MULTI-OBJECTIVE EXTRACTIVE TEXT SUMMARIZATION

Report submitted to the SASTRA Deemed to be University as the requirement for the course

CSE400: PROJECT WORK

Submitted by

AKSHAYA ROHITH M
(Reg No: 224003008, B.Tech., CSE)

HARIHARAN SUBRAMANIAN
(Reg No: 224003032, B.Tech., CSE)

MAY 2024

SRINIVASA RAMANUJAN CENTRE
KUMBAKONAM, TAMIL NADU INDIA - 612001

SRINIVASA RAMANUJAN CENTRE

KUMBAKONAM – 612 001

Bonafide Certificate

This is to certify that the report titled "**Multi-Objective Extractive Text Summarization**" submitted as a requirement for the course, **CSE 400: FINAL PROJECT** for B.Tech. is a bonafide record of the work done by Mr. AKSHAYA ROHITH.M(224003008, B.Tech.,-CSE), Mr. HARIHARAN SUBRAMANIAN(224003032, B.Tech.,-CSE) during the academic year 2023-24, in the Srinivasa Ramanujan Centre, under my supervision.

Signature of Project Supervisor:

Name with Affiliation : Smt. S. Hemamalini, AP-II/CSE/SRC/SASTRA

Date :

Final Project *Viva voce* held on 07.05.2024

Examiner 1 **Examiner 2**

SASTRA
ENGINEERING · MANAGEMENT · LAW · SCIENCES · HUMANITIES · EDUCATION
DEEMED TO BE UNIVERSITY
(U/S 3 of the UGC Act, 1956)
THINK MERIT | THINK TRANSPARENCY | THINK SASTRA
THANJAVUR | KUMBAKONAM | CHENNAI

SRINIVASA RAMANUJAN CENTRE

KUMBAKONAM – 612 001

Declaration

We declare that the report titled "**Multi-Objective Extractive Text Summarization**" submitted by us is an original work done by us under the guidance of **Smt. S. Hemamalini, AP-II/CSE/SRC/SASTRA, SASTRA Deemed to be University** during the final semester of the academic year 2023-24, in the **Srinivasa Ramanujan Centre**. The work is original and wherever we have used materials from other sources, we have given due credit and cited them in the text of the report. This report has not formed the basis for the award of any degree, diploma, associate-ship, fellowship or other similar title to any candidate of any university.

Signature of Candidate (s) **:**

Name of Candidate (s) **: Akshaya Rohith M (224003008)** **Hariharan Subramanian (224003032)**

Date **:** 07.05.2024

ACKNOWLEDGEMENTS

We pay our sincere obeisance to the God Almighty for his grace and infinite mercy and for showing on us his choicest blessings.

We would like to thank our honorable Chancellor **Prof. R. Sethuraman,** Vice Chancellor **Dr. S. Vaidhyasubramaniam** and **Dr. S. Swaminathan**, Dean – Planning and Development for the encouragement and strategic support at every step of our college life.

We extend our sincere thanks to **Dr. R Chandramouli**, Registrar, SASTRA Deemed to be University for providing the opportunity to pursue this project.

We express our deepest thanks to **Dr. V. Ramaswamy**, Dean and **Dr. A. Alli Rani**, Associate Dean, Srinivasa Ramanujan Centre for their constant support and suggestions when required without any reservations.

We express our gratitude to HOD in-charge **Dr. V. Kalaichelvi,** ACP/CSE for his constant support and valuable suggestions for the completion of the project.

We exhibit our pleasure in expressing our thanks to **Smt. S. Hemamalini,** AP-II/CSE, our guide for her ever-encouraging spirit and meticulous guidance for the completion of the project.

We would like to place on record the benevolent approach and painstaking efforts of guidance and correction of **Dr. J. Sangeetha and Smt. D. Rekha,** the project coordinators and all department staff to whom we owe our hearty thanks forever.

Without the support of our parents and friends this project would never have become reality.

We dedicate this work to our well-wishers, with love and affection.

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ECA	:	Evolutionary Computing Algorithm
ETS	:	Extractive Text Summarization
MOABC	:	Multi Objective Artificial Bee Colony Algorithm
NSGA	:	Non-dominated Sorting Algorithm
TF-ISF	:	Term Frequency Inverse Sentence Frequency
NDS	:	Non-Dominated Solution
ROUGE	:	Recall-Oriented Understudy for Gisting Evaluation
BCTSO	:	Binary Crowded Tournament Selection Operator
RCR	:	Ranking and Crowding based Replacement
HDP	:	Hierarchical Dirichlet Process
SBERT	:	Sentence Bidirectional Representation Transformer

ABSTRACT

Extractive summarization is a technique for automatically creating summaries of text documents by selecting the most important sentences from the original document. Redundancy reduction minimizes repetition in the summary, preventing information overload and improving conciseness. Coverage ensures that the summary faithfully conveys the main points and key concepts of the original text.

Current systems approach text summarization within a single-objective modelling framework, that are known to only capture homogenous facets of content and simple relationships within the text, potentially affecting the overall effectiveness and richness of the generated summaries. Multi-objective optimization strategies have shown to effectively tackle conflicting objectives such as the above. The Multi-Objective Artificial Bee Colony algorithm (MOABC) emerges as a promising approach, effectively handling both objectives through its population-based exploration. The repair operator is revised to a less computationally intensive repair operator. Then, some slight modifications in the mutation and replacement operators are proposed and its performance in the search process is recorded. SBERT transformers as a possible sentence embedding structure is also investigated and finally topic modelling is incorporated. In this project, ROUGE scores are used as the performance evaluation criteria.

Keywords: Extractive Summarization, Redundancy Reduction, Coverage, Multi-Objective Artificial Bee Colony Algorithm, ROUGE.

Table of Contents