

UNIVERSITY OF
WESTMINSTER[⌘]



INFORMATICS
INSTITUTE OF
TECHNOLOGY

Automated Mathematical Answering System for Ordinary Level Questions

THESIS

Ainkaran Biranav

w17157843 / 2018280

Supervisor: Guhanathan Poravi

Date: 9th July 2023

Department: Software Engineering

Key Words: Natural Language Processing, machine learning

Abstract

In Sri Lanka, statistics show that every year nearly half of the students fail in their GCE O/L mathematics examination. GCE O/L determines their future hence every student must have enough knowledge in this subject. The other major problem is that mathematics resources are scarce. In the current context of all other sectors becoming digitalized, education cannot be left out. Through digitalization we can make the education system more accessible to the underprivileged students.

Hence I came up with a solution that will fulfil the student's thirst to excel in their education. This system will act as a mathematical guide to everyone who is willing to learn. For this Question and Answering system, I have built up a system using Python as the core language and Natural Language Toolkit, Regular Expressions, Python libraries. The project has three types of questions as input into the system proceeding by classifying the questions, validating, mapping it with the corpus, extracting the necessary information, generating the equation and the output of the system will provide the final solution. In essence, this system design consists of two algorithms to understand the questions. 'Sentence questions' use regular expression and rule based algorithms to split the question so as to make it understandable to the machine and Name Entity Recognition and Part of Speech to extract the information from the question. 'Sentence questions' type uses machine learning techniques to classify the question. Hence this system will be a platform in providing easy access to education for the underprivileged students of our country.

The testing results were a success as the question classification accuracy was at about 78%. Accuracy of the information extraction module -97.58%. Accuracy of the generating equation and generating solution module - 99%.