

FedCSIS_2023_paper_8340.pdf

A Review of Automatic Text Summarization Techniques

Anjana Kishanpuri¹, Ph.D. (I.L.E.), Dr. Harshalata Petkar², Assistant Professor

Centre for Information and Language Engineering

Mahatma Gandhi Antarrashtriya Hindi Vishwavidyalaya, Wardha, Maharashtra.

E-mail - anjanakishanpuri@gmail.com, harshalata.p@gmail.com

Abstract: A key of the Automatic Text Summarization system for Indian and Foreign Languages has been presented in this paper. Text Summarization is one of the central areas of Natural Language Processing. A text Summarization system plays a vital role in facilitating a meaningful summary of a large text in any language. There is a wide availability of data in almost all languages on the Internet. So here, there is a need for text Summarization for reading more in less time and we can also decide whether to read it or not. Automatic Text Summarization has become the necessity of many applications for example- Shortest News, Search Engines, Business Analysis, Legal Document Summaries, etc. Summarization Approaches are primarily classified into two categories: extractive and abstractive. Extractive techniques aim to select salient phrases, sentences, or elements from the text while Abstractive techniques focus on generating summaries from scratch without the constraint of reusing phrases from the original text. In this paper, we will provide a survey in detail on the various methods and approaches of Text Summarization that are mostly used in the present scenario.

Keywords – Automatic text summarization, Extractive, Abstractive, Machine Learning, Information Extraction.

1. Introduction of Automatic text summarization: With the dramatic growth of web data, information overload has become a serious problem for humans and machines. Automatic text summarization systems provide an indispensable solution to mitigate the problem of information overload. That system creates short summaries of documents to present the most salient content in a condensed form to suit the needs of the task and the user. Thus, helping users in assimilating maximum information in minimum time.

Automatic text summarization is one of the Extensive Applications in Natural Language Processing. Automatic text summarization helps people from different locations to understand and create a Summary of the text of any language without the help of manual summarization by a human. Its Functionality and Speed are more than manual Summarization.

Summarization methods are primarily classified into two categories: extractive and abstractive. Extractive methods aim to select salient phrases, sentences, or elements from the text while abstractive techniques focus on generating summaries from scratch without the constraint of reusing phrases from the original text. Sentence extraction is a crucial step in the extractive summarization system.

Automatic text summarization has a wide range of applications viz Shortest News, Search Engine, Social Media Summarization, Domain-Specific Summaries, Business Analysis, Legal Document Summaries, Review of content, etc.

This Research Paper has been divided into 4 parts – (1) Introduction of Automatic Text Summarization. (2) Various Approaches and Methods of Text Summarization have been introduced in detail. (3) The Research Work done on Text Summarization so far has been discussed based on languages along with their characteristics, domain, etc. (4) and finally, the paper is concluded.

2. Classification of Automatic Text Summarization Techniques:

The earliest research on TS began in 1958, including the recurrence of a specific word, the positioning of the sentence, and the presence of keywords. Since then, technology has come a long way and has consequently provided huge improvements in the field of automatic summarization. There are now multiple methods used to achieve accurate summaries [1].

In the previous section, we provided definitions of automatic summarization systems. In this section, we aim to provide different categorizations of text summarization techniques. There are several classifications of summarization based on the function of summary, type of summary, genre of document, number of documents, and supported media types [2, 3]. We list summarization categories as follows:

(1) Summarization categories based on the type of generated Summary:-

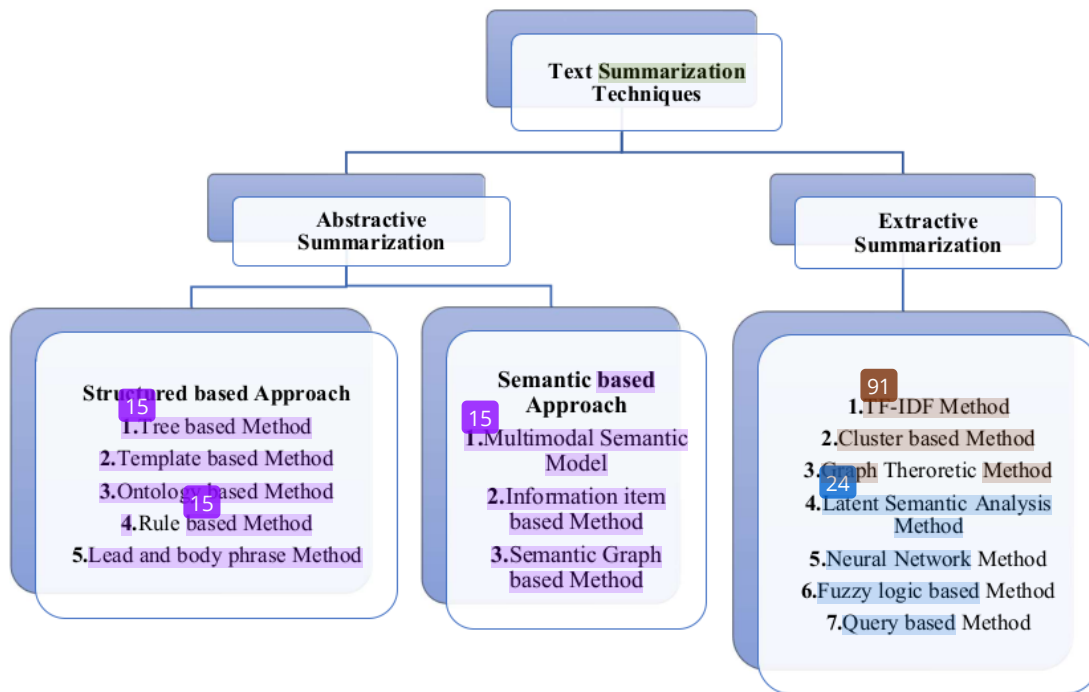


Fig.1.Text Summarization Techniques

Abstractive Summarization: Abstractive summarization process consists of understanding the original text and re-telling it in fewer words. In abstractive summarization semantic analysis of the documents is done on basis of which a summary of the document is generated. In this type of summarization interpretation of each of the sentences is done and may be represented in a different style from the original one [3].

Abstractive Summarization Approaches: Structure-based Methods

The structured-based approach encodes the most important information from the document through cognitive schemes such as templates, extraction rules, and other structures such as a tree, ontology, lead, and body phrase structure [4,5].

(a) Tree-based Method - This technique uses a dependency tree to represent the text/contents of a document. The technique uses either a language generator or an algorithm for the generation of a summary. The major strength of this approach is that the use of language generators significantly improved the quality of resultant summaries i.e. reducing repetitions and increasing fluency. The problem with this approach is that context of the sentence was not included while capturing the intersected phrase [4].

(b) Template-based Method - This technique uses a template to represent a whole document. Linguistic patterns of extraction rules are matched to identify text snippets that will be mapped into template slots. These text snippets are indicators of the summary content. The templates are filled with important text snippets extracted by the Information Extraction systems. A significant advantage of this approach is that the generated summary is highly coherent because it relies on relevant information identified by the IE system. This approach works only if the summary sentences are already present in the source documents. It cannot handle the task if multi-document summarization requires information about similarities and differences across multiple documents [5].

(c) Ontology-based Method - Most documents on the web are domain related because they discuss the same topic or event. Each domain has its knowledge structure and that can be better represented by ontology. In this method, domain ontology for news events is defined by the domain experts. The next phase is the document processing phase. Meaningful terms from the corpus are produced in this phase. The meaningful terms are classified by the classifier on basis of events in the news. A membership degree is associated with various events of domain ontology. A membership degree is generated by fuzzy inference. The limitations of this approach are it is time-consuming because domain ontology has to be defined by domain experts. The advantage of this approach is it handles uncertain data [5].

(d) Rule-based Method - In this method, Documents to be summarized are represented in terms of categories and a list of aspects. It has the potential for creating summaries with greater information density than the current state of the art. The drawback of this methodology is that all the rules and patterns are manually written, which is tedious & time-consuming [5].

(e) **Lead and Body Phrase Method** - This method is based on the operations of phrases (insertion and substitution) that have the same syntactic head chunk in the lead and body sentences to rewrite the lead sentence. It is good for semantically appropriate revisions for revising a lead sentence. The limitation of this method is parsing errors degrade sentential completeness such as grammaticality and repetition. It focuses on rewriting techniques and lacks a complete model which would include an abstract representation for content selection [5].

Abstractive Summarization Approaches: Semantic-based Methods

11 In the Semantic-based approach, the semantic representation of the document is used to feed into the natural language generation (NLG) system. This method focuses on identifying noun phrases and verb phrases by processing linguistic data [4,5].

(a) Multimodal Semantic Model - In this method, a semantic model, which captures concepts and relationships among concepts, is built to represent the contents (text and images) of multimodal documents. The important concepts are rated based on some measure and finally, the selected concepts are expressed as sentences to form a summary. An important advantage of this framework is that it produces an abstract summary, whose coverage is excellent because it includes salient textual and graphical content from the entire document. The limitation of this framework is that it is manually evaluated by humans [5].

(b) **Information Item-based Method** - In this method, the contents of the summary are generated from an abstract representation of source documents, rather than from sentences of source documents. The abstract representation is the Information Item, which is the smallest element of coherent information in a text. From this method, a short, coherent, information-rich, and less redundant summary can be formed. Despite so many advantages, this method has also many limitations. While making grammatical and meaningful sentences, many important information items get rejected. Due to this, the linguistic quality of the resultant summary gets reduced [5].

