



ISSN: 2350-0328

**International Journal of Advanced Research in Science,  
Engineering and Technology**

**Vol. 10, Issue 4, April 2023**

# **An Improved Automatic Article Summarization System Using TextRank Algorithm**

**Akinade Abigail. O, Akinsola Adeniyi. F, Oladiboye Olasunkanmi. E, Ogundele Israel. O**

Department of Computer Technology, Yaba College of Technology, Yaba, Lagos  
Department of Computer Technology, Yaba College of Technology, Yaba, Lagos  
Department of Computer Technology, Yaba College of Technology, Yaba, Lagos  
Department of Computer Technology, Yaba College of Technology, Yaba, Lagos

**ABSTRACT:** The internet is loaded with trillions of information that are scattered around the globe. Researchers find it difficult to gather relevant information from the numerous articles available on the web. It can become more tasking and tedious when analyzing to get important data manually by summarize the text on their own from the large database which will in turn reduce efficiency, waste resources and time. Automatic article summarization is needed to overcome the challenges being faced with the manual process of summarizing article papers when reading through existing literature which can be very enormous and cumbersome to getting the right understanding from the work. Previous literature was reviewed based on earlier techniques used for summarization, challenges and limitation of the system developed and how to make better the system we want to develop. The paper focused on the development of an improved automatic article summarization system using TextRank algorithm that will provide researcher with a quick and efficient way to understand the main ideas and arguments presented in a long piece of article papers. The system is developed using python programming language, NodeJs for the implementation of the backend for routing, mongoose framework was used to implement the database of the system with other software like HTML, CSS and Herokuapp. The system will therefore help to achieve the set-back of the previous system, add to researchers' knowledge and will make summarization of article papers (Journals, thesis, conferences, documents etc) easier and faster thereby decreasing understanding time and improving efficiency for production of academic papers.

**KEY WORDS:** TextRank Algorithm, Automatic Summarization, Article Paper, Researcher, Machine Learning

## **I. INTRODUCTION**

The Web is a wealth of knowledge on a variety of subjects. The Web lacks a centrally controlled content structure, in contrast to traditional collections like libraries. This has opened the opportunity for users to benefit from the available data in many interesting ways [1]. Users typically access data on the Web by browsing and keyword searching, which are simple methods of doing so. It is impossible to analyze and extract relevant data from big data bases manually; therefore, automated extraction techniques are needed to get user-requested data from the internet's billions of pages and find pertinent information. In the past, people would independently summarize texts, but as data volumes have increased, it has been increasingly challenging for people to keep up with the massive amounts of information. Researchers all over world needs tools that can help them enhance their research process, most especially when reading through article of existing literature which can be very enormous and cumbersome to getting the right understanding from the work. To overcome this challenges article summarization is needed to extract meaningful content from the large text. Article summarization is the process of condensing an article or document into a shorter version that captures the main points and ideas [2,3]. The goal of article summarization is to provide readers with a quick and efficient way to understand the main ideas and arguments presented in a longer piece of writing. This can be useful for researcher who want to quickly scan through multiple articles or documents to gather information, or for those who have limited time to read lengthy articles. Extractive and abstractive summarizing are the two basic methods of article summarization. Extractive summarization involves selecting important sentences or phrases from the original text and using them to create a summary. Abstractive summarization involves generating a summary in the form of new sentences that capture the essence of the original text [4].



ISSN: 2350-0328

# International Journal of Advanced Research in Science, Engineering and Technology

Vol. 10, Issue 4, April 2023

Automatic summarization is the use of software and algorithms to generate a summary of a longer piece of text, such as an article or document. The system developed for article summarization can be used for variety of techniques to generate summaries including statistical methods, machine learning algorithms and deep learning models [5,6]. One of machine learning algorithm used here is the TextRank algorithm for the article summarization. It is an unsupervised, graph-based algorithm for automatic summarization and keyword extraction of natural language text [7]. The TextRank algorithm is based on PageRank algorithm used by Google to rank web pages in search results for article [8]. The system can be trained on large amounts of text data and can be fine-tuned to improve their accuracy and effectiveness. Automatic summarization can be applied in several practical applications including academic research and business intelligence. It can save time and resources by providing a quick and efficient way to understand the main ideas and arguments presented in a longer research paper [9].

The significance of this project is to help users extract summarized data from a large pool of data without having to read through very large amounts of articles [4]. Users can read less while still getting the most important information needed to draw reliable conclusions. In order to decrease the amount of time it takes researchers to read through online journals, articles, theses, and other documents while still producing academic papers quickly, the research paper will focus on the development of an improved automatic article summarization system using the TextRank algorithm.

## II. LITERATURE STUDY

The act of presenting the most important fact or idea about a person or object in a succinct and recognizable structure is known as summarization. In plain English, an outline identifies the most important idea in a work of literature, shows how to omit the redundant details, and demonstrates how to combine the central idea in a relevant way. Rundown is a crucial component of article evaluation, and it's crucial for students to learn how to perform it because there will come a time when they'll need to condense an essay, an article, a research paper, or another piece of original work. Understudy is influenced by a synopsis in the following ways:

- It aids in the selection of key concepts and the fusion of those vital ideas that support them.
- It enables students to concentrate on memorable and significant phrases and expressions from the topic being taught.
- It demonstrates to learners how to reduce complex information to the most important points for comprehension that is more concise.

### A. The Need for Article Summarization

Article summarization arises from the fact that there is an overwhelming amount of information available on the internet and other sources, and it can be difficult and time-consuming to find the most relevant and important information within a large amount of text. It also provides a way to quickly and efficiently understand the main ideas and arguments presented in a longer piece of writing [4,10]. The vast majority of the data currently flooding the digital world is unstructured text data. Therefore, it is necessary to provide automatic article summary technologies that enable users to readily gain insights from them. Currently, having access to a wealth of information can be enjoyable. The implementation of summarization can enhance the readability of documents, reduce the time spent in researching for information that allow for more information to be fitted in a particular area [2]. Article summarization can help to improve comprehension by highlighting the most important information and ideas in a text [11]. It can help readers to focus on the key points and avoid getting bogged down in irrelevant details.

