Mega-Project Synopsis
On
"Tool To Summarize Texts"

A Mega – Project synopsis submitted in partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering

(Seventh Semester)
In
COMPUTER SCIENCE & ENGINEERING
Session 2023-2024

Prescribed By
Dr. Babasaheb Ambedkar Technological University, Lonere (DBATU)

Submitted By
Sneha Wankar(CSEC740)
Gudiya Prasad (CSEC711)
Achal Ragit (CSEC703)
Achal Waghmare (CSEC702)

Guided By
Prof. Anand Donald

Project Incharge
Dr. Manisha Pise

DEPARTMENT OF INFORMATION TECHONLOGY

RAJIV GANDHI COLLEGE OF ENGINEERING RESEARCH & TECHNOLOGY

Abstract

Text Summarization is the process of creating a condensed form of text document which maintains significant information and general meaning of source text. Automatic text summarization becomes an important way of finding relevant information precisely in large text in a short time with little efforts. Text summarization approaches are classified into two categories: extractive and abstractive. This paper presents the comprehensive survey of both the approaches in text summarization.

Automatic text summarization tool is basically summarizing the given paragraph using natural language processing and machine learning. There has been an explosion in the amount of text data from a variety of sources. This volume of text is an invaluable source of information and knowledge which needs to be effectively summarized to be useful. In this review, the main approaches to automatic text summarization are described. We review the different processes for summarization and describe the effectiveness and shortcomings of the different methods. Two types will be used i.e.- extractive approach and abstractive approach The basic idea behind summarization is finding the subset of the data which contains the information of all the sets. There is a great need to reduce unnecessary data. It is very difficult to summarize the document manually so there is a great need for automatic methods. Approaches have been proposed inspired by the application of deep learning methods for automatic machine translation, specifically by framing the problem of text summarization as a sequence-to-sequence learning problem.

Table of Content

1. Introduction

Text summarization is the technique for generating a concise and precise summary of voluminous texts while focusing on the sections that convey useful information, and without losing the overall meaning.

Automatic text summarization tool is the holy grail for people battling information overload, which becomes more and more acute over time. Hence it has attracted many researchers from diverse fields since the 1950s. However, it has remained a serious challenge, especially in the case of single news articles. Text summarization is the process of generating a short, fluent, and most importantly accurate summary of a respectively longer text document. The main idea behind automatic text summarization is to be able to find a short subset of the most essential information from the entire set and present it in a human-readable format. As online textual data grows, automatic text summarization methods have the potential to be very helpful because more useful information can be read in a short time. The method of extracting these summaries from the original huge text without losing vital information is called Text Summarization. It is essential for the summary to be fluent, continuous, and depict the significance.

2. Literature Survey

2.1 Introduction :

Literature surveys provide brief overviews or a summary of the current research on topics.Literature surveys are used in ensuring that the used experiments, methodologies and experiments offer reliability and validity in the research being conducted.They are useful in validating or providing proof and also provides a base of moving a research idea forward on what researchers have done and exciting avenues that it opens for investigation during future work in the field. In this literature survey research papers and articles on automatic text summarization were referred through for the project.

2.2 Literature Survey :

The following research papers and reports were used for the topic of Text Summarization System for English Language :

1. Automatic Text Summarization Using Natural Language Processing :

Authors Pratibha Devihosur, Naseer R implemented an automatic text summarization mechanism based on an unsupervised learning system.The significance of the generated summary was assessed with the assistance of Simplified Lesk calculation along with an online semantic lexicon WordNet.Based on their evaluation the algorithm provides best summarized outcome ranging from 25-50 percent with respect to the source data. In this project they also focussed on ambiguous words because a specific word may have distinctive significance in various setting.Hence they tried to inculcate the principle of word sense disambiguation to decide the right feeling of a word utilized as part of a specific setting.

2. Text Summarization using Natural Language Processing :

In this paper authors Ankit Kumar, Zixin Luo and Ming Xu created an end to end web application which can take an article as input and generate a summary.The model was trained using deep learning approach and trained on Juniper's datasets.Juniper is a corporate organization that develops and markets networking devices.In order to provide a better customer experience, Juniper Networks maintains large datasets of articles wherein each of these articles can be long and verbose.Hence these datasets were used to train the text summarization model.The model built used abstractive summarization technique and significantly generated excellent human readable sentences from given inputs.

However, it did not always generate summaries capturing all the important information in the input documents.

3. Text Summarization Techniques: A Brief Survey :
In this survey Mehdi Allahyari, Elizabeth D. Trippe, Saeid Safaei and others study the main approaches to automatic text summarization and also review the different processes for summarization and describe the effectiveness and shortcomings of different methods.Topics like the impact of context in summarization and semantic analysis are also mentioned.

4. NLP Based Text Summarization Using Semantic Analysis:
In this project paper authors Harsh Desai, Dhairya Pawar, Geet Agrawal reviewed the different methods for text summarization and provided a novel technique generating the summarization of domain specific text by using Semantic Analysis for text summarization.

5. Text Summarization:An Overview :
In this research paper author Samrat Babar provides an analysis about the meaning of text summarization in natural language processing and their types along with the technical and mathematical analysis of text summarization in detail. This paper basically is a documentation for all the information required to study as well refer through on the topic of automatic text summarization.

2.3 Literature Survey Summary :

The detailed summary of the above referred paper are mentioned in the table below :

SN	Techniques	Author & Year of Publication	Characteristics
1.	Automatic Text Summarization Using Natural Language Processing	Pratibha Devihosur , Naseer R.2017	1. Text summarization model based on unsupervised learning model. 2. Includes word sense disambiguation to determine the sentiment of the word in that particular setting.

