



Abstractive Summarization of Documents Leveraging BART Transformers and Streamlit Integration

Dr.S.Shargunam,
Assistant Professor,
dept. of Computer Science and
Engineering
Kalasalingam Academy of Research
and Education
krishnankoil, India.
s.shargunam@klu.ac.in

D Gangadhar Reddy
dept. of Computer Science &
Engineering
Kalasalingam Academy of Research
And Education
krishnankoil, India.
Email : 99220040834@klu.ac.in

Byreddy Venkata Sandeep Reddy
dept. of Computer Science &
Engineering
Kalasalingam Academy of Research
And Education
krishnankoil, India.
Email: b.v.sandeepreddy21@gmail.com

Ch Manikanta Raghava
dept. of Computer Science &
Engineering
Kalasalingam Academy of Research
And Education
krishnankoil, India.
Email : 99220040818@klu.ac.in

G Venkata Sai Vignesh
dept. of Computer Science &
Engineering
Kalasalingam Academy of Research
And Education
krishnankoil, India.
Email : vigneshsai378@gmail.com

Abstract—There is an urgent need for effective tools to summarize lengthy documents and extract key insights due to the rapid growth of digital information. A promising solution to this problem is abstractive summarization, which generates human-like summaries by interpreting and paraphrasing content. Abstractive summarization of lengthy text documents is made possible by this novel, user-friendly application that makes use of BART (Bidirectional and Auto-Regressive Transformers). The application, developed with Streamlit, makes it easy to perform on-demand summarization by letting users directly enter text or upload PDF files.

Long-form content summarization tasks are well-suited for our pre-trained model, which makes use of the facebook/bart-large-cnn architecture. To get around transformer-based models' input length constraints, we implement a chunking mechanism that breaks up large documents into sections and makes it possible to process them. A summary is created and each segment is tokenized for maximum coherence, completeness, and relevance. The application, which is hosted by Streamlit and can be used for research, education, and professional purposes, has a user-friendly interface. This tool is useful for a wide range of document types because the summarization outputs have been evaluated and demonstrate high fidelity and clarity. This study demonstrates how an accessible interface for large-scale document summarization and robust abstractive capabilities of BART can be combined.

I. INTRODUCTION

The need for sophisticated tools that are capable of efficiently and accurately summarizing large text volumes has been highlighted by the exponential growth of digital information, particularly in the form of lengthy documents. Since traditional extractive summarization methods primarily identify and extract sentences directly from the text without rephrasing or synthesizing content, they frequently fail to capture nuanced insights. A promising alternative, on the other hand, is abstractive summarization, which aims to produce concise, coherent summaries by interpreting and paraphrasing the original content. However, it is still difficult to create abstractive summarization tools that are scalable and easy to use, especially when dealing with a lot of input text that goes beyond typical model limits.

A novel application is presented in this paper that combines a user-friendly Streamlit interface with the BART (Bidirectional and Auto-Regressive Transformers) model for abstractive summarization. The application is accessible to users in a variety of fields who require quick and accurate document summarization because it lets users manually enter text or upload PDF documents. When dealing with a lot of textual input, the facebook/bart-large-cnn model is used because of its effectiveness in summarization tasks. This implementation uses a chunking mechanism to break up the text into manageable chunks, keeping each section within the model's maximum token limit because transformer-based models have limitations on input length. After that, each chunk is processed separately, and the summarized outputs are combined to create a cohesive summary at the end.

Utilizing BART's pretrained capabilities and Streamlit's interactive features, this system addresses the difficulties of processing and summarizing lengthy documents. The model produces concise, readable summaries that retain essential details while reducing verbosity by fine-tuning parameters like summary length and beam search configurations. Researchers, professionals, and educators who frequently work with extensive documentation can benefit from this paper's discussion of the architectural design, implementation process, and practical applications of this tool. Through this project, we show how abstractive summarization can make valuable information more manageable and accessible by simplifying content consumption.

II. LITERATURE SURVEY

This literature survey explains recent advances in abstractive summarization, an important area in NLP, that generates summary texts by rephrasing the text. The author categorizes techniques, highlights challenges, and discusses some future directions for this area. Tools and evaluation methods used are also discussed. Sources included in the literature survey are Elsevier, ACM, and IEEE. [1]

The field of text summarization has actually shifted from the extractive to the abstractive method, largely because of the

development in neural models that were first designed to be used in translation-related tasks. This neural capability allows for producing summaries with coherence, fluency, and closeness to a human's actual expression. Hui Lin and Vincent Ng outline recent advancements in neural-based approaches, where their impact in the growth of abstractive summarization is pointed out. [2]

