IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11507

"Summer-Riser: A novel abstractive based text summarization tool"

Nethranand P S[1], Shruthi K[2], Pavan Kalyan N[3], Kandra Akash[4]

[1, 2, 3, 4] Student, Artificial Intelligence and Machine Learning, K.S. Institute of Technology, Bengaluru, India

Abstract: This paper proposes the implementation of a system called Summer-Riser, which is an automated text summarization bot using Natural Language Processing (NLP) approaches. This paper provides a complete explanation of how the system works and gives responses to its use. The bot works based on taking the URL as user input, extracting the text, preprocessing the text, and providing the answer to the question asked by the user by summarizing the text. The paper will help both commoners and technically skilled people and researchers to understand how the NLP approaches can be used to develop a text summarization model. The authors have provided an in-depth explanation regarding the models like langchain etc. used to develop the system and various current trends in the field of NLP.

Keywords: Text Summarization, Natural Language Processing (NLP) and langchain

INTRODUCTION

In the realm of natural language processing, condensing extensive texts remains a subject of ongoing exploration. Text summarization serves as a tool for software to distill a lengthy document into a concise overview or synopsis, accentuating its crucial elements. Such summaries vary based on the input type, ranging from simple models tailored for individual documents to more intricate approaches for multiple documents, posing greater complexity. Summarizers are categorized into general or domain-specific models; the latter processes input impartially, leveraging domain knowledge to craft precise summaries grounded in facts. Operating on a query-based mechanism, domain-specific models furnish summaries that exclusively contain answers to pertinent questions about the input text.

Output types further segment summarizers into two categories: Extracted summaries, which select key phrases from the input text to generate summaries akin to human-generated ones, and abstract summaries, a more intricate task than extraction, notwithstanding recent strides in neural network technologies propelled by neural machine translation and sequence models. Despite advancements, summarizers still lag behind human-level proficiency. Applications of text summarization encompass media monitoring, search engine marketing, internal document workflows, financial research and marketing, social media marketing, and aiding individuals with disabilities.

Abstractive summarization, a method for condensing lengthy texts to extract their core meanings, holds promise for various practical applications. However, current research primarily focuses on testing models with different types of data, resulting in only incremental improvements and limited practical implementation. Notably, abstractive summarization remains underutilized in social media research, where it could enhance opinion and topic mining, addressing the complexities unique to social media data analysis. Reddit serves as a common platform for evaluating new neural text summarization models, predominantly in English and on large datasets, neglecting to test on smaller real-world data or across different languages and platforms.

The recent surge in online and offline data has significantly elevated the status of natural language processing. Extractive text summarization, a technique that selects sentences from a corpus to generate concise summaries, has gained prominence. However, existing methods often overlook contextual information and inter-sentence relations in feature engineering. To address this limitation, Tis papere proposes the use of clustered Transformer models for text summarization, aiming to enhance contextual representation and capture linguistic nuances. Our framework not only leverages surface features to better understand word and sentence elements but also incorporates a hierarchical attention mechanism to capture contextual relations across word and sentence levels. Furthermore, the introduction of clustering post-transformer modeling will help to identify the most relevant sentences, thus improving the quality of extractive text summarization.

RELATED WORKS

This chapter focuses on recent developments in area of NLP. The first study focuses on generative text summarization by fusing Multidimensional Semantic Information using Abstractive Text Summarization (ATS) approach via Transformers Model [1]. This paper proposes a complete overview regarding automatic text summarization for smoothing out message content into fundamental data towards speedy understanding of pivotal data.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11507

