



# **iJRASET**

International Journal For Research in  
Applied Science and Engineering Technology



---

# **INTERNATIONAL JOURNAL FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 10    **Issue:** IX    **Month of publication:** September 2022

**DOI:** <https://doi.org/10.22214/ijraset.2022.46769>

**[www.ijraset.com](http://www.ijraset.com)**

**Call:** ☎ 08813907089

**E-mail ID:** [ijraset@gmail.com](mailto:ijraset@gmail.com)

# A Comparative Analysis of Automatic Text Summarization (ATS) Using Scoring Technique

Akshat Gupta<sup>1</sup>, Dr. Ashutosh Singh<sup>2</sup>, Dr. A. K Shankhwar<sup>3</sup>

<sup>1,2,3</sup>Department of Electronic Engineering HBTU Kanpur-208002 India

**Abstract:** The millions of words or sentences on the Internet and in literally hundreds of archives—literary works, published research, legal documentation, as well as other data—have enabled Automatic Text Summarization (ATS), which has become exceedingly well known over the past 10 years. Word-based summarizing is time-consuming, affordable, and unsustainable when interacting with vast volumes of literary texts. It's why we prefer utilizing textual data snippets: it saves time, we acquire accurate data rapidly, and computer scientists have been attempting to create ATS replacements since the 1950s. In this research, we will compare alternative rating systems and offer the best summary based on the scoring methodology implemented. We will also clearly define a wide range of methods, techniques, and scoring algorithms. Using both resource extraction-based ATS and abstractive-based text summarization, hybrid text summarization-scoring techniques are developed. We compare the methods based on both the data and information as well as the performance of the actual content of the summary in accordance with its virtualization using term-based techniques, latent semantic analysis (LSA), and lex rank and clustering-based algorithms. The research findings, based on a study entitled "Automatic Text Summarization in Scoring Growth Strategy," revealed the best prediction accuracy and backup systems. Importance, relevance, and predictability suggest the most appropriate technique among the miscellaneous scoring systems. Multi-document summarization records are unique, and upgrading mastering methods is complicated. We offer a unique method for multi-document concatenation. On the basis of phrase score systems, we try exploring ranking algorithms and introduce a technique for multi-document summarization.

**Index Terms:** Introduction to ATS, approaches to ATS, construction of the NLTK model, SUMY, evaluation of automatic text summarization

## I. INTRODUCTION

It takes a bit of time and a typo to accurately capture a vastly large amount of data. Thirdly, the potential outcomes of such summarizations may potentially provide more in-depth information on a certain text. Text summarization processing has gained importance due to the information and task advancement that has proceeded so rapidly. Automatic language summarization aims to disseminate the original content in a concentrated form that recognizes meaning. A strategy for compressing lengthy passages of text is machine translation summarization. The objective is to reduce the text's main points into a coherent, accessible summary. An information retrieval strategy in the field of natural language processing is machine translation summarization. It automatically breaks up a long text file or document into many phrases without degrading it significantly. automatically separates complex textual data or files into many phrases without compromising the essential concepts. Because of its rapid improvements in technology, today it is clear how significant the information generated in the modern era is, playing a significant role in both the formal and informal spheres. Because of the large amount of information that is generated every minute, machine learning that can improve sentence construction and reduce sentences is essential in the digital age. Leveraging text summaries accelerates the research process more than virtually anything else, actually reduces computer time, and increases the number of outstanding, sustainable knowledge sets being developed in the field. Artificial Language Reconstruction has been shown to be ineffective in summarizing lengthy texts; farm work is critical. The overall objective of an ATS system is to create a summary that compresses the essential topics of the input documents into simple language while eliminating redundancy. Individuals can easily and rapidly fully understand the essential topics of the initial information according to the ATS process [1]. The highlights that are published automatically will benefit the users and save them a considerable amount of time and effort. For a particular user (or users) and task (or tasks), an excellent description should "try to extract the most significant evidence received from a source (or resources) to construct an extended version of certain original material" (May Bury). "A summary may be broadly summarized as a text," the summary continued.

