A HYBRID APPROACH FOR EXTRACTIVE TEXT SUMMARIZATION

‘SUMZBOT’

Vidushika Dilshani Ruwanpathirana

A dissertation submitted in partial fulfillment for the requirement for Bachelor of Engineering (Honours) degree in Software Engineering

Department of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

University of Westminster, UK

2021

Abstract

The automated text summarization is an actively researched area in NLP and several techniques have been discovered. With the rapid growth of the data on the internet, automated summarization has been applied to many domains. Recently online reviews have become a very popular method that people using to express their experience and opinions on products and services out there. With the growth of the e-business, online reviews were also increased and became more complicated to handle. So the SumzBot approach introduced an enhanced hybrid extractive summarization approach for review summarization with the aim of overcoming the current accuracy in extractive summarization. The proposed system used semantic based summarization by combining Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA) algorithms. The proposed approach has shown results that can compete and outperform the existing systems. The Sumzbot approach was evaluated using the ROUGE toolkit and the approach was able to achieve a high precision score compared to existing works.

Keywords: Extractive Summarization, Semantic-based Approach, Latent Semantic Analysis, Latent Dirichlet Allocation, Review summarization