

Submission Summary

Conference Name

6th International Conference on Electrical Information and Communication Technology 2023

Track Name

Artificial Intelligence, Computer Networks, and Security

Paper ID

28

Paper Title

Utilizing Optical Character Recognition to Summarize Bengali Text

Abstract

Optical Character Recognition (OCR) has proven to be a valuable tool for extracting textual information from images, converting physical documents into digital formats that can be easily edited, searched, and shared. However, OCR technology is not without its limitations, particularly when it comes to recognizing and processing non-Latin scripts such as Bengali. Over 230 million people speak Bengali, making it one of the most widely used languages in the world. Summarizing Bengali material is one area where OCR could be especially helpful, saving time and effort for scholars and journalists who are working with a big corpus of text. In recent years, there have been several attempts to develop OCR-based text summarization techniques for Bengali. One challenge it struggles with is the complex nature of the language, its rich vocabulary, its highly inflected grammar and the inclusion of a number of ligatures in Bengali Script and diacritics that can significantly complicate the OCR process. Overall, the development of OCR-based text summarization could save time and effort, while also providing a more objective and systematic approach to analyzing large volumes of text. It is expected that in the future, Bengali text summarizing methods based on OCR will become more advanced and precise as technology advances.

Created on

5/14/2023, 9:34:26 PM

Last Modified

5/14/2023, 9:34:26 PM

Authors

Abhijit Saha (BRAC University) < abhijit.saha@g.bracu.ac.bd> ✓

Faiza Bushra (BRAC University) < faiza.bushra@g.bracu.ac.bd> ✓

Ahmed Anwar (BRAC University) < ahmed.anwar@g.bracu.ac.bd> ✓

Rubaba Rashid (Brac University) < rubaba.rashid@g.bracu.ac.bd> ✓

Ehsanur Rahman Rhythm (BRAC University) < ehsanur.rahman.rhythm@g.bracu.ac.bd> ✓

Md Sabbir Hossain (Brac University) < md.sabbir.hossain1@g.bracu.ac.bd> ✓

Annajiat Alim Rasel (BRAC University) < annajiat@gmail.com> ✓

Submission Files

Utilizing Optical Character Recognition to Summarize Bengali Text.pdf (283.3 Kb, 5/14/2023, 9:33:52 PM)
