Sentence Selection Using Latent Semantic Analysis for Automatic Question Generation in E-Learning System

G. Deena, K. Raja

Abstract*: The Current scenario of the educational system is highly utilizing computer-based technology. For the Teaching-Learning process, both the learners and teachers are highly preferred the online system i.e, E-Learning because of its user-friendly approach such as learning at anytime and anywhere. In the Online educational system, the E-Content plays a major role so the critical importance has to be provided in generating the E-Content. Currently, a large number of study materials are dumped into the internet which has reached the highest limit. The enormous amount of content with high volume leads the learner to skim or frustration in learning. Learners have to spend too much of time to understand their concept from the selected web page. The Tutor also faces the challenges in setting the question paper from this high volume of learning content. We have proposed the computer-assisted system to summarize the learning content of the material using Machine Learning techniques. The Latent Semantic Analysis reduces the size of the content without changing their originality. Finally, the singular value decomposition is used to select the important sentences in order to generate the Multiple Choice Questions (MCQ) to assess the knowledge level of the learner.

***Index Terms*: E-Learning, Latent Semantic Analysis, Machine learning, Singular Value Decomposition, Text Summarization.**

I. INTRODUCTION

The great development of the World Wide Web (WWW) has also increased the troubles on the internet. The main issue in the usage of the internet is the problem of loading the content. Loading the high volume of content occupies more memory space where the learners are not comforted in selecting the quality information. It is not so easy to discover the exact material from a massive amount of content loaded in the website. It needs more time to create the highest quality of E-Learning content which comprises only the relevant information. Text summarization has to be converted into the essence of many applications in different fields such as email, data analytics, and online education. Text summarization is a process of mining or accumulates the important information from the original text and presents the information in the form of the summary without changing any meaning Text summarization [1, 2]. It helps to summarize a large amount of transcript in the online educational system. It presents the summary of the selected portion of the content which consumes time taken to read the documents [3]. Text summarization has the objectives of the decent summary, to produce at least a technical document. According to Radef et al. [4] the availability of E-Content needs the exhaustive research to summarize or to shorten the documents. Automatic summarization is classified as a single document and multi-document. Single document method accepts only one document as the input and a process to summarize. The multi-document method accepts more than one document as input and process to summarize. The output of both the methods is either abstract format or summary format [5]. Automatic Text Summarization collects the records from numerous different documents and produces the result as the single document that is very shorter and holds the original meaning of the documents. Hidden Patterns are extracted from the document by analyzing the document. It detects the relationship between the words, text, and sentences to summarize the content. In Existing system, the sentence extraction uses the internal methods of the input document [6-11] which consider internal information such as the position of the sentences, the importance of the word, phrases of the cue, the length of the sentences, the features of the ranking system [12-14]. This summarization is one way to measure the understanding level of the learner. Text summarization educates the learner to identify the important concept, ideas and also helps to eliminate unnecessary data and information [15]. Here section 2 discusses the related works of the summarization method. Section 3 discusses the concept of text summarization. Section 4 gives the idea of the proposed system. Section 5 gives the experimental result of the proposed system. Finally the last section 6 describes the conclusion.

II. RELATED WORKS

In general, the summarization is defined as finding the significance of the sentences in the whole document and reduces the overall document into half. The automatic summarization of natural language text has been widely investigated by researchers [16].

The author Implements the intelligent preprocessing techniques and proposes an algorithm to classify and rank the E-content for the software engineering domain [17]. E-content is the most important segment in the E-Learning System. The ranking algorithm is used to

Revised Manuscript Received on July 05, 2019

G. Deena, Research Scholar, Department of Computer Science & Engineering, Sathyabama Institute of Science and Technology, Chennai, TamilNadu, India

Dr. K. Raja, Principal, Department of Computer Science and Engineering, Dhaanish Ahmed College of Engineering, Chennai, TamiNadu India

identify the important sentences in the material.

The author proposed the importance of teaching-learning process in an E-Learning system. The system has a component for the learning material. To increase the learning skill of the learner the learning style is identified then and the content will be supplied to them [18].

