

# Extraction of Text Summary Using Latent Semantic Indexing and Information Retrieval Technique: Comparison of Four Strategies

Abdelghani Bellaachia\*, Anand Mahajan\*

\*Computer Science Department  
George Washington University  
Washington DC, 20052  
bell@gwu.edu  
<http://seas.gwu.edu/~bell>

**Abstract.** In this paper, we present four generic text summarization techniques. Each technique extracts a text summary by ranking and extracting sentences from an original document. The first method, SUMMARIZER 1, uses standard information retrieval (IR) methods to rank sentences. The second method, SUMMARIZER 2, uses the Latent Semantic Analysis (LSA) technique to identify semantically important sentences, for summary creations. The third method, SUMMARIZER 3, uses a combination of the latent semantic analysis technique, reduction and relevance measure. The fourth method simply uses the TF\*IDF (Term frequency \* Inverse Document frequency) weighting scheme. Evaluations of the four methods are conducted using Document Understanding Conferences (DUC) datasets from NIST. We have compared the summary of each method with the manual summaries. Summarizer 4, with its lowest overhead, has comparable performance to summarizer 1. Analysis shows that a combination of LSA technique and the relevance measure (Summarizer 3) has the best performance on an average.

## 1. Introduction

The speed and the scale of information dissemination have dramatically increased with the explosive growth of the worldwide web. Using conventional information retrieval (IR) techniques to find relevant information effectively in a vast sea of accessible text documents on the Internet, has become more and more insufficient. Text search engines serve as information filters that sift out an initial set of relevant documents. Their keyword-based approach retrieves millions of hits by which the user is overwhelmed. Hence, there is a need for techniques to quickly identify the most relevant documents. Text summarizers can be used to help users identify final set of relevant documents. Text search and summarization are two essential technologies that complement each other. Presenting the user with a summary of each document greatly facilitates the task of finding the desired documents.

The goals of text summarizers can be categorized by their intent, focus and coverage [MCDONALD ET AL.]. Intent refers to the potential use of the summary. Firmin and