(c) **Semantic Graph-based Method** - This method is used to summarize a document by creating a semantic graph called Rich Semantic Graph (RSG) for the original document, reducing the generated semantic graph. It produces concise, coherent, less redundant, and grammatically correct sentences. This method is limited to single document abstractive summarization.

5 **Extractive Summarization:** An extractive summarization method only decides, for each sentence, whether or not it will be included in the summary. In the extractive summarization system, different weights are assigned to each sentence of the document on which sentence is selected to get added to the summary. Weights can be assigned to the sentences according to the position of the sentence in the document i.e. sentences in the beginning and at the end are assigned more weight as they are supposed to contain more valuable information. Weights can also be assigned according to the type of information they contain. For example, if a sentence contains the name of a person, and the date of the event that occurred then more weight is given to that sentence than those which do not contain any Named Entity [3].

Classic Methods of Extractive Text Summarization

(a) **Term Frequency-Inverse Document Frequency Method** - It is a word distribution method in which two measures namely term frequency (TF) and document frequency (DF) are calculated for each non-stop word (w) in the document. Term frequency (TF), indicates the number of times a word appears in the text which measures the salience of the word within that document. Document frequency (DF) indicates the number of documents in which the word appears. The frequent occurrence of a word in a document is treated as an informative word, which is calculated by document frequency measure. Thematic words are obtained by comparing the ratio between two frequencies, referred to as (TF*IDF) measure. Once the (TF-IDF) score has been computed for each word the next step is to calculate the number of such thematic words per sentence. With this value sentences in the Input text are ranked and the highest-scored sentences are picked to be part of the summary. The redundancy of information is extremely high in this method [4].

(b) Cluster-based Method – It is an unsupervised machine learning method. It is intuitive to think that summaries should address different “themes” appearing in the documents. If the document collection for which the summary is being produced is of totally different topics, document clustering becomes almost essential to generate a meaningful summary. Several clustering techniques are also available for text categorization like K-means, Suffix tree Clustering, Label Partitioning (LIPART) Algorithm, and Semantic Hierarchical Clustering (SHOC). sentence selection is based on the sentences’ similarity to the cluster’s theme (Ci). The next factor is the location of the sentence in the document (Li). The last factor is its similarity to the first sentence in the document to which it belongs (Fi). $S_i = W_1 * C_i + W_2 * F_i + W_3 * L_i$ Where, W1, W2, and W3 are weightages for inclusion in summary [4].

(c) Graph-Theoretic Method - Representation is an extractive summarization model, which provides a method to identify themes in the document. Pre-processing steps, namely, stop word removal and stemming are done before, to obtain a graphical view of the documents. Graphic representation yields partitions indicating distinct topics covered in the documents. Identification of nodes with high cardinality forms higher preferred sentences to be included in the summary.

This system uses statistics of the frequency of words and a term positional and weight-age calculation by string pattern for scoring sentences [4].

(d) Latent Semantic Analysis Method – It is a supervised machine learning method. Latent Semantic Analysis (LSA) is an algebraic statistical method that extracts the meaning of words and the similarity of sentences using the information about the usage of the words in the context. It keeps the information about which words are used in a sentence while preserving information about common words among sentences. The more common words between sentences mean that those sentences are more semantically related. The LSA method can represent the meaning of words and the meaning of sentences simultaneously. The LSA method uses Singular Value Decomposition (SVD) for finding out semantically similar words and sentences. SVD is a method that models relationships among words and sentences. It has the capability of noise reduction, which leads to an improvement in accuracy [6].

(e) Neural Network Method – A Machine Learning technique i.e. an artificial neural network is implemented to generate a text summarization system. This method involves training the neural networks to learn the types of sentences that should be included in the summary. Such a thing is accomplished by training the network by using several test paragraphs which identify each sentence whether to be included in the summary or not. These test paragraphs were framed by human readers. By using this training the neural networks inherit the pattern of the sentences that should be included in the summary. Among the other Neural Network methods, the universal function approximation technique called three-layered feed-forward network is used [7].

(f) Fuzzy Logic-based Method – A Design of a Fuzzy logic system usually involves selecting membership functions and fuzzy rules. The four main components of Fuzzy Logic were: the zipper, the inference engine, the de-fuzzifier, and the fuzzy knowledge base. In the fuzzifier section, snappy inputs are translated into linguistic values, using a membership function, which is to be used for the input linguistic variables. After fuzzification, to derive the linguistic values, the inference engine refers to the rule base containing fuzzy IF-THEN rules. Finally, the de-fuzzifier converts the output linguistic variables from the inference to the final crisp values using the membership function which represents the final sentence score. The output membership function in the de-fuzzification step is divided into three membership functions: Unimportant, Average, and Important, which are used to convert the inference engine result into a crisp output to obtain a final score for each sentence [7].

(g) Query-based Method - In a Query-based text summarization system, the sentences in a given document are scored based on the frequency counts of terms. It uses Vector Space Model. The sentence containing the query phrases are given higher scores than the ones containing single query words. Then, the sentences with the highest scores are incorporated into the output summary together with their structural context. Portions of the text may be extracted from different sections or subsections. The resulting summary is the union of such extracts [5].

(2) Summarization categories based on the number of documents to summarize:

i. Single Document Summarization - Single Document Summarization system takes on one text document as an Input and transforms it into a condensed, shorter text in which the relevant information is preserved. Sentence ranking is an important step in such a system.

ii. Multi-Document Summarization – Multi-Document Summarization is a process of representing a set of documents with a short piece of text by capturing the relevant information and filtering out the redundant information.

(3) Summarization categories based on Content/Function of summary:

i. Indicative Summary - This type of summary categorizes the document's topics and characteristics such as length, writing style, etc. This sort of summary is required for writing an abstract for a less-structured document like an essay, editorial, or book. It helps to decide whether a user wants to read the document or not [8].

ii. Informative Summary - This summary is for writing an abstract for a strictly structured document like an experiment, investigation, survey, etc. It contains generally longer content and reduces the length of other original text by 70-80%. It includes facts and information which can replace the original text [8].

iii. Evaluative Summarization - It aims to capture the opinion or views of the author on a given topic/subject/product. Sometimes, it is also referred to as review or Opinion based summarization [9].

(4) Summarization categories based on input and output language:

i. Mono-lingual summarization - This type of summarization includes the input documents and the target document being in the same language. Example: English Summarization System [8].

ii. Multi-lingual summarization - When the source document is in several languages like English, Hindi, and Punjabi and a summary is also generated in these languages, then it is termed a multi-lingual summarization system [8].

iii. Cross-lingual summarization - This type of summary includes a source document in one language and a summary generated in another language [8].

3. Literature Review

The review of research papers on text summarization is important because summarizing extractive techniques has become a very broad research topic and is heading toward maturity [10]. Now research has shifted towards abstractive summarization and real-time summarization [10]. This is because abstractive summaries are more complex and complicated than extractive summaries. So extractive summaries are easier to give expected and better results than abstractive summaries [11,12,13].

1. **Review Method**-This review research on text summarization was conducted with the Rapid Review method.

	Research Question	Motivations
RQ1	In which Journal/Conferences paper is the Text Summarization available?	Identify the most significant Journal/Conferences papers on Text Summarization.
RQ2	What Journal/Conferences paper about Text Summarization?	Identify the research topic/trend of Text Summarization.
RQ3	What Features are available in Text Summarization?	Identify what features in Text Summarization.
RQ4	Which Techniques are available in Text Summarization?	Identify techniques that are often used in Text Summarization.
RQ5	Which Methods are available in Text Summarization?	Identify the method used in Text Summarization.
RQ6	Which Challenges/Problems are available in the Text Summarization?	Identify the challenges/problems in the Text Summarization.
RQ7	What evaluation techniques are used in text summarization?	Identify what evaluations are carried out in the text summarization

Table1. Research Question and Motivation.

2. **Research question and motivation** - RQ is constructed to facilitate the review process to be more focused and consistent.

The research question and motivation in this review paper are described in Table1. RQ3 through RQ7 are the main study research questions, while RQ1 and RQ2 are assigned to help evaluate the context of the main study. RQ3 shows the features used and RQ4 to RQ7, namely, techniques, problems, methods, and evaluations. RQ1 and RQ2 provide a synopsis of certain areas of research in summarizing texts [14].

Property	Research Question
Publication	RQ1
Research topic or trend	RQ2
Text summarization feature	RQ3
Text summarization technique	RQ4
Text summarization method	RQ5
Text summarization problem	RQ6
Text summarization evaluation	RQ7

Table2. Data Extraction.

3. **Data Extraction** - The data extraction stage is a process that functions in gathering data from the main study to answer research questions. Table 2 The following is the data extraction table used.

4. Result –

1. Publications – research paper studies

There are discussed 75 research papers on text summarization from 2001 to 2020. Fig. 2 shows a graph of the establishment of the number of paper publications for twenty years, From the graph, it appears that research on text summarization is still relevant. Our study discusses text summarization techniques and methods used in Hindi/Indian languages and foreign languages.

A large number of research on text summarization is in 2020 with 10 publications. From the graph, it can be seen that from 2001 to 2012 research on this study was of little interest, namely 1,2 or 3 publications. The research of this study experienced significant growth in 2016, with 10 publications out of 75 selected publications. The beginning of this research went quite well in 2013 until now, when, and it had 10 publications in 2020.

