From Primer to Core : Performing Text Summarization with NLP and spaCy

Ms Varsha, Chhama Sharma, Shikha Bhardwaj

Computer Science and Engineering

Raj Kumar Goel Institute of Technology, Ghaziabad, UP, India

varsharani5699@gmail.com, sharmachhama18@gmail.com shikhabhardwaj422@gmail.com

Abstract: *Automatic Text Summarization (ATS) is becoming increasingly essential because of the overwhelming growth of word-based information found online and across various domains such as news archives, academic publications, and legal documents. Manually summarizing this vast amount of content is time-consuming, labor-intensive, and often not feasible. Efforts to enhance ATS methods have been ongoing after the 1950s. ATS can be categorized into three main types: Generative, Non-Generative, and hybrid. The Non-Generative approach identifies and pick crucial sentences directly out of actual content to form a summary. In contrast, the abstractive approach generates new sentences by interpreting the meaning of the source content. The hybrid approach integrates elements from both extractive and abstractive methods. Although many techniques have been proposed, automatically generated summaries often fall short of the quality produced by human summarizers. Most existing research has been centered around extractive summarization, while abstractive and hybrid methods remain relatively underexplored. This work outlines key aspects of ATS, including its methodologies, underlying techniques, datasets, evaluation metrics, and directions for future research, offering a clear understanding of the field for those interested in further development and exploration.*

Keywords: Automated Text Summarization; Python NLP; spaCy; Text Processing ; Natural Languauge Processing ; Evaluation Metrices