TELLMESQL - AUTOMATION OF NATURAL LANGUAGE QUERY TO SQL QUERY GENERATION USING NLP

Rahal Sanjula Rajapakse

A dissertation submitted in partial fulfilment of the requirement 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

The use of having a good database is crucial to any company, organization or personal needs to store the information needed in an organized manner. To access the database to save, receive and perform other actions, a computer language should be used in order to communicate with the database. Many people have very less knowledge about these languages, the most commonly used language is the Structured Query Language (SQL) which is a domain specific programing language used with database management systems. However many people find it difficult in retrieving the necessary data (filter data) from the database by writing the appropriate SQL query. Therefore this task will be very stressful and time consuming when a user has all the required information in the database but doesn't have the adequate knowledge in writing the proper SQL query to receive them. This problem which is faced by many individuals will be the aim of this research to solve.

To overcome the problem being addressed the proposed solution (TellMeSQL), which is a system that converts the natural language query being entered by the user to the appropriate SQL query, can be used. This research uses Natural Language Processing (NLP) for the process of text classification in order to analyze the natural language input query by removing the unnecessary words, identifying the conditional clauses and keywords in the query. Then from the classified text, the system will further process the identified clauses in the query and form the final SQL query. A simple web User Interface (UI) is being used to make the application more simple and easy to understand. The system was tested and evaluated and obtained higher results of accuracy and the recall value when compared with other existing similar systems. The evaluation was conducted by having proper evaluation criteria's defined.

Key words:

Natural Language Processing, Text to SQL, Database, SQL, Data Dictionary, Natural Language Query, Syntactic, Semantic