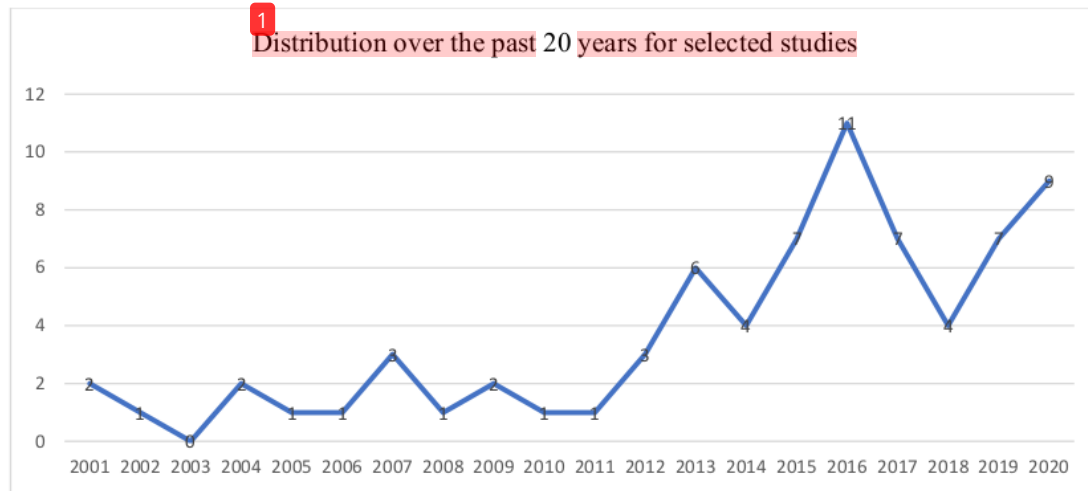


Fig.2 Partition over the past Twenty years for Text Summarization.

3 Publication journal papers or important conferences based on literature reviews are shown in fig.3.1 and fig.3.2. there are some publication journal papers or conferences that are based on Hindi/Indian languages and some of them are based on foreign languages.

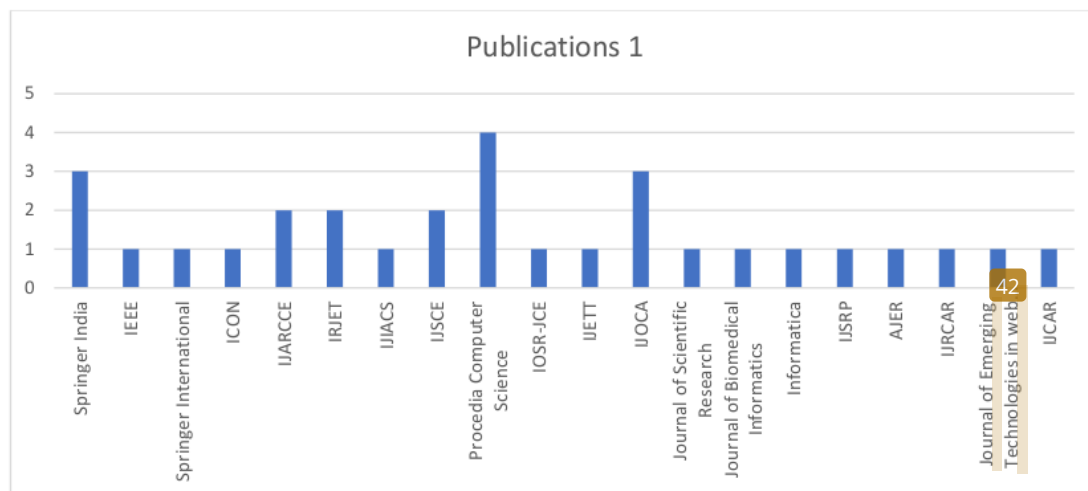


Fig.3.1 Publications and Conferences based on Text Summarization of Hindi/Indian Languages.

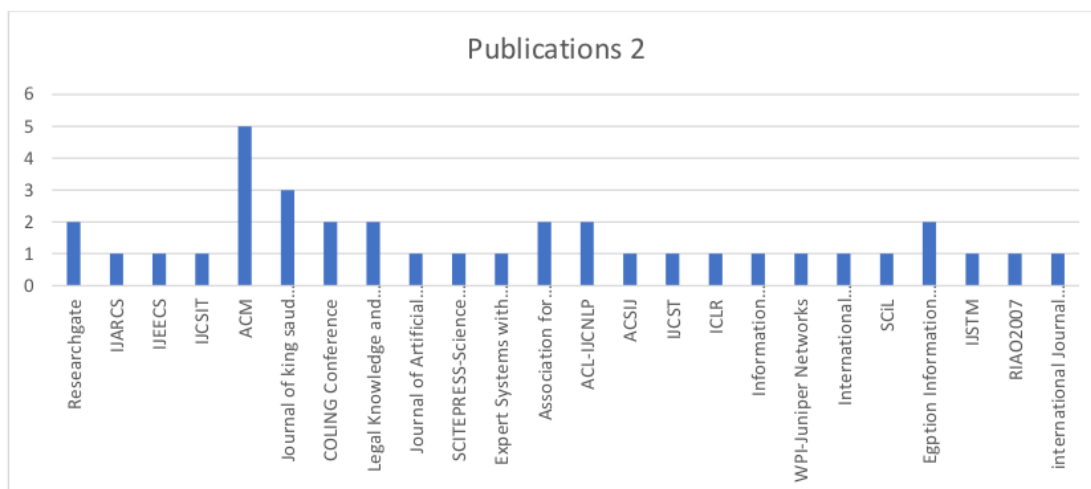


Fig.3.2 Publications and Conferences based on Text Summarization of Foreign Languages.

2. Topics and Trends Research

Automatic text summarization research includes several research topics or trends. Over the last 20 years, there have been 6 research topics or trends on text summarization, that is, abstractive, extractive, single document, multi-document, domain-specific, and machine learning. The dissemination of research topics or trends in text summarization is shown in fig.4.

A great number of favorite research topics or trends of the last 20 years are Domain-based research, which is 40%. Domain-specific research on text summarization is more challenging because the domain is classified into many parts i.e. language-based, subject-based, etc.

After domain-specific topics, abstractive text summarization and machine learning-based text summarization are in trend.



Fig. 4 Dissemination of Topic or Trends Research

Table 3. Detail Topics or Trends in Text Summarization.

Topic/Trends	Research
Extractive	k. Vimal and Yadav(Jan 2015, Sep 2015), Gulati and Sawarkar(2016), Thaokar and Malik(2013), Gupta et al.(2012), M. Bhanu and et al.(2007), Ajmal and Haroon(2015), Sarkar(2012), De ⁸⁵ nde et al.(2013), Sarwadnya and Sonawane(2018), Rathod(2018), Kallimani et al. (20 ⁹² Mohamed and Hariharan(2016), Roy et al.(2017), Megala et al.(2014), Gong and Liu(2001), Erkan and Erkan(2004), Ozsoy et al.(2010), yeh Jen yuan et al.(2005)
Abstractive	Subramaniam and Dalal (2015), Gaikwad et al.(2016), Kallimani et al.(2014), Sudha and latha(2020), Malviya and Tiwary(2016), Timofe ¹ y and Choi(2017), Brazilay and Elhadad(1997), Ramezani et al.(2015), Jingqine et al (2019), Wonjin ⁹³ (2020), Bhargava et al.(2016), kai Qiang song et al. (2019), Bansal et al. (2018), Lei Li et al. (2020), Fei Liu et al. (2019)
Machine Learning	⁸⁴ ashi and Yuji (2001), Dalal and Malik (2017), Vipul and Latesh (2017), Patel et al. (2007), J. Anitha et al. (2015), Khanam et al. (2016), Wang and Li Tao (2009), Rossiello (2016), Nallapati et al. (2016), Gallinari (2002), won jin et al. (2020), Ambati (2020)
Domain Hindi	Domain Hindi Sargule and kagalkar(2016), k. Vimal and Yadav(Jan 2015, Sep 2015), Gupta(2013), Dalal and Malik(April2017, July2017), Subramaniam and Dalal(2015), Patil et al.(2015), Gupta and Garg(2016), Gulati and Sawarkar(2017, 2019), Thaokar and Malik(2013), Reddy et al.(2015)
Panjabi	domain Panjabi Gupta et al(2012)
Tamil	domain Tamil Mohamad et al.(2016), Bhanu et al.(2007)
Malayalam	domain Malayalam Ajmal and Haroon(2015)
Bangali	domain Bengali News Sarkar(2012),
Telugu	domain Telugu Khanam et al.(2016), domain Telugu Sports Sudha and Latha(2020)
Marathi	domain Marathi news article Rathod(2018), domain Marathi Sarwadny and Sonawane(2018)
Kannada	¹ main Kannada Kallimani et al.(2012,2014)
Turkish	domain Turkish Ozsoy et al.(2010),
Arabic	domain Arabic Imam et al.(2013),
Persian	domain Persian Ramzani et al.(2015),
Hebrew	domain Hebrew Litvak et al.(2016), domain Multi-lingual MUSEEC Vanetik et al.(2016),
Gujrati and Urdu	domain Gujrati and Urdu Alkesh et al (2007),
Egyptian	domain Egyptian K. Nandhini (2013),
Legal Document	domain legal document Megala et al.(2014), domain legal Jurix Farzindar et al.(2004)
Neural document	domain Neural document Sandeep et al (2019)
Single document	⁵⁸ Sargule and kagalkar (2016), Sarkar et al.(2013), Imam et al.(2013), Alkesh et al (2007), M. Bhanu et al (2007), kallimani et al (2012), Litvak Marina et al (2016), yeh jen Yuan et al (2005), Elena and manuel(2010), Rasim M. Alguliev et al (2011), Subramaniam et al (2015)
Multiple documents	Gulati and Sawarkar(2017, 2019), Gupta (2013), Sarwadnya and Sonawane(2018), Patel et al.(2019), wang Dingding et al (2008,2009), Rasim M. Alguliev et al (2011), Sudha and Latha (2020)

3. Features in Text Summarization

Properties of text or features of text used in text summarization research from 2001 to 2020 are described in fig. 5. features of the text in text summarization are significant characteristics to be extracted to produce a 18 size summary. In the last twenty years of text summarization research are often used approximately eleven features of the text.

