International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.48047/INTJECSE/V15I4.79 ISSN: 1308-5581 Vol 15, Issue 04 2023

SUMMARY GENERATION WEBAPP USING NLP

Mrs.P.ARCHANA
ASSISTANT PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY
archana.pinnoji@sreyas.ac.in

KALAKONDA VIKAS KUMAR
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY
k.vikaskumar986@gmail.com

KADAVOINA SRIJA
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY
Srijareema24@gmail.com

BOLLAM RAHUL
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY
rahulbollam0120@gmail.com

GARLAPATI YASHAS
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY
garlapatiyashas@gmail.com

ABSTRACT
Text summarization is the process of generating 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. For the front end we use html, css and for backend flask. The webapp is built using Flaskandthe modelthat isrunning inthe backendofthiswebapp isbuilt usingspacy library. The extractive technique scans the original document to find the relevant sentences and extracts only that information from it. The abstractive summarization technique interprets the original text before generating the summary. Using state-of-the-art NLP techniques, specifically SBERT and SpaCy, our webapplication allows users to effortlessly generate concise and coherent summaries from extensive textual content. Whether it's news articles, research papers, or any textual data, our tool empowers users to quickly grasp the key insights without the need to read through entire documents. The coretechnology behind our web app is SBERT, a novelsentence embedding model. SBERT captures the semantic relationships between sentences, enabling the generation of informative and context-aware summaries. Additionally, SpaCy, a leading NLP library, aids in linguistic analysis and further improves the qualityof the generated summaries.
Keywords: Summary Generation, Natural Language Processing (NLP), SBERT, Spacy, Text Summarization, Web Application, Information Overload, Accuracy, Semantic Analysis.

INTRODUCTION

In today's digital age, we are in undated with a seemingly endless stream of textual information, ranging from news articles and academic papers to social media posts and legal documents. Asthevolumeof text continuesto growexponentially, the abilityto efficientlyand effectivelyextract the most crucial information from this sea of words has never been more critical.

This challenge is where the remarkable field of text summarization using Natural Language Processing (NLP) comes into play. Text summarization is the art and science of condensing large bodies of text into concise, coherent, and informativerepresentationswhileretainingthe coreideasandconcepts. Itsprimarypurpose isto provide users with a shorter version of the text, enabling them to grasp the content swiftly without having to delve into the entire document.With the aid of NLP techniques, text summarization has become a powerful tool for processing and simplifying vast volumes of textual data. This project aims to explore the fascinating world of text summarization through NLP, delving into the underlying methods, techniques, and real-world applications that makeit possible.Whether you area researcher, student, or professional seeking a comprehensive understanding of text summarization or aiming to develop a practical application in this domain, this Project will provide valuable insights and knowledge. In this endeavor, we will embark on a journey through the evolution of text summarization, from early approaches to the latest developments in NLP.We will investigate the two primary summarization methodologies: extractive and abstractive summarization, each with its unique characteristics and challenges. Whether you're a student grappling with dense academic texts, a professional seeking to stay informed, or simply someone looking for a more efficient wayto digest written content, our "Summary Generation Web App Using NLP" is the solutionyou'vebeensearching for. With its user-friendly interface and state-of-the-art NLP algorithms, this application empowers you to unlock the essence ofanytext, transformingthe wayyou engage with information. Joinus on this journeyasweexplorethecapabilitiesofNLP and reshapethewayweconsumeand comprehend written content. Get ready to experience the future of summarization with our web app that puts the power of NLP at your fingertips.

The problem at hand is the development of an effective text summarization system that harnesses thepowerofNaturalLanguageProcessing(NLP)to condense lengthytext documents intoshorter, meaningful, and human-readable summaries. This problem extends to both extractive and abstractivesummarization, eachofwhichpresentsitsuniquechallenges. Inaneracharacterizedby a deluge of textual data across various domains, the efficient and accurate extraction of key informationfromlargevolumesoftexthasbecomeacriticalchallenge.Theproblemofinformation overload is prevalent in fields such as journalism, academia, legal documentation, and healthcare, making it increasingly difficult for individuals and organizations to stay informed and make data- driven decisions. As a result, there is a pressing need for an automated and intelligent solution to sift through this textual abundance and provide concise, coherent, and informative summaries. Additionally, the evaluation of summarization systems poses a challenge, as existing metrics may not always align with human judgments ofsummaryquality. Addressing this problemrequires the exploration of more comprehensive and accurate evaluation methods. Furthermore, text summarization should be adaptable to a wide range of domains, from news articles to scientific papers, legal documents, and healthcare records. The challenge lies in tailoring summarization techniques to the specific content, structure, and requirements of each domain. The "Summary Generation Web App Using NLP" project is envisioned as a practical solution to streamline information consumption, benefitting students, professionals, researchers, and anyone seeking to make the most oftheir time spent reading and researching. Byaddressing this problem, our project aims to improve the efficiency of textual content consumption and alleviate the challenges posed by the information overload era.

