



Informatics Institute of Technology

In Collaboration with

University of Westminster, UK

ResearchPro

Research Paper Summarization and Recommendation by Keywords Using Computational Linguistic Techniques

A dissertation by

Ms. Gayali Hettiarachchi

W1761100 / 2019418

Supervised by

Mr. Deshan Sumanathilaka

April 2023

Submitted in partial fulfillment of the requirements for the
BEng (Hons) Software Engineering degree at the University of Westminster.

ABSTRACT

Research papers have been a massive source of information in the scholarly world for centuries. Due to advancements in technology, academic materials have become more accessible and prevalent on the internet. Numerous academic publishers, universities, and online journals release new material at a very fast pace. Hence, finding good research papers and reading them has become a quite hassle for the researcher due to the overload and the saturation of information.

A system that automatically summarizes and recommends articles to users using NLP and data science is a solution to this problem. The summarization component finds the important sentences using TextRank, and lexical and syntactic simplification is used to further simplify the document. Recommendations are done in a content-based approach, similar papers are recommended by extracting keywords from papers and ranking them by calculating cosine similarity between vectors created from a Word2Vec model trained with a research paper dataset.

Through experiments, it was determined that the system compared with similar approaches, performs rather well in terms of accuracy and effectiveness and gives better results.

Keywords: Research Papers, Recommendation Systems, Text Summarizing, Natural Language Processing

Subject Descriptors:

- Information Retrieval Systems → Machine Learning → Natural Language Processing → Text Recommendation Systems
- Computational Linguistic Technologies → Natural Language Processing → Summarization Systems
- Information Systems → Applications → Summarizing and Recommendation System