

GENERATING AUTO TEXT SUMMARIZATION FROM DOCUMENT USING CLUSTERING

JAYA D. KAPOOR¹ & KAILAS K. DEVADKAR²

¹Department of Computer Engineering, Alamuri Ratnamala Institute of Engineering and Technology, Shahpur, India

²Department of Information Technology, Sardar Patel Institute of Technology, Andheri, India

ABSTRACT

Auto text summarization is a method of reducing the size of the text document with a software program in a way to generate a summary that retains the most important points of the original document. Interest in the automatic summarization has increased due to the occurrence of information overload, and tremendous growth in quantity of data. Coherent summary can be made using technologies such as considering account variables such as length, writing style and syntax. Google is the one of the good example of the use of summarization technology.

The two technologies viz. Extraction and Abstraction are used for auto text summarizations. Extraction methods work by selecting a content of existing sentences, phrases, or words from the original textual document to form the summary. Unlike, abstractive methods generate an internal semantic of content and then use natural language generation techniques to create a summary that can be related to what a human may generate. Such a summary which may contain words not explicitly present in the original text.

KEYWORDS: Extraction and Abstraction, Coherent Summary, Nearest Neighbour (NN)