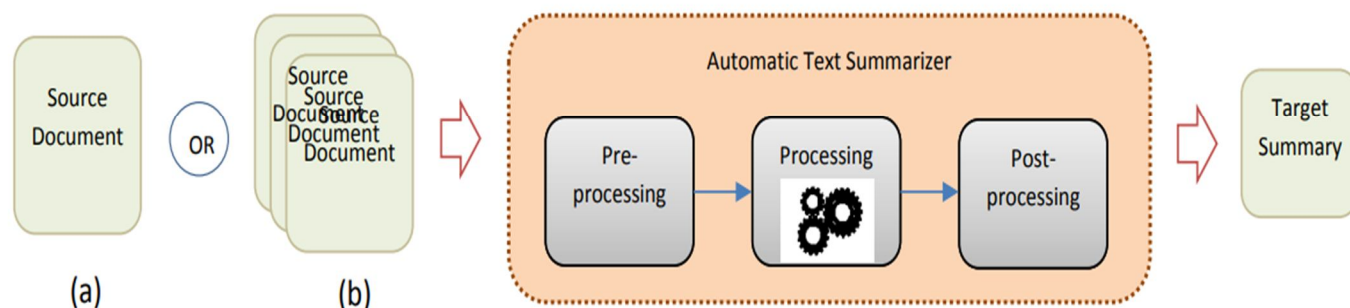


Figure no.1 (single or multiple file)

ATS systems are classified as either single-document or multi-document. The former generates a summary from a single document, while the second generates a summary from a group of documents. Section 2: As illustrated in Fig. 1, the overall look of an ATS system incorporates the following tasks [3]

- 1) *Pre-processing*: constructing and placing the material of the original text using different grammatical procedures such as sentence segmentation, word tokenization, stop-word removal, part-of-speech tagging, stemming, and so on.
- 2) *Processing*: employing one or perhaps more textual summarizing algorithms to transform the input document(s) into the summary. The completely separate ATS methodologies are explained in Section 3, and alternate methods and internal structures for establishing an ATS system are reviewed in Section 4.
- 3) *One Of The Most Difficult Jobs In Natural Language Processing (Nlp) And Artificial Intelligence (Ai) In General Is Post-Processing*: resolving evolution through collaboration among oneself and reorganising the selected sentences before constructing the final summary. ATS research began in 1958 with Luhan's work (Luhn,1958), which electronically generates abstracts from feature articles and conceptual framework-based ATS presents significant issues to the scientific world, including: summarization of multi-documents (Hahn & Mani,2000).

Evaluation of the computer-generated summary without the need for the human-produced. summary to be compared with summary. development of an abstraction summary equivalent to a human-produced summary (Hahn & Mani, 2000) Researchers continue to want a comprehensive ATS system that can generate a summary that: 1) covers all of the significant topics in the input text, 2) does not include redundant or repetitive material, and 3) is understandable and relevant to users. Since the beginning of ATS research in the late 1950s, researchers have worked and are still trying to develop methodologies and procedures for developing summaries so that computer-generated summaries can rival human-made highlights.

#### A. Need for Text Summarization

- 1) The world of documents containing text is huge and expanding every day on the World Wide Web. The majority of the data is in the form of natural language text. Most often, we keep on reading lengthy documents to get some useful information.
- 2) In recent times, text summarization has gained in importance due to the overflowing data on the web and other documents. This information overload increases the demand for more reusable and dynamic text summarization.
- 3) It finds importance because of its variety of applications, like summaries of newspaper articles, books, magazines, stories on the same topic, events, scientific papers, weather forecasts for stock markets, news, resumes, books, music, plays, films, and speeches, which ultimately save time and storage space too.
- 4) Text summarization is used for analyzing high-volume text documents and is currently a major research topic in natural language processing.
- 5) An automatic text summarization is a mechanism to extract a concise and meaningful summary of text from a document.
- 6) With such a large amount of data circulating in the digital space, there is a need to develop machine learning algorithms that can automatically shorten longer texts and deliver accurate summaries that can fluently pass the intended messages.
- 7) Additionally, using a text summary shortens reading duration, speeds up the information search process, and boosts the number of details that may be represented.



## II. APPROACHES OF AUTOMATIC TEXT SUMMARIZATION

Single and multi-document summarising systems are two types of automatic text summarization systems. One of the text summarizing methods is used in the creation of automatic text summarization systems.

- 1) Summarization based on abstraction
- 2) Summarization based on extraction
- 3) Summarization that is hybrid-based

These methods transform the summarization difficulty into a supervised sentence-level classification issue.

### A. Summarization Based on Abstraction

It utilises NLP to produce a summary that recognises the interesting concepts in sustainable systems. This technique requires the ability to create new sentences, instead of simply using terminology from the original text to form the summary. This goes against extractive summarizing, which suggests a precise definition to generate a summary. Automatic text summarization focuses on the most crucial features in the initial survey of sentences in order to develop a span of time for phrases for the summary. An additional clause may have been added to the original sentence.

