



**INFORMATICS INSTITUTE OF TECHNOLOGY**

**In Collaboration with**

**UNIVERSITY OF WESTMINSTER (UOW)**

BEng/BEng (Hons) in Computer Science Specialized in Web and Mobile

Final year Project 2018/2019

**An English-Sinhala Mixed-Language Comment  
Analyzing System for Facebook Pages**

By

**Hettigodage Chamod Oshan Hettigoda**

IIT No: 2015300 UOW No: w1608490

Supervised by

**Sudharshana Welihindha**

6<sup>th</sup> May 2019

.....  
Chamod Oshan

.....  
Sudharshana Welihindha

## **Abstract**

In this modern revolutionized tech world, social media platforms contribute a large portion of internet users daily. Contradictory multifarious problems have been evolving in the social media domain during the past decade as a result of mixed cultural, social and language barriers around the world. Interlingual or mixed language use has become a major issue in several domains by causing code-mixing and code-switching in the textual content. This research focus on the aforementioned problem by referring to Facebook Sinhala- English mixed language use in Sri Lanka. A comment analyzing system has been developed focusing on Clothing industrial online businesses in Sri Lanka. The system categorizes extracted comments into Inquiries, maybe inquiries and not inquiries while performing a language detection. A Number of textual data processing methods and a self-developed logical algorithm was used to obtain more than 80% of accurate results. Feedback from industrial people justify the usability and effectiveness towards the improvement of communication between merchant and consumer.

### **Keywords:**

**Text Categorization, Text Analysis, English-Sinhala Mixed Language, Interlingual Text Analysis, Facebook Comment Analysis, Inquiry Identification**