			3. Generated summary ranges from 25-50 percent with respect to the source data.
3.	Text Summarization Techniques: A Brief Survey	Mehdi Allahyari, Elizabeth D. Trippe, Saeid Safaei and others.2017	1. Provides detailed description about the main approaches on automatic text summarization. 2. Reviews the different processes for summarization and also describes the effectiveness and shortcomings of different methods. 3. Also includes topics like the impact of context in summarization and semantic analysis.
4.	NLP Based Text Summarization Using Semantic Analysis	Harsh Desai, Dhairya Pawar, Geet Agrawal.2016	1. Mentions the different methods used in text summarization. 2. Creates a .novel technique generating the summarization of domain specific text by using Semantic Analysis.
5.	Text Summarization: An Overview	Samrat Babar.2013	**1.** Comparatively provides a very detailed report about the process of text summarization in natural language processing. **2.** The paper includes details about the meaning of summarization , their types and also includes the technical and mathematical analysis of the different approaches used.

3. Problem Statement And Objective

The objective of the project is to understand the concepts of natural language processing and create a tool for text summarization. The concern in automatic summarization is increasing broadly so the manual work is removed. The project concentrates on creating a tool that automatically summarizes the document.

1. Automatic text summarizing by providing top sentences with the highest score in the document to save time.
2. Helping users to provide a proper caption or description for the image.

4. System Requirements

Software-

Python IDE and its libraries,Jupyter-Notebook,Google Colab,VScode.

Hardware-

- Modern Operating System:
  1)Windows 7 or 10
  2) Mac OS X 10.11 or higher, 64-bit
  3) Linux: RHEL 6/7, 64-bit (almost all libraries also work in Ubuntu)
- 4 GB RAM
- x86 64-bit CPU (Intel / AMD architecture)

5. Modules to be developed

T5 is a transformer-based model that uses text-to-text transfer learning to perform various NLP tasks, including summarization. GPT-3 is a large-scale language model that employs deep learning and self-attention to produce natural and fluent texts.

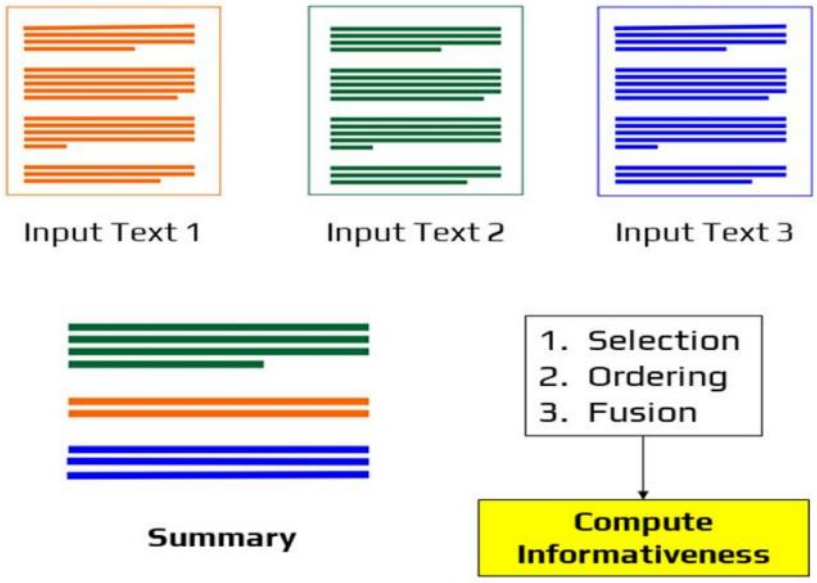

Fig. Module Architecture

6. Conclusion

Automatic text summarization is an old challenge but the current research direction diverts towards emerging trends in biomedicine, product review, education domains, emails, and blogs. This is due to the fact that there is information overload in these areas, especially on the World Wide Web. Automated summarization is an important area in (Natural Language Processing) research. It consists of automatically creating a summary of one or more texts. The purpose of extractive document summarization is to automatically select a number of indicative sentences, passages, or paragraphs from the original document. Text summarization approaches based on NLP have, to an extent, succeeded in making an effective summary of a document. Both extractive and abstractive methods have been researched. Most summarization techniques are based on extractive methods. As with time the internet is growing at a very fast rate and with it data and information are also increasing. It will be difficult for humans to summarize large amounts of data. Thus there is a need for automatic text summarization because of this huge amount of data.

7. References

1. E. Lloret and M. Palomar, "Text summarisation in progress: a literature review" in Springer, Springer, pp. 1-41, 2012.
2. U. Hahn and I. Mani, "of Automatic Researchers are investigating summarization tools and methods that", IEEE Computer33. 11, pp. 29-36, November 2000.
3.K. Spärck Jones, "Automatic summarising: The state of the art", Information Processing & Management, vol. 43, pp. 1449-1481, nov 2007.
4. V. Gupta and G. S. Lehal, "A Survey of Text Summarization Extractive techniques", Journal of Emerging Technologies in Web Intelligence, vol. 2, no. 3, pp. 258-268, 2010.
5. A. Khan and N. Salim, "A review on abstractive summarization methods", Journal of Theoretical and Applied Information Technology, vol. 59, no. 1, pp. 64-72, 2014.
6. R. Ferreira, L. De Souza Cabral, R. D. Lins, G. Pereira, E Silva, F. Freitas, et al., "Assessing sentence scoring techniques for extractive text summarization", Expert Systems with Applications, vol. 40, no. 14, pp. 5755-5764, 2013.