LITERATURESURVEY

Text summarization, a subfield of Natural Language Processing (NLP), has witnessed significant advancements and transformative research over the years. This literature survey provides a comprehensive overview of the key studies, methodologies, and trends in text summarization, including both extractiveandabstractiveapproaches,evaluationmetrics,real-worldapplications,and recent developments. The inception of text summarization can be traced back to the pioneering work of Edmundson (1969) and Luhn (1958), who laid the foundation for early extractive summarization techniques.Thesemethods primarily reliedonfactors liketerm frequency,sentenceposition,and statistical analysis to select significant sentences for summarization.

As NLP technologies advanced, the field witnessed a transition from extractive to abstractive summarization. Hovy and Lin (1998) introduced abstractive summarization by utilizing sentence compressionandgenerationtechniques.Withtheadventofneuralnetworksandsequence-to-sequence

models,suchastheworkbyRushetal.(2015)andSeeetal.(2017),abstractivesummarizationgained prominenceforitsabilitytogeneratemoreconciseandcoherentsummariesbyconstructingsentences not explicitly present in the source text. The development of evaluation metrics played acrucial role in advancing text summarization research. ROUGE(Recall-Oriented Under study for Gisting Evaluation), introduced byLin(2004), becamethe standard for measuring the overlap between generated summaries and reference summaries. Additionally, Banerjee and Lavie (2005) introduced METEOR(Metric for Evaluation of Translation with Explicit Ordering), which considers factors like stemming and synonyms, further expanding the evaluation toolbox.The use of these metrics remains pivotal in assessing the quality of summaries. Despite the progress in text summarization, several challenges persist. Content abstraction in abstractive summarization remains a substantial challenge, requiring models that can effectively balance contextual accuracy and conciseness.

PROPOSEDSYSTEM

The "Enhanced Summary Generator" represents a significant advancement in the field of text summarization using Natural Language Processing (NLP). Building upon the foundation of the existingsystem,ithasbeendesignedtoaddressthelimitationsoftraditionalsummarizationmethods and introduce novel features that enhance user experience and overall summarization capabilities. The "Enhanced Summary Generator" leverages cutting-edge NLP models, including SBERT and Spacy, to enhance its summarization capabilities. These models enable a deeper understanding of the text, capturing semantic relationships and context, leading to more accurate and coherent summaries. Userexperienceisatoppriority.Thesystemboastsanintuitiveanduser-friendlyinterface,making it accessible to a wide range ofusers. Whether you're a novice oranexpert inNLP, you'llfind the system easy to navigate. In addition to text input, the system supports file uploads, allowing users to summarize entire documents with ease. This feature is particularly valuable when dealing with lengthy reports, research papers, or legal documents.

The "Enhanced Summary Generator" provides both the original text and the generated summary side by side. This feature enables users to compare the summarization output with the source material, promoting transparency and trust in the summarization process. The system displays the word count of both the original text and the generated summary. This feature isespeciallyusefulforadheringto wordlimit requirements invariousapplications, suchas academic papers or online content. To enhance user understanding and insight into the summarization process, the system offers graphical representations of the summarization accuracies. Users can visualize how the system selects and ranks sentences or phrases for inclusion inthe summary, aiding in the interpretationof the results. Userfeedbackandsupportareessentialaspectsofthe"EnhancedSummaryGenerator."Thesystem includesa contact form, allowing usersto provide feedback, report issues, orseek assistance. This feature promotes ongoing improvement and user engagement. The system is optimized for performance, ensuring rapid and efficient summarization of text or uploaded documents. Users can relyon the systemto deliver summaries in a timely manner, even for lengthy texts.