In 2017, Author Kiriti proposed the concept by preferring the simple term frequency, inverse document frequency, position of the sentences, Length of the sentences and similarity method to summarize the document. Here the author concentrated on the single document [19].

An Author Manju Baskar proposed the paperwork in Adaptive learning based on genetic algorithm. This paper includes the concept of context-aware E-learning with different parameter values measured for conditions to satisfy the learner with learning scheme [20].

Barzilay and Elhadad [21] describe a summarization system based on lexical chains of words to extract the words from the paragraph the semantic relationship plays a major role. The semantically related words are lexically connected with each other.

Dragomir R. Radev et al. implemented the text summarization on multi-document MEAD which uses the cluster based centroids on topic summarization and to perform the evaluation on the value of the sentences. This assessment system is subjected to both single and multiple documents. In the end, multi-document summarization elaborates the two user studies that test the models of multi-document summarization.

Gurmeet Singh and Karun Verma, [22] implements the text summarization for the Punjabi document. He preferred the data pre-processing as the first stage to remove the stop words and to find out the root words. Finally, he calculated the weight associated with each sentence for the summarization.

Our proposed system focuses on the single document which calculates the important sentences using the latent semantic analysis and uses the identified sentence for automatic question generation.

III. TEXT SUMMARIZATION

Text summarization is defined as a content that is produced from an original document that contains a significant portion or generates a synopsis of the original. One of the applications of text mining or information retrieval is Text summarization (TS). The process of TS is to squeeze the input into a smaller version that covers the overall meaning of the original document. In the World Wide Web (WWW), multiple or many documents are available in a large size that sorts the information which may or may not be practical in real life. Text summarization is applied in many articles as to how it is summarized the news in the newspaper or on the internet; news present on mobiles will be produced in a short message format. Natural Language Processing (NLP) is a well-known familiar technique to summarize the content automatically. There are two types of summarization techniques which are in practice, Abstractive and Extractive summarization are two different approaches used in text summarization. Single document summarization is the process of creating a summary from a single text document. Multi-document summarization is the method of shortening the document from the collection of multiple related documents into a single summary report and the multi-document summary possesses some notable merits over a single-document summary. It offers a domain summary of a document based on their topic set which represents the identical information from the several documents.

A. E-Content

E-Content is a package known as digital content or Electronic content which comprises of data, information of the particular domain transmits through electronic devices connected in a network. In the teaching-learning process, the central core part is the development of E-Content. Contents are uploaded once on the internet and used various times by different learners. The course comprises of content such as a unit of audio, video, and text that can be reused by different learners. This type of content is preferred in online education for various subjects with thousands of volumes. To the single learning subject or to a domain many different contents are available on the internet with high density. E-content has multiple pages or size which threatens the learners to continue the course or to discontinue with study the material. This content is not only to learn but also helps to assess the learners understanding ability level. So the critical care has to be taken in generating the learning content. The high density of content is to shorten which in turn have the core and highly preferred sentences of the material. Hence the learner will easily understand the concept, feel comfort in learning the new subject, and have the interest to continue the learning. Proficiency is made in the E-learning system to assess the learner by conducting the test for which the question will be generating automatically from the summarized learning material. From this, the learners understanding level are identified. The original learning material cannot be used for the assessment for that the size of the content has to be reduced without changing the originality of the paragraphs is known as content or text summarization.

B. Abstractive Summarization

Abstractive the procedure uses the natural language processing techniques to summarize the document which gives the output similar to human thinking. A deep study on the document is needed for this type of summarization. The output of this method produces a summary that eliminates the redundancy text [23]. Abstractive is similar to human thinking form that is of merging the sentences, determine the synonyms of text or rephrase the sentence. This procedure is to realize, the most important concept to brief in short. Abstractive summarizers used to constrict the original text. Abstractive summaries fabricate the significant information about the document at a higher level of abstraction and include information which is not explicitly present in the original document [24]. It also generates new sentences with the redundancy of the sentences. Abstractive summarization method uses semantic methods to examine and interpret the text and then to find new concepts and expressions to illustrate it by generating a new brief text that conveys the most important information from the original document as an outcome.