The most popular features are sentence length and sentence position. In the sentence length feature, long sentences contain more important or relevant information. That means short sentences do not cover any relevant information, so short sentences are considered unimportant or ignored.

Sentence position is the most studied feature in extractive text summarizing [15,16,17,18,19,20]. Various methods have been used to feature the sentence position, one of which is weighting with the reverse sentence order. Its use for example with a text of 30 sentences, the first sentence will be 30 times more important than the last. So it's almost impossible the last sentence will appear in the summary.

Recent research that uses quite complete features are word features (title word, proper noun word, thematic word, keywords using TF-IDF, numerical data), sentence features (sentence length and sentence position), and scoring features using the fuzzy logic system [21]. Combining the statistical approach with fuzzy-based, namely the TF-IDF method and fuzzy logic with multi-document text input, multi-document summarization (MDS) produces extractive summaries. The advantage of the approach taken is to use sentence assessment in a better tuning process. The disadvantage is the problem of sorting sentences in summary. Ordering sentences is a very difficult task but very important in summarizing documents. One way that can be done in the future is to improve the MDS form by adding semantic and linguistic features to produce a more coherent summary [21].

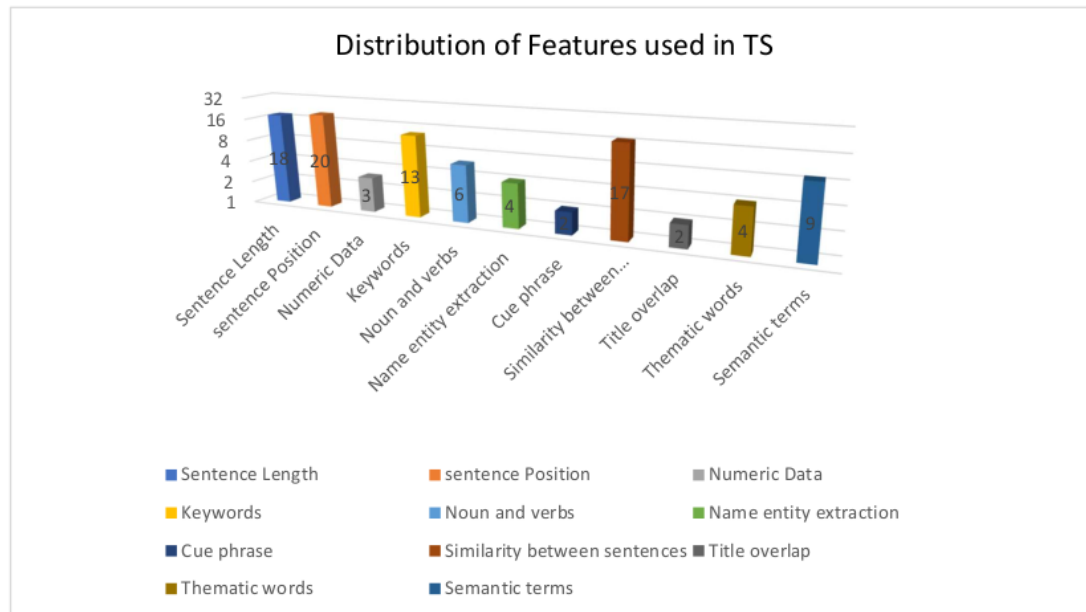


Fig. 5 Description of features in Text Summarization

4. Approaches Techniques for Text Summarization

In the literature studies that have been obtained over the last twenty years, there are six approaches or techniques used in text summarization, that is fuzzy-based, machine learning, statistical, Graph-based approach, thematic approach, and rule-based. To find out the distribution of approaches and techniques for text summarization in the past twenty years, it can be shown in Fig.6.

The most popular approach technique used in text summarization is machine learning. The machine learning performance is automatic and learns to improve from experience without being explicitly programmed.

The less popular approach in research on text summarization in the last 20 years is rule-based. The rule-based approach has the advantage of being applied to a simple domain, so the rule-based is easy to verify and validate but has weaknesses when applied to a domain with a high level of complexity, if the system cannot recognize the rules, then no results are obtained [22]. Besides, if there are too many rules, it will become difficult to maintain the system's performance.

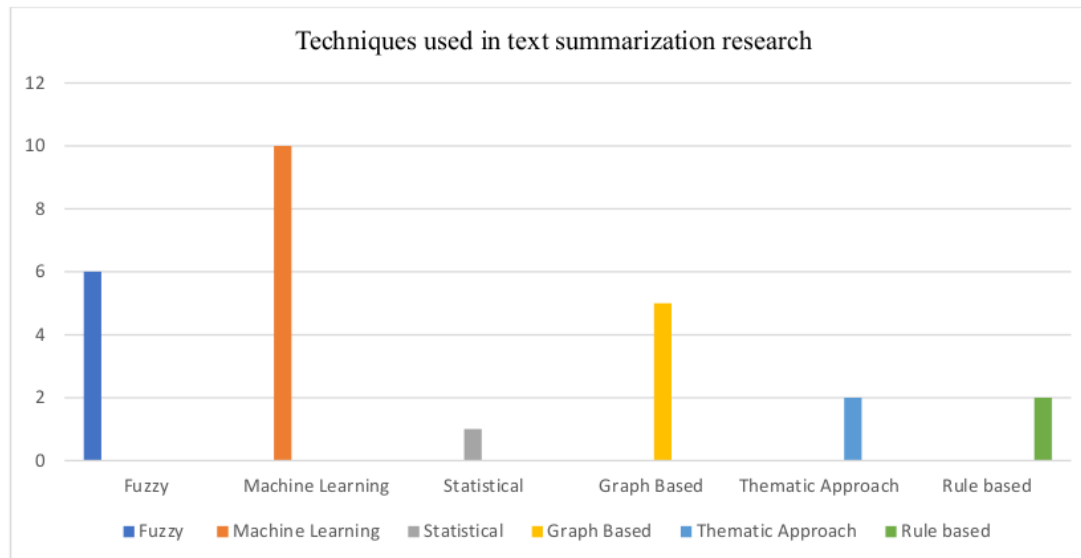


Fig.6 Distribution of techniques used in text summarization

5. Methods for Text Summarization

The method of text summarization research that is used there are various kinds of various approach techniques. The distribution of methods from 2001 to 2020 for this research is described in Fig.7. Based on literature studies, the most widely used method in text summarization over the past twenty years is the neural network and fuzzy logic. A neural network is trained to learn the relevant features of sentences that should be included in the summary of the article. The network observes the pattern of the dataset and gets trained. For each neuron, the network calculates what the target output is and a scaling factor, how much lower or higher the output must be adjusted to match the target output. The network performs better if it is trained more. Fuzzy logic is often used to extract or determine the final value of words or sentences included in the summary. The fuzzy logic approach can prevent data contradiction because it involves the role of humans to be able to examine sentences and reach an agreement on the choice of certain sentences to produce summary sentences [23].

The most widely used methods for summarizing texts are TF-IDF. TF-IDF is a method with a statistical approach technique. Some text summarization studies that use this method are [24,2,6,25,26,27] etc. This method is one algorithm that can be used to analyze the relationship between a phrase/sentence with a collection of documents. The concept is to calculate the TF value and IDF value [28]. For a single document, each sentence is considered a document. TF is the frequency of occurrence of the word (t) in the sentence, which shows how important the word is in the document. While IDF is the number of sentences where a word (t) appears, which shows how common the word is. Word weight will be greater if it often appears in a document and smaller in many documents. To calculate the weighting of words in a document is to multiply the TF value by the IDF value.

