International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211
Volume 9, Issue 3, March-2021, Impact Factor: 7.429, Available online at: www.ijaresm.com

Mahatma Education Society's
Pillai College of Arts, Commerce & Science - Autonomous
New Panvel

NAAC Re-Accredited 'A' Grade (3rd Cycle)
ISO 9001:2015 Certified

Department of B.Com. Financial Markets & Department of B.Com. Accounting & Finance Organizes

2 DAYS INTERNATIONAL CONFERENCE
on
"TREND & IMPACT OF E-TECHNOLOGY ON BUSINESS IN 21st CENTURY"
DATE: 27th Saturday & 28thSunday March 2021

Electronics technology (E technology) is the application of scientific theories and principles in the design, production, installation, testing, service, use, and control of electrical and electronic parts, equipment, and systems. Electronics technology is used across all businesses.

We are in the 21st Century and it'seasier to set up a business now than ever before. E-technology plays a vital role in enhancing business by promoting their business on the web, on media, and many other sources that technology created. It has numerous positive effects on modern business in this period of time. E-technology adds various new techniques of promoting businesses whether it is a small level business or business of high level.

International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211
Volume 9, Issue 3, March-2021, Impact Factor: 7.429, Available online at: www.ijaresm.com

Use of Natural Language Processing in the process of "Translating Natural Language to SQL Queries"

Aditi Save

Prahladrai Dalmia Lions College, aditi.s@dalmialionscollege.ac.in, 9987598935

----------**********----------

ABSTRACT

Natural Language Processing holds great promise for making computer interfaces easier to use as it facilitates people to talk to computers in their own languages. This is essential in ever evolving current business era. NLP can be used to express programming ideas, thus increasing the accessibility of programming to non-expert users. This paper is developed to create easy understanding of the different steps involved in Natural Language Processing. This paper talks about how we can communicate with the computer in Natural Language instead of using SQL queries to retrieve or update operational or relational databases widely used in daily business transactions. It also describes each method used to translate Natural Language to SQL Statements like Parsing, use of CFG Grammar, Regular Expressions.

INTRODUCTION

In today's technologically enriching world, human efforts are getting replaced by advanced computer systems. In the era of Artificial Intelligence, revaluation is taking place in the various business domains. Advancement of applications involves various computer systems that understands natural or regional languages. Text based or speech based CHATBOT is the popular example for the same.

Natural Query Language provides a facility to type SQL commands in simple English Language.
NQL is based on the current research topics like Artificial Intelligence and Natural Language Processing. It also uses some part of Compiler theory to validate input given by the user. Regular Expressions are used to validate the input given to the NQL.

The aim of the NQL is to translate Natural Query to its equivalent SQL Query. This paper primarily focuses on translating four major queries Select, Update, Delete and Insert. Other complex SQL queries can be considered on the same ground.

ARTIFICIAL INTELLIGENCE

AI has several definitions.
The study of Computer systems that attempt to model and apply the intelligence of the human mind. AI is a field of computer science that seeks to comprehend and implements computer-based technology that can simulate characteristics of human intelligence include learning, adapting, reasoning, self-correction and much more.

AI is the study of mental faculties using various computational models. Search, Pattern Recognition, Machine Learning, Deep Learning, Genetic Algorithms are some of the branches of the AI. Game Playing, Understanding Natural Language, Computer Vision, Heuristic Classification are some of the applications of Artificial Intelligence.

NATURAL LANGUAGE

Natural Language is any language that humans learn from their environment and use to communicate with one another. The study of structure of language units and their relationships is called *syntax.* It is easy for humans to speak and understand natural language.

However, in case of the computers to understand language like humans is difficult as linguistics is also concerned with semantics, which is the study of relationship between the language structure and the meaning intended. For E.g. **The tree sang the chair**, is not a meaningful sentence though its grammatically correct. It is not enough for a sentence to be in the correct form, it must also make sense. It means that the sentence should be correct both *syntactically* as well as *semantically*.

NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP) is an area of research that explores how computers can be used to understand and process the natural language. It can be text or speech both. NLP aims to design a computer system on the same way in which human beings understand and use language. Various tools and techniques can be developed to make computer systems understand and process natural languages to perform the desired tasks.

The practice of building computer programs that understand natural language involves three major problems: Thought Process, Representation and linguistic input and World knowledge.
There can be various levels in NLP system. But generally, it has following main levels.

Word Level:
To determine the morphological structure or nature of the word such as part-of-speech, meaning.

Sentence Level:
To determine the word order, grammar, sentence structure, meaning of the entire sentence.

Context and overall Environment or Domain Level:
A given word or a sentence may have a specific meaning or connotation from a given context or domain and may be related to many other words and/or sentences in the given context.

V. NLP TEXT PROCESSING SYSTEM [7]
NLP Text Processing System involves mainly following three phases:

Information Extraction:
Manipulation of text for knowledge extraction, for producing text in a desired form, has been recognized an important research area in NLP.

