Informatics Institute of Technology

In Collaboration With

The University of Westminster, UK

The University of Westminster, Coat of Arms

News Snap

Text Summarization for Online News Articles

A dissertation by

Ms. Hamra Imam

W1761272 | 2019292

Supervised by

Ms. Krishnakripa Jayakumar

May 2023

Submitted in partial fulfilment of the requirements for the

BSc (Hons) Computer Science degree at the University of Westminster.

ABSTRACT

In the era of digital information, the overload of online content, particularly news articles, has made information gathering a challenging and time-consuming task. This problem is exacerbated for readers with short attention spans. This research presents an automatic text summarization tool that generates concise, relevant, and accurate summaries from online news articles, providing solutions for users with limited time and attention.

The research project aims to develop a tool that can automate abstractive text summarization, alleviating the manual process of filtering through vast amounts of data. The focus lies on the detection of the most significant aspects, producing high-quality summaries tailored to the user's interest. The proposed solution introduces a novel approach to aspect-based summarization. It employs a sophisticated blend of NER and aspect embedding sentence similarity for precise aspect extraction and DL methods, focusing on accurate extraction of relevant aspects from news articles.

The novelty of this work is in developing a system capable of generating both generic and aspect-based summaries, accurately identifying and extracting relevant aspects from news articles, an area uncharted in previous research in NLP. The work also contributes to the research domain by fine-tuning existing DL models. The research concludes with promising results, as the designed system effectively outperforms existing solutions in benchmarking, showcasing significant improvements in performance. The application of Hierarchical Attention Mechanism shows noticeable impact on the model's performance, improving the quality of the generated summaries.

Keywords: Natural Language Processing (NLP), Named Entity Recognition (NER), Deep Learning (DL)

Subject Descriptors:

- Computing Methodologies → Artificial Intelligence → Natural Language Processing → Natural Language Generation
- Computing Methodologies → Artificial Intelligence → Deep Learning → Transformer Models
- Natural Language Processing → Information Extraction → Aspect-Based Text Analysis

Hamra Imam | w1761272