While neural methods improve the fluency of abstractive summarization, they have difficulty choosing what's relevant. Sebastian Gehrmann, Yuntian Deng, and Alexander M. Rush introduce a content selector that is data efficient. The two-step method introduced focuses the model to ensure more accurate compression in its content. The system thus outperforms previous systems by achieving higher ROUGE scores on CNN-DM and NYT datasets given very limited training data and proves suitable for new domains. [3]

Most of the document summarization models do not support user preferences over summary length, style, or focus areas. Researchers presented a neural model that allows users to set these attributes, thus making summaries suitable to their needs. The adaptive model is also very good at being automatically selected, and with better F1-ROUGE1 scores and positive human feedback on the CNN-DailyMail dataset than state-of-the-art prior systems in terms of quality and customization. [4]

New Approach for Abstractive Summarization Yixin Liu, Pengfei Liu, Dragomir Radev, and Graham Neubig. Presenting a new training objective for abstractive summarization that allows probabilities for candidate summaries to be weighted according to their quality - the approach reaches state-of-the-art ROUGE-1 scores on CNN/DailyMail and XSum. [5]

Fei Liu et al. present an abstractive summarization framework based on Abstract Meaning Representation (AMR) graphs. The proposed method parses text into AMR graphs, transforms them into summary graphs, and generates text from these summaries, with promising results across various domains. [6]

Researchers present GSum, a guided summarization framework, that improves the coherence and faithfulness of neural models. GSum attains state-of-the-art ROUGE scores on various datasets by using different forms of external guidance, allowing for more controllable and reliable summary generation. [7]

Asli Celikyilmaz et al. come forward with a new architecture of encoder-decoder with deep communicating agents for the enhancement of abstractive summarization of long documents. In this approach, the encoding task is divided among several cooperative agents who encode parts of the input text and have access to the same decoder. The experimental results depict that this architecture yields much better quality summaries than traditional models using a single encoder or non-communicating encoders. [8]

Bogdan Gliwa et al. present the SAMSum Corpus, a dataset for abstractive dialogue summarization. Their results show

model-generated dialogue summaries have ROUGE scores that surpass news article summaries but human evaluation differences point to this area where tailored models and quality measures are necessary. This corpus constitutes a valuable resource that future work in the area will find useful. [9]

Arman Cohan et al. present a new model for abstractive summarization of longer documents, like research papers, using a hierarchical encoder and attentive decoder. Their experiments show significant improvements over existing state-of-the-art models on large datasets. [10]

III. PROJECT SCOPE

The project focuses on the development of an application in order to achieve abstractive summarization of long documents-an application which, in fact, may turn out to be pretty handy in research. This tool is supposed to produce abstractive, coherent, and meaningful summaries over a PDF uploaded by the user or directly entered text with the help of state-of-the-art models of neural networks-the Bidirectional and Auto-Regressive Transformers, that is, BART. This application is impressive in that it can get text from PDF documents and makes the whole process so smooth for users to get their documents summarised.

The application will use BART to produce abstractive summaries of good quality retaining fluency and informative content; the application will also work based on the specifications specified by the user regarding the length and style. The interface will be designed using Streamlit, with users able to upload files or input text manually. The resulting summaries will clearly be presented for easy reading. To accommodate the input size limitation of the model, the application will implement chunk processing, which will enable effective handling of longer texts without affecting the quality of summaries.

The target users of this application will be research people who want short summaries of long academic papers; students requiring effective study material to supplement their course text and lecture notes; professionals seeking rapid insights from the comprehensive reports or documentation. The app will support the following: uploading PDF files; a text input field to enter the information manually, and easy display of the created summaries within the main interface.

Results from the system are expected to be an application that will accurately and efficiently summarize long documents. Contributions to the NLP field will be brought by the system through practical provision of a tool in text summarization. Improvement may be in the way of more innovative user controls to customize summaries, addition of further NLP tasks like sentiment analysis or keyword extraction. It will gather feedback from the user and keep on developing constantly to satisfy the user demand and the new development regarding summarization techniques. In its entirety, this project aspires to meet the rapidly increasing needs for an effective document summarization process, in turn, facilitating smoother research and information handling among users.

IV. OBJECTIVES

The key intention of this project is the development of an application, which should be usable with users being able to better access information by abstracting into summary lengthy documents easily. The application takes in utilizing advanced methods of natural language processing: more precisely the BART model, called Bidirectional and Auto-Regressive Transformers, that make use of which will create short, coherent, meaningful summaries from uploaded PDFs as well as from any text input by the user.

This would be an important goal, where the text extraction from PDF documents will be very effective, allowing users to transform complex and detailed content into easily digestible summaries. The application is designed for longer texts using a chunking technique, processing inputs within the model's token limit, thereby ensuring the quality and relevance of the generated summaries.

This project will also provide options to customize, specifying the length and style of a summary, to fit in the requirements of a person. That would benefit students, researchers, and professionals requiring quick answers to long materials.

