ABSTRACT

Information content available for use in various real world scenarios and applications has greatly increased recently and continues to grow. Specifically, the huge growth of textual information available in the Internet and intranets has made finding relevant text information on time as a major technical challenge. Document summarization is a reductive transformation of input texts into a summary text and can help in comprehending the abundant textual information.

A good document summarization system should preserve the salient information content in the texts. Document summarization can be classified into various types based on the context factors of the summarization problem. Many document summarization techniques exist. However, the effectiveness of these systems, in terms of quality of the summaries created, is found lagging when compared to the quality of the summaries created by human beings. Very few document summarization techniques make use of human information processing behaviour and even such systems have severe limitations and strict assumptions. Human cognition and intelligent information processing behaviour are described by cognitive psychology models. Computational cognitive models transform the cognitive psychology models to adapt them for real world applications. Computational cognitive models for document summarization built from cognitive psychology models could address the gap between the effectiveness of system generated summaries and manually created summaries.

The objective of this research work is to improve the effectiveness of the document summarization process to create high quality summaries, using

human text understanding and summarization behaviour. Hence, this research work explores how human information processing behaviour, structures and underlying processes can be used in document summarization using computational cognitive models. This research work also analyses the effectiveness of the designed models and their applicability in various document summarization contexts and tasks.

The proposed computational cognitive model, Causality-Relation Using Girju Phrases EI Computational Cognitive Model (CR-GP), adapts the event-index cognitive psychology model of human reading and understanding for document summarization. This model applies human text understanding and processing behaviour, and related cognitive processes for extractive, single-document summarization. Though the CR-GP model closely mimics the human reading and understanding process, it does not use the human memory and knowledge storage and retrieval processes.

The proposed computational cognitive model, Knowledge Based Event-Index Computational Cognitive Model (KB-EI), enhances the CR-GP model by incorporating human memory architecture and human knowledge retrieval processes in the single document summarization task. This addition of a knowledge base with context information and its use enables KB-EI to closely follow the human summarization behaviour and human knowledge structures. However, as KB-EI model does not use human attention, it is not effective for multi-document summarization.

The proposed computational cognitive model, Human Text Comprehension Cognitive Model For Summarization (HTC-CMS), adapts human text summarization behaviour described in the landscape cognitive psychology model. Like the KB-EI model it uses human memory architecture

and human knowledge retrieval processes. In addition, it employs human attention behaviour and redundancy reduction, and hence is suitable for multi-document summarization tasks.

The effectiveness of the designed computational cognitive models has been analysed based on the quality of summaries created by these models using standard document summarization datasets and evaluation metrics. The validity and consistency of the obtained results have been verified using statistical measures. These experimental results have proved that the document summarization models designed in this research work achieve significantly better results than prevalent document summarization systems. The effectiveness of the computational cognitive models for document summarization designed in this thesis work makes them applicable in varied scenarios and highly relevant to current, real world applications.