SQL Translation Based on Query Process for Training School

Khin Aye Mar
Lecturer
Department of Computational Mathematics
University of Computer Studies (Bamaw), Myanmar
Corresponding Author's email id: *khinayemar@gmail.com*
DOI:- *http://doi.org/10.5281/zenodo.3752599*

Abstract

Querying in Natural language is more convenient for the user than querying in SQL. This paper is intended to construct the query translation system from Natural Language to SQL. Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications. This paper supports users to understand SQL statement and then generates the SQL statement and query from the database. It simply accepts query sentences inputs and then inference through the lexicon database in order to produce the SQL string outputs. This system is intended to query information in Training school. They can get efficient result by using English Language query without the knowledge of SQL language. This paper is specialized for developing query and can make analysis of query sentences as far as possible.

Keywords:- *Natural Language Processing, SQL language, Tokenization, Lexicon*

INTRODUCTION

In these days, Natural Language Processing is necessary for every language processing system. The Purpose of Natural Language is to convert human Language into a formal presentation that is easy for computers to manipulate [2].Natural Language Processing systems range from machine translation to information retrieval, information extraction, question

answering, text summarization, term extraction, grammar and spelling checking.

Structure Query Language (SQL) is almost universally used with relation universally used with relational database systems to make queries and manipulate data. SQL is a database computer language designed for the retrieval and management of data in relational database management systems (RDBMS), database schema creation and modification, and database object access control management. This paper aims to develop query sentence [2].

If the users want to know about the information of Training Center, users can type query in the English Language. The users can retrieve information about students, teachers, subjects, and grades in the school. Users can query details of student's data and teacher's data, total number of students and teachers.

Users can know data by relationship from database to another database immediately. Users who want to know data easily can query by using Natural Language. By using Microsoft Access database, user can view only data. But using SQL query [3], users can select data easily and immediately. Users can view data by saving much time and effort. The administrator analyzes query sentences and checks the spelling correct or incorrect by using lexicon database that is stored words. If spelling error is found, the system displays error message. Furthermore, the administrator can manage data to update, insert and delete to know user easily. The aims of this paper are to improve query process completely, to enable to know data.

BACKGROUND THEORY

Translating an arbitrary SQL query into a logical query plan (i.e., a relational algebra expression) is a complex task [4]. Natural language Processing is the use of computers to process written and spoken language for some practical, useful, purpose to translate languages, to get information from the web on text data banks so as to answer questions, to carry on conversations with machines, so as to get advice about salary, pensions and so on.

English Sentence Parser

The core of any NLP system is the parser. The job of the parser is to examine each word in a sentence and create the parse tree that identifies all of the words and puts them together in the right way [1].

The parser in this system generates a parse tree of English sentence.

The program split the input English sentence looking for spaces and punctuation marks to identify the individual words. It also identifies its parts of speech of each word by working together with the lexicon. The parser works closer with the lexicon in doing semantic analysis. The parser and lexicon works together to pick apart a sentence and then create the parse tree [1].

Tokenization

A text needs to be segmented into words and sentences. This process is called tokenization. Tokenization divides the character sequence into sentences and the sentences into tokens. Not only words are considered as tokens, but also numbers, punctuation marks, parentheses and quotation marks. Each sentence is partitioned into a list of words and removes the stop words. Stop words are frequently occurring, insignificant words that appear in a database record, article or propositions, etc [6]. In this process, the input is an English sentence. The tokenizer tokenizes the sentence into word by word and counts the number of tokens.

Lexicon

The lexical entry represents a distinct sense of word and contains two main attributes of “Word” and “Meaning”. The example of the structure is:(See Table :-1)

