



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Text Summarization Using NLP

Mr. Ashish Shinde¹, Mr. Dhananjay Shinde², Mr. Onkar Naram³, Mr. Suyash Sathe⁴,
Mrs. Sushma Ghose⁵

Department of Computer Engineering, Siddhant College of Engineering, Sudumbare, Pune India^{1,2,3,4,5}

ABSTRACT: Data mining is a field that has seen significant evolution in recent years as a result of enormous breakthroughs in software and hardware technologies. As technology advances, more types of data become available, which is especially useful in the case of text data. The software and hardware platforms that power social networks and the internet have aided in the rapid production of massive data stores. Structured data is typically maintained by a database system, whereas text data is typically managed by a search engine due to the lack of structures. With the help of a keyword query, the search engine allows the online user to find the essential information from the gathered works. Text summary is the practice of extracting the most relevant information from a source document in order to create an abridged version for a specific job.

KEYWORDS: NLP(natural language processing), extractive, abstractive, encoding, decoding

I. INTRODUCTION

In today's world, a massive amount of data is being generated on the internet every day. As a result, a better mechanism for extracting important information quickly and effectively is required. Text summarization is one of the strategies for finding the most significant and meaningful information in a document or set of linked documents and compressing it into a shorter version while maintaining the overall meaning. It cuts down on the time it takes to read an entire page and solves the space issue that comes with keeping big amounts of data. There are two techniques to Automatic Text Summarization. 1) Abstractive text summarising and 2) Extractive text summarization are two types of text summarization. The term "extractive text summarization" refers to the extraction of key information or sentences from a text file or source document. An extractive text summarizing approach selects interesting informative sentences based on linguistic or statistical criteria. An abstractive text summary will attempt to comprehend the input or original file and re-generate the output in a few words by recognizing the input file's core concept. That extracted text has been mentioned in a number of academic studies. Text summarization is widely utilized in a variety of sectors, including science, medicine, law, and engineering. Researchers have concentrated on creating summaries of doctor's prescriptions, which have proven to be extremely beneficial to patients. Long news stories have also been summarized so that readers can get a lot of information on a variety of topics in a short amount of time. For the past five years, we've discussed the numerous strategies utilized in text summarizing in this document. Machine learning (ML), neural networks (NNs), reinforcement learning, sequence to sequence modeling, and fuzzy logic were determined to be the most popular methods. Similarly, for the goal of text summarization, numerous optimization methods have been applied to maximize the specified objective function.

II. RELATED WORK

[1] Text mining and text summarization have a lot more in common than you would think. Established summarizing systems should be built and classified based on the kind of input text, based on the differences in requirements summary with respect to input text. The concept of text mining and its relationship with text summarization are discussed initially in this work. Following that, a study of some of the summary methodologies and their key parameters for extracting dominant phrases was conducted, as well as the main stages of the summarizing process and the most critical extraction criteria. Finally, the most basic proposed evaluation methods are taken into account.

[2] Text summarizing techniques have been altered by the use of linguistics to advanced machine learning models; this work investigates summarization approaches as well as contemporary state-of-the-art models for single and multi-document summary. This survey aims to conduct a comprehensive investigation using machine learning, modern graph and evolutionary based methods, from feature representation to sentence selection and summary creation. The whole study will assist researchers in properly handling enormous amounts of data while developing effective Natural Language Processing apps. Finally, this research identifies common abstractive mechanisms and observations that will be useful in the research.

[3] A novel statistical strategy for extracting text summarization on a single document is demonstrated in this research. The method of sentence extraction is provided, which gives the idea of the input text in a concise way. Sentences are graded by assigning weights to them and then ranking them according to those weights. Highly scored sentences are taken from the input document, allowing it to extract essential sentences that lead to a high-quality summary of the input material, which may then be saved as audio.

