

Suleiman D., Awajan A. 2019. Deep Learning Based Extractive Text Summarization: Approaches, Datasets and Evaluation Measures. Proceedings of Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS). Granada, Spain. October 22-25, 2019. 10.1109/SNAMS.2019.8931813. Pages 204-210.

## **Abstract**

In recent years, the volume of textual data has rapidly increased, which has generated a valuable resource for extracting and analysing information. To retrieve useful knowledge within a reasonable time period, this information must be summarised. This paper reviews recent approaches for abstractive text summarisation using deep learning models. In addition, existing datasets for training and validating these approaches are reviewed, and their features and limitations are presented. The Gigaword dataset is commonly employed for single-sentence summary approaches, while the Cable News Network (CNN)/Daily Mail dataset is commonly employed for multisentence summary approaches. Furthermore, the measures that are utilised to evaluate the quality of summarisation are investigated, and Recall-Oriented Understudy for Gisting Evaluation 1 (ROUGE1), ROUGE2, and ROUGE-L are determined to be the most commonly applied metrics. The challenges that are encountered during the summarisation process and the solutions proposed in each approach are analysed. The analysis of the several approaches shows that recurrent neural networks with an attention mechanism and long short-term memory (LSTM) are the most prevalent techniques for abstractive text summarisation. The experimental results show that text summarisation with a pretrained encoder model achieved the highest values for ROUGE1, ROUGE2, and ROUGE-L (43.85, 20.34, and 39.9, respectively). Furthermore, it was determined that most abstractive text summarisation models faced challenges such as the unavailability of a golden token at testing time, out-of-vocabulary (OOV) words, summary sentence repetition, inaccurate sentences, and fake facts.