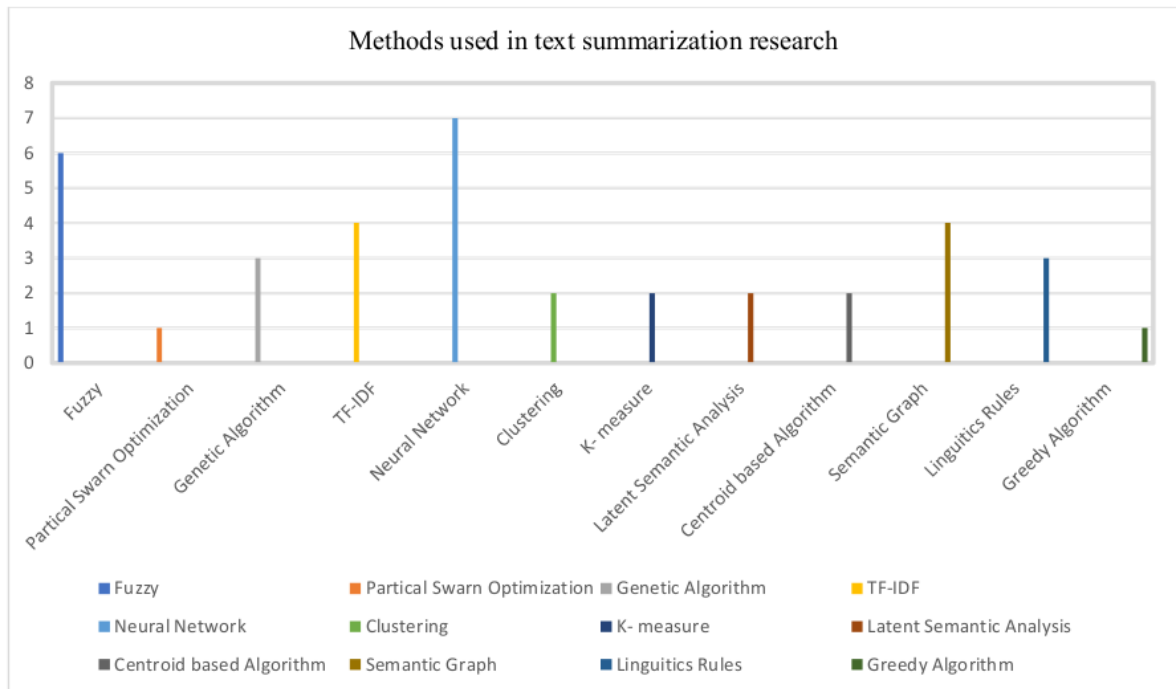


Fig. 7 Distribution of Methods used in text summarization

6. Challenges in Text Summarization

Based on a text summarization research study from 2001 to 2020, several problems have become challenges that have been tried to be resolved. Details of the problems in the text summarization in the last twenty years can be shown in Fig. 8. the problem most often done in text summarization over the past ten years is extraction because extraction is a fairly challenging problem. In-text summarization, extraction is the process of taking data from a data source (structured or unstructured data) to be processed to produce a summary. [14] problems that have rarely been done in the last 20 years are word frequency, sentiment analysis, sentence relevancy, overlapping, clustering and sentence ranking, etc. Word frequency and sentence ranking are less challenging problems because these problems can be solved by a statistical approach. While sentiment analysis is one of the problems that arise because it is part of the benefits of text summarization that is used to determine the sentiment analysis of the contents of the text. A less desirable but quite challenging problem in text summarization is determining new features to produce a summary.

One other problem of redundancy and similarity is the same, that is, looking for sentences that are the same or similar to finally be trimmed, or discarded. So get summary results without changing the essence of the document.

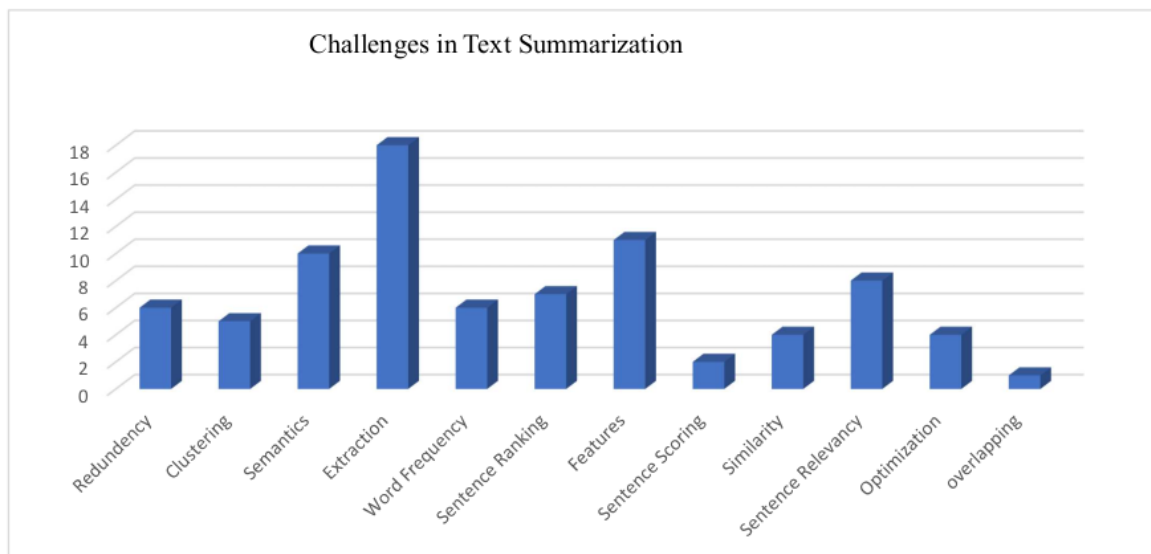


Fig. 8 Distribution of Challenges in text summarization

7. Evaluation

Evaluation is an identification process to measure/assess the performance of a method or tool. In text summarization research, the evaluation of text quality is often assessed by human annotators [29]. The annotator sets the value of the scale that has been determined for each summary. Based on studies over the past 20 years, there have been various approaches to evaluating the results of machine summaries, namely evaluation in terms of essential, extracting sentences, content-based, and task-based.

Evaluation with the essential approach is often done by comparing the results of the machine summary with an ideal summary from the expert. Essential evaluation is done by measuring terms of grammatical, non-redundancy, and coherence. This measurement is very rarely done because it is very rare research that raises grammatical and coherent problems. Evaluate in terms of extracting sentences that are by finding how many ideal sentences are in an automatic machine summary. The evaluation was carried out using precision, recall, and f-score/f-measure measurement methods. For evaluation in terms of content, compare the actual words in a sentence, not the whole sentence. The evaluation was carried out by the measurement method of Rouge, pyramids, and cosine similarity. The advantage of the evaluation is that it can compare human extracts and automatic summarizing machines with human abstracts that contain new sentences or paraphrases. Another evaluation is a task-based approach, which measures the performance of an automatic summarizing machine by using summaries for specific tasks, for example, question answering and document categories [29].

from the literature studies of the last 20 years, the most evaluation approach taken is in terms of sentence extract and content-based. In terms of sentence extracts, measures that are often performed are precision, recall, and f-measure/f-score. While in terms of content-based, the measure that is often done is N-Gram matching (Rouge), which is used in approximately 21 studies of literature, then pyramid and cosine similarity. For details of the distribution can be seen in Fig. 9.

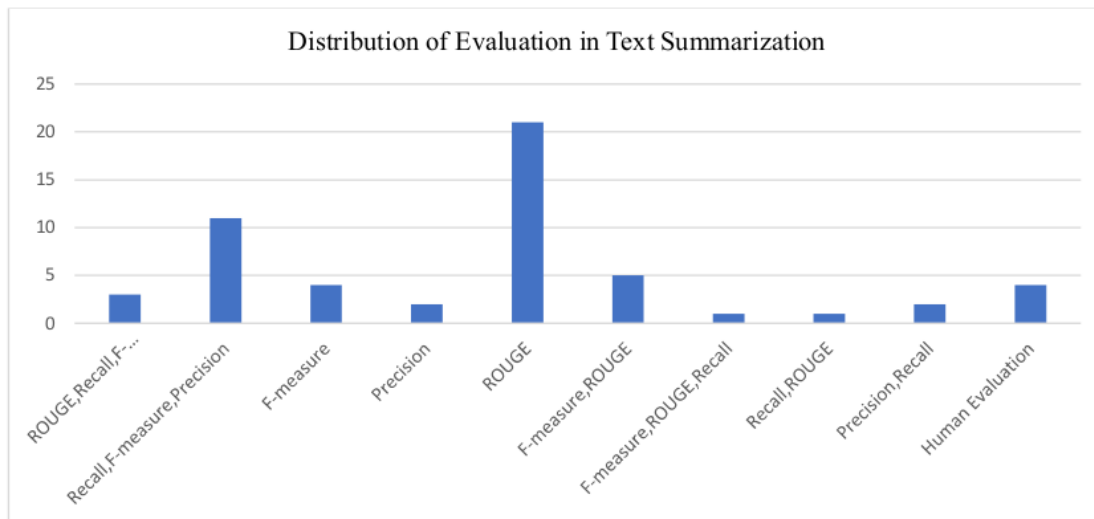


Fig. 9 Distribution of Evaluation in Text Summarization

Conclusion: The above review leads to the conclusion that a lot of research is being done in the area of Natural Language Processing and every view of the research in Indian and Foreign Languages. Automatic Text Summarization is an emerging topic that will give rise to the possibility of further research like Corpora based approach, and cross-lingual summarization, and also decide the future course of action. In the present paper, a brief description of the research work done in the past related to the creation of the Text Summarization System and the various methods, techniques, and approaches used in them has been presented. This will help in the formulation of appropriate and effective methods for Hindi Text Summarizer for various domains and languages.

CONFLICT OF INTEREST

We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere. We have no conflicts of interest to disclose.

AUTHOR CONTRIBUTIONS

Anjana Kishanpuri and Dr. Harshalata petkar were the main manuscript text. Anjana Kishanpuri analysed the data. Anjana Kishanpuri prepared all figures 1-9 and tables 1-3. All authors reviewed and approved the final version of the manuscript.

ACKNOWLEDGMENT

The author thanks her research guide, Dr. Harshalata Petkar, for her encouragement and many fruitful discussions on this research. She suggested the problems and helped in clarifying the solutions. Without her, this work would not have started, progressed, or ended.