III. PROPOSED ALGORITHM

A] 1: procedure PROFILE (text File)

2: tokens = tokenize (textFile);

3: for all tk \in tokens do

4: score = TF-DF(tk);

5: wordVector.put(tk, score);

6: categories Vecotrs = load Vectors();

7: for all ca in cate Vecotrs do

8: similarity = cosineSimilarity(ca, wordVector);

9: profile.put(ca.name, similarity);

10: save profile;

B] **TFIDF**

1: procedure PROFILE(text File)

2: tokens = tokenize(textFile);

3: for all tk \in tokens do

4: score = TF-IDF(tk);

5: word Vector.put(tk, score); \triangleright calculate ifidf for all words,

6: categories Vecotrs = loadCategory Vectors();

7: for all ca \in categories Vecotrs do

8: similarity = cosineSimilarity(ca, wordVector);

9: profile.put(ca.name, similarity)

IV. SYSTEM ARCHITECTURE

In implementation phase of our project we have implemented various module required of successfully getting expected outcome at the different module levels. With inputs from system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality which is referred to as Unit Testing. Test Summarization allows the user to examine the contents of a text for accurate, short, and exact information, which increases productivity. As a result, the tool relieves the user of work by reducing the size of the text and increasing productivity by allowing the user to focus their energy on more important tasks.

