines](http://www.webopedia.com/TERM/S/search_engine.html) and other [enterprise](http://www.webopedia.com/TERM/S/enterprise.html) searching and indexing platforms. Stop words are natural language words which have very little meaning, such as "and", "the", "a", "an", and similar words.[8]. In the University of Adelaide, study has been performed for 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 entropymeasure LSA vectors are weighted[7]. P.Y.Hui, and H.Y.Meng, used LSA for semantic explanation of a multimodal language with speech and gestures[9]. V.Balakrishnan and E.Lioyd-Yemoh, compared the information retrieval performance using stemming and lemmatization techniques[7]. Stemming and lemmatization improves the language model[1].The process is used in removing derivational suffixes as well as inflections(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 Lovin’s stemmer [6].In the Paice/Husk stemmer, a file is created which holds a set of rules, and these rules are read by an array which implements it until a final stem is achieve[7].The Lovins stemmer is a single pass, context-sensitive algorithm which 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].

The Porter’s stemmer is one of the widely used stemmers in information retrieval[6].Firstly it will remove all the stop words,this are the words that frequently occur in ouranswers like ‘and’,’the’,etc. The next step will be to remove endings that make the keyword plural (e.g. -s, -es), past tense (-ed), and continuous tenses (-ing)[7].The stemmer then moves on to check and convert double suffices to single suffice. Other suffices such are -ic, -full, -ness,-ant, -ence[1]. Lemmatization also helps to match synonyms by the use of a thesaurus so 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 SelfOrganizingMap(SOM) clustering technique for passage information[2]. SOM used to projects the documents. K.Appiah, A.Hunter, A.Lotfi, C.Waltham, and P.Dickinson [2] used SOM for mechanically categorize the hidden location of a moving object in the covered surroundings[4]. T.Kohonen used SOM for data investigation in linguistics, finance andindustryand clustering problems[4].Cosine similarity is used to measure the similarity between two vectors. It will generate a value that tells how two answers are related by looking at the angle. This survey discussed the methods which is suitable for assessment of descriptive type answers. In this paper, detailed form of answers are assessed with Latent Semantic Analysis and Self-Organizing Map[2]**.**

Software Requirements:

1. Eclipse
2. JDK 6.0
3. Apache Tomcat
4. MySQL
5. Windows 8 Operating System

Hardware Requirements**:**

1. **RAM 2 GB**
2. **HDD : 80 GB**
3. **Procrssor i3**

**4.System Design**

**4.1 Database Design**

GIVE QUESTIONS ALONGWITH ANSWERS & KEYWORDS

SOLVE TEST

**Database**

**DESCRIPTIVE EVALUATION SYSTEM**

**STUDENT**

**FACULTY**

GIVE QUESTIONS

GET QUESTION

PROVIDE DATA

GET DATA

**REPORT GENERATION**

Figure 3.1: Dataflow diagram for online evaluation system of descriptive answers

The DFD diagram explains that the faculty will provide questions alonwith the answers and keywords that will be stored in the database. The questions will be sent to the online evaluation system. The system will provide the questions to the student which the staff have set. Student’s answers will be given to the evaluation system where it will be compared with the data in the database and the system will check the answers and it will be passed to the report generation where the marks will be calculated and the result will be sent back to the system which will then will mailed to the students.

**4.2 Proposed Functionality**

The proposed work is an educational based system. In this system the online exam will be descriptive unlike other objective online exams. The system will be administered by the exam system admin. The teaching staff will conduct exams and students will be involved in solving the tests.

* **Describe Main Objective.**
* To develop descriptive type of answer checker filter .
* To develop auto checking system for descriptive types of answer
* To reduce the time consumption in results generation for university exams
* Provide efficient and fast way for descriptive types of examination

**5.Input & Output Design**

**1.Input Design**

Input design is the process of converting the user originated input to a computer based format. The design decision for handling input specify hoe data are accepted for computer processing .Input design is a part of overall system design that needs careful attention.

**2.Output Design**

One of the most important features of a system for user is the output it produces. Output design should improve the systems relationship with the user and help in decision making Computer output is a process that involves designing necessary output that have to be given to various users according to their requirements. Efficient, intelligible output design should improve the system relationship with the user and help and in decision making. A major form of output is the hardcopy from the printer. The output devices are selected by considering the response time requirements, print quickly etc. The print formats and editing for the final printout are very much considered during output design. The objective of output design is to define the controls and format of all printed documents and reports and screens that will be produced by the system. The output is the most important and direct source of information to the user. For many end users output is the main reason for developing the system and the basis on which they will evaluate the usefulness of the application. Output generally refers to the system results. The output of the system is designed so as to include a number of reports. Reports reflect the output design. Output design is an ongoing activity, which start during study phase itself. Output generally refers to the results and information data are generated by the system. It can be in the form of operational documents and reports

**6. Advantages**

* **Security and confidentiality:** Prepared exams need to be securely kept. Any leakage will definitely compromise the standard of the exam and may result to a cancelation or a retake. All these features are well addressed using an online system because not only is the content of the exam safely locked away in a database, access to the database is only possible with an authorized personnel.
* **Time Management:** Online evaluation systems make use of computers that helps in saving time. Manual evaluation of descriptive answers requires more efforts and more time so automatic evaluation makes it easier.
* **Cost Saving:** The cost of paper, copying, and distribution expenses are all reduced or eliminated. Even the cost of scanning the papers and then distributing among the staff for evaluation is eliminated as the students will directly give online test and it will be checked automatically by the machine.

**Disadvantages**

* In the proposed system, dictionary will be provided for identifying the similar words of the given keywords. The words written by the student other than the words mentioned in the dictionary will not be evaluated.

**7.Conclusion & Future Scope**

**Conclusion**

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

**Future Scope**

Although a lot of research work has already been done, many modifications are still required. Such as in near future we can make use of lemmatizer instead of stemmer, because it takes care of additional analysis that stemmer does not support. For instance- Lemmatizer looks at the synonym of word unlike stemmer. In future we can use Explicit Semantic Analysis (ESA) instead of LSA that uses Wikipedia as a source of knowledge and creates feature for each text where each feature vector corresponds to a Wikipedia article.

**7. 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] Self Organizing Map https://www.cs.hmc.edu/~kpang/nn/som.html

[6] : 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 SaarelaIEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 11, NO. 3, MAY 2000

[7] H.Ahonen, O.Heinonen, M.Klemettinen and A.I.Verkamo, “Applying data mining techniques for descriptive phrase extraction in digital document collections”; Proc. IEEE International Forum, Digital Libraries, 1998, pp.2-11.

[8] <https://www.techopedia.com/definition/13698/tokenization>

[9] 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.

[10] L.D.Lee, H.Chuang and K.Seamons, “Document ranking and the vector-space model. Software”, IEEE, Volume 14, Issue 2, 1997, pp.67-75.

[11] 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, pp.9325–9333.

[12] 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, pp.3-17.

[13] 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, pp.558-562.

[14] J.Vesanto and E.Alhoniemi, “Clustering of the Self- Organizing map”, Neural Networks, IEEE Transactions, Volume 11,Issue 3, 2000, pp.586-600.

[15] P.Willett, “The Porter stemming algorithm: then and now”, Program, Volume 40, Issue 3, 2006, pp.219-223

[16] 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

[17] Ankita Patil and Prof. Achamma Thomas,” A Survey of Effective Techniques for SubjectiveTest Assessment” International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 4 Issue: 3 400 – 403

[18]Meena,K and Lawrance,R ,” Semantic Similarity Based Assessment of

Descriptive Type Answerstment “2016 IEEE International Conference

[19] 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 Science and Software Engineering