- 1) *Advantages of summarization based on abstraction:* Better summaries with new terminologies not included in the original text are generated by utilising more versatile phrases based on summarising and compressing. Its strategy will take advantage of current deep learning technology. because integrating the literary text into it requires that the operation be looked at as a cascading translation activity.
- 2) *Disadvantages of Summarization based on Abstraction:* In practice, generating a high-quality abstractive summary is very difficult.

### B. Summarization Based on Extraction

In this process, we focus on the vital information from the input sentences and extract that specific sentence to generate a summary. There is no generation of new sentences for summary; they are exactly the same as those present in the original group of input sentences selects the most important sentences from the input text and uses them to generate the summary. As an illustration, Data Flair is a self-paced, online technology school that offers students an interactive learning experience. The greatest projects for students to learn how to use machine learning with data flair are those that allow them to work on actual problems. Online learning is available to students all over the world at Data Flair. Data Flair provides line projects, quizzes, and lifetime support. The finest initiatives to learn about data-driven machine learning are those that include real-world problems.

- 1) *Advantage of Summarization Based on Extraction:* The extractive strategy is quicker and more straightforward than the abstraction approach, and it delivers in that it allows full control due to the immediate extraction of sentences, which permits users to comprehend the summary using the precise terminology that occurs in the original.
- 2) *Disadvantage Of Summarization Based On Extraction:* The extractive approach is far from the method that human experts write summaries.
- 3) *Drawbacks of generated extractive summary*
  - a) Redundancy in some summary sentences.
  - b) Lack of vocabulary and integrity in summary sentence co-reference relationships.
- 4) *It depends mainly on three independent steps described below*
  - a) *A Characterization first at Intermediate stage is Formed:* A characteristic first The text that has to be summarised is introduced in an intermediate form, either through a specific topic format or indicator representation. each type of representation or indication of representation. Each type of representation has a varied degree of complexity and may be achieved using only a range of techniques.
  - b) *Assign a Score to Each Sentence:* The sentence score directly implies how important the sentences are to the text.
  - c) *Select Sentences for the Summary:* The most essential k sentences for the summary are determined based on several parameters such as deleting duplication, viscerally satisfying the context, and more.

### III. COMMON MODEL OF EXTRACTION BASED SUMMARY

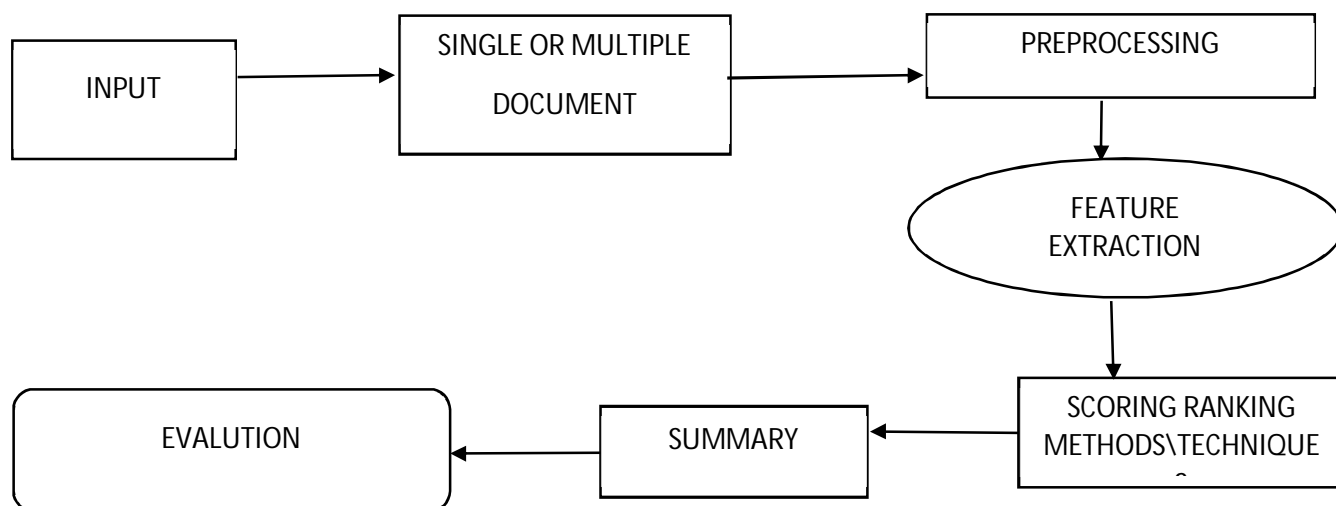


Figure no.2 common NLTK model [11]

#### A. Algorithm

INPUT: A text format of the data is taken as input.

OUTPUT: An appropriate summarized output text is generated, which is shorter than the original document. This is the extracted, summarised output.