C. Extractive Summarization

The extractive procedure gives a summary of the original document by selecting the subset from the sentences. It uses different methods to allocate a score to a different portion of the text input and returns the highest score. The score is allocated to each sentence based on the term frequency and the highest score has the high priority which set as the topmost in summarized output. The output is obtained by selecting the most important content from the document [25]. It organizes and arranges the most important sentences and covers them into a summary. The extraction method is a selection of sentences or a phrases or a paragraphs of the original text generally presented to the user in the crisp order. The length of the sentences is based on the compression rate. An extractive summarization method is further classified into two methods, one by the sentence based and another by keyword based. The sentence based represents the selection of a subset of sentences from the document (Mittal et.al 2000, Carenini et.al 2007). Sentence based method will not have to preprocess further to get a summarized output. The keyword type method will produce only a set of words as output.

The appropriate learning content with the learning style helps the teaching-learning process to succeed [26]. Extractive summarization is applied to both the word and sentence level. Word level representation is given in content of word by selecting the most important keyword from the document which determined by the Term Frequency and inverse document frequency (TF-IDF). Another level of word feature is by cue-phrase identification. The cue-phrase related words are used to connect the two sentence by means of 'because', 'then', 'while', which highly used to summarize the document.

Sentence level features are also used in extractive summarization by the means of finding the location of the sentence in the material, length of the sentence in which the normalized length highly related for summarization because a short and long sentence will not produce the exact information of the content **[**27**]**. Cosine similarity method is used to the identify similarity between the sentences to rank the sentences. The sentence similarity will used for the question preparation.

IV. PROPOSED SYSTEM

The main objective of the proposed system is to implement the unsupervised learning methods of Latent Semantic Analysis to summarize the learning content in the E-Learning course. The important sentences are determined and frame a summary of the sentences. The summarized sentences are in turn generating the assessment questions automatically to determine the ability skill level of a learner for the supplied material. Our system consists of module extraction, preprocessing of the learning material, selecting the heuristics content to generate the MCQ. In order to extract the abstraction scheme, the verbs, and nouns which having similar meanings are determined with syntactic positions. The architectural diagram of our proposed is shown in Figure 1. The

A. PREPROCESSING OF DOCUMENT

Preprocessing is the essential steps involved in text summarization to remove the dead words, unwanted words from the input file. The words filtered out from the sentences followed by the sentence filter. The original paragraphs are filtered into individual sentences using the sentence tokenization. Now, the sentences are split into separate words using word tokenization002E

B. STOP-WORD REMOVAL

The non-linguistic or the lexical words will not produce any semantic and relevant information about the document. They are in the form of punctuations, conjunctions and prepositions word which has to remove from the learning material. These set of words are highly repeated in the learning content. To perform any text related analysis or any mining process it is advised to remove the stop-words from the material. There are more than five hundred stop-words seen in the English language. We also add our own stops to the existing list.