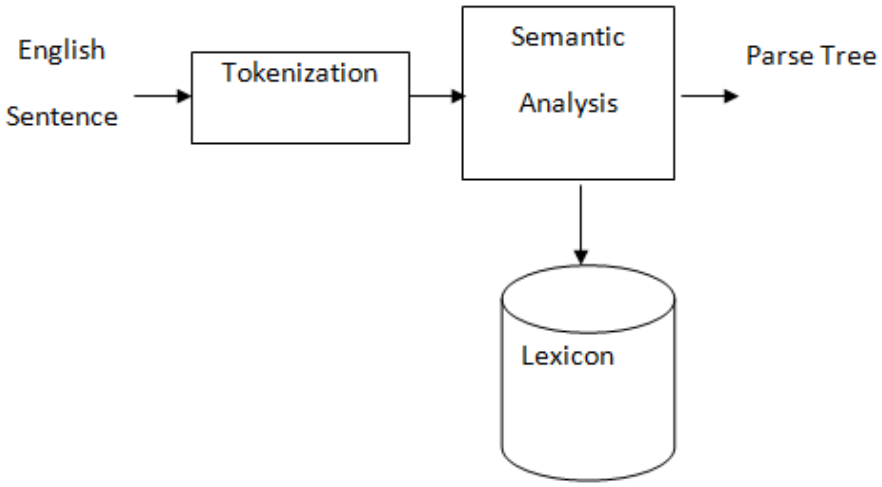

Fig 1. Parsing process for English Sentence

Table 1. Lexical database table

Word	Meaning
Show	SELECT
List	SELECT
Display	SELECT
How Many	SELECT
Describe	SELECT
Find	SELECT
Subject	tblSubject
Student	tblStudent
Teacher	tblTeacher
Class	tblClass
Time	tblTimetable
Fee	tblFee
Grade 1	Grade =1
Grade 2	Grade =2
Grade 3	Grade =3

Any language knows thousand of words. All of this knowledge is contained in a component of the grammar called the lexicon. The lexicon contains all the words and morphemes in our vocabulary and can be thought of as our "mental dictionary" [6]. The Lexical database consists of a list of the translation and contexts already presented in the human dictionary [6].

Understander and Knowledge Base

The Purpose of the understander is to use the parse tree to refer to the knowledge base. Thus the understander may answer a question using the knowledge base [4]. The Knowledge base can be divided into two parts, general knowledge base and domain/task specific knowledge base.

Generator

A generation system must be able to choose appropriately from among the different possible constructions, based on knowledge of the context. The generator would write a program in a query language to begin a search for specific information. The main points to understand are roughly what happens at each stage of analysis that fits the sentence [4, 6].

Semantic Analysis

The example of the parse tree using semantic grammar is as follow.

For a sentence:

Total number of Student.

table-name

The equivalent SQL statement:

SELECT COUNT FROM tblStudent

<QUERY> → <TOTAL>

<TOTAL> → Total number

<PP><table-name>

<PP> of <tblStudent>

<tblStudent> <TOTAL> Student

Total number <PP>

of

<tblStudent>

Student

Semantic analysis interprets the description of the meaning of each word and combines those meanings to form a meaning for the whole sentence. In the semantic investigation, the distinguished arrangement of words will be given as input.

The parse tree is produced through parser and subject, object and verb present in the arrangement of words is recognized. The output of this investigation will be the accumulation of distinguished words. A semantic grammar applies knowledge about classifications of concepts in a specific domain to the interpretation of sentences in order to pass a sentence according to its meaning [4].

Relational Database

A database management system (DBMS) is computer software designed for the purpose of managing databases. A DBMS is a complex set of software programs that controls the organization, storage, management, and retrieval of data in a database.

DBMS are categorized according to their data structures or types. A database is a computerized system whose overall purpose is to store information and to

allow users to retrieve and update that information on demand [9].

A relational database is a database that groups data using common attributes found in the data set. The resulting "clumps" of organized data is much easier for people to understand.

For example, a data set containing all the real estate transactions in a school can be grouped by the year the transaction occurred; or it can be grouped by the students total, teachers total of the transaction; or it can be grouped by the teachers' name, students' name ; and so on. Such a grouping uses the relational model. Hence, such a database is called a "relational database." [2, 7].

SQL (Structured Query Language)

SQL is a popular query language for relational database management systems (RDBMS) according to ANSI (American National Standards Institute). In the course for beginner, two types of SQL usually introduce to learners, i.e., DDL and DML [4, 5].

Structured Query Language is a database computer language designed for retrieval and management of data in relational database management systems, database schema creation and modification database object access control management.

SQL can be used to create base tables, using the CREATE TABLE statement. We then gave some examples of the SELECT, INSERT, UPDATE, and DELETE statements. SQL translation can be simplified in the well defined target language is well defined. A translator interpreted the attribute grammar generates the target language as a by-product of parsing an SQL statement [4, 5].

OVERVIEW OF THE SYSTEM

Figure 2 illustrates the Major Process of Query Process translation system.