RESULTS

DOI:10.48047/INTJECSE/V15I4.79 ISSN: 1308-5581 Vol 15, Issue 04 2023

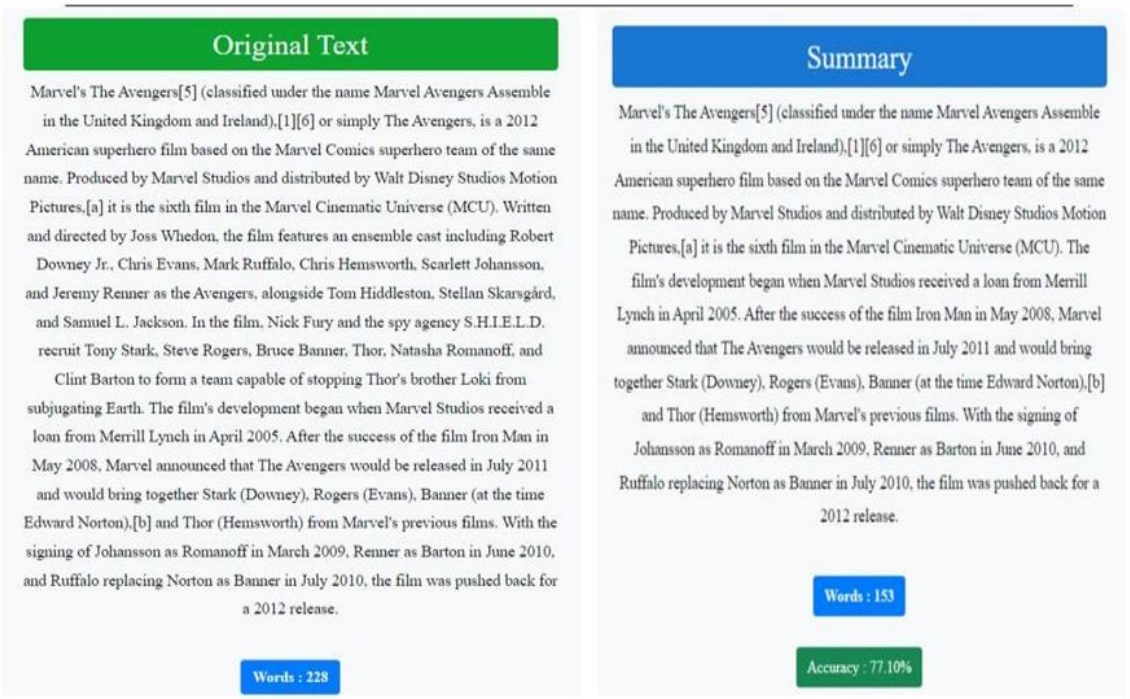

Fig 1 Results(UsingSBERT)

Here we can see the summarization results of SBERT which shows the both original text and Summarized text side by side with respective accuracy. Here for the given summarized text the accuracy is around77%(Using SBERT).