References

- [1] Sanjan S Malagi and Dr. D V Ashoka, "An Overview of Automatic Text summarization Techniques", *International Journal of Engineering Research & Technology*, (ISSN: 2278-0181) Volume 8, Issue5, NCAIT-2020 conference proceedings, Special issue – 2020.
- [2] Wang Dingding and Li Tao et al, "Multi-Document Summarization using Sentence-based Topic Models", *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 297–300, 4 August 2009, Suntec, Singapore.
- [3] Jhatta Nikita, Bhatla Ashok Kumar, "A REVIEW PAPER ON TEXT SUMMARIZATION OF HINDI DOCUMENTS", *INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS*, ISSN 2320-7345, Vol.3 Issue.5, Pg.: 89-92, May 2015.
- [4] Shimpikar Sheetal and Govilkar Sharvari, "A Survey of text Summarization Techniques for Indian Regional Languages", *International Journal of Computer Applications*, (ISSN: 0975 – 8887) Volume 165 – No.11, May 2017.
- [5] Saranyamol C S, Sindhu L, "A Survey on Automatic Text Summarization", *International Journal of Computer Science and Information Technologies*, Vol. 5 Issue 6, 2014.
- [6] Ozsoy Makbule Gulcin, Cicekli Ilyas and Alpaslan Ferda Nur, "Text Summarization of Turkish Texts using Latent Semantic Analysis", *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 869–876, August 2010, Beijing.
- [7] Megala S. Santhana et al, "Enriching Text Summarization using Fuzzy Logic", *International Journal of Computer Science and Information Technologies*, Vol. 5 (1), 863-867 (IJCISIT) 2014.
- [8] Sarkar Kamal, et al "Automatic Single Document Text Summarization Using Key Concepts in Documents", *J Inf Process Syst.*, Vol.9, No.4, pp.602-620, December 2013.
- [9] Singh Kumar Abhishek, "Neural Approaches towards text summarization", *International Institute of Information Technology, Hyderabad*, 2018.
- [10] Gupta, S., Gupta, S.K., "Abstractive summarization: An overview of the state of the Art", *Expert Syst. Appl.* 121, 49–65. <https://doi.org/10.1016/j.eswa.2018.12.011>, 2019.
- [11] Elrefa'iy, A., Abas, A.R., Elhenawy, I., "Review of recent techniques for extractive text summarization", *J. Theor. Appl. Inf. Technol.* 96, 7739–7759, 2018.
- [12] Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J.B., Kochut, K., "Text summarization techniques: A brief survey", *Int. J. Adv. Comput. Sci. Appl.* 8. 2017.
- [13] Mishra, R., Gayen, T., "Automatic lossless-summarization of news articles with abstract meaning representation", *Procedia Comput. Sci.* 135, 178–185. <https://doi.org/10.1016/j.procs.2018.08.164>. 2018.
- [14] A. P. Widyassari, S. Rustad, G. F. Shidik, et al., "Review of automatic text summarization techniques & methods", *Journal of King Saud University–Computer and Information Sciences*, <https://doi.org/10.1016/j.jksuci.2020.05.006>, 14 May 2020.
- [15] Gupta Vishal, "Hybrid Algorithm for Multilingual Summarization of Hindi and Punjabi Documents", *R. Prasath and T. Kathirvalavakumar (Eds.): MIKE 2013, LNAI 8284, pp.717–727, 2013. © Springer International Publishing Switzerland 2013*.
- [16] Kadam Deepali P, Patil Mrs. Nita, and Gulathi Mrs. Archana, "A Comparative Study of Hindi Text Summarization Techniques: Genetic Algorithm and Neural Network", *International Journal of Innovations & Advancement in Computer Science (IJACS)*, ISSN 2347 – 8616 Volume 4, Special Issue, March 2015.

- [17] Naik, S.S., Gaonkar, M.N., "Extractive Text Summarization by Feature-based Sentence Extraction using Rule-based", *proceeding in 2017 2nd IEEE International Conference on Recent Trends in Electronics Information & Communication Technology (RTEICT)*. IEEE, India, pp. 2–6. 2017.
- [18] Babar, S.A., Patil, P.D., "Improving performance of text summarization", *Procedia Comput. Sci.* 46, 354–363. <https://doi.org/10.1016/j.procs.2015.02.031>, 2015.
- [19] Thakkar Chetana and Malik Latesh, "Test Model for Summarizing Hindi Text using Extraction Method", *IEEE Conference on Information and Communication Technologies*, 2013.
- [20] Gulati, Sawarkar, "Comparative Analysis of Hindi Text Summarization for Multiple Documents by Padding of Ancillary Features", *Performance Management of Integrated Systems and its Applications in Software Engineering* © Springer 11 September 2019.
- [21] Patel, D., Shah, S., Chhinkaniwala, H., "Fuzzy logic based multi-document summarization with improved sentence scoring and redundancy removal technique", *Expert Syst. Appl.* 134, 167–177. 2019.
- [22] Gupta Manisha, Garg Naresh Kumar, "Text Summarization of Hindi Documents using Rule-Based Approach", *International Conference on Micro-Electronics and Telecommunication Engineering* © IEEE 2016.
- [23] Abbasi-ghalehtaki, R., Khotanlou, H., Esmailpour, M., "Fuzzy evolutionary cellular learning automata model for text summarization", *Swarm Evol. Comput.* 1–16. <https://doi.org/10.1016/j.swevo.2016.03.004>, 2016.
- [24] Maaviya Shrikant and Tiwary Uma Shanker, "Knowledge-Based Summarization and Document Generation using Bayesian Network". Published by Elsevier B.V. This is an open-access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). © 2016 The Authors. Peer review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016.
- [25] Amini Massih-Reza and Gallinari Patrick, "The Use of Unlabeled Data to Improve Supervised Learning for Text Summarization". Copyright 2002 ACM 1-58113-561-0/02/0008, SIGIR'02, Tampere, Finland, August 11-15, 2002.
- [26] Kumar K. Vimal and Yadav Divakar, "An Improved Extractive Approach to Hindi Text Summarization", *Information Systems Design and Intelligent Applications, Advances in Intelligent Systems and Computing* 339, © Springer India 2015 J.K. Mandal et al. (eds.), DOI 10.1007/978-81-322-2250-7_29, September 2015.
- [27] Dalal Vipul and Malik Dr. Latesh, "Automatic Summarization for Hindi Text Documents using Bio-inspired Computing", *International Journal of Advanced Research in Computer and Communication Engineering*, ISO 3297:2007 Certified Vol. 6, Issue 4, April 2017.
- [28] Robertson, S., 2004. "Understanding inverse document frequency: On theoretical arguments for IDF", *J. Doc.* 60, 503–520. <https://doi.org/10.1108/00220410410560582>.
- [29] Steinberger, J., Jezek, K., "Evaluation measures for text summarization", *Comput. Inform.* 28, 1001–1026, 2009.
- [30] Kumar K. Vimal, Yadav Divakar and Sharma Arun "Graph-Based Technique for Hindi Text Summarization", *Information Systems Design and Intelligent Applications, Advances in Intelligent Systems and Computing* 339, DOI 10.1007/978-81-322-2250-7_29, © Springer India 2015 J.K. Mandal et al. (eds.), January 2015.
- [31] Sargule Swati, Kagalkar Ramesh M, "Strategy for Hindi Text Summarization using Content-Based Indexing Approach", *International Journal of Computer Sciences and Engineering*, vol 4, issue 6, 30/Sep/2016.
- [32] Subramaniam Manjula, Dalal Prof. Vipul, "Test Model for Rich Semantic Graph Representation for Hindi Text using Abstractive Method", *International Research Journal of Engineering and Technology (IRJET)*, Volume: 02 Issue: 02, May 2015.
- [33] Dalal Vipul and Malik Latesh "Data Clustering Approach for Automatic Text Summarization of Hindi Documents using Particle Swarm Optimization and Semantic Graph", *International Journal of Soft Computing and Engineering (IJSCE)*, ISSN: 2231-2307, Volume-7 Issue-3, July 2017.
- [34] Gulati, Sawarkar S.D., "A novel Technique for multi-document Hindi Text Summarization", *International Conference on Nascent Technologies in Engineering (ICNTE)* © IEEE 2017.