The expected outcome is a working application that can summarize long documents both accurately and efficiently and make a contribution to the area of NLP in providing a practical tool for summarizing text. Future work may be extending the advanced user controls for customizing summaries and examining possibilities of integration with other NLP tasks like sentiment analysis or keyword extraction, besides getting user feedback in order to iteratively improve the summarization with user needs and technological advances in summarization techniques. It is an effort towards answering the ever-increasing need for the effective summarization of documents, hence enhancing research and information processing efficiency for its users.

V. EXISTING METHOD

The text summarization methods are extractive and abstractive with tools and limitations currently. Extractive summarization involves directly extracting important phrases and sentences from a source text in order to create the summary. Traditional extractive summarizers rely on statistical and heuristic techniques, such as TF-IDF, whereby the weight of terms is computed based on ranking content. Graph-based ranking algorithms are one of the most popular types of extraction methods. Such algorithms constitute the backbone models like TextRank and LexRank. However, these generally lack flexibility in creating fluid coherent summaries due to piece by piece verbatim content copying from the document.

VI. PROPOSED METHODOLOGY

This is the proposed methodology where the BART transformer-based neural network model is integrated into Streamlit to create an interactive abstractive summarization application that can handle long documents. The interface is useful for uploading PDFs or manual inputting of text, which

can be passed to the backend summarization model to create good summaries. First, PDF content will be extracted from the uploaded ones using a library such as PyMuPDF for multi-page documents to handle it efficiently. Text content or text typed in will then be processed with the BART model for summary generation.

This has the input token limits so that the proposed system also integrates chunking, which is breaking the text down into chunks that can be fitted inside the processable form for processing with the model so that even documents of many pages may be efficiently managed. Now, each text block would then be fine-tuned over a BART model with a number of hyperparameters so that it was quite fluently and coherent in order to let improvements arrive at the quality of the summaries it could generate. Hyperparameters that are the length penalty will help in giving optimum improvements, and in return, every detail about the matter were passed while keeping clarity or a minimum amount of redundancy on the topic but at the same time kept it readable.

Streamlit is used by this methodology for a fluid, interactive interface that the end user can upload their file, view the extracted text, and generate a summary all within the same setting. This way, users can easily engage with the summarization process and get results in a timely manner. With handling PDF, summarization based on BART, and accessible interfaces for users, the system proposed here shows a very robust approach for abstractive summarization applicable for long documents that remains practical and adaptable in different types of user inputs.

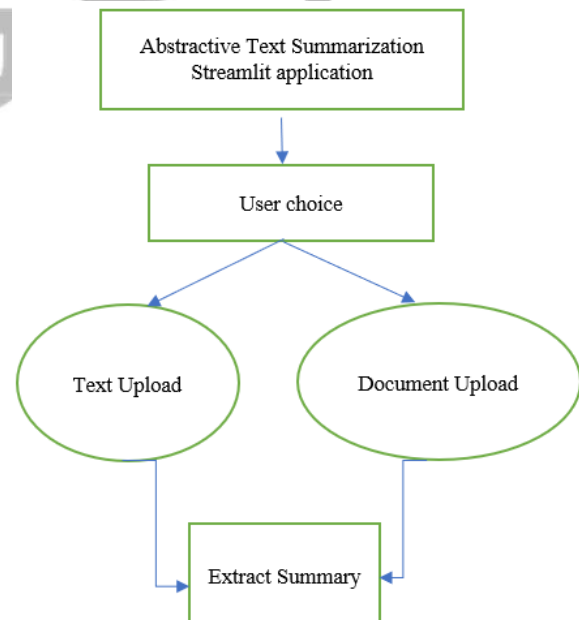


Fig 1: Flowchart Of Abstractive Text Summarization.

VII. RESULTS AND DISCUSSIONS

- 1) The user has the option to either upload a PDF file or manually enter text into the provided input area.