The next study focuses on issues of the seq2seq model and proposes a programmed text strategy for consolidating multi-faceted semantic data. The model depends on the pointer age organization, that extricates worldwide semantic data and removes nearby semantic data with neighbourhood convolution extractor [2].
The next study [3] captures automated analysis of electronic text. This study highlights text rundown, language interpretation, feeling classifier and title age of news stories as a portion of the well-known methods of text handling. The act of changing composed text from one language into another was effectively perceived using language interpretation.
The next study performs automatic Thai text summarization using keyword-based abstractive approach [4]. This paper explains the preparation period of abstractive text rundown by contributing two arrangements of number groupings. The first set, is the source text, and the second set is the reference summary's words. The model encodes and decodes these parts effectively.
Another study focuses on medical reports summarization by using Text-To-Text transformer [5] using deep learning and large language models (LLM). Summarization of medical reports is done by the model in this study. The results achieved from the study were encouraging.
The next study [6] describes the same concepts of LLM which was discussed in previous studies.
The next study [7] explains the previous study on text summarization and the results are satisfactory.
The next study [8] uses the NLP-based text representation approaches for supporting Requirements Engineering (RE) tasks. These tasks are automated using NLP approaches towards language content extraction, and transformations, for handling feature lists or vector representations based on embeddings. These NLP approaches are typically utilized as inputs for rule-based or ML algorithms.
The subsequent study [9] also describes about the NLP approaches which was described previously and obtaining satisfactory results.
The preceding study [10] describes about the evaluation techniques which was discussed in deep in [13].
The next study [11] explains the traditional text summarization methods and Hand Written Answer Evaluation (HAES) methods which was explained more clearly in [14].
Another study focused on text summarization using the transformer Model [12]. This paper gives a detailed information on text summarization using online comments and review mechanisms. Text summarization methods have been applied here to news and scientific articles. A text summarizing technique based on the Text-to-Text Transfer Transformer (T5) model was designed in this work.
The next study provides an overview of AutoEval – an Automatic Test Evaluation System [13], which is a computerized system for automatically evaluating exam papers. The evaluator's mood fluctuations and the rapport between the student and the evaluator are the two aspects that are assessed by the model effectively.
The next study [14] highlights on a common method that is used to assess student achievement in the education sector with reference to written test. The human labour required for the assessment is quite significant, which is dependent on number of variables, including teacher's expertise, application-level comprehension, grading criteria, and time allocated. The design and usage of Handwritten Answer Evaluation (HAES) method for student exam papers are presented in this study.
Yet another study [15] focuses on extractive and abstractive text synopsis towards scientific categorization of ATS (Automated Text Summarization) space. This study explores the text synopsis approaches that reflects on potential recuperation strategies, including element extraction, execution estimation procedures, along with the complications in ATS space. This paper concisely proposes past, present, and future examination in ATS space.
This section is concluded by Table-1, which highlights contribution of researchers in the field of NLP.

METHODOLOGY

1. **Langchain**: The use of langchain helps in better building text summarization applications as it has tools to interact with large language models (LLMs) effectively. LangChain offers a cutting-edge framework that enhances the creation of text summarization applications by effectively engaging with large language models (LLMs). This innovative tool provides a diverse set of approaches, such as the map-reduce method, which excels in handling lengthy texts. In the map-reduce process, LangChain intelligently dissects the text into smaller, manageable sections, summarizes each segment using the LLM, and seamlessly merges these summaries into a comprehensive and concise overview. This streamlined approach not only simplifies the summarization process but also optimizes the output by leveraging the full potential of language models

2. **Data Collection**: This section details the implementation of the data collection process in LangChain, which involves inputting news articles or other article URLs via a user-friendly web interface. The UnstructuredURLLoader component is a key feature of LangChain that facilitates the retrieval of data from provided URLs. This component is part of a larger suite of document loaders within LangChain, each designed to handle specific types of data sources.

DOI: 10.17148/IARJSET.2024.11507

3. **Data Preprocessing**: Here, This paper describe the text preprocessing steps involved in preparing the retrieved data for further analysis. The preprocessing stage includes segmenting articles into manageable chunks using the RecursiveCharacterTextSplitter. This component is responsible for dividing the retrieved data into smaller, more manageable chunks to facilitate further analysis as follows:

   A. tokenization.text_splitter = RecursiveCharacterTextSplitter(...) sets up a text splitter object, configured to:
   B. Split text based on newlines, periods, and commas.
   Limit each chunk to a maximum size of 1000 characters

4. **Embedding Generation**: In this subsection, This paper will discuss the process of generating embeddings from text data using Gemini's language model in LangChain. First, This paper will explain the concept of embeddings and their role in representing textual information. Embeddings are vector representations of text data that capture semantic meaning and context. They are used to represent text in a numerical format that can be processed by machine learning algorithms. Next, This paper will discuss the technical aspects of creating and storing embeddings efficiently. This will include a description of the embedding generation process using Gemini's language model, as well as best practices for storing and managing embeddings to optimize performance and minimize resource usage. Finally, This paperwill provide examples and use cases to illustrate the practical applications of embeddings in LangChain, including text classification, clustering, and recommendation systems.