Parsing:
Parsing transforms input text into a data structure, usually a tree, which is suitable for later processing and which captures the implied hierarchy of the input. Parsing plays an important role in translation. It is the process of identifying structures in data. To process, it requires a set of Grammar to be defined.

Translation:
The central task for natural language text processing systems is the translation of potentially ambiguous natural language queries and texts into unambiguous internal representations on which matching and retrieval can take place.

Translation involves two main steps: Decoding the meaning of the source text; Re-encoding this meaning in the target language.

COMPILER

Grammars:
The most common way to represent grammar is a set of production rules. Although details of the forms that are allowed in rules vary, the basic idea remains the same.
In NLP, we generally use Context-Free Grammar (CFG).CFG is a formal grammar in which every production rule is of the form
$V \rightarrow w$
where V is a no terminal symbol and w is a string consisting of terminals and/or non-terminals.

A CFG G can be defined as a 4-tuple:
G = (Vt, Vn, P, S) where

- Vt is a finite set of terminals
- Vn is a finite set of non-terminals
- P is a finite set of production rules
- S is an element of Vn, the distinguished starting non-terminal

Production Rules (PR) Defined for NQL:
PR for Natural Select Query:
S -> ADBC | ADCB
A -> VP ADJ DET
B -> PRP DET noun

C -> ADV phrase | Epsilon
D -> phrase | Epsilon
VP -> Verb | Verb me
VERB -> show | give | list | return | select | retrieve
NOUN -><tblnm> | <colnm>
DET -> the | Epsilon
PRP -> of | from
ADV -> where | having | with
ADJ -> all | Epsilon
PHRASE -><valid character set>

PR for Natural Update Query:
S -> ABC | ACB
A -> VP noun | VP phrase
B -> PRP noun
C -> ADV phrase | Epsilon
VP -> Verb
VERB -> change | set | update
NOUN -><tblnm> | <colnm>
PRP -> of | from | in
ADV -> where | having | with
PHRASE -><valid character set>

Grammar for Natural Delete Query:
S -> ABC | ACB
A -> VP ADJ DET noun
B -> PRP DET noun
C -> ADV phrase | Epsilon
VP -> Verb
VERB -> delete | drop | remove | truncate
NOUN -><tblnm> | records
DET -> the | Epsilon
PRP -> of | from
ADV -> where | having | with
ADJ -> all | Epsilon
PHRASE -><valid character set>

Grammar for Natural Insert Query:
S -> ABCC | ACCB
A -> VP | VP noun
B -> PRP noun
C -> noun phrase | Epsilon
VP -> Verb
VERB -> insert | add | put
NOUN -><tblnm> | <colnm> | vals | values | cols | columns
PRP -> in | into | to
PHRASE -><valid character set>

Regular Expressions (RE):
A regular expression, often called a pattern, is an expression that describes a set of strings. They are usually used to give a concise description of a set, without having to list all elements. In most formalism, if there is any regular expression that matches a particular set then there are an infinite number of such expressions.

Regular Expressions are created for various types of Natural queries to validate the input accepted by the user.

REs for Natural *Select* Query:
1. (?i)(^(show | give | list | return | select | retrieve) \s ([\D\S\W]+) \s (from | of) \s ([\D\S\W]+)\S$)
2. (?i)(^(show | give | list | return | select | retrieve) \s ([\D\S\W]+) \s (from | of)\s([\D\S\W]+)\s(where | having | with)\s([\D\S\W]+)\S$)
3. (?i)(^(show | give | list | return | select | retrieve) \s ([\D\S\W]+) \s (where | having | with)\s([\D\S\W]+)\s(from | of)\s([\D\S\W]+)\S$)

REs for Natural *Update* Query:
1. (?i)(^(change | set | update) \s ([\D\S\W]*) \s (from | of | in) \s ([\D\S\W]+)\S$)
2. (?i)(^(change | set | update) \s ([\D\S\W]*) \s (from | of | in) \s ([\D\S\W]+) \s (where | having | with) \s ([\D\S\W]+)\S$)
3. (?i)(^(change | set | update) \s (from | of | in) \s ([\D\S\W]+) \s (where | having | with) \s ([\D\S\W]+) \s (as) \s ([\D\S\W]+)\S$)
4. (?i)(^(update | set | change) \s ([\D\S\W]+)\s (where | having | with) \s ([\D\S\W]+) \s (from | of | in) \s ([\D\S\W]+)\S$)

