



**INFORMATICS
INSTITUTE OF
TECHNOLOGY**

**UNIVERSITY OF
WESTMINSTER**

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Modifying the SNLIDB

(Sinhala Natural Language Interface to Database)

A dissertation by

Mahdi Refaideen

Supervised by

Ms. Fathima Farhath

**Submitted in partial fulfillment of the requirements for the
BEng in Software Engineering degree at the University of Westminster**

1st May 2021

**© The copyright for this project and all its associated products resides with the
Informatics Institute of Technology**

Abstract

The database is a collection of data that is stored in an organized way so that it can be accessed and manipulated easily. Every system requires a database to store its data and everybody interacts with the database in some way, with or without knowingly.

To access the data from the database directly, it is required to know the formal query language that can communicate with the database. A person with no knowledge of formal query language will find it difficult to work with the database. To overcome this problem, NLIDB was introduced. Initially, the NLIDB systems were developed for the English language and over time, people started developing the NLIDB systems for their native languages too.

The Sinhala language is the most widely spoken language in Sri Lanka. Developing NLIDB systems for Sinhala languages will be helpful for Sinhala peoples to communicate with the database. SNLIDB is a system that allows a user to get data from the database via getting the user input in the Sinhala language. Since it was in its initial stage of development, it is only capable of performing SELECT statements and required some major upgrades such as performing basic SQL commands and improving the accuracy and performance to produce a better NLIDB system.

So our proposed solution is to modify the SNLIDB system, which is capable of generating SQL queries for SELECT, INSERT, UPDATE and DELETE statements, and designed a new architecture that can result in higher accuracy and performance.

Keywords: Sinhala Natural Language Interface to Database (SNLIDB), Structured Query Language (SQL), Natural Language Processing (NLP), Syntax, Analysis, Semantic analysis