### B. Automatic Article Summarization

Automated article summarizing is a problem in information science and natural language processing that aims to create a brief, understandable report from a lengthy one. The amount of content that is available on the internet is always growing, thus rundown tactics are highly recommended because they allow us to consume information more quickly. Filtering through mountains of documentation may be time-consuming and difficult. Without a dynamic or outline, it might take minutes to understand what a certain paper or report is about. To put it another way, just picture going through a number of archives. Representatives, examiners, legal assistants, students, and specialists must sort through vast amounts of documents every day in order to advance, and a significant portion of their time is consumed by determining which documents are pertinent and which ones are not. It is possible to quickly determine whether a document is worth reading by removing important sentences and creating thorough outlines.

Understudies and content developers might also benefit from program content planning. Imagine having the ability to build a dynamic base for your exam paper or a section of a book in a sensible and condensed manner that is dedicated to



ISSN: 2350-0328

# International Journal of Advanced Research in Science, Engineering and Technology

Vol. 10, Issue 4, April 2023

the primary source material. The way a computer-generated rundown of material operates is by first analyzing the word density for the entire content record. The most well-known words are then tucked away and sorted. The number of thick words in each sentence is then scored; the thicker the word, the higher the value of the sentence. Finally, based on their placement in the first content, the top N sentences are selected and arranged. Programmed content outline makes things straightforward and broadly useful, the programmed content synopsis calculation can work in various circumstances that different usage may battle with, for example, reports containing unknown dialects or one-of-a-kind word affiliations that aren't found in Standard English language corpuses [12].

The best challenge in content summary may be to extract important information from provided basic sources, such as internet pages, any archive, and databases. Content summary systems must offer a good overview in a shorter amount of time with fewer repetition. Because all of the rules and examples are physically written, the rule-based strategy's output has a high data density but is exceedingly tedious to complete. Ontology's technique allows for the consideration of dubious data, which is absurd in conventional space metaphysics. The problem with this approach is that it can be laborious for lone field experts to characterize the cosmology of the universe. Due to the use of a language generator, the nature of the outline is improved in the Tree-based technique. The one drawback to this approach is that, in focusing on the intersection of expressions, the essential structure of the sentences is overlooked. The Multimodal Semantic Model approach generates a theoretical breakdown that includes both printed and graphical information, and as a result, offers excellent results. Issue with this strategy is that assessment is to be done physically [13]. The selection of useful data is completed in the information thing-based approach. The sentences and outlines are developed using the selected data item. This process provides a concise, thoughtful, and information-rich outline. The problem with this method is that occasionally useful information items are disregarded while significant and syntactically sound sentences are being developed, which lessens the phonetic quality of rundown. The Semantic diagram strategy, Sentences framed are less repetitive just as linguistically right [14]. In any event, this approach is only applicable to one record. Despite the fact that the system of programmed summary is an established test, experts are now leaning more towards abstractive rundown techniques than extractive outline techniques. This is because abstractive rundown techniques result in increasingly comprehensible, less repetitious, and data-rich summaries. It can be difficult to create dynamic using abstractive synopsis techniques because it necessitates extensive semantic and etymological research. The examination of abstractive outline techniques appears to be getting more and more useful for the aforementioned reasons.

Tools for automatic article summaries are required because:

- Summaries shorten reading time.
- Outlines facilitate decision-making when looking for records.
- Compared to human summarizers, automatic outline computations are less slanted.
- Because they provide tailored data, personalized outlines, also known as note frameworks, are valuable.
- By utilizing self-loading or programmed outline frameworks, business theoretical administrations can expand the number of writings they can process.

## C. Single Vs. Multiple Document Summarization

Rather than using numerous report synopses that aim to remove rundown from distinct material or articles that elaborate on the same subject, a single archive outline manages abridging a single record. Individual clients can quickly familiarize themselves with the information in a large collection of records thanks to the next summary report. In a manner various report synopsis framework assists client with adapting to data over-burden [15].

The main benefit of using various archive rundowns is that they provide data reports that are both concise and thorough, with clear hypotheses being formed and described, and each theme is depicted from a variety of points of view inside a single record. The goal of a rundown is to simplify data search and reduce time by selecting the most important source report, but an exhaustive list of all available archives should already contain the necessary information, reducing the need to approach the first records in circumstances where clarification is needed. Programmed summaries contain information that has been algorithmically deleted without any article contact, making it objective.

The numerous record outline is more mind boggling than abridging single archive [16]. An excellent outline strategy aims to select and combine the main topic with coherence, understandability, and speed. The issue is caused by connected randomness among a huge collection of records. The optimum multiple record synopsis structure keeps the introductory content brief and delivers information organized around the main angle to speak to various viewpoints.

A successful multiple document summary contains the following characteristics:

- Easily readable



ISSN: 2350-0328

# International Journal of Advanced Research in Science, Engineering and Technology

Vol. 10, Issue 4, April 2023

- A clear structure with a primary substance framework that makes it simple to explore the remaining content divisions.
- Progressive shifts from broadly connected views to increasingly specific ones.
- Text within segments is divided into key components.

## D. Techniques to Achieving Automatic Summarization

The Abstractive and Extractive approaches are the two main methods for developing an autonomous summarization system. The two methods are briefly discussed below:

- **Abstractive Approach:** The abstractive approach will generate new phrases or use new words that were not in the original text to generate the summary [4]. Naturally abstractive approaches are harder, as the model has to first understand the documents and then try to express that understanding using new words and phrases to create a perfect summary. The abstractive approach is barely used as it deals with semantic problems and produces less effective summary than the extractive approach. This method uses advanced natural language processing and deep learning which is a growing field in itself.
- **Extractive Approach:** In this approach the automatic system takes out sentences from the entire collection, without altering the sentences themselves. This approach will check out the important sentences and use these sentences to form the summary, there are different techniques and algorithm that is used to check out weight of the sentences and then rank those sentences depending on how similar and important they are. Extractive outlines frequently give preferable outcome over abstractive rundowns, this is on the grounds that abstractive synopsis techniques adapt to issues like semantic portrayal, surmising and characteristic language age which is generally harder than information driven methodologies like sentence extraction [17]. There are numerous systems accessible to produce extractive synopsis, with the end goal of this exploration work we will utilize solo learning approach of the extractive methodology.

## E. Machine Learning for Summarization

Machine learning approaches model the summarization as a classification problem. [18] is an early research attempt at applying machine learning techniques for summarization. Kupiec et al. create a classification function, the naive-Bayes classifier, using a training set of documents and their extractive summaries to distinguish between sentences that are summary sentences and non-summary sentences based on the attributes they possess. The Bayes' rule is used to statistically learn the classification probabilities from the training data.