5. **Embedding and Vectorstore Creation**: In the process of creating embeddings and a Vectorstore in LangChain, the code snippet embeddings = OpenAIEmbeddings() initializes an object for generating gemini's embeddings, which are numerical representations of text. Subsequently, vectorstore_openai = FAISS.from_documents(docs, embeddings) is used to construct the Vectorstore by utilizing the FAISS library to index the embeddings generated from the provided documents. This approach efficiently organizes and indexes the embeddings for quick retrieval and similarity search operations within LangChain's text processing framework.

   A. embeddings = OpenAIEmbeddings() creates an object for generating Gemini's embeddings (numerical representations of text).
   B. -vectorstore_openai = FAISS.from_documents(docs, embeddings) constructs

6. **FAISS vectorstore**: In LangChain's text processing framework, the FAISS vectorstore is used for efficiently storing and retrieving embeddings based on similarity. To create the vectorstore, the following steps are taken:

   1. `main_placeholder.text("Embedding Vector Started Building...✅✅✅")` displays a status message indicating that the embedding vector creation process has begun.
   2. `time.sleep(2)` is used to add a 2-second delay, which can be optional and is used for visual feedback or to comply with API rate limits.

      A. By using the FAISS vectorstore, LangChain can efficiently store and retrieve embeddings based on similarity, which is essential for various text processing tasks such as semantic search, clustering, and recommendation systems.

7. **Saving FAISS Index:** In LangChain's text processing framework, the FAISS vectorstore is saved for future use by following these steps:

   1. `with open(file_path, "w") as f:` opens a file in write mode for saving the FAISS vectorstore.
   2. `json.dump(vectorstore\_openai.as\_json(), f)` saves the FAISS vectorstore to the JSON file.

      A. Additionally, a text input field is created for users to enter their question about the loaded news articles:

   1. `query = main_placeholder.text_input("Question: ")` creates a text input field labeled "Question:".
      B. By saving the FAISS vectorstore as a JSON file and providing a user query input, LangChain enables efficient and effective text processing and retrieval for a wide range of applications.

RESULT

In this section, This paper will evaluate the performance of Summer Riser using various datasets and methodologies. This paper cover the following:

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁑ Peer-reviewed & Refereed journal ⁑ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11507

1. **Datasets and Methodologies:** This paper will describe the datasets used for the evaluation and the methodologies employed to assess the experiments. This includes the selection of diverse corpora spanning a wide range of topics to ensure comprehensive coverage.

2. **Experimentation Process:** This paper will outline the experimentation process, which involves refining parameters for optimal performance. This includes adjusting hyperparameters and testing different configurations to ensure the best possible results.

3. **Comparison with Other Works:** Finally, This paper will compare our results with those of other similar works to provide context and demonstrate the effectiveness of Summer Riser. This will help establish the strengths and weaknesses of our approach and identify areas for future improvement.

By evaluating Summer Riser using diverse corpora and methodologies, This papercan ensure that it performs well in a variety of scenarios and is competitive with other similar solutions.

Qualitative evaluation:

In this evaluation, our objective is to measure user satisfaction with the generated summaries. For this purpose, This papercarried out a qualitative evaluation by inviting ten English-speaking individuals to rate our summaries. This paperadopted the same qualitative evaluation method outlined in [38]. To illustrate, while a 3-level scale might include the categories "low," "medium," and "high," a 5-level Likert scale provides varying degrees to gauge agreement on a specific matter, ranging from "strongly agree," "agree," "neither agree nor disagree," "disagree," to "strongly disagree." Specifically, the asked questions are:

- Q1: The summary reflects the most important issues of the document.
- Q2: The summary allows the reader to know what the article is about.
- Q3: After reading the original summary provided with the document, the alternative summary is also valid.