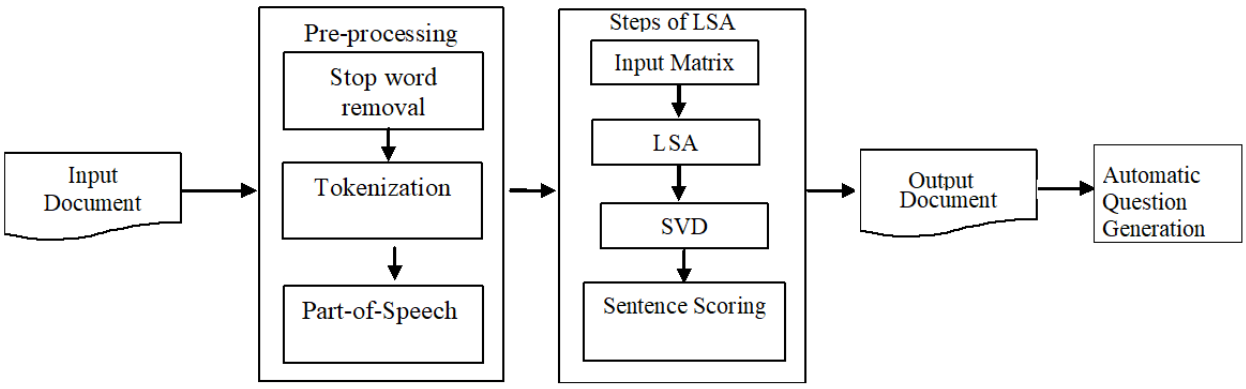

Fig.1. Architecture of Proposed System

C. TOKENIZATION

Tokenization is the method of breaking the text into small units as tokens. The tokens may be the words, the numbers, the special symbols, and the punctuation mark. The paragraph can be tokenized into either in small sentences or as the text into small individual units. When a paragraph is tokenized into sentences, they are termed as sentences tokenization. When the text from the sentences is tokenized then it is termed as word tokenization. In the English language, the white spaces are set as a delimiter to split the word individually.

D. PART-OF-SPEECH TAG

Parts of speech are used to find the tag of the individual word in the paragraph. In the English language, the tags of the words are given in eight different forms namely Noun, verb, adjective, adverb, pronoun, interjection, preposition and conjunction. The nouns are termed as person name, place name, animals name or any name of the organization. In text processing or text analytics, the noun tag words are highly preferred.

E. SIMILARITY OF THE DOCUMENT

To conduct the test first we need to determine the most important sentences from the given learning material so that the similarity between the sentences has to identify. The unwanted sentences will not be involved in question generation. Here, we select the most important sentences for the assessment by implementing the Latent Semantic Analysis (LSA) and Singular Value Decomposition (SVD). There are different steps involved to find the similarity between the sentences and the words from the given input paragraph.

F. LATENT SEMANTIC INDEXING

Latent Semantic Indexing (LSI) was first developed to automatically determine the relevant content of the learning material. This is one of the statistical ways to find sentence similarity. It is one of the unsupervised types of the algorithm to summarize the document without any prior training or knowledge. This type of summarization is preferred in E-Learning material and in the assessment part. The input file consists of the learning material, this LSA fetches the frequently repeated linguistic words and the most commonly repeated word and unique word from the input. The number of the common words in the sentences is representing the relation between the sentences and the word from the sentences is identified by the Singular Value Decomposition (SVD) [28]. It reduces the noise level to maintain a high level of accuracy from the sentences by removing the stop words, the proposition.

G. INPUT MATRIX

The matrix normally has row and column. The column vector is given by M_i where the vector of the sentences is generating. The given input is a sparse matrix with m * n representation of row and column where all the words are not present in the sentences, we have the 'n' number of sentences with 'm' words in the learning material. The Singular value decomposition without generality loss is given by the form.

$A=U\sum V^T$.

Where U & V are representing the orthogonal matrix of the left, right singular matrix and $\sum$ denotes the diagonal matrix. When the determinant is zero it is termed as a singular matrix. When all the entries are zero except diagonal are termed as a diagonal matrix. An original row represents the matrix U. The complex matrix of the vector is represented by V [29]. Before entering the cell value the unwanted words are removed from the sentences such as 'is', 'was', 'then', 'it', 'are', 'be' etc which will not make an impact on the summary. The removal of stop-words will help to reduce the size of the matrix. The cell value can be given by a number of ways but here followed by the frequently repeated words from the sentences. The Term Frequency (TF) representation is given in the binary format where '1' represents the presence of words in the sentence and '0' represents the absence of the word in the sentence. The row represents the words and the column represents the sentences from the paragraph. Input sentences are taken from the computer science book of grade V under the central board of Secondary Education. Input paragraph dealt about the Charles Babbage invention given in the figure 2. The Binary representation of the input is shown in Table 1.

Table1. Input Matrix of the Paragraph

Key terms	Sentence 1	Sentence 2	Sentence 3	Sentence 4
Charles Babbage	1	1	0	0
Machine	1	1	0	0
Difference Engine	1	0	0	0
Automatic	0	1	0	0
Maths Table	0	1	0	0
Calculation	0	1	0	0
General Purpose	0	0	1	0
Computer	0	0	1	0
Analytical Engine	0	0	1	0
Element	0	0	0	1
Modern computer	0	0	0	1
Input	0	0	0	1
Output	0	0	0	1
Memory	0	0	0	1

It had the same basic elements as the modern computer input, output, and memory.
In the 19th century, Charles Babbage invented the machine called the difference engine.
This was an automatic calculating machine that was used to prepare mathematical tables and carry out complex calculations.
He also invented the first general purpose computer known as the analytical engine.