The sentence is rated according to how likely it is that it will fit into the outline. The selected classifier performs the function of a sentence scoring capability. The placement of sentences inside the text, the length of phrases, the proximity of capitalized words, the proximity of the sentence to the title of the archive, and other factors are some of the successive capabilities employed in rundown. Rundown made extensive use of AI systems to provide examples. Machine learning approaches have been widely used in summarization by [19, 20, 21], to name a few. The most well-known framework learning techniques used for synopsis are probably Naive Bayes, decision trees, support vector machines, Hidden Markov models, and Conditional Random Fields. The requirement that phrases included in the summary be freely resolved is one of the fundamental differences between classifiers. It appears that techniques expressly accepting the reliance among sentences which incorporate Hidden Markov form [22] and Conditional Random Fields often outflank different systems [23]. The necessity for a large number of tutoring reports (organized information) to create the classifier, which may not be consistently without difficulty accessible, is one of the key problems with using controlled aging techniques for outlining.

Experts have suggested a few options to address the problem:

- **Creation of annotated corpora:** By producing annotated corpora for summary, researchers can more easily evaluate various summarization techniques because there will be more publicly available benchmarks. Moreover, it lessens the possibility of overfitting with scant data. Ulrich et al. [24] introduce a publicly available annotated email corpus and its creation process. It takes a lot of effort to create an annotated corpus, and more importantly, there is no consensus on the sentences to be used, thus various persons may choose different phrases to build the summary.
- **Semi-supervised methods:** Training a classifier using a semi-supervised method. We use the unlabeled data for training in semi-supervised learning. Typically, there is a huge volume of unlabeled data and a limited amount of labeled data. For complete overview of semi-supervised learning, see [25]. [26] suggested a semi-supervised extraction summarization technique. To take use of unlabeled data, they iteratively co-trained two classifiers. The two classifiers are trained on the



ISSN: 2350-0328

# International Journal of Advanced Research in Science, Engineering and Technology

Vol. 10, Issue 4, April 2023

fresh training data in each iteration, and the unlabeled training examples (sentences) with the highest scores are added to the labeled training set.

Machine learning methods have been shown to be very effective and successful in single and multi-document summarization, specifically in class specific summarization where classifiers are trained to locate particular type of information such as scientific paper summarization [25] and biographical summaries [26, 27].

## F. Techniques of Extractive Approach

There are several methods for the extractive approach, only a few of which will be briefly addressed.

- **TextRank Algorithm:** It is similar to PageRank calculation and is an unaided learning strategy to extractive synopsis. TextRank is a globally beneficial chart-based positioning calculation for NLP. The main idea behind TextRank is to score each sentence in a book, select the top n sentences, and then arrange them according to appearance frequency to create a pre-programmed list. A chart is developed by making a vertex for each sentence in the record, the edges between sentences depend on some type of semantic closeness [28].
- **Cosine Similarity:** Sentences are spoken to as a pack of vectors; hence the cosine similitude identifies comparability between sentences. Cosine likeness is a proportion of proximity between two non-zero vectors of an internal item space that calculates the cosine of the edge between them.
- **Sub-modular set function:** In mathematics, a submodular set has a capacity whose value has the quality that creates a difference in the gradual estimation of the capacity a single component makes when added to a set of information. Rundown submodular works typically serve as models for ideas of diversity, data representation, and inclusion. For instance, in a report summary, the rundown may need to include very important facts from the report; this is an instance of set spread in submodular capacities.
- **LexRank Algorithm:** LexRank is an unsupervised method for text summarization that rates sentences' centrality on graphs. The key concept is that sentences "recommend" other sentences that are similar to them to the reader. As a result, if a sentence sounds a lot like many others, it probably has a lot of significance. lexrank is very similar to textrank. Lexrank and Textrank are extremely similar.

## G. Related works

In the existing literature reviewed, we discussed that various techniques of machine learning and others have been used for text summarization which has proving to be efficient and effective. These techniques have also been applied to different areas of human endeavors such as in education, healthcare, News, languages, embedded wording etc. TextRank algorithm have also been used for text summarization but most papers have not fully considered the area of article summarization of multiple uploads of documents and downloading of multiple summarized documents.

[29] focus on using enhanced Genetic Algorithm for Single Document Extractive summarization to improve the quality of extractive text summary. The research conducted an experiment to using Daily Meal dataset on the proposed model to imitate the process of natural evolution with the process of solution representation, fitness function, population initialization, parent selection, crossover and mutation. The CNN was used for turning the fitness function and also for the evaluation of the effectiveness of the GA. The GA was compared with other techniques of Machine learning, such as TextRank, SummaRunNer algorithm and these shows that the choosing algorithm performs better than others. The limitation of the research was that it can only be applied for a single-document summarization and did not consider the summarization of article papers.

[30] developed a model for summarizing text documents using TextRank algorithm and term frequency. It combines the use of TF-IDF and TextRank algorithm to solve summarization problem. Here, the set of English text were considered for article summarization and the evaluation was done sing the Rogue metric. The performance evaluation shows that the hybrid model produced a better result for summarization which can in turn help researcher to speed up their work. The research did not consider using multiple-document summarization from the large article paper researchers have to summarize daily.

[31] the paper discussed on the structured based and semantic based approaches for generating a huge text-documents summarization. The study makes use of various datasets to tests the summaries of the developed model which gives different text summarization in length and content. The research did not compare with other techniques to evaluate the algorithm that produce better based on NLP.

[32] the paper proposed a method based on natural language processing (NLP) for single Arabic document summarization. The suggested technique is based on extracting the most important data from the document. The study sample was a collection of articles taken from Wikipedia (<https://ar.wikipedia.org/>). where 33 Arabic articles in various subjects, such



ISSN: 2350-0328

# International Journal of Advanced Research in Science, Engineering and Technology

Vol. 10, Issue 4, April 2023

as astronomy, biology, chemistry, etc., were chosen at random from the pool. the method is based on extractive type for a single document summarization. The study did not consider the used of multiple-document and this is only based on Arabic language for article summarization.

[33] the researcher proposed a system for extractive summarization of single-multiple Telugu text documents using TextRank algorithm. For summarizing, a number of Telugu text documents are provided as input. Any news item or other type of text document that has been trained for summary may be used as the source material. The TextRank algorithm is considered as best for summarizing the Telugu document coming from the support of India NLP library. The limitation of the paper is that it cannot handle summarization for English language text-document but for only Telugu text-documents.

