INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER (UOW)

Summarizer- An Intelligent Mailing List Summarizer of Open Source Projects

A dissertation by

By

John Sriskandarajah-2013224

Supervised By

Mr. Cassim Farook

Submitted in partial fulfillment of the requirements for the

BSc (Hons) Software Engineering Degree

Department of Computing

May 2018

Abstract

Open source projects, products and the open source community embraces and encourages open exchange, collaborative participation, rapid prototyping, transparency, meritocracy, and community-oriented development. A mailing list is one of the main communication medium that is used in open source projects. The initial step of contributing to any open source project is to first join a mailing list. In such a mailing list, one could find mass amount of conversations taking place. Developers who join these mailing list find it difficult to catch a glimpse of what the actual project is about or what should be done to solve the issue. Most new developers spend time on just going through the whole mailing list just to read through what has been discussed and find out what should be there part in the development. When bug fixers are done, the developer who is given the responsibility needs to find out the previous fixers that we done, most developers go through the mailing list which takes out value time that could be spent in development. This could be achieved by summarizing the contents of the mailing list, which leads to one of the main research domains that could be found that is text summarization. Many researches have been carried out in the text summarization domain. Most text summarization are focused on a sentence scoring mechanism, which is a traditional method. As part of this research, a semantically based text summarization approach is carried out to get more accurate results. An intelligent mailing list summarization system is proposed as part of this research.

Subject descriptors:

Artificial intelligence---Natural language processing, Natural language processing---information extraction, Search methodologies--- Heuristic function construction.

Key words:

Mailing List, Summarization, Open source projects.