Fig.2. Input File

H. SINGULAR VALUE DECOMPOSITION (SVD)

Singular value decomposition is the matrix that represents the relationship between the words, sentences, and phrases of the learning material. The original input matrix is d decomposed into three different matrices namely U, V and ∑. Next, to the input matrix, the query matrix is constructed by the available nouns keywords. The nouns words are most important to summarize the material whereas another part of speech tagger will not produce the semantic representation. Query matrix is given by 'Q', which consists of the major noun related keys represented as nk_i.

$Q= [nk_1, nk_2 \ldots nk_n]$

The sentences of the 'U Matrix' is the given in Table 2. Similarly, the matrix V, Transpose of V and ∑ are given in Table 3, 4 and 5 respectively.

Table 2: Matrix U

Key terms	Sentence 1	Sentence 2	Sentence 3	Sentence 4
Charles Babbage	- 0.289	0.000	0.000	0.500
Machine	-0.577	0.000	0.000	0.000
Difference Engine	-0.577	0.000	0.000	0.000
Automatic	-0.289	0.000	0.000	0.500
Maths Table	-0.289	0.000	0.000	-0.500
Calculation	-0.289	0.000	0.000	-0.500
General Purpose	0.000	0.000	-0.577	0.000
Computer	0.000	0.000	-0.577	0.000
Analytical Engine	0.000	0.000	-0.577	0.000
Element	0.000	-0.447	0.000	0.000
Modern computer	0.000	-0.447	0.000	0.000
Input	0.000	-0.447	0.000	0.000
Output	0.000	-0.447	0.000	0.000
Memory	0.000	-0.447	0.000	0.000

Table 3: Matrix ∑

Key terms	Sentence 1	Sentence 2	Sentence 3	Sentence 4
1	2.449	0.000	0.000	0.000
2	0.000	2.236	0.000	0.000
3	0.000	0.000	1.732	0.000
4	0.000	0.000	0.000	1.4.14

The new query vector 'Q' is generated from the most important noun key terms which are manually selected in the sentences.

Q = [Charles Babbage, Maths Table, Analytical Engine, Modern Computer]

Transpose of the Query Vector is given by,

Q^T = [0 1 0 0 0 0 1 0 0 1 0 1 0 0 0]

$q = q^T U_k S_k^{-1}$

The 2-dimensional space is given for the vector coordinates of the U matrix and the inverse of the ∑. The vector value of the final query matrix is determined as Q.

Q= [-10.781, -5.469]

I. SENTENCE SCORING

Sentence scoring is calculated after finding the similarity between the given sentences. To find the sentence scoring first apply the Cosine Similarity which is to compute the similarity between the two non-zero vectors. It applies the product space of the two vectors in the positive space of high dimensional value. The value in each dimension is assigned by the number of occurrence of the words in the sentences. The Cosine similarity applies to the sparse vector. For each individual sentences the vector coordinates are calculated by using the mathematical formula

$$\text{similarity}(qn, dn) = \frac{q.d}{|q||d|}$$

The individual sentences will have two vector values, for all the four sentences they are given by

d1 (-0.707, 0.000)
d2 (-0.707, 0.000)
d3 (0.000, 0.000)
d4 (0.000, -1.000)

Similarity is given for d1 as

$$\text{sim}(q,d1) = \frac{(-10.781)(-0.707) + (-5.469)(0.000)}{\sqrt{(-10.781)^2 + (-5.469)^2}\ \sqrt{(-0.707)^2 + 0^2}}$$

= 0.89

Similarly calculate the cosine similarity of d2, d3 and d4 sentences and the sentences were arranged according to the top score value. The top score sentence has the highest priority and the least score sentences has the low priority. According to this score, the sentences are arranged. Manually we fix a limit to include the number of sentences for the assessment. The important sentences are identified which will be further used for the question generation to conduct the assessment.