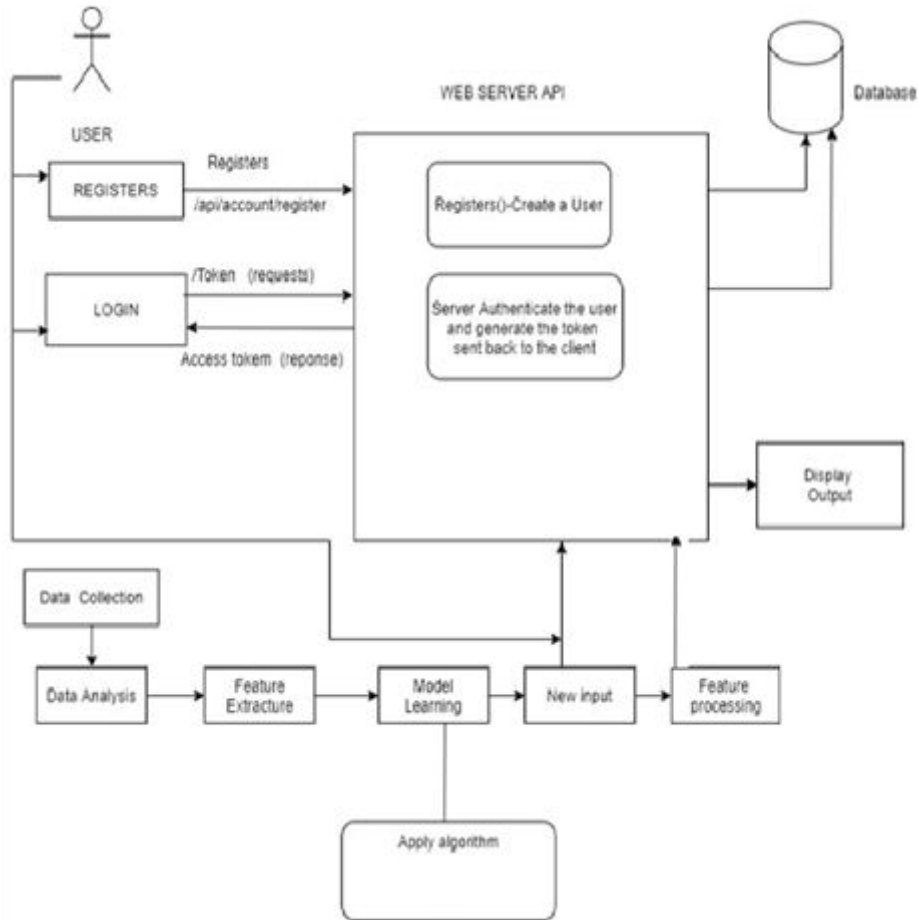


Fig.1. Sysytem Architecture

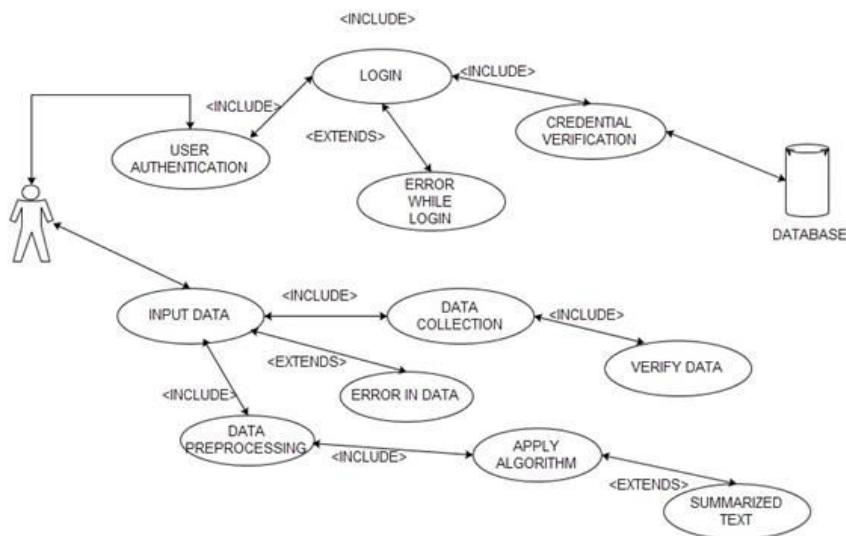


Fig. 2. User Case

V. CONCLUSION AND FUTURE WORK

Automatic text summarization is a method that allows individuals to achieve a quantum leap in productivity by reducing the sheer volume of information they encounter on a daily basis. This not only allows people to reduce the amount of reading they must do, but it also allows them to read and comprehend previously neglected literary works. Text summarising is a rapidly growing area, with specialized tools being created to handle increasingly concentrated summary jobs. Users are expanding the use case of this technology as open-source software and word embedding packages become more widely available.

Because the suggested approach does not require a knowledge basis, it can be used to summarize articles from a wide range of topics, including politics, sports, current events, and finance. It does, however, result in a tradeoff between domain independence and a knowledge-based summary, which would provide facts in a more human-friendly format. This work could be used to make data available on the go on a mobile network by shortening the words generated by our algorithm and then shortening them again. To achieve this goal, a variety of NLP-based algorithms can be deployed. As a result, we'd first create a summary by extracting sentences from multiple texts, and then use abstractive approaches to condense those sentences. This will guarantee that the summary presented is the most compact form possible in the mobile sector.

REFERENCES

1. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean, "Dist ribut ed Representations of Words and Phrases and their Compositionality," arXiv:1310.4546v1 [cs.CL], 2013.
2. Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, "Efficient Estimation of Word Representat ions in Vector Space," arXiv:1301.3781v3 [cs.CL], 2013.
3. Shi Ziyang " The Design and Implementation of Domain-specific Text Summarizat ion System based on Co-reference Resolution Algorithm" 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery.
4. Paul Gigioli, Nikhita Sagar, Anand Rao, Joseph Voyles " Domain-Aware Abst ract ive Text Summarizat ion for Medical Document s" published during 2018 IEEE BIBM.
5. Niladri Chat terjee, Amol Mittal and Shubham Goyal's " Single Document Ext ract ive Text Summarizat ion Using Genetic Algorithms" (2012)
6. Amol Tandel, Brijesh Modi, Priyasha Gupta, Shreya Wagle and Sujata Khedkar's "Mult i-document text summarization - A survey" 2016 Internat ional Conference on Data Mining and Advanced Comput ing (SAPIENCE).
7. Adit ya Jain, Divij Bhatia, Manish K T hakur's "Extractive Text Summarizat ion using Word Vect or Embedding"(2017).
8. Canasai Kruengkrai and Chuleerat Jaruskulchai "Generic T ext Summarizat ion Using Local and Global Properties of Sentences"(2003).
9. Nithin Raphael, Hemanta Duwara and Philemon Daniel "Survey on Abstractive Text Summarization" 2018 International Conference on Communicat ion and Signal Processing (ICCSP).
10. Yang Wei and Yang Zhizhuo "Query based Summarization using topic background knowledge" 2017 13t h International Conference on Natural Computat ion, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD).



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

 **doi**[®]
cross **ref**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details