Abstract ID: P49

AUTOMATED TEXT SUMMARIZATION OF SCIENTIFIC DOCUMENTS

M.A.C. Akmal Jahan*, R.D.R.M. Gunathilaka

Department of Computer Science, Faculty of Applied Sciences,
South Eastern University of Sri Lanka, Sammanthurai, Sri Lanka.

**akmaljahan@seu.ac.lk*

Abstract

Text summarization plays a major role in natural language processing, especially in scientific communities like researchers, students, and so on. Due to the quantity of scientific publications available online is rapidly rising, it takes too much time to identify the most appropriate, quality and relevant materials for their search out of thousands. Therefore, there should be an alternative way to sort out and simplify the search and get a quality and appropriate document based on our search. The aim of this work is to generate an online platform for a digital library that provides a good summary for any scientific document which is subscribed by the library of the institution. Therefore, we need to find an appropriate and best suitable text summarization algorithm out of some state-of-the-art text processing algorithms such as Text Rank algorithm, TF-IDF algorithm, and K-Means algorithm, which have been used in different text processing scenarios. To evaluate and select the best suitable algorithm, we used publicly available scientific dataset and manually generated summary from the dataset. From the experiments processed, the Text Rank algorithm performed better than the other algorithms.

Keywords: *text summarization, text-rank, TF-IDF, K-Means*