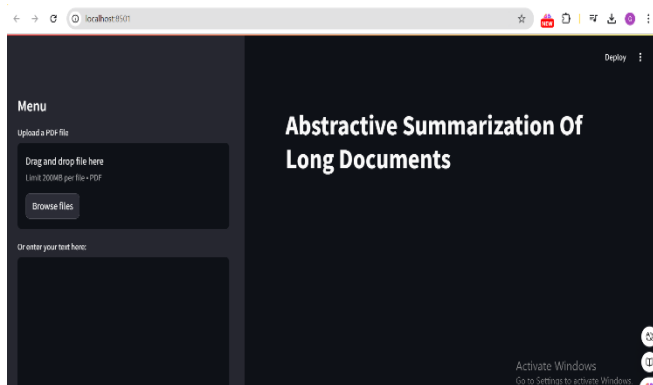


Fig 2

- 2) Extracted text from pdf

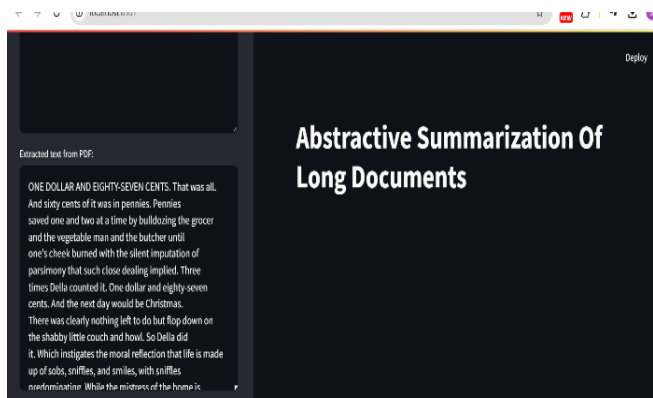


Fig 3

- 3) Extracted summary

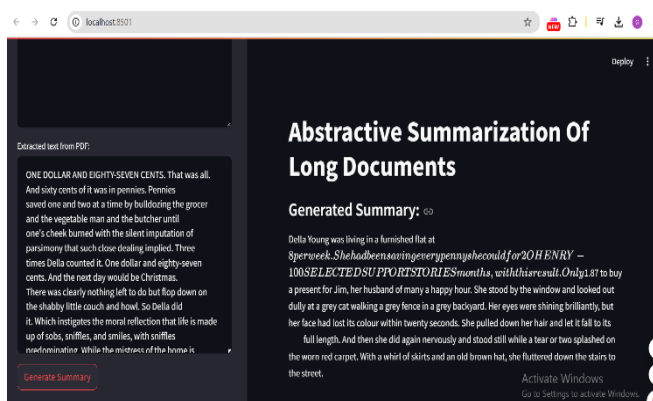


Fig 4

VIII. CONCLUSION

This project provides a very intuitive way of summarizing long texts by using natural language processing to make highly detailed content more accessible. The BART neural network model transforms complex information into clear, concise summaries for users to understand the basics of large documents. It allows users to upload PDF files or input text for summarization needs. This approach demonstrates how it's possible for neural networks to produce coherent summaries as well as address the difficulty that is usually posed when faced with long documents - they segment text and provide the summarization in levels. This project finally culminates into a successful instrument which could be used in order to summarize lengthy texts that pinpoint key information for subsequent progressions in text summarizing techniques.

REFERENCES

- [1] Som Gupta a, S. K Gupta b "Abstractive summarization: An overview of the state of the art" Expert Systems with Applications Volume 121, 1 May 2019, Pages 49-65
- [2] Lin, H., & Ng, V. (2019). Abstractive Summarization: A Survey of the State of the Art. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 9815-9822. <https://doi.org/10.1609/aaai.v33i01.33019815>
- [3] Sebastian Gehrmann, Yuntian Deng, Alexander M. Rush "Sebastian Gehrmann, Yuntian Deng, Alexander M. Rush" arXiv:1808.10792 [cs.CL]
- [4] Angela Fan, David Grangier, Michael Auli "Controllable Abstractive Summarization" arXiv:1711.05217 [cs.CL]
- [5] Yixin Liu, Pengfei Liu, Dragomir Radev, Graham Neubig "BRIO: Bringing Order to Abstractive Summarization" arXiv:2203.16804 [cs.CL]
- [6] Fei Liu, Jeffrey Flanigan, Sam Thomson, Norman Sadeh, Noah A. Smith "Toward Abstractive Summarization Using Semantic Representations" arXiv:1805.10399 [cs.CL]
- [7] Zi-Yi Dou, Pengfei Liu, Hiroaki Hayashi, Zhengbao Jiang, Graham Neubig "GSum: A General Framework for Guided Neural Abstractive Summarization" arXiv:2010.08014 [cs.CL]
- [8] Asli Celikyilmaz, Antoine Bosselut, Xiaodong He, Yejin Choi "Deep Communicating Agents for Abstractive Summarization" [Submitted on 27 Mar 2018 (v1), last revised 15 Aug 2018 (this version, v3)] arXiv:1803.10357 [cs.CL]
- [9] Bogdan Gliwa, Iwona Mochol, Maciej Biesek, Aleksander Wawer "SAMSum Corpus: A Human-annotated Dialogue Dataset for Abstractive Summarization" [Submitted on 27 Nov 2019 (v1), last revised 29 Nov 2019 (this version, v2)] arXiv:1911.12237 [cs.CL]
- [10] Arman Cohan, Franck Dernoncourt, Doo Soon Kim, Trung Bui, Seokhwan Kim, Walter Chang, Nazli Goharian "A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents" arXiv:1804.05685 [cs.CL]