This system consists of three main components:

- English Sentence Parsing process
- English to SQL transfer process
- SQL Language Generator

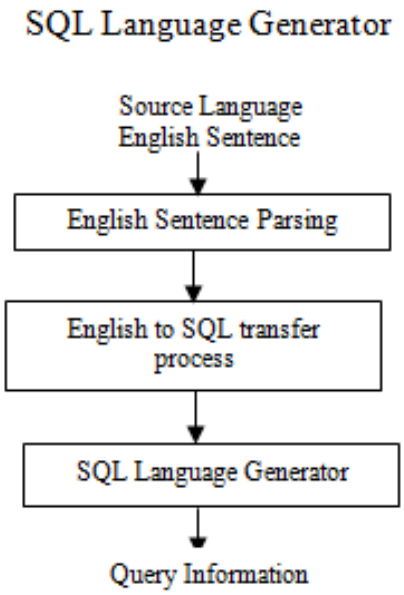

Fig 2. Major process of query process translation system

All information is store in a database and there are numerous approaches to interface with the database to access our information. A user needs some specialized knowledge to extricate information from the database. They have to utilize Structured Query Language (SQL) for information definition, information manipulation, or information control.

SYSTEM DESIGN AND IMPLEMENTATION

In this section, we present design of the proposed system in Figure 3. The modularity of SQL change was additionally appeared. The algorithm naturally maps NL inquiries to SQL questions, which could be executed on web table or RDBS to find the solution. The implementation system is explained in next section. In this system, user need to process login. After that, user can enter input statement (For example Total number of student.) to the system. The system breaks the English sentence into individual's words and extracts the keywords. The words that are tokenization phase will be checked by semantic analysis and using the lexicon database that has stored words. The keywordthat the system extracts are words that refer to table name and field's names. Query process orders the keywords and indicators and then produces SQL statement. Finally, the system executes SQL statement and produce result data to the user.

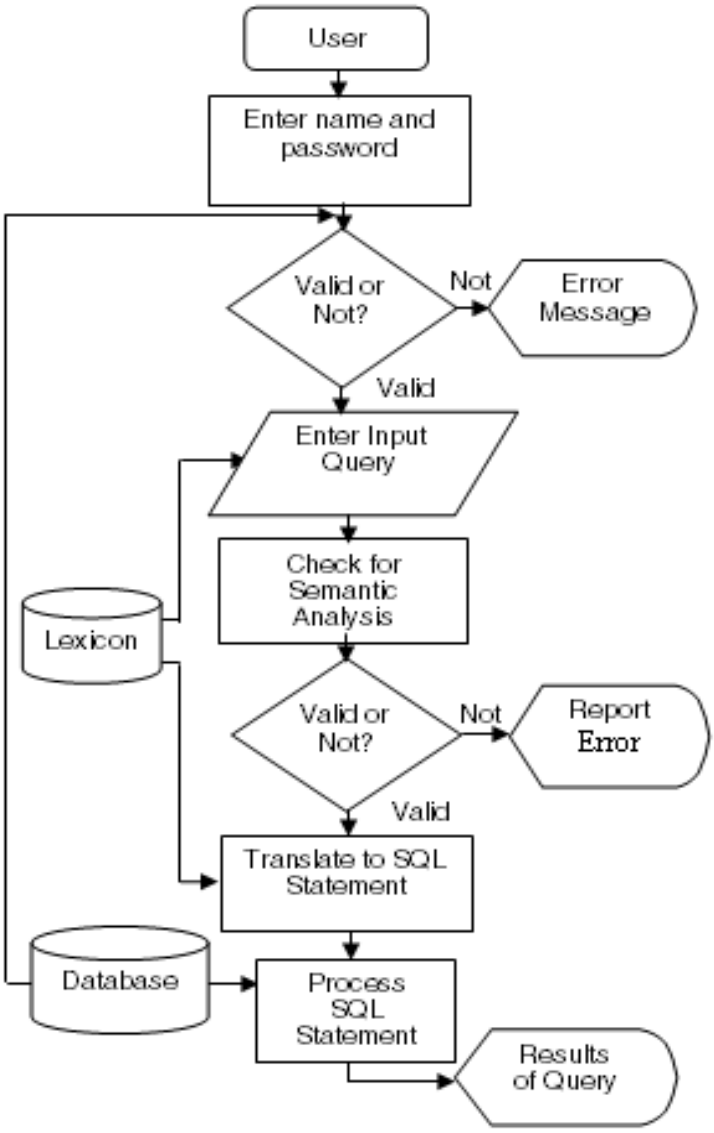

Fig 3. System design of the proposed system

Implementation of the System

This system is developed by using VB.net. It accepts the English sentence as an input and produces the results of SQL statement and produce result data as an output. The system has the following components:

- Translation from the English Language to SQL Query - The system accepts the English language query sentence as an input text and translates results of SQL query statement and result data as an output.

