AUTO TITLE CRAFT

A.Shivani Reddy,B.Sneha,K.Shivani Reddy

Under the esteemed guidance of
Dr P.Kayal
Associate Professor

VISHNU
UNIVERSAL LEARNING

Bachelor of Technology
Department of Information Technology
BVRIT HYDERABAD College of Engineering for Women

November 16, 2023

Contents

Introduction

- A title generator is a tool or model that generates titles or headlines for text-based content.

- Title generators can be useful for content creators, journalists, and marketers to come up with catchy and informative titles.

Literature Survey

S.NO	Title	Journal	Author	Year	Description
1	Automatic Title Generation for Text with Pre-trained Transformer Language Model	International Civil Service Commission (ICSC)	Prakhar Mishra, Chaitali Diwan, Srinath Srinivasa	2021	This paper introduces an innovative title generation approach using the GPT-2 Transformer model, involving modules for generation, selection, and refinement, with promising results and human validation.
2	Sheershak: An Automatic Title Generation Tool for Hindi Short Stories	International Conference on Advances In Computing Communication Control and Networking (ICACCCN)	Leena Jain, Prateek Agarwal	2018	This paper presents an automatic title generation system for Hindi stories, using various approaches like noun-adjective combinations, proverbs, and nouns, with satisfactory results, aiding Hindi language promotion.

Literature Survey (Contd..,)

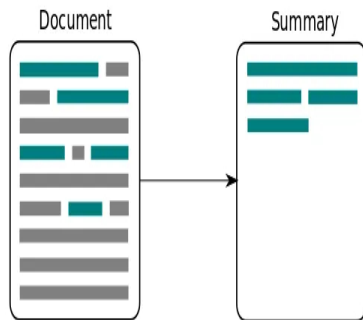| S.NO	Title	Journal	Author	Year	Description
3	Keyword Extraction from Scientific Articles in Bahasa Indonesia using TextRank Algorithm	Electrical, Telecommuni cation and Computer Engineering (ELTICOM)	Dani Gunawan, Fanindia Purnamasari, Ranti Ramdhiana	2020	This research employs TextRank for keyword extraction in three stages: pre-processing, TextRank application, and post-processing, improving recall with increased assigned keywords. Future work can optimize extraction time.
4	Automatic Text Summarization of News Articles in Serbian Language	International Symposium INFOTECH-JAORINA	Dijana Kosmajac, Vlado Keselj	2019	This paper explores using TextRank, a graph-based method, for extractive text summarization on Bosnian news data to enhance information retrieval.

Problem Statement

- The proposed project aims to implement a computer program using advanced language understanding and artificial intelligence that can generate catchy and appropriate titles for any collective digital text.
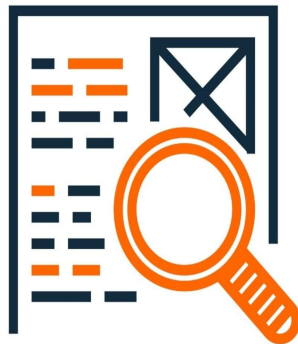
Proposed Method

- **TEXT SUMMARIZATION:-** BERT-extractive-summarizer is used to generate concise text summaries. It leverages the Bert model, extracting key information. The summary aids in presenting information more efficiently and effectively from the user given text.
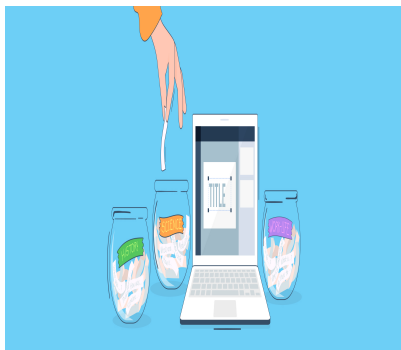

Document   Summary

Proposed Method(Contd..,)

- **KEYWORD EXTRACTION AND RANKING:-** Keywords are identified and extracted from the summary, facilitating efficient retrieval of essential information and enhancing content un-derstanding. YAKE library assigns a "score" to keywords based on deduplication algorithm, specified by 'dedupFunc' in 'Keyword Extractor' instantiation.

Proposed Method

- **TITLE GENERATION:-** YAKE library uses a detailed method for extracting keywords. It assigns scores to keywords using a specific deduplication algorithm, which helps avoid repetition and ensures a nuanced approach to identifying important terms.

Implementation and Result

User_Text : Coffee is a globally beloved beverage that has woven its rich aroma and bold flavor into the fabric of countless cultures. One of the most widely consumed stimulants, coffee owes its popularity not only to its ability to awaken the senses but also to the ritualistic and social aspects that surround its consumption. Whether enjoyed alone in quiet contemplation or shared among friends in a bustling café, coffee has a unique ability to bring people together. The nuances of each step contribute to the diverse array of coffee varieties, each with its own distinct flavor profile.

Summary : Coffee is a globally beloved beverage that has woven its rich aroma and bold flavor into the fabric of countless cultures. Whether enjoyed alone in quiet contemplation or shared among friends in a bustling café, coffee has a unique ability to bring people together.

Implementation and Result (contd..,)

```
Keywords:
0.0050460870971503115 (Score: globally beloved beverage)
0.019395584690361505 (Score: countless cultures)
0.028743132570503215 (Score: globally beloved)
0.028743132570503215 (Score: beloved beverage)
0.028743132570503215 (Score: woven its rich)
0.028743132570503215 (Score: rich aroma)
0.028743132570503215 (Score: aroma and bold)
0.028743132570503215 (Score: bold flavor)
0.028743132570503215 (Score: fabric of countless)
0.09692139183006357 (Score: Coffee)
```

Implementation and Result

Sno	Modules	Status
1	Problem identification	Completed
2	Literature Survey	Completed
3	Data Acquisition (https://www.kaggle.com/datasets/thejas2002/titlegen)	Completed
4	Text Summarization	Completed
5	Keyword Extraction Ranking	Completed
6	Title Generation	Completed

Conclusion

- Auto Title Craft, a versatile tool, extracts titles from text using summarization and keyword identification, enhancing content representation.
- Ideal for authors and content creators, this streamlined process ensures concise, impactful titles, contributing to effective communication and engagement with the audience.

References

- P. Mishra, C. Diwan, S. Srinivasa and G. Srinivasaraghavan, "Automatic Title Generation for Text with Pre-trained Transformer Language Model," 2021 IEEE 15th International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 2021, pp. 17-24, doi: 10.1109/ICSC50631.2021.00009.
- L. Jain and P. Agrawal, "Sheershak: an Automatic Title Generation Tool for Hindi Short Stories," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCN), Greater Noida, India, 2018, pp. 579-584, doi: 10.1109/ICACCN.2018.8748377.

References (Contd..,)

- D. Gunawan, F. Purnamasari, R. Ramadhiana and R. F. Rahmat, "Keyword Extraction from Scientific Articles in Bahasa Indonesia using TextRank Algorithm," 2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), Medan, Indonesia, 2020, pp. 260-264, doi: 10.1109/ELTICOM50775.2020.
- D. Kosmajac and V. Kešelj, "Automatic Text Summarization of News Articles in Serbian Language," 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 2019, pp. 1-6, doi: 10.1109/INFOTEH.2019.8717655.

Thank you