

SQL Generation and PL/SQL Execution from Natural Language Processing

Swapnil Kanhe

Computer Engineering, Late G.N. Sapkal College of Engineering, Nashik, India.

Pramod Bodke

Computer Engineering, Late G.N. Sapkal College of Engineering, Nashik, India.

Vaibhav Udawant

Computer Engineering, Late G.N. Sapkal College of Engineering, Nashik, India.

Akshay Chikhale

Computer Engineering, Late G.N. Sapkal College of Engineering, Nashik, India.

Abstract— In this paper we propose a method of executing query with the databases by using a natural language interface. This is a major issue in the area of database management is to provide accurate results for the persons who don't know about language processing actually mean end user with less knowledge of Database and programming language. Normal users don't know about the formal language like PL/SQL. The issue is that how they interact with the database system and retrieve the accurate results. This paper presents an interface module that converts user's query given in natural language like English into a corresponding PL/SQL command and gets actual results. We propose the architecture for translating English Query into PL/SQL using Semantic Grammar. Natural language processing is becoming one of the most active areas in Human-computer Interaction. In the previous system only SQL query is fired on database. This system architecture describes the PL/SQL query fire on database (For Example, Cursor, Procedure, Store Procedure, Etc.). A natural language query is translated to an equivalent PL/SQL query after processing through various stages. This project proposes translating English statement into PL/SQL Query using Semantic Grammar.

Keywords— Databases, Structured Query Language (SQL), Natural Language Interface for Databases (NLIDB), Natural Language Processing (NLP), Database Management System (DBMS), Procedural Language – Structured Query language (PL/SQL).

I. INTRODUCTION

Natural Language is the language that is used by almost all human beings for communication in the real world. The term real world makes the problem much more difficult. While the term Natural Language (NL) known to be very convenient language and spoken by human beings, Natural Language Processing (NLP) is part of Artificial Intelligence (AI) that deals with systems and programs and communicate in natural language. Systems that are suitable for the processing and understanding natural language bridge between the man-machine communication barriers to a great extent.

The main purpose of NLP is to enable communication between human and computers without execution of complex Commands and procedures. NLP is the techniques

that can make the computer to understand the natural languages used by humans. Now a day's requirement of commercial system is to extracting data from a Database Management System such as MS Access, Oracle and others. Today one of the most targeted problems in the field of artificial intelligence (Computer Science) is to make machine this much interesting so then it can almost behave like a human being. The behaviors of human beings have been accomplished during machine implementation e.g. now days machines can hear with the use of microphone, speak by producing sound, see with the use of cameras, but still there are some areas where this machine development is not completely successful and some of them are to understand natural language, learning from experience and making autonomous decisions in real time environment etc.

II. RELATED WORK

A. LUNAR: LUNAR (1972) [4] is the system that uses two Databases for answering the questions like chemical Analysis and chemical References but there is some limitations on the lunar system, that Requires ATN Parser. ATN stands for the Augmented transition Network. There is some limitation on lunar system such as ATN network and also Wood's Procedural Semantics.

B. PHILIQA: PHILIQA (1977) [2] is also system known as the best system for retrieving the results from the questioning, which uses a syntactic Analysis that runs on a separate pass from the semantic understanding passes. This system mainly deals with the problems of the semantics by using of the three separate layers of semantic understanding, these layers are called as the English Formal Language as the "World Model" or "Data Base Language".

C. LIFER/LADDER: For retrieving the information of US Navy ships LIFER/LADDER (1978) [7] was designed as natural language interface to database and it is also known to be a first good database NLP systems.

D. CHAT-80: CHAT-80 (1980) [3] is developed in the prolog language which is best known as Natural Language Interface to Database in the early of eighties. In the CHAT-

80 system English text is converted into prolog expressions, which were executed against the Prolog database. The code of CHAT-80 was broadcast widely and formed the basis of several other experimental Natural Language Interface to Database.