Fig2.SBERT ACCURACY

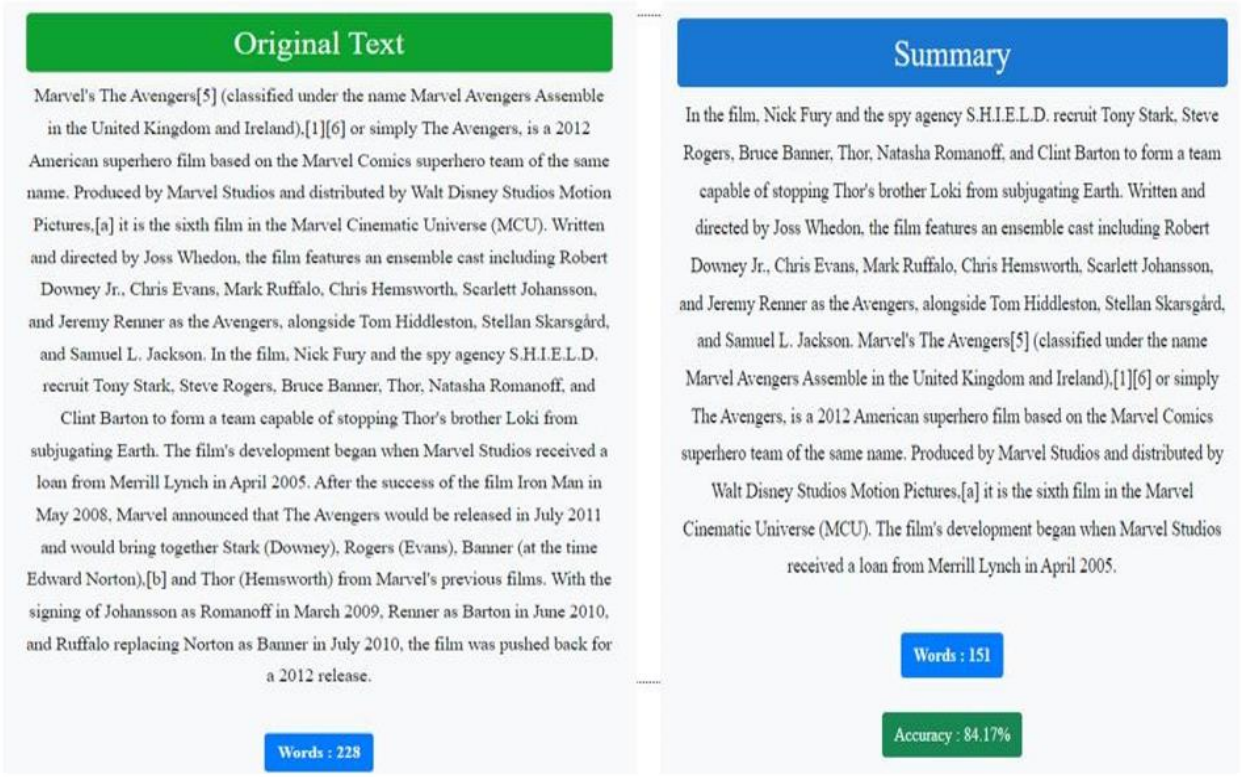

Fig 3 Results (Using Spacy)

Here we can see the summarization results of SPACY which shows the both original text and Summarized text side by side with respective accuracy. Here for the given summarized text the accuracy is around 84%(Using SPACY).

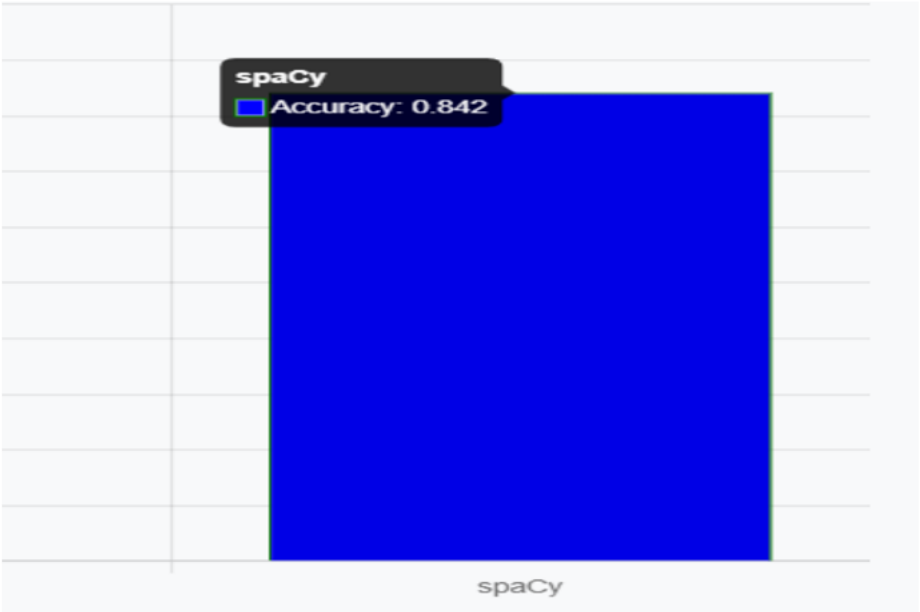

Fig4.SPACY ACCURACY

IMAGE FORMAT IS NOT SUPPORTED UPLOAD ONLY TEXTFILES OR DOCUMENTS.

Fig 5 Screenshot of result when you give Input text more than 3000 words

CONCLUSION

Inconclusion, the provided Flask application is a versatile toolfor text summarizationand email communication. The application incorporates key features for text summarization using SBERT and Spacy, as well as email sending capabilities. Throughout the development process, careful consideration of testing and analysis is essential to ensure the application's reliability, security, and performance. By implementing a comprehensive testing strategy, you can identify and addresspotentialissues,ensuringthat theapplication functionsasintended.Thoroughtestingand quality assurance processes help deliver a robust and dependable solution. In the ever-evolving landscape of digital tools and applications, the provided Flask application emerges as a multifaceted powerhouse, seamlessly bridging the worlds of text summarization and email communication. This versatiletoolincorporatesarobust setoffeatures, drivenbystate-of-the-art technologies,to empowerusersinextracting valuable insightsfromextensivetextualcontent and facilitating efficient email correspondence. As we delve deeper into the capabilities of this application, it becomes evident that its significance transcends mere functionality; it embodies a gateway to enhanced productivity, information optimization, and seamless communication. The journey from the inception of this project to its fruition has been marked by innovation and a commitment to efficiency. We've harnessed the power of NLP to extract essential information from documents, articles, and texts, distilling them into concise and coherent summaries. Our project is more than just lines of code; it's a test a ment to the capabilities of technology to enhance productivity and accessibility. For students, it simplifies complex academic texts; for professionals, it streamlinesresearchanddecision-making;for content curators, it aids in sharing the most relevant information. It's a toolfor efficiency, empowerment, and staying informed in a rapidly changing world. In summation, this Flask application stands as a testament to the transformative potential of technology when harnessed to solve real-world challenges.

