

A MAJOR PROJECT - II REPORT

On

**ANALYSIS OF AUTOMATIC TEXT SUMMARIZATION
TECHNIQUES**

Submitted to
Delhi Technological University, Delhi

In partial fulfillment of the requirements for the award of the degree of
MASTER OF TECHNOLOGY

In
SOFTWARE ENGINEERING

By

HEMANT KUMAR
(Roll No. 2K14/SWE/06)

Under the guidance of

DR. AKSHI KUMAR
Assistant Professor

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY, DELHI**



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY, DELHI, 110042, INDIA
SESSION 2014-2016**



Computer Science & Engineering Department
Delhi Technological University
Delhi-110042
www.dtu.ac.in

CERTIFICATE

This is to certify that the work contained in this dissertation entitle “**Analysis of Automatic Text Summarization Techniques**” submitted in the partial fulfilment, for the award of degree of M. Tech in Software Engineering, Department of Computer Science & Engineering at **Delhi Technological University** by **Hemant Kumar**, Roll No. **2K14/SWE/06**, is carried out by him under my supervision. The matter embodied in this project work has not been submitted earlier for the award of any degree or diploma in any university/institution to the best of my knowledge and belief.

Date:

Dr. Akshi Kumar

(Project Guide)

Assistant Professor

Department of Computer Science Engineering

Delhi Technological University



Computer Science & Engineering Department
Delhi Technological University
Delhi-110042
www.dtu.ac.in

ACKNOWLEDGMENT

I take this opportunity to express my gratitude and regards to my guide **Dr. Akshi Kumar** for her exemplary guidance, monitoring and constant encouragement throughout the course of this project. I am extremely grateful to her for her valuable guidance and suggestions.

I am obliged to the Head of Department **Prof. (Dr.) O.P. Verma** and faculty members of the Department of Computer Science & Engineering at Delhi Technological University, Delhi for the valuable information provided by them in their respective fields. I am grateful for their cooperation during the period of my assignment.

I also take this opportunity to express a deep sense of gratitude to my friends for their support and motivation which helped me in completing this task through its various stages.

Lastly, I thank my parents for their constant encouragement without which this assignment would not have been possible

Hemant Kumar
Roll No. 2K14/SWE/06

ABSTRACT

It's been 58 years since the publication of Luhn's original paper on automatic summarization. Automatic Text Summarization (ATS) is the process of reducing a text document with a computer program in order to create a summary that retains the most important points of the original document. As the problem of information overloading has grown and the quantity of data has increased, subsequently the interest in automatic summarization has also flourished. It is very difficult for human beings to manually summarize large textual documents. Text Summarization methods can be classified into extractive and abstractive summarization. Extractive summarization technique includes choosing vital sentences, paragraphs of text and linking them along in a shorter form. Because of the presence of this technique, simplicity, high-speed report, prices and study time reduction may be mentioned. One disadvantage of this technique is that the extracted sentences could also be too moderate. Vital and relevant information might also be distributed between sentences and extractive technique cannot establish them. The overall method of extractive summarization is performed in two steps: pre-processing and processing. In the pre-processing step, identifying the end of sentences, removing terms that do not have any meaning and finding the word roots are performed. In the process step, the effectiveness and connection of the sentences to the subject area unit known and appointed a weight to every sentence after this by use of PageRank algorithm to ranking them according to their weights. At last, sentences with the minimum scores are chosen to be in final summarized text. In this project, we present some techniques for generating text extraction based summaries including Pattern matching, K-means clustering, Reduction, TextRank algorithm and LexRank algorithm. In frequency based technique obtained summary makes more meaning. But in k-means clustering due to out of order extraction, summary might not make sense.

Keywords:

ATS, MDS, Text Summarization, Clustering, extractive, abstractive, semantic, frequency based, K-means algorithm, TextRank, LexRank, Reduction PageRank.