[34] the research presented a text-document summarization using word embedding. The goal of the project is to create high-quality summaries by retaining the semantics of text, which is a key component of summarizing a document. The suggested model makes use of additional stylistic and statistical features as well as the semantics of text as a feature. used main task benchmark dataset from Document Understanding Conference<sup>3</sup> (DUC-2007) collected from ACQUAINT corpus [35] The paper presents a summarization technique based on the distributional hypothesis to capture the semantics of the text for producing better summaries using text summarization process. The system was compared to baselines that are considered to be state-of-the-art, and the results reveal that the suggested approach outperforms the baselines. The limitation to the system is that the focus of the summarization is on the text-document but not specify on the literature for summarization.

[36] the paper presented an automatic summary generation using TextRank based extractive text summarization technique. The primary goal of the research is to locate the text's most pertinent sentence fragments and deliver them to the user via extractive text summarization. When the TextRank algorithm used was compared with other summarization techniques, it shows clearly that TextRank does the job of the text summarization better. The work did not consider the article summarization for researcher to use when reviewing existing literature to advance their work and save time.

## IV. METHODOLOGY

At a high level of abstraction, the system essentially enables the user to examine summaries of previously posted documents by registered users on the website. Text summary is used in this to automatically generate an abstract of the uploaded material that website visitors can read. The system forces such visitors to sign up in order to give them the privileges of other signed-up users, such as the ability to download the entire document and the ability to upload your own documents. Interested users may want to get the full document, but this is only possible for registered users. The option to automatically generate your document abstract using the system's integrated summary engine is another useful feature for registered users.

### A. Extractive Summarization Using TextRank

The extractive method of creating an automatic summarization system stacks a number of sentences or phrases to create a summary. The extractive and unsupervised text summarization method known as TextRank. The following steps would be taken using the TextRank algorithm to obtain the necessary summary.

- i. All the texts contained in an article will be concatenated.
- ii. The texts will be spilt into individual sentences.
- iii. Vector representation for each sentence will be found
- iv. Similarities between sentence vectors are then calculated and stored in a matrix
- v. The similarity matrix is then converted into a graph for sentence rank calculation
- vi. Finally, a certain number of top ranked sentences form the final summary

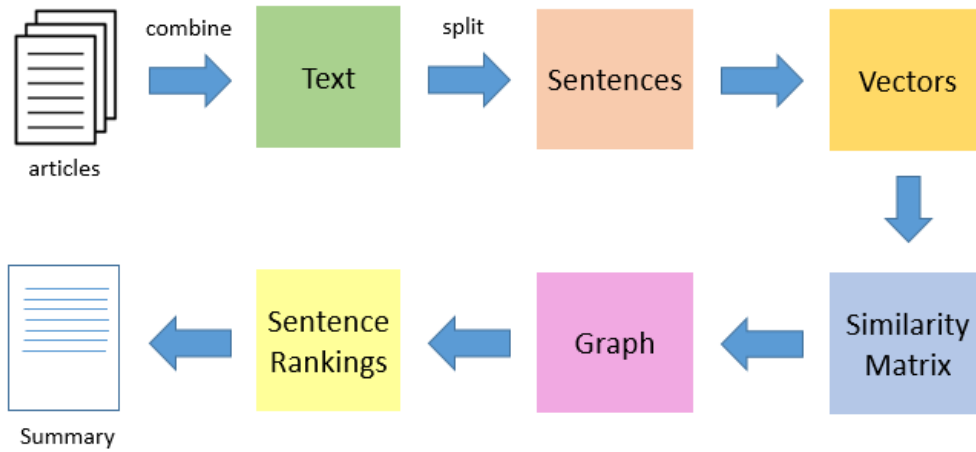


Figure 1: Source: TextRank algorithm process [36]

### B. Proposed FrameWork for Article Summarization System

Figure 2 is the proposed model for the article summarization system. This is based on extractive summarization using Textrank algorithm to summarize meaning and relevant content from different article of researchers.