E. ASK: ASK (1983) [1] was another system that permits the end-users to keep the system with new words and concepts in the interaction at the any stage by using the ASK system the dictionary of the database could be maintain. For providing own built-in database and to deal with the number of external database ASK was used as a complete information management system. ASK was used by the end users by requesting a Natural language.

F. TEAM: In the 1987 TEAM [6] is widely used system, it is developed over many issues like portability. TEAM is designed as the database administrator, having no limitations like there is not necessary of having knowledge of the natural language interface to the database systems.

G. RENDEZVOUS: There is also RENDEZVOUS (1977) [1] system, users can retrieve databases via natural language. In this Codd's system, special focuses on query. Paraphrasing and keep users busy in understanding dialogs when there is some problems occurred in parsing user input.

III. SYSTEM DESCRIPTION

We consider a database ORACLE or SQL Server. Within this database placed certain tables which are completely normalized. If the end user wants to retrieve the data from the table, he/she has to be technical knowledge of the PL/SQL language to make a query for the SQL Server database. Proposed system remove this part and enables the end user to access the tables in his/her language.

So if we want to view information of all employee from Employee_dtl and Employee_Mst table then we are use the following SQL query:

```
SELLECT Employee_Mst.*, Employee_dtl.* FROM
Employee_Mst INNER JOIN Employee_dtl ON
Employee_Mst.nID = Employee_dtl.nEmployeeID;
```

But the User who doesn't know Microsoft SQL or ORACLE, will not be able to retrieve the database unless he/she knows the syntax and semantics of a query to the database. However using NLP, this process of accessing the database will be much easy. So the above query will be rewritten using NLP.

IV. SYSTEM ARCHITECTURE

This paper describe the use of NLP for interacting with the database using Natural language. In this work we use English language for providing the input. In this system we assume a Database SQL Server 2008 and Default Table is use which is properly arranged. A system is developed that reduce the problem end user to interact with the database with query language PL/SQL. The end user are able to access data by provide query in simple English language. The system is developed in .NET Framework (C#)

language. The different methods used for preparing the software can be given by following step [5].

- A. *Morphological Analysis:* Separate words are analyzed into their components and non-word tokens such as punctuation are separated from the words. Morphology is the structure of word. It is also concerned with derivation of new words from existing ones, Ex. Lighthouse (formed from light house).
- B. *Synthetic Analysis:* Synthetic analysis structure indicates how the words are related to each other. Synthetic analysis is concerned with construction of sentences [5].
- C. *Semantic Analysis:* Semantic analysis is concerned with the meaning of the language. In this stage assign the meaning of structure which is produce by synthetic analysis.
- D. *Parsing:* Determine the parse tree (grammatical analysis) of a given sentence. The grammar for natural languages is typical sentences have multiple possible analyses.

This methods will be used for solving the problem [2].

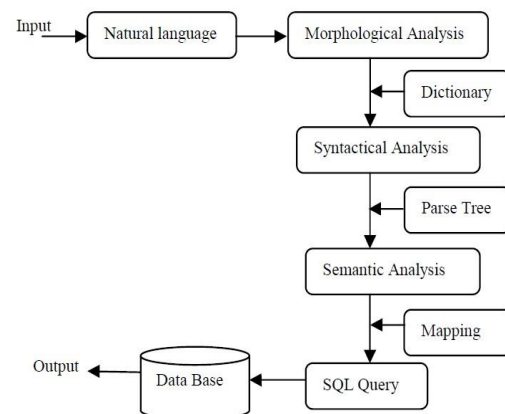


Fig 1: Structure of the System

Existing system includes the following modules:

- *GUI:* Designing the user interface where the user will enter the query in Natural Language (English).
- *Parsing:* Obtain the meaning of the Natural Query given by the user and parses it in its logical form [8].
- *Query Generation:* After the successful scanning of the statement entered by the user, the system generates a query for the statement which is entered by the user in PL/SQL and then it gives corresponding information to the back end database.
- *Data Collection:* This module obtain the output of the PL/SQL statement which is entered by the end user and display it in the User Interface Screen in the form of result.