- [35] Patel Alkesh, Siddiqui Tanveer and Tiwary U. S. "A language-independent approach to multilingual text summarization", *Conference RIAO2007*, Pittsburgh PA, U.S.A. May 30-June 1, 2007.
- [36] J. Anitha, Reddy Prof. P.V.G.D. Prasad and Prasad Bab of. M.S., "A NOVEL FRAMEWORK FOR HINDI TEXT SUMMARIZATION IN ANDROID MOBILE DEVICES", *International Journal of Science, Technology & Management*, www.ijstm.com, Volume No.04, Issue No. 04, April 2015.
- [37] Gupta Vishal and Lehal, Gurpreet Singh "Automatic Punjabi Text Extractive Summarization System", *COLING 2012*, December 2012, Mumbai.
- [38] M. Bhanu, C. Kartika, and P. Sudarmani, "Tamil Document Summarization using Semantic Graph Method", *Conference on Computational Intelligence and Multimedia Applications*, Volume: 2, 2007.
- [39] Mohamed Syed Sabir and Hariharan Shanmugasundaram, "A Summarizer for Tamil Language Using Centroid Approach", *International Journal of Information Retrieval Research*, January 2016.
- [40] Ajmal E.B. and Haroon Rosna P., "An Extractive Malayalam Document Summarization Based on Graph-Theoretic Approach". *Fifth International Conference on e-Learning (econf)*, IEEE, 2015.
- [41] Sarkar Kamal, "An Approach to Summarizing Bengali News Documents". DOI: 10.1145/2345396.2345535, August 2012.
- [42] Deshpande Anjali R. and Lobo L. M. R. J. "Text Summarization using Clustering Technique", *International Journal of Engineering Trends and Technology (IJETT)*. V4(8):3348-3351, ISSN:2231-5381, Volume-4 Issue-8, Jul 2013.
- [43] Gaikwad Deepali K., Sawane Deepali, and Mahender C. Namrata, "Rule-Based Question Generation for Marathi Text Summarization using Rule-Based Stemmer", *IOSR Journal of Computer Engineering (IOSR-JCE) 2016*, e-ISSN: 2278-066, p-ISSN: 2278-8727.
- [44] Sarwadnya Vaishali V. and Sonawane, "Marathi Extractive Text Summarizer Using Graph-Based Model". *Fourth International Conference on Computing Communication Control and Automation (ICCUBE)*, 2018.
- [45] Rathod Yogeshwari V., "Extractive Text Summarization of Marathi News Articles", *International Research Journal of Engineering and Technology (IRJET)*, e-ISSN: 2395-0056 Volume: 05 Issue: 07, July 2018.
- [46] Khanam Dr.M.Humera, S.Sravani, "Text Summarization for Telugu Document", *OSR Journal of Computer Engineering (IOSR-JCE)*, e- ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 18, Issue 6, Nov.-Dec. 2016.
- [47] Kallimani Jagadish S, Srinivasa K G, and Reddy B Eswara, "A Comprehensive Analysis of Guided Abstractive Text Summarization" *IJCSI International Journal of Computer Science*, e-Issues, Volume 11, Issue 6, No 1, November 2014.
- [48] Sudha D Naga and latha Y Madhavee "Multi-Document Abstractive Text Summarization through Semantic similarity matrix for Telugu Language", *International Journal of advanced science and technology*, Vol. 29, no.1, 2020.
- [49] Kallimani Jagadish S., Srinivasa K. G., and Reddy B Eswara "Summarizing News Paper Articles: Experiments with Ontology-Based, Customized, Extractive Text Summary and Word Scoring", *CYBERNETICS AND INFORMATION TECHNOLOGIES, BULGARIAN ACADEMY OF SCIENCES*, Volume 12, No 2, Sofia • 2012.
- [50] Roy Nirmalya et al, "Extraction and Classification of Linguistic Text for Text Summarization", *International Journal of Electronics, Electrical and Computational System IJECS*, ISSN 2348-117X Volume 6, Issue 10, October 2017.
- [51] Gong Yihong and Liu Xin, "Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis", *SIGIR'01*, Copyright 2001 ACM 1-58113-331-6/01/0009, New Orleans, Louisiana, USA, September 9-12, 2001.
- [52] Erkan Gunes and Erkan Dragomir R, "LexRank: Graph-based Lexical Centrality as Salience in Text Summarization". *Journal of Artificial Intelligence Research* 22 (2004) 457-479, 2004.
- [53] Timofeyev, A. Choi, B. "Knowledge-based Automatic Summarization", *SCITEPRESS – Science and Technology Publications, Lda. In Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2017)*, Volume 1: KDIR, pages 350-356 ISBN: 978-989-758-271-4 Copyright © 2017.

- [54] Barzilay Regina and Elhadad Michael, "Using Lexical Chains for Text Summarization". *Volume: Intelligent Scalable Text Summarization*, <http://www.aclweb.org/anthology/W97-0703>, 1997.
- [55] Litvak Marina, Vanetik Natalia, et al, "MUSEEC: A Multilingual Text Summarization Tool", *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics—System Demonstrations*, pages 73–78, Berlin, Germany, August 7-12, 2016.
- [56] Rossiello Gaetano, "Neural Abstractive Text Summarization", gaetano.rossiello@uniba.it, Nov 2016.
- [57] Kamezani Majid, Mohammad-Reza Feizi-Derani, "Ontology-Based Automatic Text Summarization Using FarsNet", ISSN: 2322-5157, Vol. 4, Issue 2, No.14, *ACSIJ Advances in Computer Science: An International Journal*, March 2015.
- [58] Imam Ibrahim et al, "Query Based Arabic Text Summarization", *International Journal of Computer Science And Technology (IJCSIT)*, Vol. 4, Issue Spl - 2, April - June 2013.
- [59] Ramesh Nallapati, Bing Xiang & Bowen Zhou, "SEQUENCE-TO-SEQUENCE RNNs FOR TEXT SUMMARIZATION". *Workshop track - ICLR 2016*.
- [60] Takamura Hiroya and Okumura Manabu, "Text Summarization Model based on Maximum Coverage Problem and its Variant", *Proceedings of the 12th Conference of the European Chapter of the ACL*, pages 781–789, Athens, Greece, 30 March – 3 April 2009.
- [61] Yeh Jen-Yuan et al, "Text summarization using a trainable summarizer and latent semantic analysis", *Information Processing and Management* 41 (2005) 75–94 May 2004 Elsevier Ltd.
- [62] Conroy John and O'Leary Dianne P., "Text Summarization via Hidden Markov Models and Pivoted QR Matrix Decomposition", *SIGIR'01: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 406-407, September 2001.
- [63] Farzindar Atefeh and Lapalme Guy, 'LetSum, an automatic Legal Text Summarizing system' *proceeding in T. Gordon (ed.), Legal Knowledge and Information Systems, Jurix 2004: The Seventeenth Annual Conference. Amsterdam: IOS Press, pp. 11-18, 2004,*
- [64] M. Saravanan, B. Ravindran and S. Raman, "Improving Legal Document Summarization Using Graphical Models". *Proceedings of the 2006 conference on Legal Knowledge and Information System: JURIX 2006: The Nineteenth Annual Conference*, June 2006.
- [65] Das D., Martins, A.F., "A survey on automatic text summarization", *Tech. Rep. Literature Survey for the Language and Statistics II course at CMU* (2007).
- [66] Gaikwad Deepali K. and Mahender C. Namrata, "A Review Paper on Text Summarization", *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 5, Issue 3, March 2016.
- [67] Dhawale Apurva D., Kulkarni Sonali B., and Kumbhakarna Vaishali, "Survey of Progressive Era of Text Summarization for Indian and Foreign Languages Using Natural Language Processing", *Innovative Data Communication Technologies and Application*, pp. 654–662, January 2020. https://doi.org/10.1007/978-3-030-38040-3_74, © Springer Nature Switzerland AG 2020.
- [68] Torres-Moreno J.M. and EBSCOhost. "Automatic text summarization", *ISTE Ltd; John Wiley Sons, Inc London: Hoboken, NJ.*, 2014.
- [69] Chettri Roshna and Chakraborty Udit Kr., "Automatic Text Summarization", *International Journal of Computer Applications*, (ISSN: 0975 – 8887) Volume 161 – No 1, March 2017.
- [70] Sarkar Kamal, et al "Automatic Single Document Text Summarization Using Key Concepts in Documents", Vol.9, No.4, pp.602-620, *J Inf Process Syst.*, December 2013.
- [71] Verma Pradeepika and Verma Anshul, "Review on Text Summarization Techniques", *Journal of scientific research*, volume 64, issue1, 2020.
- [72] Sahoo Deepak et al, "Hybrid approach to Abstractive Summarization", *International Conference on Computational Intelligence and data science (ICCIDIS)*, Elsevier, 2018.
- [73] K. Nandhini and S.R. Balasundram, "Improving readability through extractive summarization for the learner with reading difficulties", *Egyptian Information Journal*, Elsevier, Oct Oct 2013.

- [74] Shai Erera et al, "A summarization system for scientific documents", *EMNLP 2019 and IJCNLP (System demonstration)* pages 211-216, Hong Kong, China, ACL. Nov 2019.
- [75] Wonjin Yoon et al, "Learning by semantic similarity makes abstractive summarization better", *arXiv:2002.07767v1[Cs.CL]*. 18 Feb 2020.
- [76] Bhatt Ambati Rajeev et al, "Read, highlight and summarize: A hierarchical neural semantic Encoder based approach", *ICLR 2020*, *arXiv:1910.03177v2*. 2020.
- [77] Yen Chun Chen and Bansal Mohit, "Fast abstractive summarization with reinforce selected sentence rewriting", *arXiv:1805.11080v1[Cs.CL]*. 28 May 2018.
- [78] Bhargava Rupal, "AISSI: Abstractive text summarization using sentiment infusion", *12th International multi-conference on information processing (IMCIP-2016)*, 2016.
- [79] Elena Iolvet and Manuel Palomar, "Challenging issues of automatic summarization: Relevance detection and Quality based evaluation", *Informatica*, 2010.

