Informatics Institute of Technology

In collaboration with

University of Westminster, UK

BEng. (Hons) in Software Engineering

A Neural based NLP approach for Tamil Text Summarization

A dissertation by

Miss. P.Vinojini

Supervised by

Mr.Prasan Yapa

Submitted in partial fulfillment of the requirements for the

BEng(Hons) Software Engineering degree Department

of Computing

May 2021

Abstract

Text summarization is one of the most important fields of NLP, which is in immense growth since decades. Though there is a tremendous g4owth in text summary for the languages such as English and Spanish in the languages with rich morphology such as Tamil, the research is more diminutive. Though there are some models for Tamil text summarization, the generated summary is not efficient as the features considered for the summary generation are statical features. The main drawback of the existing models for the summary generation of Tamil texts is that the semantics of words are not considered in the summary generation.

The approach considers the semantics of words when generating the summary by considering the sentences with the critical sequence of words are the most important sentences. As the sequence of words carries the meaning of words, the semantics will be considered in the generation of summary, improving accuracy. So the readability of the generated summary will be increased. The proposed approach suggests novel criteria to sort the sequences of words according to their importance.

The consideration of semantics in the generation of summary increases the accuracy and readability of the summary.

Keywords: Tamil text summarization, Sequential pattern mining, NLP,RBM