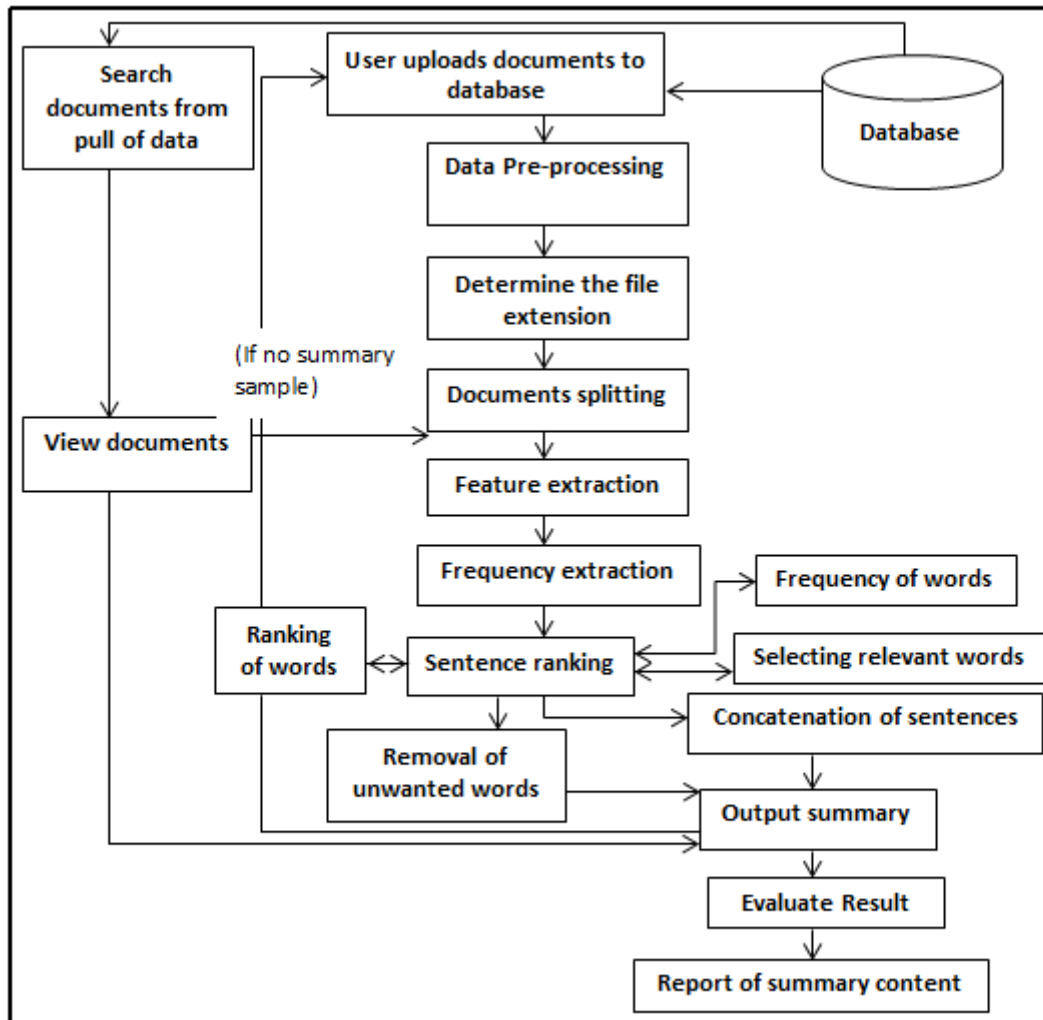


Figure 2: Adapted from Source [36]

C. Use Case Diagram

In figure 3 shown below is a use case diagram that allows users to perform some functions, such as to visit the platform, upload document to be summarize, download summary and to view the document.



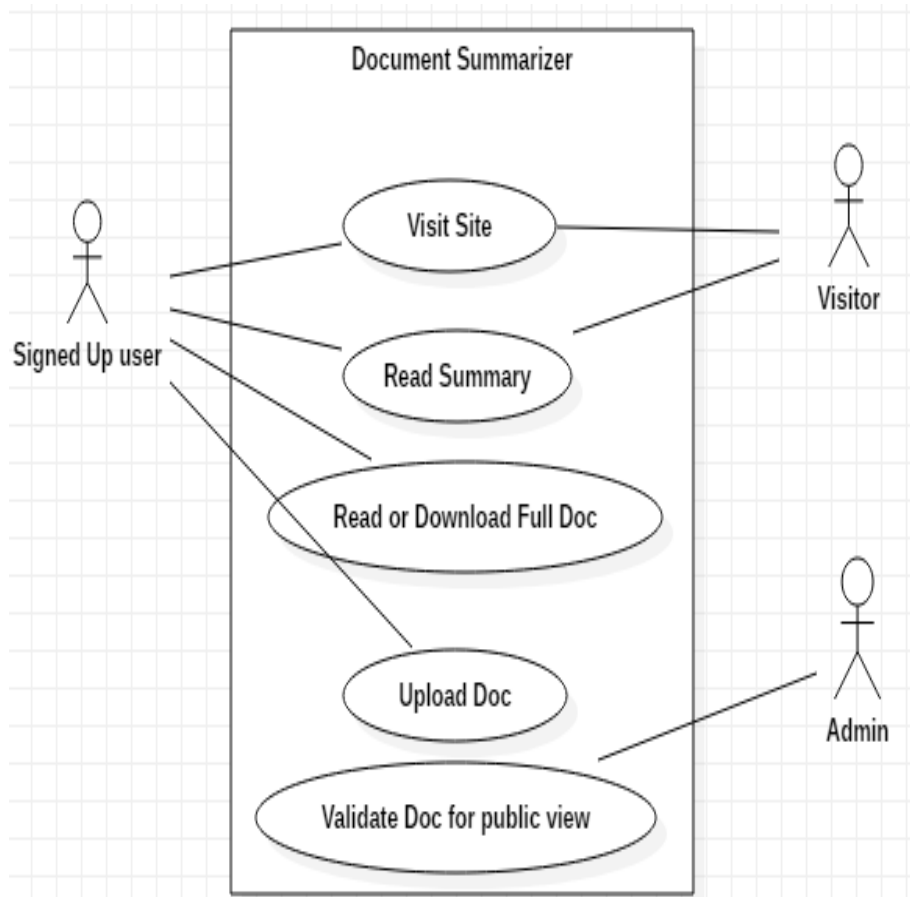


Figure 3: use case diagram for the automatic article summarization

## V. DISCUSSION AND RESULT

The system was created to support computer-produced summaries, which are summaries that are generated by computers. In order to summarize a document using this technique, one must first copy and paste the text of the argument, then paste the website's URL, and then upload the target document. Python was the dominant language used to implement the TextRank algorithm for text summarization, Mongoose was used for the database, Nodejs was used for the backend, and scripting languages like HTML and CSS were used for the frontend. The system was developed using a variety of programming languages to meet the needs of this system. Text is summarized by the system by requesting text strings from the user, which are then converted into a summary form. The system was also created to achieve document summarizing; to do this, upload the document to be summarized. We went ahead and implemented summary of websites by requesting users to put the link of the website to be summarized into a text field supplied on the index page of the system.



