



INFORMATICS INSTITUTE OF TECHNOLOGY (IIT)

In Collaboration with

UNIVERSITY OF WESTMINSTER (UOW)

An Intelligent Conversation Summarization System

A dissertation by

N. D. Wageesha Ishani

Supervised by

Rathesan Sivagnanalingam

Submitted in partial fulfilment of the requirements for the

B. Eng (Hons) in Software Engineering

Department of Computing

Abstract

In the modern world, over the phone calls and VoIP calls are often used for many important and professional conversations. But capturing the important points of conversions without missing any is still a challenging task because of the unavailability of proper methods to do that. Since years ago people use the less accurate, inconvenient and time consuming manual methods to get the job done. When it comes to business conversations this has become a burning problem.

A research was carried out to find a solution for the aforementioned problem. The requirement identified by the quantitative research was an accurate, efficient and convenient method to capture important information from an over the phone or over the internet conversation. And also through the qualitative research Natural Language Processing and Machine learning approaches were identified to give a solution to this problem. The proposed solution is an intelligent conversation summarization system. The features of the system are conversion voice to text, subject-based filtration and summarization. ML feature extraction technique and NLP Term Frequency-Inverse Document Frequency (tf-idf) technique were used to implement the subject based filtration and summary creation features. Then a domain-specific conversation corpus was trained using feature extraction technique with the purpose of subject-based filtration.

As a result of this research a system was built using extractive summarization approach which is capable of summarizing a voice conversation based on its subject with a high accuracy by providing user to convenience and efficiency.

Keywords:

Natural language processing, machine leaning, feature extraction, term frequency