REs for Natural *Insert* Query:
1. (?i)(^(insert | add | put | fill) \s (to | in | into) \s ([\D\S\W]+) \s(cols |columns)\s([\D\S\W]+)\s(values | vals)\s*([\D\S\W]+)\S$)
2. (?i)(^(insert | add | put | fill) \s (cols | columns) \s ([\D\S\W]+) \s (values | vals) \s([\D\S\W]+)\s(to | in | into)\s([\D\S\W]+)\S$)
3. (?i)(^(insert | add | put | fill) \s (to | in | into) \s ([\D\S\W]+) \s (values | vals)\s ([\D\S\W]+)\S$)
4. (?i)(^(insert | add | put | fill) \s (values | vals) \s ([\D\S\W]+) \s(to | in | into) \s ([\D\S\W]+)\S$)

REs for Natural *Delete* Query:
1. (?i)(^(delete | drop | remove | truncate) \s (((all)\s(the)\s(records)) | ((all)\s(records))|((the)\s(records))|(records)|(all))\s(where | having | with) \s([\D\S\W]+)\s(from | of)\s([\D\S\W]*|[a-z])\S$)
2. (?i)(^(delete | drop | remove | truncate) \s (((all)\s(the)\s(records)) | ((all)\s(records))|((the)\s(records))|(records)|(all))\s(from | of) \s ([\D\S\W]*|[a-z])\s(where | having | with)\s([\D\S\W]+)\S$)
3. (?i)(^(delete | drop | remove | truncate) \s (((all)\s(the)\s(records)) | ((all)\s(records))|((the) \s (records))|(records)|(all)) \s (from | of) \s ([\D\S\W]*|[a-z])\S$)
4. (?i)(^(delete | drop | remove | truncate) \s (((all)\s(the)\s(records)) | ((all)\s(records))|((the)\s(records))|(records)|(all))\s(from | of) \s ((the){0,1})\s*([\D\S\W]*|[a-z])\S$)
5. (?i)(^(delete | drop | remove | truncate) \s (from | of) \s ((the){0,1})\s*([\D\S\W]*|[a-z])\S$)
6 .(?i)(^(delete | drop | remove | truncate) \s (from | of) \s ((the){0,1})\s* ([\D\S\W]*|[a-z]) \s (where | having | with) \s ([\D\S\W]+)\S$)
7. (?i)(^(delete | drop | remove | truncate) \s (where | having | with) \s ([\D\S\W]+)\s(from | of)\s((the){0,1})\s*([\D\S\W]*|[a-z])\S$)

Parse Tree:
A Parse Tree is a tree that represents the syntactic structure of a string according to some formal grammar. Parse trees may be generated for sentences in natural languages,as well as during processing of computer languages, such as programming languages.

A parse tree for a grammar G is a tree where,
- The root is the start symbol for G.
- The interior nodes are the non-terminals of G.
- The leaf nodes are the terminal symbols of G.
- The children of a node T (from left to right) correspond to the symbols on the right-hand side of some production for T in G.

Every terminal string generated by a grammar has a corresponding parse tree; every valid parse tree represents a string generated by the grammar (called the *yield* of the parse tree).

IMPLEMETATION OF NLP

Source Language: English
Target Language: SQL Statement
To implement NQL can using NLP, we need several important concepts of compiler.

In NQL, user can give input in natural language. User's input is transformed to SQL query by using several phases of translation. It first creates the tokens from the input. It then matches each token with the inbuilt vocabulary. Thus, English Vocabulary is the Knowledge base for NQL. If a match is found, the token is replaced by appropriate SQL keyword(s). The string or query is created after token matching and is the *Intermediate Representation* (IR) of the query. This IR then undergoes various transformation functions to formulate a valid SQL Query. To increase the scope of the paper we need to increase the knowledge base i.e. vocabulary and translation for other SQL queries.

CONCLUSION

This paper describes a prototypical NLP tool for Natural Query Language. It puts forward a mechanism to translate Natural Query into an SQL query which is then fired on the given database and produces output in an appropriate format. Although this tool is promising, the effectiveness and the user satisfaction will have to be evaluated further. Such tools are defiantly useful in the daily business transactions, as end user doesn't need to remember the complete SQL queries to perform various database operations.

REFFRENCES

[1]. RajendraAkerkar, *Introduction to Artificial Intelligence*, Prentice Hall of India, 2005
[2]. Rich Knight, Artificial Intelligence, 2nd Edition, 15th chapter
[3]. Jurafsky and Martin, *Speech and Language ProcessingAn Introduction to Natural Language Processing, Computational Linguistic and Speech Recognition*
[4]. Ivan Bayross, *SQL, PL/SQL: The Programming Language With Oracle*
[5]. Diksha Khurana1, Aditya Koli1, KiranKhatter and Sukhdev Singh, *Natural Language Processing: State of The Art, Current Trends and Challenges*
[6]. Mohammed Bahja, *Natural Language Processing Applications in Business*, Open access peer-reviewed chapter - ONLINE FIRST
[7]. Chowdhury, G. (2003) *Natural language processing. Annual Review of Information Science and Technology*, 37. pp. 51-89. ISSN 0066-4200