TRENDS AND CHALLENGES IN TEXT SUMMARIZATION, A LITERATURE SURVEY

Karari Kinyanjui[a], Kariuki Paul[b], Maina Sammuel[c], Malanga Ndenga[d]

[a]School of Computer Science, Dedan Kimathi University of Technology, Private Bag, Nyeri, 10143, Kenya, E-mail: efantusk@gmail.com

[b]School of Computer Science, Dedan Kimathi University of Technology, Private Bag, Nyeri, 10143, Kenya, E-mail: memekariuki@gmail.com

[c]School of Computer Science, Dedan Kimathi University of Technology, Private Bag, Nyeri, 10143, Kenya, E-mail: samicho4@gmail.com

[d]School of Computer Science, Dedan Kimathi University of Technology, Private Bag, Nyeri, 10143, Kenya, E-mail: malangalanga@dkut.ac.ke

Keywords: Deep Learning, automatic text summarization, Query oriented algorithm, Extractive summary, ROUGE

Interest in text summarization is heightened by the rapid growth in content especially content available in the World Wide Web. There are a myriad types of summarization and approaches to text summarization that have evolved courtesy of relentless efforts by various researchers. Researchers have also encountered several challenges in automatic text summarization and effort to address these challenges have resulted to multiple algorithms. This survey aimed at providing a single reference for the types and approaches to text summarization and has included the challenges, attempted solutions and key recommendation by researchers.

This survey used systematic approach to review research articles already published and available from the various journals [1]. The main search terms were deep learning and automatic text summarization. Relevant data was extracted from the reviewed articles and where necessary standardization of the metrics was done to ensure a fair comparison in performance.

The survey revealed various types of summaries, approaches to automatic text summarization, challenges of text summarization and their solutions. It sheds light into machine learning and specifically deep learning as the preferred contemporary approach towards automatic text summarization. Further, some studies have indicated that unsupervised deep learning results to solutions that are easily deployable in the industry since they require no labelled data. Hybridization is recommended by researchers for its ability to exploit the strength of each composite algorithm and reduce the challenges of automatic text summarization [2].

[1] Tawfik, G. M., Dila, K. A. S., Mohamed, M. Y. F., Tam, D. N. H., Kien, N. D., Ahmed, A. M., & Huy, N. T. (2019). A step by step guide for conducting a systematic review and meta-analysis with simulation data. Tropical medicine and health, 47(1), 1-9

[2] Kirmani, M., Hakak, N. M., Mohd, M., & Mohd, M. (2019). Hybrid Text Summarization: A Survey. In Soft Computing: Theories and Applications (pp. 63-73). Springer, Singapore.