

Research Article

Web-Based News Straining and Summarization Using Machine Learning Enabled Communication Techniques for Large-Scale 5G Networks

Amita Arora ¹, **Ashlesha Gupta**,¹ **Manvi Siwach**,² **Pankaj Dadhech** ³,
Krishnaveni Kommuri,⁴ **Majid Altuwairiqi**,⁵ and **Basant Tiwari** ⁶

¹Department of Computer Engineering, J.C. Bose University of Science and Technology, YMCA, Faridabad, India

²Department of Computer Application, J.C. Bose University of Science and Technology, YMCA, Faridabad, India

³Department of Computer Science & Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan (SKIT), Jagatpura, 302017, Jaipur, Rajasthan, India

⁴Department of ECM, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India

⁵Department of Computer Science, College of Computers and Information Technology, Taif University, Saudi Arabia

⁶Department of Computer Science, Hawassa University, Awasa, Ethiopia

Correspondence should be addressed to Basant Tiwari; basanttiw@hu.edu.et

Received 31 March 2022; Revised 10 May 2022; Accepted 16 May 2022; Published 23 June 2022

Academic Editor: Mohammad Farukh Hashmi

Copyright © 2022 Amita Arora et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent times, text summarization has gained enormous attention from the research community. Among the many uses of natural language processing, text summarization has emerged as a critical component in information retrieval. In particular, within the past two decades, many attempts have been undertaken by researchers to provide robust, useful summaries of their findings. Text summarizing may be described as automatically constructing a summary version of a given document while keeping the most important information included within the content itself. This method also aids users in quickly grasping the fundamental notions of information sources. The current trend in text summarizing, on the other hand, is increasingly focused on the area of news summaries. The first work in summarizing was done using a single-document summary as a starting point. The summarizing of a single document generates a summary of a single paper. As research advanced, mainly due to the vast quantity of information available on the internet, the concept of multidocument summarization evolved. Multidocument summarization generates summaries from a large number of source papers that are all about the same subject or are about the same event. Because of the content duplication, the news summarization system, on the other hand, is unable to cope with multidocument news summarizations well. Using the Naive Bayes classifier for classification, news websites were distinguished from nonnews web pages by extracting content, structure, and URL characteristics. The classifier was then used to differentiate between the two groups. A comparison is also made between the Naive Bayes classifier and the SMO and J48 classifiers for the same dataset. The findings demonstrate that it performs much better than the other two. After those important contents have been extracted from the correctly classified newscast web pages. Then, extracted relevant content is used for the keyphrase extraction from the news articles. Keyphrases can be a single word or a combination of more than one word representing the news article's significant concept. Our proposed approach of crucial phrase extraction is based on identifying candidate phrases from the news articles and choosing the highest weight candidate phrase using the weight formula. Weight formula includes features such as TFIDF, phrase position, and construction of lexical chain to represent the semantic relations between words using WordNet. The proposed approach shows promising results compared to the other existing techniques.