



INFORMATICS INSTITUTE OF TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

CSLS

(CEFR Based Lexical Simplification)

Project Specification Design and Prototype Design document by

Mr. Harindu Amarasena

Supervised by

Ms. Sulochana Rupasingha

Submitted in partial fulfillment of the requirements for the BEng(Hons) in Software Engineering degree at the University of Westminster.

Abstract

Common European Framework of Reference for Languages (CEFR) is an international standard for describing language ability. There are six levels in the CEFR and most of the English text in the world can be classified into one on these six levels. And everyone who knows English in the world can be classified into one of those levels as well. So there can be a scenario where the persons level and the level of the text that they are reading might be mismatched and the person might have a harder time understanding what they are reading.

So the solution that is mentioned in the paper wishes to solve this issue by creating a system where the CEFR level of a given text is translated in to a higher or a lower level making it easier for the reader to understand it. The developed system works by taking a paragraph that the user enters, calculates its CEFR level, takes a CEFR level from the user and translates the paragraph to that level without changing the original meaning.

Keywords : CEFR, python, sentence difficulty