Step 1: taking .txt file which is in having N number of document.

Step 2: Tokenize the words and characters in a text file using an extractive approach.

Step 3: Executing the Possession Procedures

Step 4: Feature extraction approach by using different scoring techniques

Step 5: Individual term ranks are calculated.

Step 6: Develop the NLTK modal (desired summary) using different scoring techniques.

Step 7: assess their accuracy in terms of words and time, as well as characters vs. time.

Step 8: We'll figure out which scoring methods work best for the desired summary.

#### B. Prepossessing Techniques in ATS

Frequent prepossessing is carried out in order to eliminate the harsh, unprocessed text. Messages and interactions including errors in the content, including terminology or discarded terminology, are recognised as "additive noise" and "uncompressed text." The following list of prospective gains seems to include a few of the most frequently used benefit techniques:

- 1) *Parts of Speech (pos) Tagging*: Speech tagging is a technique for arranging text words into categories based on speech categories like parts of speech.
- 2) *Stop Word Filtering*: Stop words are dropped either prior to or following textual criticism, depending on the circumstances. Stop words that may be spotted and omitted from plain text provide a, an, and by.
- 3) *Stemming*: It eliminates speech patterns and derivative forms from a group of words known as primary or root forms. Text stemming transforms words to examine additional work forms by applying language's methods, such as acknowledgment.
- 4) *Named Entity Recognition (NER)*: **Words** in the input text are recognised as names of items.
- 5) *Tokenization*: It is a text processing method that links text streams into tokens, which can be words, phrases, symbols, or other crucial factors. The technique's purpose is to examine specific words in a document.
- 6) *Noise Reduction*: Special characters and punctuation make up the vast majority of textual content's numerous other characters. Key punctuation and special characters are also used. While special characters and significant punctuation are essential for human perception of publications, they can indicate a problem with segmentation systems.

Others are like slang and abbreviations and capitalization, etc.

#### IV. METHODOLOGY

FEATURE EXACATION: scoring technique in Automatic Text Summarization

- 1) Term-Based Technique
- 2) Cluster- Based Technique
- 3) LEX Rank-Based Technique
- 4) LSA - Based Technique
- 5) Statistical Technique; KL-sum

##### A. Term-based Scoring Technique

Term-based techniques always conduct the bag-of-words (BOW) model to determine the frequency of a term using the tf-isf (term frequency-inverse sentence frequency) model, tf-Idf model, and many variations of this scheme. Oliver et al. [4] compared the performance of 18 shallow sentence salience-scoring techniques used in both single and multi-document summarization. Worldwide research and experimental studies are frequently conducted in order to assess the effectiveness of experimental sentence-scoring algorithms. The sentence scoring methods should include different factors, including terms specific to the heading, the order, length, and centrality of the sentence; an appropriate noun; good communication; statistic knowledge; nouns and verbal expressions; entities by names; computational complexity in terminology; cue words; eventual frequency; and bushy pathway. These scoring methods, along with K-nearest Neighbors (IBK), the Multi-layer Learning algorithm, Multi-normal Logistic Regression (Logistic), Naive Bayes, Random Forest, and Random Forest (RF), were utilised as sample points for various machine-learning algorithms [10]. This was predicated on the assumption that when paired with snipping techniques, those same techniques provide site quality, but one's advanced computing capability is reduced. Then, to boost the uniqueness of the summary, the essential sentences are constructed. The strategy contains summaries of living things to acquire final grading. Once the computer has established the significant due, the performance is reviewed using the user-defined summaries [9]. The recommended sentence sequencing method reviews standardised sentences to begin creating the overall ranking. The operational efficiency overhead associated with projects that require wide-ranging ordering is eliminated by using a greedy research methodology to successfully complete this. This technology is suitable for comprehensive text summarizing computers because that just makes it simpler to rapidly renew sentences in extended summaries.