V. AUTOMATIC QUESTION GENERATION

In our previous work, we have considered all the sentences to frame the Multiple Choice questions (MCQ). In general, all the sentences will not have importance in producing the questions. Hence, select the highest priority sentences by using Latent Semantic Analysis (LSA). This LSA used to prioritize the sentences in the given paragraph. The sentences are arranged according to the score. When the paragraph size is larger then use the average to have the number of sentences to frame the MCQ type questions. If the paragraph size is smaller than all the sentences are involved to frame the questions [30]. The subsequent set of questions is generated from the sentences using a rule-based approach.

A. QUESTION PATTERN

The templates are framed using the rule-based approach in blooms taxonomy. The noun words are filtered out from the selected sentences which in turn involved generating the blanks spaces in Multiple Choice Questions. The set of rules are involved to frame the template by using the Noun, Pronoun, Adjective and the Verb. In our given paragraph, a randomized algorithm is used to generate questions dynamically to get the different set of questions repeatedly. Only the most important sentences are used to test the skill of the learner.

VI. EXPERIMENTAL RESULT

The input paragraph for this concept is taken from the computer science domain of Grade V in Central Board of Secondary Education. The topic deals with the inventory by Charles Babbage. This input document has four different sentences which talk about the analytical engine and the computer Sentence. In manual, the question setter confused in selecting the question for the given paragraph. Similarly in automatic question generation all the sentences were involved to generate the question. The sentences may or may not have importance for question generation. So here, we applied the LSI concept to prioritize the sentences where the prioritized sentences are involved in generating the multiple choice questions. Our system was implemented and tested with 15 different sets of the paragraph in the computer science domain. The paragraph size varied with the inputs, arranged in descending order and manually fixes the limitations to select the number of sentences. By this, the less prioritized sentences are not involved in question generation. The input file has number of sentences, after implementing our system the important sentences are selected for MCQ which is given in figure 3. Preferred sentences represents the important sentences are selected for the question generation.

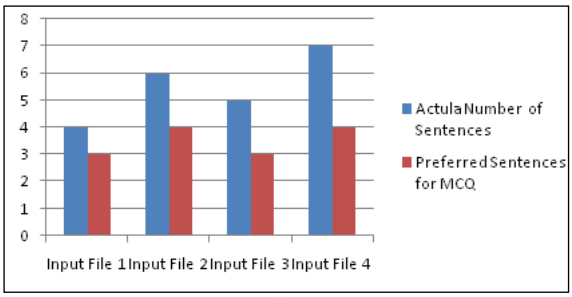

Fig. 3. Representation of input and output sentences

VII. CONCLUSION AND FUTURE ENHANCEMENT

In our proposed system, the important sentences were identified using the LSI method to generate the multiple choice question. Hence, the efficiency levels of the question set are improved when it compared with the manual setting. The learners were able to remember the learned concept. In future, the number of sentences for question generation will be fixed automatically without any human intervention.