ISSN: 2350-0328

# International Journal of Advanced Research in Science, Engineering and Technology

Vol. 10, Issue 4, April 2023

## D. Samples of the Implementation Screenshot

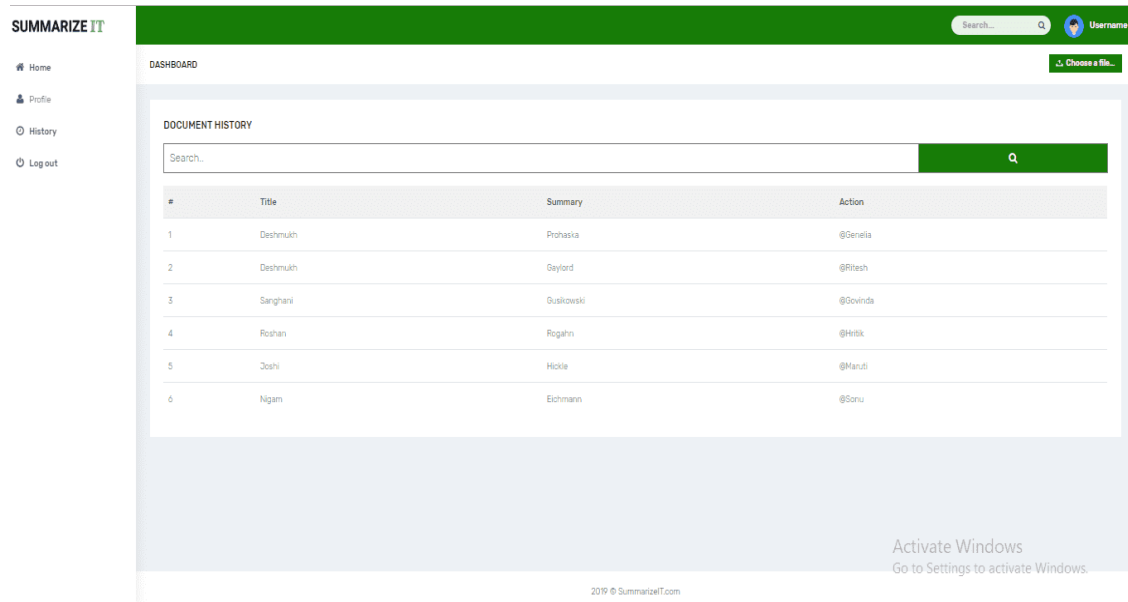


Figure 4: User Dashboard of the system

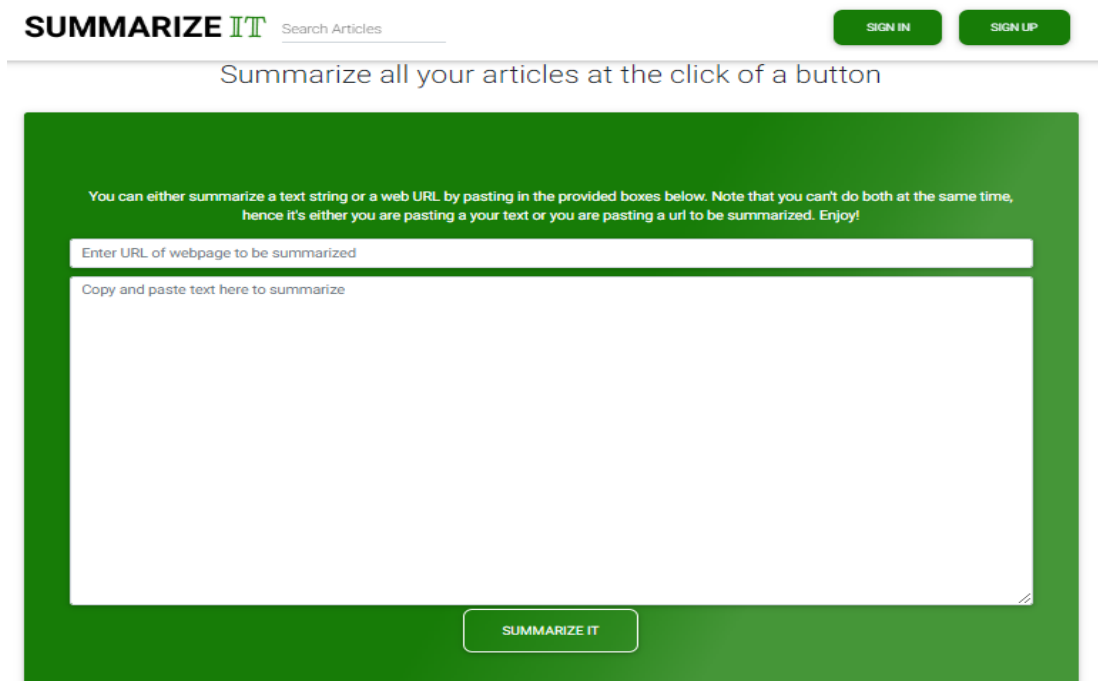
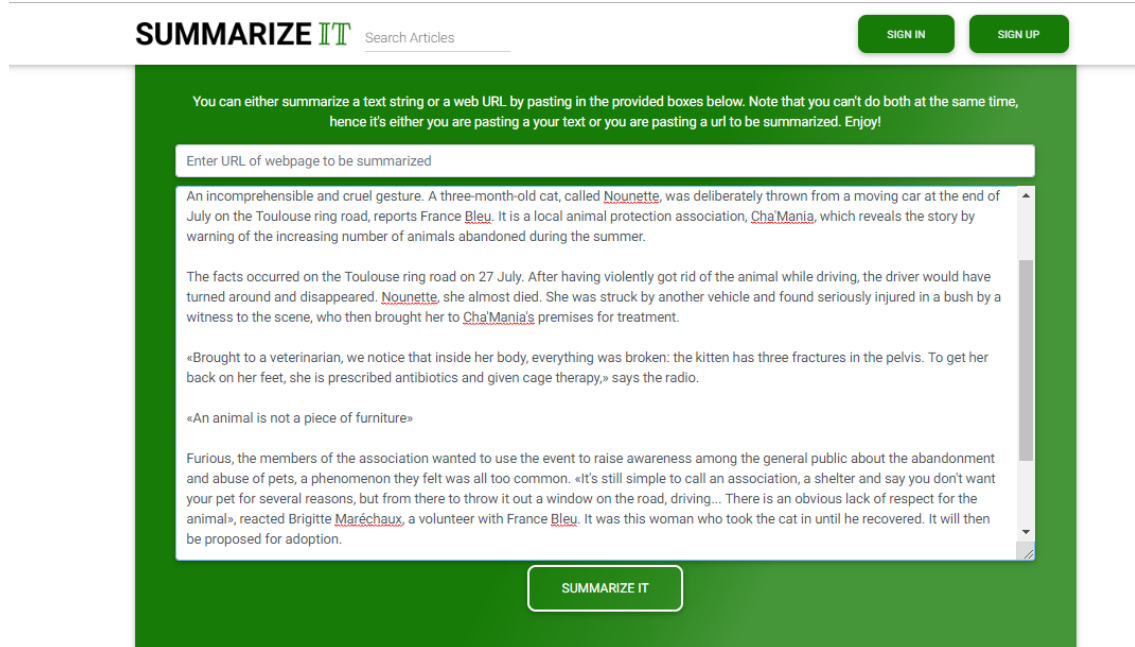


Figure 5: Index Page for Pasting Argumentative Text



**SUMMARIZE IT** Search Articles SIGN IN SIGN UP

You can either summarize a text string or a web URL by pasting in the provided boxes below. Note that you can't do both at the same time, hence it's either you are pasting a your text or you are pasting a url to be summarized. Enjoy!

