SQLFINDER: HUMAN UNDERSTANDABLE LANGUAGE INTERFACE FOR GENERATING SQL QUERIES

Peduru Hewa Duneesha Suloshini

A Dissertation Submitted in Partial Fulfilment of the Requirements for Bachelor of Engineering (Honours) Degree in Software Engineering

Department of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

University of Westminster, UK

2021

Abstract

Data is the heart of the decision making process in every business, every organization, every government office. But at the same time working with data stored in the databases is very time consuming and requires special technical skills like Structured Query Language (SQL). Mainly, SQL is used to communicate with relational databases. It is an identical problem that the non-technical people facing difficulties while retrieving data from the databases. In Sri Lanka, the literary of English language is very low because the native language of the people is Sinhala, who constitute nearly 70 percent of Sri Lanka's population, which equivalents thirteen million people. As a result, the demand for Sinhala language, as opposed to English language, is high. A considerable amount of e-governance applications in Sri Lanka use databases. Therefore, to manipulate data from such database applications easily, non-technical users who are more confident with the Sinhala language, need a solution to agrees with a simple sentence in Sinhala and generate a valid SQL query.

Due complexity of the structure of Sinhala than English, a few amounts of studies have been accompanied in the field of developing natural language interfaces for databases. Therefore, literature analysis exposes, that there is a solid research gap in the field of generating formal SQL queries using Sinhala natural language sentences. Based on the literature review, there is no comprehensive tool or past work regarding this proposed research topic.

In this paper, author discusses the architecture of a system which translate Sinhala sentence into SQL query using intermediate representation based approach. The implemented system is able to get the input in Sinhala language and provide the results in same language.

The suggested solution, SQLFinder provides novel and user-friendly way to access relational databases. It was designed, implemented and tested according to establish testing criteria before being evaluated. This solution can be considered as a successful study based on the gained accuracy and precision rates.

Key words: Natural Language Processing (NLP), Relational Databases, Structured Query Language (SQL)