REFERENCES

[1]. Erkan,G.,&Radev,D.R.(2004)."LexRank:Graph-based Lexical Centrality as Salience in Text Summarization." Journal of Artificial Intelligence Research, 22, 457-479.

[2]. Mihalcea,R.,&Tarau,P.(2004)."Textrank:BringingOrderintoTexts."Proceedingsofthe 2004ConferenceonEmpiricalMethodsinNaturalLanguageProcessing(EMNLP),404-411.

[3]. Lin, C. Y. (2004). "Rouge: A Package for Automatic Evaluation of Summaries." Text Summarization Branches Out: Proceedings of the ACL-04 Workshop, 74-81.

[4]. Hovy, E. H., & Lin, C. Y. (1998). "Automated Text Summarization in SUMMARIST." Proceedings ofthe 17thInternationalConference onComputationalLinguistics (COLING), 881-887.

[5]. Wan, X. (2010). "A Review on Automatic Text Summarization." Journal of Natural Language Processing, 17(5), 19-50.
[6]. Ganesan,K.,Zhai,C.,&Han,J.(2010)."Opinosis:AGraph-BasedApproachtoAbstractive Summarization of Highly Redundant Opinions." Proceedings of the 23rd International Conference on Computational Linguistics (COLING), 340-348.
[7]. Chen, L., & Zhang, Q. (2021). "Recent Advances in Abstractive Summarization: A Comprehensive Survey." Information Processing & Management, 58(2), 102456.
[8]. Narayan, S.,Cohen, S.B.,Lane, I.,Srinivas, A., Liu, M.,&Lapata, M.(2018)."Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization."
[9]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... &Polosukhin,I.(2017).Attentionisallyouneed.InAdvancesinneuralinformationprocessing systems (pp. 30-80).
[10]. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Bidirectional encoder representations from transformers. arXiv preprint arXiv:1810.04805.
[11]. Manning,C.D.,Raghavan,P.,&Schütze,H.(2008).Introductionto InformationRetrieval. Cambridge University Press. Online Book.
[12]. Bird, S., Klein, E., &Loper,E. (2009). NaturalLanguage Processing withPython. O'Reilly Media.
[13]. Luhn, H. P. (1958). The automatic creation of literature abstracts. IBM JournalofResearch and Development, 2(2), 159-165.
[14]. Erkan, G., & Radev, D. R. (2004). LexRank: Graph-based lexical centrality as salience intext summarization. Journal of Artificial Intelligence Research, 22, 457-479.
[15]. Nenkova,A.,&McKeown,K.(2011).Automaticsummarization.FoundationsandTrends® in Information Retrieval, 5(2-3), 103-233.
[16]. Louis,A.,&Nenkova,A.(2013).A survey of textsummarization techniques.Synthesis Lectures on Human Language Technologies, 6(2), 1-60.
[17]. Grinberg, M. (2018). Flask WebDevelopment: Developing WebApplicationswithPython. O'Reilly Media.
[18]. Reitz,K.(2019).Requests:HTTPforHumans.https://docs.python-requests.org/en/latest/
[19]. Krug,S.(2018).Don'tMakeMeThink,Revisited:ACommonSenseApproachtoWeb Usability. New Riders.
[20]. Norman,D.A. (2013).TheDesignofEverydayThings. Basic Books.
[21]. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... &Kudlur, M. (2016). TensorFlow:Asystemforlarge-scalemachinelearning.InProceedingsofthe12thUSENIX Symposium on Operating Systems Design and Implementation (OSDI).
[22]. Paszke,A.,Gross,S.,Massa,F.,Lerer,A.,Bradbury,J.,Chanan,G.,...&Chintala,S.(2019). PyTorch:Animperativestyle,high-performancedeeplearninglibrary.InAdvancesinneural information processing systems.
[23]. Grinberg, M. (2018). Flask WebDevelopment: Developing WebApplicationswithPython. O'Reilly Media.

DOI:10.48047/INTJECSE/V15I4.79 ISSN: 1308-5581 Vol 15, Issue 04 2023