REFERENCES

1. Detection using Bigram method and sentiment analysis on twitter Dataset, 978-1-0673-7231-2 at IEEE 2015.
2. Li Bing Keith C.C Chan, “A Fuzzy Logic approach for opinion mining on large scale twitter Data”, 978-1-4799-7881 at IEEE 7th International Conference on utility and cloud Computing, 2014.
3. K. Sarkar, Bengali text summarization by sentence extraction, Proceedings of International Conference on Business and Information Management (ICBIM), 233-245, India, 2012.
4. Dragomir R Radev, Eduard Hovy, and Kathleen McKeown. 2002. Introduction to the special issue on summarization. Computational linguistics 28, 4 (2002), 399–408
5. Kamal Sarkar, Automatic Text Summarization Using Internal and External Information, Fifth International Conference on Emerging Applications of Information Technology (EAIT). doi:10.1109/eait.2018.8470412 2018.
6. Nenkova and K. McKeown, "Automatic summarization," Foundations and Trends in Information Retrieval, vol. 5, Jun 2011,pp. 103-233
7. P. B. Baxendale, "Man-made index for technical literature-An experiment," IBM Journal of Research and Development, vol. 2, Oct. 1958, pp. 354–361
8. H. P. Luhn, "The automatic creation of literature abstracts," IBM Journal of research and development, vol. 2, Apr. 1958, pp. 159-165
9. H. P. Edmundson, "New methods in automatic extracting," Journal of the ACM (JACM), vol. 16, Apr. 1969,pp. 264-285
10. K. Sarkar, "Bengali text summarization by sentence extraction," In Proceedings of International Conference on Business and Information Management, NIT Durgapur, 2012, pp. 233-245
11. K. Sarkar, "An approach to summarizing Bengali news documents, " In Proceedings of the International Conference on Advances in Computing, Communications and Informatics, ACM, Aug. 2012, pp. 857-862.
12. D.R. Radev, H. Jing, M. Styś, and D. Tam, "Centroid-based summarization of multiple documents," Information Processing & Management, vol. 40, Nov. 2004, pp. 919-938
13. K. Sarkar, “Automatic single document text summarization using key concepts in documents”, Journal of information processing systems, 2013, 9(4), pp. 602-620.
14. K. Sarkar, “A keyphrase-based approach to text summarization for English and Bengali documents”, International Journal of Technology Diffusion (IJTD), 2014, 5(2), pp. 28-38
15. Asad Abdi, Automatic summarization assessment through a combination of semantic and syntactic information for intelligent educational systems, Information Processing and Management
16. K. Sparck-Jones, "Automatic summarising: The state of the art," Information Processing and Management: An International Journal, vol. 43, pp. 1449-1481, 2007.
17. N. Partheeban, An Intelligent Content Classification Algorithm for Effective E-Learning, International Journal of Scientific Research in Science, Engineering and Technology, 2016.
18. G.Deena and K. Raja, "A study on knowledge-based E-Learning in Teaching Learning Process", In: Proc.of International Conf. On Algorithms, Methodology, Models and Applications in Emerging Technologies, Chennai, India, 2017
19. S.S.R.K Kiriti, Gogineni Siri, ‘Automatic Text Summarization: Comparison of Various Techniques”, International Journal of Engineering Technology and Computer Research, Vol-5, Issue 2 March-2017.
20. Manju Bhaskar, Minu M Das, “Genetic Algorithm Based Adaptive Learning Scheme Generation for Context Aware E-Learning”, International Journal of Scientific Research in Science, Engineering and Technology
21. Barzilay, R., Elhadad, M.Using Lexical Chains for Text Summarization. In Proc. ACL/EACL’97 Workshop on Intelligent Scalable Text summarization, Madrid, Spain, 1997, pp. 10–17
22. Deepali K Gaikwad and C.Narmada,”A review paper on text summarization”, International journal of advanced research in computer and communication engineering. vol 5, issue 3, March 2016
23. Saranyamol C S, Sindhu L, “A Survey on Automatic Text Summarization”, International Journal of Computer Science and Information Technologies, 2014, vol. 5 Issue 6.
24. Sonia Haiduc, Jairo Aponte,”On the Use of Automated Text Summarization Techniques for Summarizing Source Code” 17th Working Conference on Reverse Engineering, 2010
25. K. Sparck-Jones, "Automatic summarising: The state of the art," Information Processing and Management:” An International Journal, vol. 43, pp. 1449-1481, 2007
26. G. Deena and K. Raja, "The Impact of Learning Style to Enrich the Performance of Learner in E-Learning System", Journal of Web Engineering, Vol.17, No.6, pp.3407-3421, 2018.
27. N.Moratanch, S.Chitrakala.” a Survey on Extractive Text Summarization” IEEE International Conference on Computer, Communication, and Signal Processing (ICCCSP-2017)
28. M.G.Ozsoy, F.N.Alpaslan, and Cicekli, "Text summarization using latent semantic analysis," Journal of Information Science, vol. 37, no. 4, pp. 405-417, 2011.
29. Scott Deerwester, Susan T. Dumais, George W. Furnas, thomas k. landauer, and richard harshman. “Indexing by latent semantic analysis.” Journal of the American Society for information science, 41(6):391–407, 1990.
30. G.Deena and K.Raja, "Implementing Bloom's Taxonomy to Determine the Cognitive Level of E-Learner through Automatically Generated Questions", In: Proc.of Second International conference of Computing and Informatics ICCI’2019, Benha University, Egypt.