Given the diverse lengths of the documents, our evaluation approach focused on utilizing 10 randomly selected documents from each of the tested datasets.

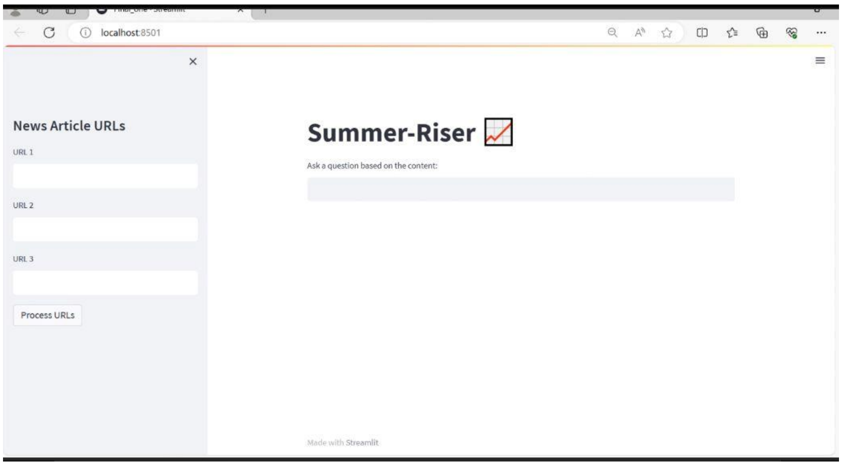

Fig 1: Homepage of text summarization tool

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11507

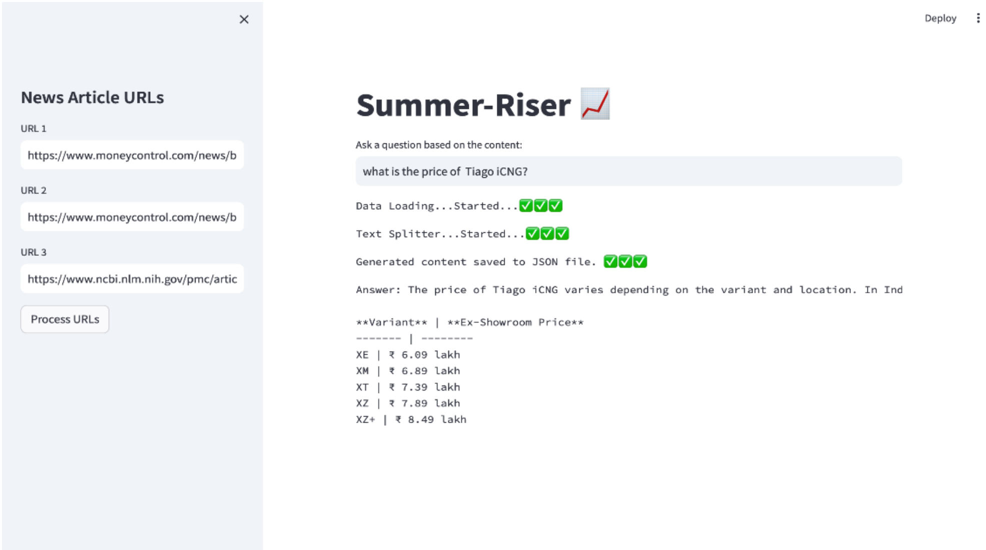

Fig 2: Answer obtained when text summarization is done

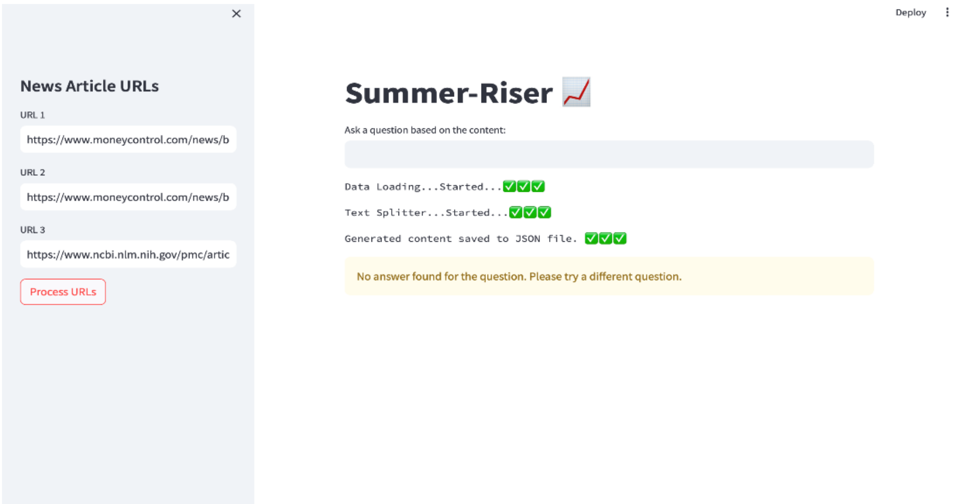

Fig 3: Warning obtained when a wrong question is asked

CONCLUSION

In conclusion, Automatic text summarization, a subject of academic inquiry, finds extensive applications in various commercial domains. Condensing vast amounts of information into concise forms, summaries serve diverse purposes such as news reporting, content aggregation, and headline generation. The prevalent aggregation algorithms primarily fall into two categories. The extraction method involves rearranging and duplicating passages from the source material, while the abstract approach generates new phrases through rephrasing or incorporating terms absent in the source text.

Due to the intricacies involved in abstract summarization, the majority of research has predominantly focused on the extractive approach. Extractive summarization ensures grammatical correctness and precision by directly lifting segments from the original document. Conversely, abstract summarization necessitates advanced skills like paraphrasing, generalization, and assimilation of real-world knowledge, essential for achieving high-quality summaries. Despite its greater complexity, recent advancements in deep learning have yielded some successes in abstract summarization.

DOI: 10.17148/IARJSET.2024.11507

REFERENCES

[1] F. Jiang and X. Zhou, "Research on Generative Text Summarization Fusing Multidimensional Semantic Information," 2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS), Chengdu, China, 2023, pp. 803-807, doi: 10.1109/ISCTIS58954.2023.10213081.

