

Real Text Summarizer Extension

Devendra Yadav¹, Vivek Sharma², Gopal Sharma³

¹Devendra Yadav CS Department & H.R Institute of Technology

²Vivek Sharma CS Department & H.R Institute of Technology

³Gopal Sharma CS Department & H.R Institute of Technology

Abstract - The advent of the World Wide Web, with a wealth of information, required easily accessible information, so databases and search engines were created to access that information. This information system is updated whenever new data is generated, since the volume of electronic information rises everyday, and no one can promise that manual analysis will provide the full picture anymore.

Accordingly, the need for automated summarization systems is high. Automatic summary reads few papers as input and produces a summarized version and hence saves both information and time. The research was performed in a single paper and produced multiple papers. In this report, we are Specifically focusing on frequency-based approach for text summarization

Key Words: Automatic summarization, Extractive, Node package manager (npm), javascript, frequency-based.

1. INTRODUCTION

Real Text summarization is the process of extracting key information from the given article or document and then can be shortened by a program. With increase in the amount of data there was increase in the problem of data overload so this interested people in capturing the text as amount of data increased. A large document is difficult to summarize manually because it is completely a human effort and time taking also.

The two primary ways of summing up text documents are :-

The concentrate on extractive summaries is in deciding on what portions, sentences, phrases, etc. are critical from the primary text and connecting them in a compact format. It is inferred that critical sentences are important by analyzing the sentences based on related analytical and semantic features.

Summary systems are usually based on sentence delivery methods and for understanding the whole document properly as well as for extracting the important sentences from the document. The technique of generating a brief description that comprises a few phrases that describe the key concepts of an article or section is known as abstractive summarization. This function is also included to naturally map the input order of words in a source document to the target sequence of words called the summary.

2. LiteratureSurvey

Summarization can be classified into three basic summaries that how can be evolved.

i — extractive (Just pick sentences from source text),

ii—abstractive(Generate new sentences after reading source text)

iii — combination of extractive and abstractive.

Extractive Automatic Text Summarization: Extractive text summarization is the process that skim over the summary from a document corpus and concatenate the summary (Rau et al., 1989).

Then, we will define the process in the framework of abstractive automatic text summarization for the given corpus where the new sentences are generate by paraphrasing (Zhang et al., 2019).

Hybrid Automatic Text Summarization: It is the combination of both extractive and abstractive. In your words, it is extracting some sentences, from a corpus, and generating a new one (Binwahlan et al., 2010).

The vast larger sizes of the research papers are also related to monolingual text summarization. It is difficult to do, especially when compared with monolingual. Teaching a machine more than one language structure requires work. SUMMARIST (Hovy & Lin, 1996) was a language independent text summarization system of abstracts based on an extraction approach, developed for six languages: English, Indonesian, Spanish, German, Japanese, and Korean. Cross Language Text Summarization (CLTS) (Linhares Pontes et al., 2020): translates learners to a target language and then generates a summary in that language. As these are some combination of text summarization along with machine translation approach. But this combination also leads to errors, thus diminishing the quality of the summaries.

By jointly analyzing information in the source and destination languages, CLTS systems may extract information that is more relevant and thus are conducive to the development of extractive cross lingual summaries. CLTS methods previous provided either a comprehensive or abstract method, but these methods depend on frameworks or tools that are only existing in few language, limiting ability in other languages.

In (Moradi et al., 2020), the authors have proposed an elegant way of handling the afore mentioned problems with graph based approaches. Continuous word representation model was trained by the min order to capture the linguistic, semantic and

Contextual relationships between the sentences. These being the Skip grams and ContinuousBagOfWords(CBOW) models by word2vec (Mikolov et al., 2013) and GloVe (Mutlu et al., 2020)(HansonandEr, 1971)on a large biomedical text corpora.

As we have seen in a previous study, supervised methods were some approaches where human-generated summaries guided us to scope for parameters or features of summarization algorithms. Nevertheless, unsupervised approaches (Nomoto &Matsumoto, 2003) incorporating (discriminative) diversity (Tsatsaronisetal.,2010)discern salient properties without any application of the expert reports. In (Yeh et al., 2005b), a trainable summarizer is implemented, which performs sentence selection driven by various factors e.g. location, positive keyword, negative keyword, centrality, and relation to the title. The key is feature weighting for which we use a genetic algorithm (GA) to train a score function which can be used to find the right balance of the combination endent feature weights. From there, it uses latent semantic analysis (LSA) to generate both a document (or corpus) semantic matrix and a semantic representation of the sentence, creating a semantic text relationship map.

A hybrid approach using diversity-based technique, fuzzylogic and swarm-based approaches (Binwahlan et al., 2010) can be resulted in good summaries. Since diverse-based method shave a characteristic of figure out the similarity in the sentences and to obtain a sentence with larger diversity, extend the diversity of prototype sentence and are concentrated on redundancy detection, swarm-based methods are likely to identify important sentence and unimportant sentence, and the swarm- based methods present obvious rough reasoning mechanism, and therefore focus on the scoring techniques of specific sentences, and the above combination can contribute the paragraph extraction as a single entity. Here comparing two- way Swarm-fuzzy based methods does better than any other diversity based methods.

