

Extractive text summarization

Nong Tien Cong

Abstract: This thesis investigates extractive text summarization, a method that generates concise summaries by selecting and concatenating the most salient sentences from a source document. The research addresses the core challenge of identifying key information while minimizing redundancy and maintaining coherence. A novel framework is proposed, integrating statistical features such as term frequency-inverse document frequency (TF-IDF) with graph-based ranking algorithms, including an adaptation of TextRank. The model evaluates sentences based on their relevance to the document's central themes, their position within the text, and their similarity to other high-scoring sentences. Experimental evaluations were conducted on standard benchmark datasets, comparing the proposed method against established baselines. Results demonstrate that the hybrid approach achieves competitive performance in terms of ROUGE scores, effectively balancing informativeness and conciseness. The findings contribute to the field of natural language processing by offering a robust and computationally efficient solution for automatic summarization, with practical applications in information retrieval and document management.

Keywords: extractive summarization, sentence extraction, text summarization, Vietnamese