[2] P. Shanmugam S, M. Sudharsan R S, V. P S and A. S S, "Abstractive Text Summarisation Using Keywords with Transformers Model," 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2023, pp. 1-9, doi: 10.1109/ICONSTEM56934.2023.10142867.

[3] P. F. Noronha, M. Bhan, M. Niranjanamurthy and D. Chandana, "Text Analysis Tool," 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), Bangalore, India, 2023, pp. 1-7, doi: 10.1109/ICAIA57370.2023.10169652

[4] J. Ranganathan and G. Abuka, "Text Summarization using Transformer Model," 2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS), Milan, Italy, 2022, pp. 1-5, doi: 10.1109/SNAMS58071.2022.10062698.

[5] K. Sharmila, N. S. Devi and R. Devi, "Survey on Sentiment Analysis using Deep Learning," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 1434-1438, doi: 10.1109/SMART55829.2022.10047158.

[6] R. D. A. Ahmed, M. AbdulHak and O. H. ELNabrawy, "Text Summarization Clustered Transformer (TSCT)," 2022 International Conference on Frontiers of Artificial Intelligence and Machine Learning (FAIML), Hangzhou, China, 2022, pp. 175-178, doi: 10.1109/FAIML57028.2022.00041.

[7] Y. Li, "Sentiment Analysis Using Neural Network: Application and Future Direction," 2022 International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACAI), Zakopane, Poland, 2022, pp. 217-222, doi: 10.1109/ICDACAI57211.2022.00051.

[8] R. Sonbol, G. Rebdawi and N. Ghneim, "The Use of NLP-Based Text Representation Techniques to Support Requirement Engineering Tasks: A Systematic Mapping Review," in IEEE Access, vol. 10, pp. 62811-62830, 2022, doi: 10.1109/ACCESS.2022.3182372.

[9] T. Shaik et al., "A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis," in IEEE Access, vol. 10, pp. 56720-56739, 2022, doi: 10.1109/ACCESS.2022.3177752.

[10] M. N. Alrasheedy, R. C. Muniyandi and F. Fauzi, "Text-Based Emotion Detection and Applications: A Literature Review," 2022 International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 2022, pp. 1-9, doi: 10.1109/ICCR56254.2022.9995902.

[11] Helwan, Abdulkader & Azar, Danielle & Uzun Ozsahin, Dilber. (2023). Medical Reports Summarization Using Text-To-Text Transformer. 01-04. 10.1109/ASET56582.2023.10180671.

[12] P. Ngamcharoen, N. Sanglerdsinlapachai and P. Vejjanugraha, "Automatic Thai Text Summarization Using Keyword-Based Abstractive Method," 2022 17th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Chiang Mai, Thailand, 2022, pp. 1-5, doi: 10.1109/iSAI-NLP56921.2022.9960265.

[13] M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan and M. M. Kabir, "A Survey of Automatic Text Summarization: Progress, Process and Challenges," in IEEE Access, vol. 9, pp. 156043-156070, 2021, doi: 10.1109/ACCESS.2021.3129786.

[14] G. Sanuvala and S. S. Fatima, "A Study of Automated Evaluation of Student's Examination Paper using Machine Learning Techniques," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2021, pp. 1049-1054, doi: 10.1109/ICCCIS51004.2021.9397227.

[15] M. Agarwal, R. Kalia, V. Bahel and A. Thomas, "AutoEval: A NLP Approach for Automatic Test Evaluation System," 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), Kuala Lumpur, Malaysia, 2021, pp. 1-6, doi: 10.1109/GUCON50781.2021.9573769.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ☵ Peer-reviewed & Refereed journal ☵ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11507

[16] S. Alhojely and J. Kalita, "Recent Progress on Text Summarization," 2020 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2020, pp. 1503-1509, doi: 10.1109/CSCI51800.2020.00278.

[17] S. Zhao and F. You, "A Topical Keywords Fusion Based on Transformer For Text Summarization," 2020 13th International Conference on Intelligent Computation Technology and Automation (ICICTA), Xi'an, China, 2020, pp. 294-297, doi: 10.1109/ICICTA51737.2020.00068.

[18] R. Parthiban, R. Ezhilarasi and D. Saravanan, "Optical Character Recognition for English Handwritten Text Using Recurrent Neural Network," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2020, pp. 1-5, doi: 10.1109/ICSCAN49426.2020.9262379.

[19] R. Boorugu and G. Ramesh, "A Survey on NLP based Text Summarization for Summarizing Product Reviews," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 352-356, doi: 10.1109/ICIRCA48905.2020.9183355.

[20] P. Ranjitha, K. Rajashekar and Shamjith, "A Review on Text Detection from Multi-Oriented Text Images in Different Approaches," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 240-245, doi: 10.1109/ICESC48915.2020.9156002.

[21] P. Yang, W. Li and G. Zhao, "Language Model-Driven Topic Clustering and Summarization for News Articles," in IEEE Access, vol. 7, pp. 185506-185519, 2019, doi: 10.1109/ACCESS.2019.2960538.

[22] L. Cagliero, L. Farinetti and E. Baralis, "Recommending Personalized Summaries of Teaching Materials," in IEEE Access, vol. 7, pp. 22729-22739, 2019, doi: 10.1109/ACCESS.2019.2899655.

[23] H. Li, J. Zhu, C. Ma, J. Zhang and C. Zong, "Read, Watch, Listen, and Summarize: Multi-Modal Summarization for Asynchronous Text, Image, Audio and Video," in IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 5, pp. 996-1009, 1 May 2019, doi: 10.1109/TKDE.2018.2848260.

[24] Z. Fu, F. Huang, X. Sun, A. V. Vasilakos and C. -N. Yang, "Enabling Semantic Search Based on Conceptual Graphs over Encrypted Outsourced Data," in IEEE Transactions on Services Computing, vol. 12, no. 5, pp. 813-823, 1 Sept.-Oct. 2019, doi: 10.1109/TSC.2016.2622697.