Enter URL of webpage to be summarized

An incomprehensible and cruel gesture. A three-month-old cat, called [Nounette](#), was deliberately thrown from a moving car at the end of July on the Toulouse ring road, reports France [Bleu](#). It is a local animal protection association, [Cha'Mania](#), which reveals the story by warning of the increasing number of animals abandoned during the summer.

The facts occurred on the Toulouse ring road on 27 July. After having violently got rid of the animal while driving, the driver would have turned around and disappeared. [Nounette](#), she almost died. She was struck by another vehicle and found seriously injured in a bush by a witness to the scene, who then brought her to [Cha'Mania's](#) premises for treatment.

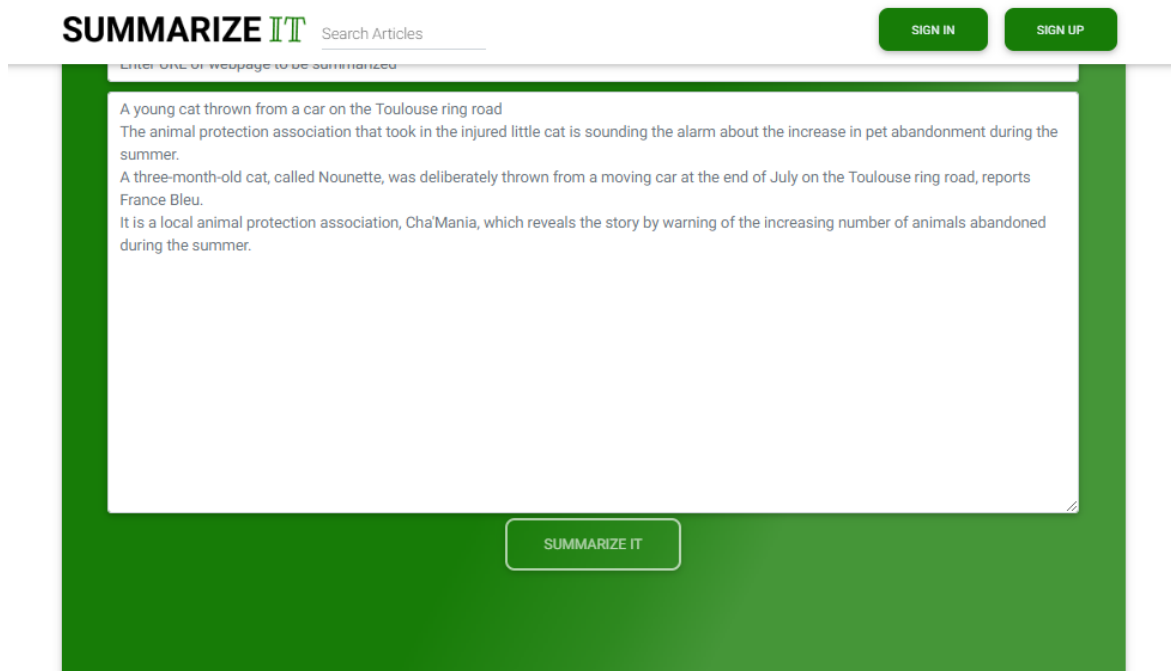
«Brought to a veterinarian, we notice that inside her body, everything was broken: the kitten has three fractures in the pelvis. To get her back on her feet, she is prescribed antibiotics and given cage therapy.» says the radio.

«An animal is not a piece of furniture»

Furious, the members of the association wanted to use the event to raise awareness among the general public about the abandonment and abuse of pets, a phenomenon they felt was all too common. «It's still simple to call an association, a shelter and say you don't want your pet for several reasons, but from there to throw it out a window on the road, driving... There is an obvious lack of respect for the animal», reacted Brigitte [Maréchaux](#), a volunteer with France [Bleu](#). It was this woman who took the cat in until he recovered. It will then be proposed for adoption.

SUMMARIZE IT

Figure 6: Sample Text to Generate Summarization



**SUMMARIZE IT** Search Articles SIGN IN SIGN UP

Enter URL of webpage to be summarized

A young cat thrown from a car on the Toulouse ring road  
The animal protection association that took in the injured little cat is sounding the alarm about the increase in pet abandonment during the summer.  
A three-month-old cat, called Nounette, was deliberately thrown from a moving car at the end of July on the Toulouse ring road, reports France Bleu.  
It is a local animal protection association, Cha'Mania, which reveals the story by warning of the increasing number of animals abandoned during the summer.

SUMMARIZE IT

Figure 7: Result Generated from Summarization

## VI. CONCLUSION

The work considered an improved automated article summarization system using Textrank algorithm to help research in reduce the numerous article papers to read when reviewing literature to gain knowledge in the area of study. The paper used extractive summarization method to select relevant phrases from the original text to create a summary without losing important content of the text. The framework shows the process to take in the implementation of the system from when



the user needs to visit the sites, get signed in, upload articles, and download the summarized article. The developed system will aid researcher to read through journals, articles, thesis, document within a limited time and we therefore boost their efficiency and increase speed in producing their own article. It is recommended for educational sector and to be use by researchers, lecturers, academician, student writing their project and other people that have interest reviewing articles.

