INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER (UOW)

Sārānshā - Generating Mind Maps in Sinhala from Extractive Text Summarization

A dissertation By

K.D. Isuru Amantha Siriwardena

Supervised By

Ms. Aloka Fernando

Submitted in partial fulfillment of the requirements for the BSc (Hons) Software Engineering degree

Department of Computing

May 2018

Abstract

When it comes to the exam time, most of the students cannot remember the things they have learned or studied. Studying or remembering things is actually a really complex procedure and researches have shown that there is a “right” and a “wrong” way to do it. Creating summaries, mind maps, flash cards and writing questions to flashback the memory are some of the ways that are considered right to remember things.

In the modern world due to lack of the time, generating an accurate and intelligent summary for a long document or text pieces has become a popular research as well as an industry problem. The field of automatic text summarization has evolved since 1950’s. In 2004 with the adaptation of Unicode technology into the internet, Sinhala language also faced the above problem. This research has proposed a solution for summarizing the text in Sinhala language by identifying the most important and relevant sentences based on linguistic and statistical features of the text, using unsupervised extractive summarization approach. In order to generate a better summary, keyword and sentences extraction is manipulated by using a graph based TextRank algorithm.

Subject Descriptors:

- Computing methodologies~Natural language processing
- Theory of computation~Design and analysis of algorithms

Keywords:

- Natural language processing
- Graph theory