##### B. Cluster based scoring techniques

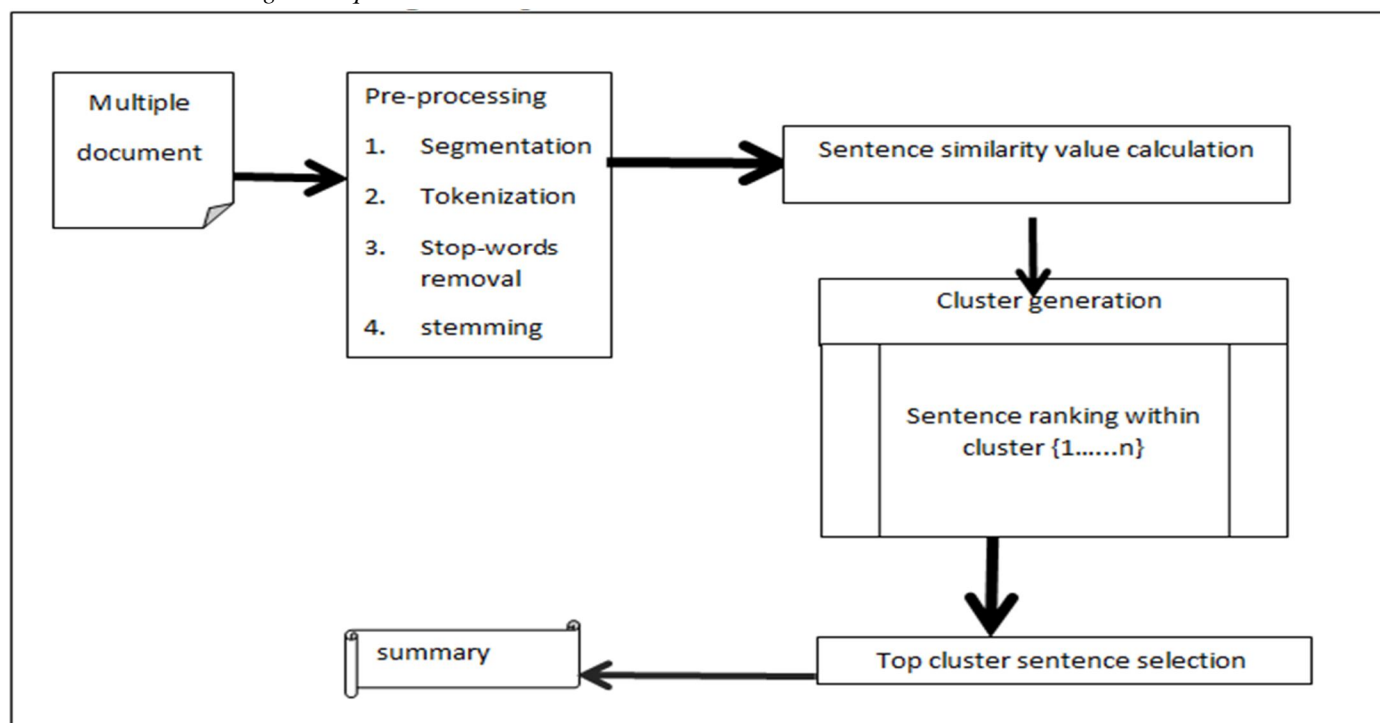


Figure no.3 (cluster based) [5]

To grade sentences, hierarchically clustering-based techniques compute their similarity, also known as the perspectives of the sentence. A clustering-based method for word extraction utilises Lex rank as an illustration of unsupervised learning in machine learning. The project is evaluated based on three parameters: positional concentration (documents contain  $N$  sentences, leading sentences are assigned a score of 1, and the score tends to decrease with a ratio of  $1/N$ ), value of centroid (the average cosine similarity between sentences and the rest of the sentences in the files), and eventually first prison terms intersect (the closest neighbors' similarity of a sentence with the first sentence in the same document). The parameters are linearly dense peak clustering (DPCS), which was suggested by Z. Hung et al. and predicted the phrases' variation and sweeping generalisation of the finding scores. It proceeds by employing the sentence similarity matrix to split the input data into sentences and eliminate any stop words. Continuing to follow that, a cosine curve illustrates the relationship using sentences as a collection of phrases. To generate new words, a word net is employed. Using keywords that are semantically virtually identical and frequently used words, sentence filtering is done for each text in the last phase. The phrases that contain frequently occurring terms and words that are close to them are taken from each page to generate the summary. Every moment, similar sentences are eliminated; a summary is designed. The map adjustment information will be collected for every significant connection resembling the same key and integrated in the collection. In the computer system, the algorithm's results are maintained. A file for every filler material an identification of the gathering of papers is gradually delivered using sentences comprising popular keywords [5]. This considerable focus on process awareness inside the MDS framework is incredibly complex and expensive too, though. Each summary has a specified limit. Hierarchical summaries of the findings of the study are provided. technical summary adds further technical data. events or any other phenomenon. Each summary already comprises coherence, which combines parent-child consistency, cross-functional coherence, and cross coherence. The viewer initially understands the information as a cause for worry since it is not very reliable. There are two main steps to the summary programming stage. Creating groups and clusters comes first. Hierarchical clustering produces the segmentation based on time. The extracted files' summaries are then chronologically executed. Recursively, the clustering method is utilised to choose the set of clusters that have been time-tagged prior to the aggregation. The sentences are parsed using the Stanford Parser. A sentence subject rating directs pages towards the topic. It highlights how different wording may have an impact. Such a regression analysis classifier, which also eliminates duplicate terms, was performed to examine the data set.