Research Methodology:

Real text summarizer extension is ultimate tool for quickly summarizing content and highlighting key insights. Its uses OpenAI’s CHATGPT technology to condense long webpage content into easy – to -understand nuggets of wisdom.

Experience the real Text extension, your ultimate content summarization tool! With the power of OpenAI’s advanced GPT technology, real Text effortlessly extracts content from your active browser tab and transforms them into concise summaries and highlights. You can also copy the summary and highlights easily to your clipboard with a single click.

Keep track of your reading history, access it whenever you need, and stay informed about your online/offline status. Plus, real text gracefully handles errors related to restricted content,

data loading issues, data changes, and other network errors. Elevate your browsing and experience with real Text.

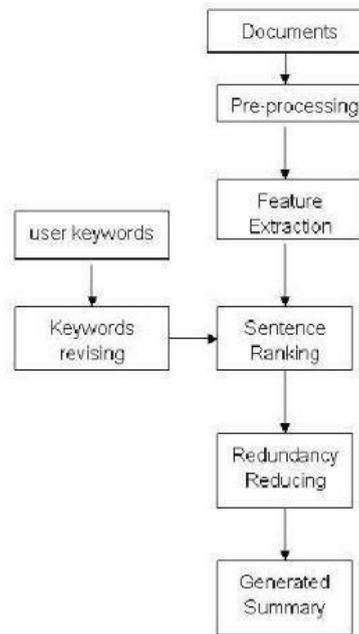


Fig.1: Figure showing the block diagram of system working

Result:

The real text summarizer extension significantly enhances information retrieval by producing concise summaries, saving time and resources through automation. It improves document management and accessibility by organizing content better. The summarizer is adaptable to various domains, ensuring relevance and accuracy. Advanced algorithms and techniques like stop words filtering and k-means clustering enhance summary quality. Customization options allow users to tailor summaries to their needs, making complex information and get the summary more accessible and manageable.

Future Scope:

This section outlines potential future extensions for the study. The current focus has been on summarizing news articles in the realms of sports and technology, but the strategies proposed exhibit flexibility across various domains.

- Develop an overview framework targeting the main topics in news articles or blogs, enhancing coherence and relevance in summaries.
- Increase the exploration of machine-dependent methods to elevate the accuracy and efficiency of summarization techniques.

3. CONCLUSIONS

Text summaries have been shown to be useful for natural language processing tasks such as question and answer or other related fields of computer science such as text classification and data retrieval and access time for information search will be improved. At the same time, sequencing enhances the effect and its algorithms are less biased than human creams. Using a text summary system, commercial capture services allow users to increase the number of texts they can process.

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this thesis work.

REFERENCES

- [1]. Binwahlan, M. S., Salim, N., & Suanmali, L. (2009a). Swarm Based Text Summarization. 2009 International Association of Computer Science and Information Technology - Spring Conference, IACSIT-SC 2009, 145–150.
- [2]. Binwahlan, M. S., Salim, N., & Suanmali, L. (2010). Fuzzy swarm diversity hybrid model for text summarization. *Information Processing and Management*, 46(5), 571–588.
- [3]. Hovy, E., & Lin, C.-Y. (n.d.). AUTOMATED TEXT SUMMARIZATION AND THE SUMMARIST SYSTEM.
- [4]. Hovy, E., & Lin, C.-Y. (1996). Automated text summarization and the SUMMARIST system.
- [5]. Linhares Pontes, E., Huet, S., Torres-Moreno, J. M., & Linhares, A. C. (2020). Compressive approaches for cross-language multi-document summarization. *Data and Knowledge Engineering*.
- [6]. Moradi, M., Dashti, M., & Samwald, M. (2020). Summarization of biomedical articles using domain-specific word embeddings and graph ranking. *Journal of Biomedical Informatics*, 107(May), 103452.
- [7]. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings, 1–12.
- [8]. Mutlu, B., Sezer, E. A., & Akcayol, M. A. (2020). Candidate sentence selection for extractive text summarization. *Information Processing and Management*, 57(6).
- [9]. HANSON ER. (1971). Musicassette Interchangeability. the Facts Behind the Facts. *AES: Journal of the Audio Engineering Society*, 19(5), 417–425.
- [10]. Nomoto, T., & Matsumoto, Y. (2003). The diversity-based approach to open-domain text summarization. *Information Processing and Management*, 39(3), 363–389.
- [11]. Yeh, J. Y., Ke, H. R., Yang, W. P., & Meng, I. H. (2005a). Text summarization using a trainable summarizer and latent semantic analysis. 41(1), 75–95.
- [12]. Yeh, J. Y., Ke, H. R., Yang, W. P., & Meng, I. H. (2005b). Text summarization using a trainable summarizer and latent semantic analysis. *Information Processing and Management*, 41(1), 75–95.