Computational & NLP Lab

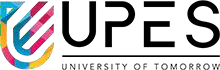
PROJECT (SRS)

For

**Title : Automatic Text Summarization**

Submitted By

|  |  |  |
| --- | --- | --- |
| **Specialization** | **SAP ID** | **Name** |
| AIML(H) | 500093671 | Garvit Sharda |
| AIML(H) | 500092974 | Devansh Aggarwal |
| AIML(H) | 500094000 | Nitika Singh |

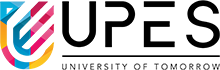


Department of Informatics School of Computer Science

UNIVERSITY OF PETROLEUM & ENERGY STUDIES,

DEHRADUN- 248007. Uttarakhand

# Submitted To : Dr. Achala Shakya



## School of Computer Science

University of Petroleum & Energy Studies, Dehradun

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Topic** | | **Page No** |
|  | Abstract | 3 |
| 1. | Introduction | 3 |
| 2. | Literature Review | 4 |
| 3. | Problem Statement | 4 |
| 4. | Methodology & Algorithm | 5 |
| 5. | SWOT Analysis | 7 |
| 6. | Applications | 8 |
| 7. | Conclusion | 9 |
| 8. | References | 9 |

**Report Title: -** Automatic Text Summarization

# Abstract

Through the use of computer approaches, automatic summarizing reduces massive amounts of data into more digestible and focused representations that highlight the most important details or primary ideas included in the original content. When handling large volumes of information, this method is very helpful in domains like business intelligence, academics, and media.

Algorithms based on artificial intelligence are crucial to automatic summarization because they make it possible to automatically extract pertinent data from a variety of sources, including text documents, articles, and multimedia content. These algorithms use cutting-edge methods including natural language processing (NLP), machine learning, and deep learning. They are designed to handle several kinds of data, including textual, numerical, and multimedia data.

# Introduction

Automated text summarizer, a cutting-edge instrument that condenses long documents, articles, or writings into readable, logical summaries. It analyzes and extracts the most crucial information using cutting-edge natural language processing algorithms, making it possible for users to quickly and effectively understand essential insights. Our summarizer makes it easy to condense research materials for students or professionals looking for concise summaries of intricate studies. It saves time and effort and makes sure you don't miss any important details. Use our automatic text summarizer now to discover the power of simplified information.

The capacity to swiftly identify pertinent content in an information-rich society is crucial. Our automatic text summarizer uses state-of-the-art algorithms that simulate human comprehension, reading texts and highlighting important points, main ideas, and arguments. It produces summaries that are clear and thorough while simultaneously condensing information while preserving context and meaning.

Our summarizer may also be tailored to a variety of businesses and demands thanks to its versatility. Our application is designed to meet the special needs of individuals working in academia, business, journalism, or research. It may help with various tasks like literature reviews, market assessments, news aggregation, and more. You can customize summaries to your tastes with its easy-to-use interface and configurable choices, guaranteeing an enjoyable reading experience. By using our automatic text summarizer, you can stay on top of trends and improve your information processing skills.

# Literature Review

The importance of automating the billing process cannot be emphasized in the fast-paced corporate environment of today, where accuracy and efficiency are crucial. While manual billing is still a common practice, it has drawbacks that can make it difficult for firms in a range of industries to run smoothly. Thankfully, the development of automated billing systems has brought about a new era of financial administration that has numerous advantages.

One of the key advantages of automating the billing process is the release of important staff time. Manual billing requires a lot of labour and might take up a sizable amount of workers' time. Human error is a factor in time-consuming tasks including applying taxes, computing invoices, and keeping track of various charges. [1]

Another crucial element that contemporary organizations should take account is the efficient payment integration inside invoicing software. The transaction process is streamlined by integrating payment processing, which also eliminates the need for manual accounting of payments and shortens customer wait times. A more effective and user-friendly experience that increases happiness and loyalty is the final outcome [2]. Additionally, integrated payment systems produce thorough, error-free transaction records, making it easier to handle financial data and guarantee accounting rules are being followed.

Automation offers benefits that go beyond the billing procedure alone. To fully utilize technology, many firms are switching to cloud-based billing solutions. By offering secure financial data storage, reliable data backup capabilities, and a superior user experience for both employees and consumers, these systems enable increased productivity and accuracy.

1. **Problem Statement :**

* Overwhelming amounts of textual data make it difficult for users to sort through and derive pertinent information.
* Time-consuming : It takes a lot of work and inefficiency to manually read and summarize lengthy documents.
* Having trouble making decisions: Inadequate synopses impede prompt and well-informed decision-making procedures

1. **Methodology & Algorithm :**

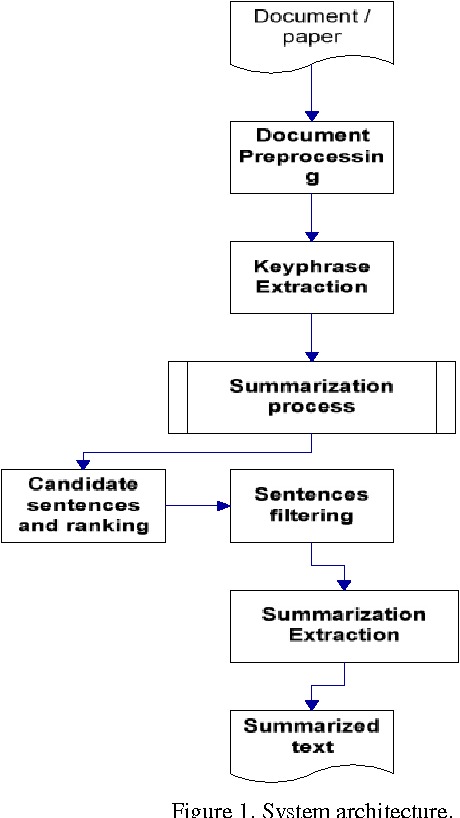
* **Importing Libraries :** Open text processing libraries (e.g., pandas, spaCy, etc.).
* **Tokenization :** Divide the text into discrete words or meaningful units.
* **Frequency Count :** Examine the frequency at which words appear to determine their significance.
* **Stop Words Removal :** Get rid of ordinary, meaningless words like "the" and "a".
* **Ordering Every Sentence** : Based on keywords and relevancy, give sentences a score.
* **Combining Sentences :** For conciseness, combine related, highly-scoring sentences.
* **Summarization :** To create the final summary, choose the sentences that rank highest.
* **Output :** Display the streamlined content as the abridged version.

