"PACEBYTE": A Research Engine Based on Automated Summarization

Mitra, Abhishek

The explosive growth of verbose content on the internet in recent years, especially of news articles and blogs, coupled with a shift towards mobile internet access, has created a case for intelligent summarization of text. Through the use of advanced summarization algorithms, it is now possible to compress verbose text to suit small screen devices as well. With this in mind, this project endeavors to create a research engine which after fetching original content from the internet, applies smart compression on the text while retaining as much of the original information as possible. Prior work done in this field has only implemented sentence elimination. Although the results showed more density of information, they fell short in flow and coherence. The most important process in generating coherent summaries is the identification of the key topics or the “flow of ideas“. Essentially, summarization is achieved through the generation of the key ideas using keyword Identification. Once the key ideas in a topic are generated, the relations between individual topics are identified using word graphs. This algorithm utilizes the process of extraction, to remove redundant sentences. The selected sentences are further refined using an abstractive algorithm. The algorithm also identifies recurring topics and relations between them, allowing for a coherent summary.

Awards Won: