Project Id: MCA2022-24PID0XX

Project Synopsis

on

-Project Title-

Submitted in the partial fulfillment of the requirement for the award of

Degree

of

Master of Computer Applications (MCA)

Submitted By

Mrtyunjy Singh (XXX900100091)

Nikhil Saini (XXX900100097)

Ram Kumar Sharma (XXX900100117)

Under the guidance of

Ms. Deepshikha Rai

(Assistant Professor)

DEPARTMENT OF MCA

ABES INSTITUTE OF TECHNOLOGY, GHAZIABAD

AFFILIATED TO

Dr A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY,

LUCKNOW,UTTAR PRADESH

(ODD SEM, 2023-2024)

Index

Abstract

This Project represents the work related to Data Summarization and also Text Summarization. In this paper, we present a framework for summarizing the huge information. The proposed framework depends on highlight extraction from the internet, utilizing both morphological elements and semantic data. Presently, where huge information is available on the internet, it is most important to provide improved ways to extract the information quickly and most efficiently. It is very difficult for human beings to manually extract the summary of a large document of text. There are plenty of text materials available on the internet. So, there is a problem of searching for related documents from the number of documents available and absorbing related information from it. In essence to figure out the previous issues, automatic text summarization is very much necessary. Data Summarization is the process of identifying the most important and meaningful information in an input document or set of related input documents and compressing all the inputs into a shorter version while maintaining its overall objectives.

INTRODUCTION

Data summary is the way of selecting important points from the provided article or a document that can be reduced by a program. As the data overload problem increased, so did the interest in capturing the text as the amount of data increased. Summarizing a large document manually is challenging since it requires a lot of human effort and is time-consuming. There are mainly two methods for summarizing the text document that can be done by using extractive and abstractive techniques. Extractive summaries concentrate on selecting important passages, sentences, words, etc. from the primary text and connecting them into a concise form. The importance of critical sentences is concluded on the basis of analytical and semantic features of the sentences.

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.

Literature survey

S.No	Title	Author Name	Conclusion	Drawback
1	TEXT SUMMARIZATION USING NLP,2022	ChetanaVaragantham1, J.SrinijaReddy 2 , UdayYelleni3 , MadhumithaKotha4 , Dr P.VenkateswaraRao5	Decreasing the input textual data to a more compact and reduced summarization.	Less Accuracy And Data lossy.
2	Text Summarizer Using NLP (Natural Language Processing), 2021	AAKASH SRIVASTAVA1 , KAMAL CHAUHAN2 , HIMANSHU DAHARWAL3 , NIKHIL MUKATI4 , PRANOTI SHRIKANT KAVIMANDAN5	access time for information search will be improved.	Large data cannot be process.
3	TEXT SUMMARIZATION USING NATURAL LANGUAGE PROCESSING AND GOOGLE TEXT TO SPEECH API,2020	SUBASH VOLETI1, CHAITAN RAJU1, TEJA RANI1 MUGADA SWETHA2	mainly focuses on providing a reliable summary.	Time constraint is less.
4	Automatic Text Summarization Approaches 2017	Ahmad T. Al-Taani	It converts the multiple documents into single document summarization.	It cannot determine/detect numeric value.
5	A Survey of Text Summarization Extractive Techniques,2010	Vishal Gupta, Gurpreet Singh Lehal	It uses the two approaches on text summarization are abstraction and extraction summarization	Less Accuracy

PROBLEM STATEMENT

In the new period, where tremendous measure of data is accessible on the Web, it is most vital to give the enhanced gadget to get data rapidly. It is extremely intense for individuals to physically pick the synopsis of expansive archives of content. So there is an issue of scanning for vital reports from the accessible archives and discovering essential data. Along these lines programmed content rundown is the need of great importance. Content rundown is the way toward recognizing the most vital important data in a record or set of related archives. What's more, compact them into a shorter rendition looking after its implications.

PROPOSED METHODOLOGY

For obtaining automatic text summarization, there are basically two major techniques i.e.-Abstraction based Text Summarization and Extraction based Text Summarization.

Extraction Based Summarization: The extractive summarization method involves selecting relevant words and sentences from a source document to create a summary. It is commonly used by search engines to generate summaries from web pages. Various mathematical formulations are used to score and select words and sentences, with the highest-scoring words being included in the summary. This approach is relatively straightforward to implement but faces challenges such as ranking, selection, and coherence.

The ranking problem involves determining the importance of words, while the selection problem involves choosing specific ranked units for inclusion in the summary. Coherence is crucial for creating understandable summaries. Several algorithms address the ranking problem, and efforts are made to enhance diversity, reduce redundancy, and select important sentences. Sentences are scored, ranked, and ordered, with the most important sentence appearing first in the summary.

Abstraction Based summarization: Abstractive summarization involves people reading content, understanding the topic, and creating their own concise summaries in their own words while retaining essential information. However, this is a challenging task for machines. Abstractive summarization aims to generate summaries using natural language processing, constructing new syntactically correct sentences. Unlike extractive methods, abstractive summarization requires a deeper semantic understanding of the text.

One major challenge in abstractive summarization is sentence combination, which can lead to inconsistency in the generated summaries due to the lack of a well-established framework. Models for abstractive summarization are typically trained on titles and captions and can be adapted for document-level summarization. Sentences are often reordered during the inference process.

CONCLUSION

- In conclusion, data summarization is a valuable tool for information extraction and condensation, addressing the information overload problem in the digital age. While challenges persist, ongoing research and technological advancements are driving the field forward, offering promising opportunities for more accurate, coherent, and context-aware summaries.

- As text summarization technology matures, it is expected to play an increasingly vital role in helping individuals and organizations extract actionable insights from vast amounts of textual data.

REFRENCES

- TEXT SUMMARIZATION USING NLP ChetanaVaragantham1, J.SrinijaReddy 2, UdayYelleni3 , MadhumithaKotha4 , Dr P.VenkateswaraRao5,2022[1].
- Text Summarizer Using NLP (Natural Language Processing), AAKASH SRIVASTAVA1, KAMAL CHAUHAN2, HIMANSHU DAHARWAL3, NIKHIL MUKATI4, PRANOTI SHRIKANT KAVIMANDAN5,2021[2].
- TEXT SUMMARIZATION USING NATURAL LANGUAGE PROCESSING AND GOOGLE TEXT TO SPEECH API, SUBASH VOLETI1, CHAITAN RAJU1, TEJA RANI1 MUGADA SWETHA,2020[3].
- Automatic Text Summarization Approaches Ahmad T. Al-Taani, 2017[4]
- A Survey of Text Summarization Extractive Techniques, Vishal Gupta, Gurpreet Singh Lehal 2010 [5],
- P. Janjanam and C. P. Reddy, "Text Summarization: An Essential Study," nternational Conference on Computational Intelligence in Data Science, 2019 [6].