**Libraries Used:**

* **SpaCy :** A library for natural language processing.
* **Collections :** The built-in library for lists, dictionaries, and sets in Python.
* **Pandas :** Pandas is a data analysis and manipulation library.

* **heapq :** A heap data structure implementation library.
* **tkinter :** A Python standard GUI toolkit.
* **ttk (a subset of tkinter) :** A tkinter themed extension for building GUIs that appear native.

**Flowchart :**



1. **SWOT Analysis :**
2. **Strengths :**

* **Efficiency:** Users can save time and effort by using automatic text summarizers, which can swiftly provide summaries of lengthy materials.
* **Reliability:** It is ensured by their consistent production of summaries, which is independent of the person interpreting the text.
* **Scalability:** These technologies are appropriate for summarizing enormous volumes of information since they can manage high text volumes.
* **Customization:** To fit their own requirements, users can frequently adjust elements like the amount of detail or the summary length.

1. **Weaknesses:**

* **Accuracy**: Automatic summarization algorithms may find it difficult to faithfully convey the text's complex meaning, which could result in summaries that are inaccurate.
* **Language Complexity**: Idiomatic phrases, complicated linguistic patterns, and terminology unique to a given field may pose challenges for summarizers.
* **Subjectivity:** Different algorithms may generate different summaries for the same text, and summarization is prone to subjectivity.
* **Lack of Context:** In the absence of context, automatic summarizers run the risk of omitting important details or misinterpreting the text's meaning.

1. **Opportunities :**

* **Technological developments:** The precision and efficacy of automatic summarization systems can be enhanced by developments in machine learning and natural language processing (NLP).
* **Integration:** To improve user experience, summarization technologies can be incorporated into a variety of applications, including news aggregators, search engines, and content management systems.
* **Multilingual Support:** A larger user base can be served by creating summarizers that can summarize content in several languages.
* **Specialized Summaries:** It is possible to create summarizing algorithms that are tailored to particular fields, such as law or science.

1. **Threats :**

* **Competition:** As the market for automatic text summarizing technologies expands, there is more rivalry amongst the service providers.
* **Privacy Concerns:** When summarizing private documents, automatic summarizers could unintentionally reveal sensitive information.
* **Regulatory Difficulties:** Summarization technologies that handle sensitive or personal data may find it difficult to comply with laws like GDPR or HIPAA.
* **User Adoption:** Because they may have doubts about the technology's reliability or correctness, users may be reluctant to rely exclusively on automatic summarization.

1. **Applications :**

* **News & Research :** An automatic text summarizer is a vital tool for journalists and researchers working in the fields of news and research. The summarizer effectively condenses articles and reports into brief summaries in the fast-paced world of news reporting, where information overload is typical. This guarantees that readers stay informed without spending undue time on lengthy texts.
* **Education :** Automatic text summary is showing promise as a useful tool for teachers and students alike. The summarizer helps students understand and retain material by condensing complicated ideas and theories into manageable summaries. This is especially helpful for students who are overwhelmed by lengthy reading assignments and research materials.
* **Business :** For professionals and decision-makers looking for practical insights from a sea of data, computerized text summarizing provides a number of benefits. Executives and analysts can use the summarizer to extract important insights from reports and financial papers, facilitating strategic planning , informed decision-making.
* **Content Creation :** Automatic text summary is a potent tool for content curation and productivity optimization for content creators on a variety of platforms, including bloggers and social media managers. Through the automated summarization of news articles, blog posts, and other content, authors can select pertinent material for their audiences and incorporate their own commentary and views.

1. **Conclusion :**

To sum up, there is a great deal of potential for the development and application of an automatic text summarizer project in a variety of fields, such as news and research, education, business, and content production. Through the utilisation of artificial intelligence and natural language processing algorithms, this project facilitates the effective extraction of crucial insights and information from substantial textual data sets, hence maximising productivity, reducing turnaround times, and promoting well-informed decision-making. The automatic text summarizer project proves to be a flexible and vital tool for managing the intricacies of the information age, whether it is used for keeping abreast of current affairs, supporting education and research, advancing business strategy initiatives, or optimizing content creation procedures.

**8.References :**

1) El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2021). Automatic text summarization: A comprehensive survey. Expert systems with applications, 165, 113679.

2) Fattah, M. A., & Ren, F. (2008). Automatic text summarization. World Academy of Science, Engineering and Technology, 37(2), 192.

3) Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: a survey. Artificial Intelligence Review, 47(1), 1-66.

4) Tas, O., & Kiyani, F. (2007). A survey automatic text summarization. PressAcademia Procedia, 5(1), 205-213.

5) Munot, N., & Govilkar, S. S. (2014). Comparative study of text summarization methods. International Journal of Computer Applications, 102(12).

6) Gupta, V., & Lehal, G. S. (2010). A survey of text summarization extractive techniques. Journal of emerging technologies in web intelligence, 2(3), 258-268