### C. Techniques for Lex Rank Score

The words have a lex rank of the file matches all words that have the same viewpoint. As a result, every lexical segment depicts A, which appears to be briefly described in the file. Instead of producing the noun to generate the notion of the term, a lexical chain is first constructed by taking the noun from the file. After the development of a lex rank, the best lexical chains are discovered and ordered from highest to lowest in overall strength [2]. Since each word in a lexical chain expresses the same concept, we select the most straightforward term from each lexical sequence to represent the subject of the lex ranking. Finally, we vote on summary phrases that include the common vocabulary. Similarly, to text rank, the Lex rank method for Nearly identical to text rank, the Lex rank approach for text categorization is a page rank sibling that also adopts a graph-based strategy. Utilizing the concept of Eigen vector centrality in a graph representation of sentences, lex rank is used to determine the relevancy of a phrase. In this theoretical framework, an inter-connectivity matrix based on single or multi-cosine similarity expresses the directed graph of phrases mathematically. The main goal of this sentence distillation procedure is to select a centroid sentence, which serves as the standard for all other sentences in the document. Then the sentences are rated according to how similar they are. The midpoint sentences, which serve as the standard for any additional sentences in the data, are what our sentence destabilises extraction aims to find. Depending on how virtually identical the sentences are, they are then graded. A central Eigen vector serves as the foundation for the visual method that we are using. Positioning sentences should be done at the graph intersections. Using the cosine similarity measure, the edges are given more weight. A technique called Lex Rank uses a graph-based phrase centrality score to categorically reject unsupervised text. The key structures are that to the viewer, sentences "recommend" other, absolutely similar phrases. Therefore, it almost always constitutes a statement of immense benefit if one conclusion is really quiet, like a number of others. Given the identify areas of improvement are that sentences "positive in the sense" to the viewer more, almost identical sentences. Therefore, if a conclusion is quite similar to numerous others, it's nearly always a phrase with considerable value. We have investigated systematically the idea and use of the Lex rank technique for text accumulation using the observations stated below. Using lex rank approaches, lex rank for SUMY implementation in NLP.

- 1) *Advantage of this Techniques:* Maintains redundancy, Improves coherency
- 2) *Disadvantage of this Techniques:* It cannot deal with dangling anaphora problem.

#### D. Latent Semantic Analysis Scoring Techniques (LSA)

A technique for maintaining summary priority in a document collection by applying latent semantic analysis (LSA). It typically generates a word and sentence matrix, with the columns showing the weighted term frequency vector of each phrase in the document collection. The mathematical terminology of singular value decomposition (SVD), which demonstrates the relationship between words and phrases on the input matrix, is only used to reconstruct the latent semantic structure. The document collection is evaluated to identify numerous themes, and during the summary, the sentences with the best average weights through all topics are adopted. In their research, Ferreira et al. extended their attempts at improving sentence construction by relying on phrase parallel processing and grammatical patterns. The significant features, in their opinion, have thus far been overlooked by the scholarly world: The significance question [4]. The representative's relevance to the number of queries and similarity to other candidates in the summary constitute the two core parts of the selection and recruitment score in MMR.

Each result of data analysis is included in these scores, and the process is concluded once all conditions have been fulfilled. The MMR method is effective [5] because lengthy articles typically have many long words. It is frequently used to extract phrases and other data from journals associated with a specific thread. It makes an informational multi-document summary less boring and predictable. Users will be able to separate text to add to the summary at each iteration. A sorted list of the available sentences is sent to the client. Through a series of difficult, activations of prior knowledge with feedback, the achievement of the summarization algorithms is analysed (CIQA).