[25] Sultanum, Nicole & Singh, Devin & Brudno, Michael & Chevalier, Fanny. (2018). Doccurate: A Curation-Based Approach for Clinical Text Visualization. IEEE Transactions on Visualization and Computer Graphics. PP. 1-1. 10.1109/TVCG.2018.2864905.

[26] X. Sun and H. Zhuge, "Summarization of Scientific Paper Through Reinforcement Ranking on Semantic Link Network," in IEEE Access, vol. 6, pp. 40611-40625, 2018, doi: 10.1109/ACCESS.2018.2856530.

[27] A. Kapelner, J. Soterwood, S. Nessaiver and S. Adlof, "Predicting Contextual Informativeness for Vocabulary Learning," in IEEE Transactions on Learning Technologies, vol. 11, no. 1, pp. 13-26, 1 Jan.-March 2018, doi: 10.1109/TLT.2018.2789900.

[28] A. Shimada, F. Okubo, C. Yin and H. Ogata, "Automatic Summarization of Lecture Slides for Enhanced Student PreviewTechnical Report and User Study," in IEEE Transactions on Learning Technologies, vol. 11, no. 2, pp. 165-178, 1 April-June 2018, doi: 10.1109/TLT.2017.2682086.

[29] K. -Y. Chen, S. -H. Liu, B. Chen and H. -M. Wang, "An Information Distillation Framework for Extractive Summarization," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 1, pp. 161-170, Jan. 2018, doi: 10.1109/TASLP.2017.2764545.

[30] B. Xu, Y. Fu, Y. -G. Jiang, B. Li and L. Sigal, "Heterogeneous Knowledge Transfer in Video Emotion Recognition, Attribution and Summarization," in IEEE Transactions on Affective Computing, vol. 9, no. 2, pp. 255-270, 1 April-June 2018, doi: 10.1109/TAFFC.2016.2622690.

[31] C. -Y. Liu, M. -S. Chen and C. -Y. Tseng, "IncreSTS: Towards Real-Time Incremental Short Text Summarization on Comment Streams from Social Network Services," in IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 11, pp. 2986-3000, 1 Nov. 2015, doi: 10.1109/TKDE.2015.2405553.

[32] X. Li, H. Xie, Y. Song, S. Zhu, Q. Li and F. L. Wang, "Does Summarization Help Stock Prediction? A News Impact Analysis," in IEEE Intelligent Systems, vol. 30, no. 3, pp. 26-34, May-June 2015, doi: 10.1109/MIS.2015.1.

DOI: 10.17148/IARJSET.2024.11507

[33] M. T. Mills and N. G. Bourbakis, "Graph-Based Methods for Natural Language Processing and Understanding—A Survey and Analysis," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 44, no. 1, pp. 59-71, Jan. 2014, doi: 10.1109/TSMCC.2012.2227472.

[34] W. Dou, L. Yu, X. Wang, Z. Ma and W. Ribarsky, "HierarchicalTopics: Visually Exploring Large Text Collections Using Topic Hierarchies," in IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, pp. 2002-2011, Dec. 2013, doi: 10.1109/TVCG.2013.162.

[35] G. Di Fabbrizio, A. Aker and R. Gaizauskas, "Summarizing Online Reviews Using Aspect Rating Distributions and Language Modeling," in IEEE Intelligent Systems, vol. 28, no. 3, pp. 28-37, May-June 2013, doi: 10.1109/MIS.2013.36.

[36] P. Goyal, L. Behera and T. M. McGinnity, "A Context-Based Word Indexing Model for Document Summarization," in IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 8, pp. 1693-1705, Aug. 2013, doi: 10.1109/TKDE.2012.114.

[37] C. -L. Liu, W. -H. Hsaio, C. -H. Lee, G. -C. Lu and E. Jou, "Movie Rating and Review Summarization in Mobile Environment," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 3, pp. 397-407, May 2012, doi: 10.1109/TSMCC.2011.2136334.

[38] C. C. Chen and M. C. Chen, "TSCAN: A Content Anatomy Approach to Temporal Topic Summarization," in IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 1, pp. 170-183, Jan. 2012, doi: 10.1109/TKDE.2010.228.