#### REFERENCES

- [1] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mob. networks Appl.*, vol. 19, pp. 171–209, 2014.
- [2] [L.M. Abualigah, A.T. Khader, M.A. Al-Betar, Multi-objectives-based text clustering technique using K-mean algorithm, in 2016 7th International Conference on Computer Science and Information Technology (CSIT) (IEEE, 2016), pp. 1–6 3.
- [3] A.M. Azmi, N.I. Altmami, An abstractive Arabic text summarizer with user-controlled granularity. *Inf. Process. Manage.* **54**(6), 903–921 (2018).
- [4] V. Gupta and G. S. Lehal, "A survey of text summarization extractive techniques," *J. Emerg. Technol. web Intell.*, vol. 2, no. 3, pp. 258–268, 2010.
- [5] I. Awasthi, K. Gupta, P. S. Bhogal, S. S. Anand, and P. K. Soni, "Natural language processing (NLP) based text summarization-a survey," in 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 1310–1317.
- [6] N. I. Altmami and M. E. B. Menai, "Automatic summarization of scientific articles: A survey," *J. King Saud Univ. Inf. Sci.*, 2020.
- [7] C. Mallick, A. K. Das, M. Dutta, A. K. Das, and A. Sarkar, "Graph-based text summarization using modified TextRank," in *Soft Computing in Data Analytics: Proceedings of International Conference on SCDA 2018, 2019*, pp. 137–146.
- [8] S. Zaware, D. Patadiya, A. Gaikwad, S. Gulhane, and A. Thakare, "Text summarization using tf-idf and textrank algorithm," in 2021 5th International conference on trends in electronics and informatics (ICOEI), 2021, pp. 1399–1407.
- [9] L. Gao, E. Chang, and S. Han, "Powerful tool to expand business intelligence: Text mining," *Trans. ENFORMATIKA, Syst. Sci. Eng.* vol. 8, ESSE 2005, Budapest, pp. 110–115, 2005.
- [10] A. Nenkova and K. McKeown, "Automatic summarization," *Found. Trends@ Inf. Retr.*, vol. 5, no. 2–3, pp. 103–233, 2011.
- [11] C. Westby, B. Culatta, B. Lawrence, and K. Hall-Kenyon, "Summarizing expository texts," *Top. Lang. Disord.*, vol. 30, no. 4, pp. 275–287, 2010.
- [12] C.-Y. Lin and E. Hovy, "From single to multi-document summarization," in *Proceedings of the 40th annual meeting of the association for computational linguistics*, 2002, pp. 457–464.
- [13] S. Sidhpurwala, S. Jain, and S. Verma, "A Hybrid Approach for Text Summarization," *Top. Intell. Comput. Ind. Des.*, vol. 2, no. 2, pp. 142–146, 2020.
- [14] D. Gaurav, S. Shandilya, S. Tiwari, and A. Goyal, "A machine learning method for recognizing invasive content in memes," in *Knowledge Graphs and Semantic Web: Second Iberoamerican Conference and First Indo-American Conference, KGSWC 2020, Mérida, Mexico, November 26–27, 2020, Proceedings 2, 2020*, pp. 195–213.
- [15] G. Magesh and P. Swarnalatha, "Analyzing customer sentiments using machine learning techniques," *Int. J. Civ. Eng. Technol.*, vol. 8, no. 10, pp. 1829–1842, 2017.
- [16] C.-Y. Lin and E. Hovy, "From single to multi-document summarization," in *Proceedings of the 40th annual meeting of the association for computational linguistics*, 2002, pp. 457–464. [1] R.
- [17] R. Nallapati, B. Zhou, C. Gulcehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence rnns and beyond," *arXiv Prepr. arXiv1602.06023*, 2016.
- [18] J. Kupiec, J. Pedersen, and F. Chen, "A trainable document summarizer," in *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, 1995, pp. 68–73.
- [19] Y. Ouyang, W. Li, S. Li, and Q. Lu, "Applying regression models to query-focused multi-document summarization," *Inf. Process. Manage.*, vol. 47, no. 2, pp. 227–237, 2011.
- [20] K.-F. Wong, M. Wu, and W. Li, "Extractive summarization using supervised and semi-supervised learning," in *Proceedings of the 22nd international conference on computational linguistics (Coling 2008)*, 2008, pp. 985–992.
- [21] L. Zhou and E. Hovy, "A web-trained extraction summarization system," in *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 2003, pp. 284–290.
- [22] J. M. Conroy and D. P. O'leary, "Text summarization via hidden markov models," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, 2001, pp. 406–407.
- [23] D. Shen, J.-T. Sun, H. Li, Q. Yang, and Z. Chen, "Document summarization using conditional random fields.," in *IJCAI*, 2007, vol. 7, pp. 2862–2867.
- [24] M. Yousefi-Azar and L. Hamey, "Text summarization using unsupervised deep learning," *Expert Syst. Appl.*, vol. 68, pp. 93–105, 2017.
- [25] V. Qazvinian and D. R. Radev, "Scientific paper summarization using citation summary networks," *arXiv Prepr. arXiv0807.1560*, 2008.
- [26] B. Schiffman, I. Mani, and K. Concepcion, "Producing biographical summaries: Combining linguistic knowledge with corpus statistics," in *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, 2001, pp. 458–465.
- [27] S. Soares, B. Martins, and P. Calado, "Extracting biographical sentences from textual documents," in *Proceedings of the 15th Portuguese Conference on Artificial Intelligence (EPIA 2011)*, Lisbon, Portugal, 2011, pp. 718–730.
- [28] N. Moratanch and S. Chitrakala, "A survey on extractive text summarization," in 2017 international conference on computer, communication and signal processing (ICCCSP), 2017, pp. 1–6.
- [29] B. T. M. Anh, N. T. My, and N. T. T. Trang, "Enhanced genetic algorithm for single document extractive summarization," in *Proceedings of the 10th International Symposium on Information and Communication Technology*, 2019, pp. 370–376.
- [30] R. V. B. Vangara, S. P. Vangara, K. Thirupathur, and B. Tavidaboina, "A Hybrid Model for Summarizing Text Documents Using Text Rank Algorithm and Term Frequency," 2020.
- [31] D. Rao, B. Harika, A. Srikanth, and Y. Vahini, "Nlp Based Machine Learning Approaches for Text Summarization," B. Srikanth, A. Vahini, Y., *Nlp Based Mach. Learn. Approaches Text Summ.* (February 18, 2023), 2023.



ISSN: 2350-0328

**International Journal of Advanced Research in Science,  
Engineering and Technology**

**Vol. 10, Issue 4, April 2023**

- [32] A. A. Bialy, M. A. Gaheen, R. M. ElEraky, A. F. ElGamal, and A. A. Ewees, "Single Arabic document summarization using natural language processing technique," *Recent Adv. NLP Case Arab. Lang.*, pp. 17–37, 2020.
- [33] K. U. Manjari, "Extractive summarization of Telugu documents using TextRank algorithm," in *2020 Fourth international conference on I-SMAC (IoT in social, mobile, analytics and cloud)(I-SMAC)*, 2020, pp. 678–683.
- [34] M. Mohd, R. Jan, and M. Shah, "Text document summarization using word embedding," *Expert Syst. Appl.*, vol. 143, p. 112958, 2020.
- [35] E. Vorhees and D. Graff, *AQUAINT-2 information-retrieval text: research collection*. Linguistic Data Consortium, 2008.
- [36] S. Upasani, N. Amin, S. Damania, A. Jadhav, and A. M. Jagtap, "Automatic summary generation using textrank based extractive text summarization technique," 2020.