- Entry/Edit Lexicon - The system allows the administrator to add, update, and delete to the Lexicon.

- Viewing Lexicon - The system shows the Lexicon table in the form of grid.

- Adding/Updating Database - The system allows the administrator to add, update and delete to the Database.

This system represents main function as query in data developing with SQL translation. It describes the information of students and teachers. In administrator portion, the system checks user name and password. If it is successes, the authorize user can update data to the system's information and lexicon database. All users can view the desired data from the database. If the users want to translate SQL statement, user needs to input the English Language Sentence. In this input query statement, the system does not supports for logical operations (and, or, not) and relational operations (=, >, <, >=, <=,! =).

After that, the system checked the tokenize words by using semantic analysis from lexicon. If the user's query sentence is valid, translate sentence by using words and meanings from the lexicon then, process the SQL statement from database.

For a sentence:

List of Teacher Who Teach English?

The equivalent SQL statement:

SELECT * FROM tblTeacher WHERE English=1

<QUERY> ⟶ <List>
<List> ⟶ List<PP><table-name><PP>
<PP> ⟶ of
<table-name> ⟶ tblTeacher
<tblTeacher> ⟶ Teacher
<PP> ⟶ Who Teach <column-name>
<column-name> ⟶ English

For a sentence:

How Many Student Who Study in Grade 8.

The equivalent SQL statement:

SELECT COUNT FROM tblStudent WHERE Grade=8

<QUERY> ⟶ <HOW MANY>
<HOW MANY> ⟶ How Many<table-name><PP>
<table-name> ⟶ tblStudent
<tblStudent> ⟶ Student
<PP> ⟶ Who Study<PP><column-name>
<PP> ⟶ in
<column-name> ⟶ <GRADE>
<GRADE> ⟶ Grade/<value>
<value> ⟶ 8

After the transaction process, the system will be displayed the SQL statement and the result of query data. If the sentence is not valid, the system displays report error to the user. By using these facts, the system can support users to develop data who want to know easily and fast.

CONCLUSION

This paper provides Natural Language Processing (NLP). This system especially emphasizes on semantical analyzing of the query sentence. In this paper, the system

helps the users who want to know data by query process. The purpose of this study is to develop query processing. The system assists as an advisor for people who have no knowledge of SQL statement and queries data easily. By using this system, it is good for those who are studying NLP because it is based on NLP theory and SQL Translation. As a result, the system can help in developing the computerized system for SQL translation in data processing. This system can be extended for data analyzing process. This system can be developed to include pass or fail result of students, percentage of pass or fail students and compare number of students and teachers year by year.

REFERENCES

I. Praveena Mydolalu Veerappa, Ajeet Annarao Chikkamannur, Syntax and Table Aware Parsing Based Naturalized Structured Query Language, International Journal of Intelligent Engineering and Systems, Vol.12, No.3, 2019
II. Jan Van den Bussche, Stijn ansummeren, Translating SQL into the Relational Algebra
III. Joseph Fong, Wilfred Ng, San Kuen Cheung, Ivan Au, Positioning-Based Query Translation between SQL and XQL with Location Counter.
IV. Hans-Petter Halvorsen, Structured Query Language, https://www.halvorsen.blog
V. S.Cannan and G.Otten: SQL -The Standard Handbook, Maidenhead, UK: McGraw- Hill International, 1993.
VI. Thierry Fontenelle, Microsoft Speech and NL Group, Identifying tokens: Is word-breaking so easy? 2005.
VII. http://en.wikipedia.org/wiki/Database Management System

Cite this Article as

Khin Aye Mar (2020) "**SQL Translation Based on Query Process for Training School**" Journal of Networking, Computer Security and Engineering, 5 (1), 11-20

http://doi.org/10.5281/zenodo.3752599