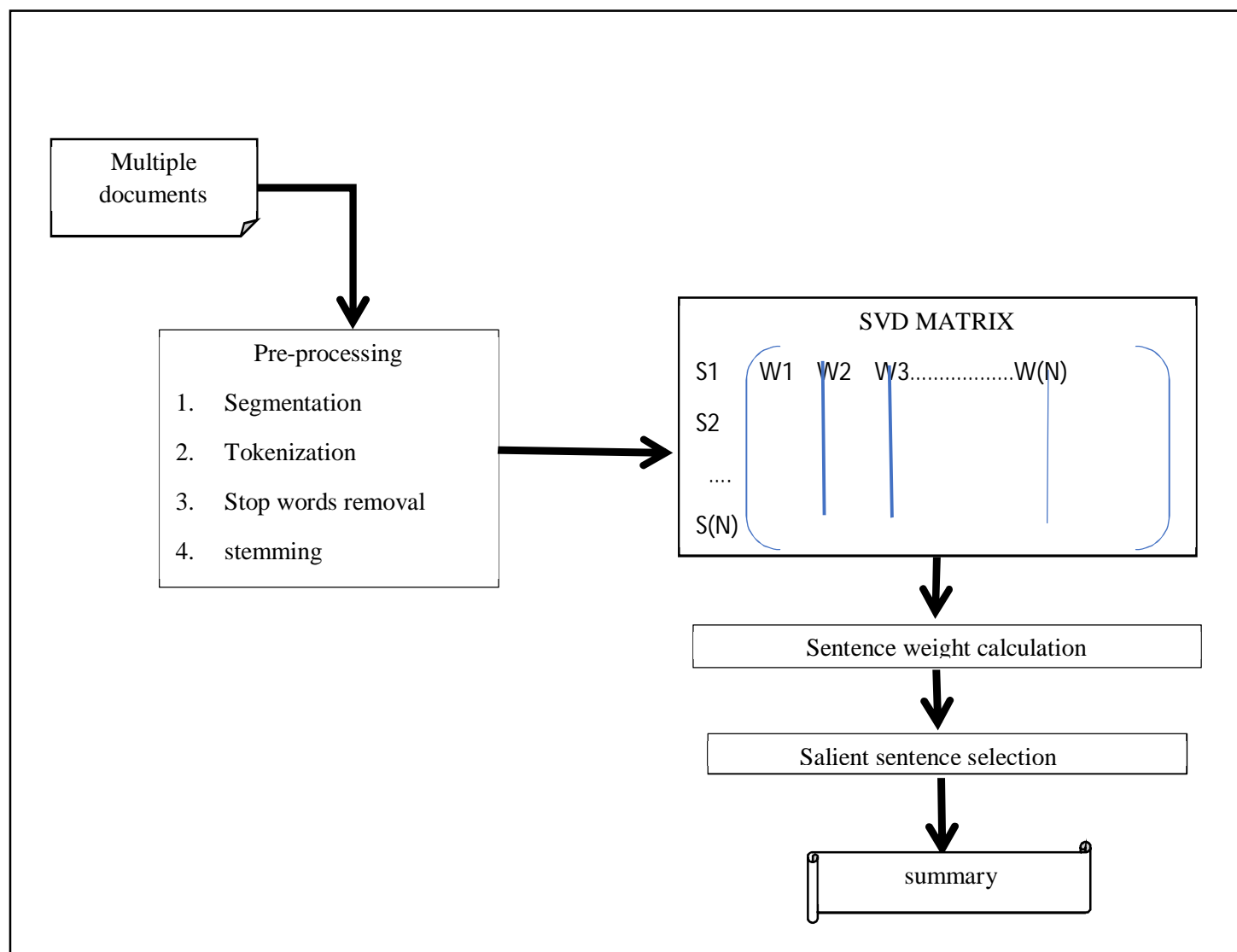


Figure no.4 Latent semantic analysis (LSA) [6]



### E. Statistical Technique

Sum Basic: Multi-document summaries are frequently prepared using this. By considering that the higher-frequency phrases in the word2vec model of the text have a higher probability of showing up in the document's summary, it makes use of the fundamental idea of probability. The chance increases as summarised sentences are generated. I'll continue to utilise the supervised learning kullback-liebler (kl) sum technique. We shall analysis the kl sum algorithm (Kullback-Liebler (kl) sum algorithm, where the summary length is fixed (L words). Even as kl divergence decreases, this algorithm eagerly seeks to add words to a summary.

1) *Introduction to KL-SUM*: In this approach, we build a collection of paragraphs that are as similar to the original material as feasible, with ranges less than L and Uni-gram composition. A continuous chain of n items from a data piece of audio or text is referred to as an "n-gram" or "Uni-gram" in the study of natural language processing (NLP). The evaluation of this method: The Kullback-Liebler divergence (relative entropy) in parameter estimation measures how different one probability distribution is from another. The summary and the document are essentially the same, with less variation and more meaning being interpreted effectively. The Kullback-Liebler (kl) divergence is described by  $KL(p||q)$ .

$$Q = \log p(w) q(w) \dots \dots \dots (1)$$

2) *Algorithm*: It uses a greedy optimization approach;

- a) Set  $s = \{ \}$  and  $d = 0$ .
- b) While  $||s||=L$  do:
- c) For  $I$  in  $[1 \dots N]$ ,  $d_i = KL(p_s | p_d)$ .
- d) Set  $s += s_i$  to the smallest  $d_i$  and  $d = d_i$  to the smallest  $d_i$ .
- e) Stop if there is no "I" such that  $d_i$

The possible limitation is to help organize the selected phrases in the major influence on the way by the value of  $p_i$ . The technique used is to compute a position for each selected phrase  $s$  from document  $D$  by following the order in the source documents.