66%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.researchgate.net Internet	1287 words — 14%
2	www.ijcaonline.org Internet	700 words — 8%
3	Adhika Pramita Widyassari, Supriadi Rustad, Guruh Fajar Shidik, Edi Noersasongko et al. "Review of automatic text summarization techniques & methods", Journal of King Saud University - Computer and Information Sciences, 2020 Crossref	626 words — 7%
4	ijcsit.com Internet	521 words — 6%
5	www.ijrcar.com Internet	196 words — 2%
6	www.ijarcce.com Internet	188 words — 2%
7	www.arxiv-vanity.com Internet	129 words — 1%
8	www.ijert.org Internet	118 words — 1%

9	etd.aau.edu.et Internet	116 words — 1%
10	hari-sriblog.blogspot.com Internet	101 words — 1%
11	cse.anits.edu.in Internet	86 words — 1%
12	link.springer.com Internet	86 words — 1%
13	Harsh Mehta, Santosh Kumar Bharti, Nishant Doshi. "Automatic Text summarization in Gujarati language", 2022 IEEE 2nd International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC), 2022 Crossref	75 words — 1%
14	www.ijeat.org Internet	59 words — 1%
15	1library.net Internet	50 words — 1%
16	www.ijarcsse.com Internet	49 words — 1%
17	www.ijssrd.com Internet	46 words — 1%
18	ijres.org Internet	43 words — < 1%
19	core.ac.uk Internet	41 words — < 1%

20	www.irejournals.com Internet	41 words — < 1%
21	www2.latech.edu Internet	41 words — < 1%
22	mafiadoc.com Internet	40 words — < 1%
23	www.igi-global.com Internet	37 words — < 1%
24	"Intelligent Data Communication Technologies and Internet of Things", Springer Science and Business Media LLC, 2022 Crossref	33 words — < 1%
25	www.ijrte.org Internet	33 words — < 1%
26	dergipark.org.tr Internet	32 words — < 1%
27	summarization.com Internet	31 words — < 1%
28	www.ijsce.org Internet	31 words — < 1%
29	repository.sustech.edu Internet	30 words — < 1%
30	svuce.edu.in Internet	30 words — < 1%
31	C. Sunitha, A. Jaya, Amal Ganesh. "A Study on Abstractive Summarization Techniques in Indian	29 words — < 1%

32 M Rahul Raj, Rosna P Haroon. "Malayalam text summarization: Minimum spanning tree based graph reduction approach", 2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Fall), 2016

Crossref

29 words — < 1%

33 in.bgu.ac.il

Internet

28 words — < 1%

34 Ruby Rani, D.K. Lobiyal. "Performance evaluation of text-mining models with Hindi stopwords lists", Journal of King Saud University - Computer and Information Sciences, 2020

Crossref

27 words — < 1%

35 ijettjournal.org

Internet

27 words — < 1%

36 ijream.org

Internet

27 words — < 1%

37 dspace.lib.uom.gr

Internet

26 words — < 1%

38 journals.scholarpublishing.org

Internet

26 words — < 1%

39 mathoverflow.net

Internet

26 words — < 1%

40 www.eurecom.fr

Internet

25 words — < 1%

-
- 41 etd.lib.metu.edu.tr 24 words — < 1%
Internet
-
- 42 "Mining Intelligence and Knowledge Exploration", Springer Nature, 2013 23 words — < 1%
Crossref
-
- 43 Aashil Shah, Devam Zanzmera, Kevan Mehta. "Deep Learning based Automatic Hindi Text Summarization", 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), 2022 23 words — < 1%
Crossref
-
- 44 sersc.org 23 words — < 1%
Internet
-
- 45 Rishabh Karmakar, Ketki Nirantar, Prathamesh Kurunkar, Pooja Hiremath, Deptii Chaudhari. "Indian Regional Language Abstractive Text Summarization using Attention-based LSTM Neural Network", 2021 International Conference on Intelligent Technologies (CONIT), 2021 22 words — < 1%
Crossref
-
- 46 preview.aclanthology.org 22 words — < 1%
Internet
-
- 47 Yazan Alaya AL-Khassawneh, Essam Said Hanandeh, Sattam Almatarneh. "Chapter 1 Extractive Text Summarization Using Syntactic Sub-graph Models", Springer Science and Business Media LLC, 2023 21 words — < 1%
Crossref
-
- 48 www.ripublication.com 21 words — < 1%
Internet
-
- 49 ijcseonline.org

20 words — < 1%

50 old.rjournals.com

Internet

20 words — < 1%

51 Chetana Badgujar, Vimla Jethani, Tushar Ghorpade. "Abstractive Summarization using Graph Based Methods", 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018

Crossref

19 words — < 1%

52 Wuying Liu, Lin Wang. "Efficient Korean text summarization based on key phrase extraction", 2017 International Conference on Machine Learning and Cybernetics (ICMLC), 2017

Crossref

19 words — < 1%

53 ijetsr.com

Internet

19 words — < 1%

54 jcreview.com

Internet

19 words — < 1%

55 vtechworks.lib.vt.edu

Internet

19 words — < 1%

56 zenodo.org

Internet

19 words — < 1%

57 Rosna P Haroon, Abdul Gafur M, Nasreen Ali, Barakkath Nisha U. "An Efficient Text Summarization Using Term and Inverse Frequency With Key Phrase Identification in Malayalam Language", 2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), 2021

18 words — < 1%

-
- 58 eprints.gla.ac.uk 17 words — < 1%
Internet
-
- 59 www.ijstm.com 17 words — < 1%
Internet
-
- 60 www.scribd.com 17 words — < 1%
Internet
-
- 61 "Innovative Data Communication Technologies and Application", Springer Science and Business Media LLC, 2020 15 words — < 1%
Crossref
-
- 62 Abdullah Alshanqiti, Abdallah Namoun, Aeshah Alsughayyir, Aisha Mousa Mashraqi, Abdul Rehman Gilal, Sami Saad Albouq. "Leveraging DistilBERT for Summarizing Arabic Text: An Extractive Dual-Stage Approach", IEEE Access, 2021 15 words — < 1%
Crossref
-
- 63 Nilantha Premakumara, Chamani Shiranthika, Chathurangi Shyalika, Surangani Bandara. "Optimized Text Summarization method based on Fuzzy Logic", 2022 2nd International Conference on Image Processing and Robotics (ICIPRob), 2022 15 words — < 1%
Crossref
-
- 64 timm.ujaen.es 15 words — < 1%
Internet
-
- 65 www.mdpi.com 15 words — < 1%
Internet

66 "Information Systems for Intelligent Systems", Springer Science and Business Media LLC, 2023
Crossref 14 words — < 1%

67 arxiv.org
Internet 14 words — < 1%

68 silp.iita.ac.in
Internet 14 words — < 1%

69 www.freepatentsonline.com
Internet 14 words — < 1%

70 www.ijirset.com
Internet 14 words — < 1%

71 www.frontiersin.org
Internet 13 words — < 1%

72 www.koreascience.or.kr
Internet 13 words — < 1%

73 www.politesi.polimi.it
Internet 13 words — < 1%

74 Jinal Tandel, Kinjal Mistree, Parth Shah. "A Review on Neural network based Abstractive Text Summarization models", 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019
Crossref 12 words — < 1%

75 Mykola Trokhymovych, Diego Saez-Trumper. "WikiCheck", Proceedings of the 30th ACM International Conference on Information & Knowledge Management, 2021
Crossref 12 words — < 1%

76	Torres-Moreno, Juan-Manuel. "Single-Document Summarization", Automatic Text Summarization, 2014. Crossref	12 words — < 1%
77	cris.bgu.ac.il Internet	12 words — < 1%
78	mmmut.ac.in Internet	12 words — < 1%
79	www.iajit.org Internet	12 words — < 1%
80	www.thinkmind.org Internet	12 words — < 1%
81	vdocuments.mx Internet	11 words — < 1%
82	Regina Merine, Saptarshi Purkayastha. "Risks and Benefits of AI-generated Text Summarization for Expert Level Content in Graduate Health Informatics", 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI), 2022 Crossref	10 words — < 1%
83	ijarcce.com Internet	10 words — < 1%
84	aclanthology.org Internet	9 words — < 1%
85	coek.info Internet	9 words — < 1%
86	ijsret.org	

Internet

9 words — < 1%

87 mitwpu.edu.in

Internet

9 words — < 1%

88 pdfslide.net

Internet

9 words — < 1%

89 turcomat.org

Internet

9 words — < 1%

90 www.jatit.org

Internet

9 words — < 1%

91 Pradeepika Verma, Sukomal Pal, Hari Om. "A Comparative Analysis on Hindi and English Extractive Text Summarization", ACM Transactions on Asian and Low-Resource Language Information Processing, 2019

Crossref

8 words — < 1%

92 aclweb.org

Internet

8 words — < 1%

93 deepai.org

Internet

8 words — < 1%

94 export.arxiv.org

Internet

8 words — < 1%

95 nozdr.ru

Internet

8 words — < 1%

96 repository.usthb.dz

Internet

8 words — < 1%

97 www.phi-base.org

Internet

8 words — < 1%

98 www.yorku.ca
Internet

8 words — < 1%

99 Md. Majharul Haque, Suraiya Pervin, Anowar Hossain, Zerina Begum. "Approaches and Trends of Automatic Bangla Text Summarization", International Journal of Technology Diffusion, 2020
Crossref

7 words — < 1%

100 Namrata Kumari, Pardeep Singh. "chapter 17 Text Summarization and Its Types", IGI Global, 2021
Crossref

7 words — < 1%

101 Gupta, Vishal, and Gurpreet Singh Lehal. "A Survey of Text Summarization Extractive Techniques", Journal of Emerging Technologies in Web Intelligence, 2010.
Crossref

6 words — < 1%

102 Jingwen Wang, Hao Zhang, Cheng Zhang, Wenjing Yang, Liqun Shao, Jie Wang. "An Effective Scheme for Generating An Overview Report over A Very Large Corpus of Documents", Proceedings of the ACM Symposium on Document Engineering 2019 - DocEng '19, 2019
Crossref

6 words — < 1%

EXCLUDE QUOTES OFF
EXCLUDE BIBLIOGRAPHY OFF

EXCLUDE SOURCES OFF
EXCLUDE MATCHES OFF