INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER (UOW), UK.

SNLIDB

(Sinhala Natural Language Interface for Database)

A comprehensive approach on developing a Sinhala Natural Language Interface for Databases by using advanced Natural Language Processing Techniques.

A dissertation By

Mr. Janath Vimukthi Don Francisco

Supervised By

Mr. Rathesan Sivagnanalingam

Submitted in partial fulfillment of the requirements for the

BEng (Hons) Software Engineering Degree

Department of Computing

May 2019

Abstract

A collection of data which is sorted out in order effectively accessed, managed, updated can be identified as a data base. In retrieving data from the database one needs to make a significant query in structured query language, since the non-technical individuals find query formulation as troublesome due to lack of knowledge in programming and structured query language.

Proposed solution enables users to access data from the database by submitting natural language query through an interface. This consists of many software systems that translate natural language queries in for diverse domains. Those tools are accessible for English language, although similar systems for Sinhala language is yet not implemented. There is various amount of attempts, implementing language dependent natural language interfaces for databases, such as Hindi, Tamil, etc. As Sinhala, the native language of Sri Lankans, is quite alike with these languages, the requirement of a comprehensive natural language interface for database is proposed.

Nevertheless, building a solid connection between Sinhala language and structured query language has been the main objective of this research. The complexity of the language caused a novel algorithm in-order to make an interpretation of Sinhala language queries into structured query languages.

The targeted audience of the project SNLIDB is the individuals who are lack of knowledge in English and formulating structured query languages such as call center operators, cashier operators etc. Non-technical people can direct their question in Sinhala and retrieve relevant data. A prototype of a translating Sinhala queries into structured queries is developed and presented here.

This system was implemented and tested under a proper testing criteria and later it was evaluated. Implemented algorithms influence this system to perform well. Accuracy and precision rates were exceptionally high and this solution can be viewed as a positive research.

Keywords – Natural Language Processing, Structured Query Language, Semantic Grammar.