The index  $p_i$  (in  $[0 \dots 1]$ ) represents the location of  $s_i$  under  $D$ . The words or phrases in the implementing the best are arranged according to the order of  $p_i$ .

### 3) Properties of k L -divergence

- a) The Kullback-Liebler divergence is often not negative.
- b) For statistical distributions, the kl continues to be well defined and is invariant, perhaps because of fluctuation.
- c) If the kl  $D(p^*q)$  in two probity mass functions  $(p_1, q_1)$  and  $(p_2, p_2)$  is unique, then the two probity mass functions  $(p_1, q_2)$  and  $(p_2, p_2)$  are convex.
- d) For independent distributions, the Kullback-Liebler divergence, like the Shannon entropy, is symmetric.

KL occasionally appear in machine learning, and a thorough understanding of what the kL divergence measures is quite beneficial. If you are interested into learning more about kl divergence software packages in statistics, I recommend reading papers on Bayesian inference. In information theory, kl-divergence has a long history. The following are success in learning. If you enjoy deep if you love deep learning, two very important concepts in the field using kl divergence right now are vase and information bottleneck.

### F. Tool Used

- 1) Libraries: NLTK (Natural Language Toolkit)
- 2) WORDNET LEMMAIZER
- 3) MATH
- 4) MATPLOTLIB
- 5) Tool: Python with NLP
- a) *NLTK*: Natural Language Toolkit, which contains packages for text processing libraries such as Tokenization, parsing, classification, stemming, tagging, character and word counts, and semantic reasoning.
- b) *Word Net Lemmatization*: It is a module in the NLTK stem. It performs stemming using different classes. It is a processing interface for removing morphological affixes, for example, grammatical role, tense, and derivation morphology, from words and learning only the word stem.
- c) *Math*: This module provides access to the mathematical function summarization that mathematical concepts like linear algebra are required for pertaining to vector spaces, matrices, etc. It is essential while calculating the frequency score.

- d) *Mat Plot Lib*: Its modules the plots the graph and structures of required the correct graph for particular topic in python .in the make the graph by this software.

## REFERENCES

- [1] AL Zuhair, A, and AL.Handheld , M., "An approach for combining multiple weighting schemes and ranking methods in graph based multi -document summarization", IEEE Access, vol 7 ,pp.120375-120386,2019.2.
- [2] Cohen A., and Goharian, "Scientific document summarization", IEEE International journal on digital libraries, vol 19(2), pp.87-303,2019.
- [3] Ashraf, M. Zaman and M. Ahmed, "To ameliorate classification accuracy using ensemble vote approach and base classifiers", emerging technologies in data mining and information security, pp.321-334, springer ,2019.
- [4] Sabahi, K. Zhang, Z.P. Nadher, "A Hierarchical Structured Self-Attentive Model for Extractive Document Summarization (HSSAS)", IEEE Access, vol. 6, pp. 205–242, 2018.
- [5] Chen, S.H. lieu and H.M., "An information distillation framework for extractive summarization", IEEE\ACM trans audio speech lang-process., vol26, pp.161-170,2018.
- [6] Naik, S.S., and Ga Onkar, M.N., " IEEE international conference on recent trends in electronics information and communication technology (RTEICT), vol. 9, pp. 156–178, 2017
- [7] Fang, s., Mud., Deng z., and Wu, z., " word sentences co-ranking for automatic extractive text summarization", In expert systems with applications, vol 72, pp.189-195,2017.
- [8] Wang, S., Zhao, X., Li, B., Ge, B., Tang, D.Q., "Integrating extractive and abstractive models for long text summarization," IEEE international congress on big data big edge data congress, pp. 305–312, 2017.
- [9] M. p. Agus., and D. Suharto no, " Summarization using Term Frequency Inverse Document Frequency (TF-IDF)" by Christen, Com. Tech., vol. 7, pp. 285-294, 2017.
- [10] Liu., C. tseng., Machan, "Incest's: towards real time incremental short text summarization on comment streams from social network service", IEEE trans, Knowle., .data Eng., vol 60, pp114,2015.
- [11] KIM, H. Moon and H. Han," A weight adjusted voting algorithm for ensembles of classifiers", Journal of the Korean statistic society, vol 40, pp.437-449, march,2011.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)