V. OVERVIEW OF THE PROCESS

First, the English input (in the form of a list) is scanned by the semantic analysis, after table and attribute names are compared by postprocessor and if match found then joins the tables if the query include number of tables. Then the post-processor can build the PL/SQL query and appropriate output for it.

The purpose of a natural language interface is to provide the user to computer in easiest way. For this purpose, we should develop a specific dictionary to keep the similar words of the columns and tables Names. The addition of similar words makes it happen for the user to express a sentence in different easiest ways.

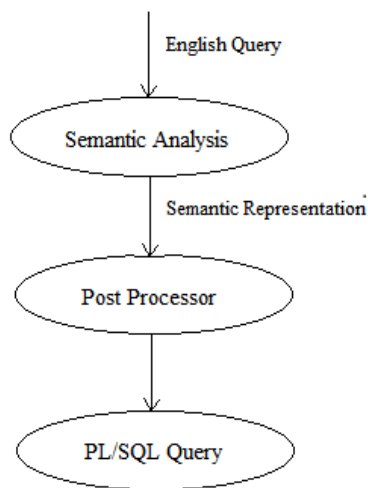


Fig 2: Database Semantics

VI. SCOPE OF THE SYSTEM

Scope of the proposed system is given below:

- To deal with any RDBMS we should know the syntax of the commands of that database software (SQL Server, Microsoft SQL, Oracle, etc.).
- Natural language processing is done on English language i.e. Input take as English Statement.
- Input from the end user is taken in the form of questions (wh- form like what, who, where, etc). Or like give etc.
- All of the possible words are included in the particular dictionary which is related to a system. Data dictionary is regularly updated when we add some information in the database.
- Data Duplication among the words will be taken care while processing the Query.

VII. ALGORITHM

- Scanning (Tokenization)
 - Break the Query in Appropriate tokens.
 - Identify the tokens and numbered it.
 - Identify the words that works as connector.

- Divide Query which emphasis on the connector tokens.
- For specify primary condition in query use connector words.
- After connector token retrieve the attributes and values.
 - For obtained attribute and resultant table map the words.
 - Change synonyms with appropriate attribute names
 - Transform English query into the Result.

VIII. RESULT

The graphical User Interface is as shown below. The user should be login and then connect to the database. A database setting is required to access the database after getting the information about host name, database name, user name and the password. When the database is ready for the use, English Query can translate very complex English queries to SQL with the capability of searching multiple tables and multiple fields.

Steps followed to get the result:-

- Type the Natural Language Statement into the Dialog box.
- Click on “Submit” button.
- The correct result of the SQL Query is display on the Screen.

Following Example describe SQL statement translated by English Query:-

The screenshot shows a web application interface. At the top, there is a "Main" header and a text input field labeled "Enter Natural Language:". Below the input field, there is a "Submit" button. The input field contains the text "Give me list of all employee". Below the input field, there is a table with the following data:

| nID | sEmpNo | sFirstName | sMiddleName | sLastName | sAddress | sPhoneNo | nD1 | nEmployeeID | sDesignation | nSalary |
|-----|--------|------------|-------------|-----------|--------------------|----------|-----|-------------|------------------|---------|
| 1 | ED001 | Kunal | T | Debmukh | Nashik Gangou... | 122456 | 1 | 1 | Junior Softwa... | 10000 |
| 2 | ED002 | Ram | N | Sharma | Nashik Road | 12245678 | 2 | 2 | Senior Softwa... | 20000 |
| 3 | ED003 | Shardul | K | Binnar | Indira Nagar N... | 12245678 | 3 | 3 | HR | 25000 |
| 4 | ED004 | Makrand | J | Kulkarni | Parijat Nagar N... | 12245690 | 4 | 4 | Team lead | 30000 |

Fig. 3: Result of query

English Query: - Give me list of all employee.

Meaning of Query: - Details of the employee with the Address, salary etc.

PL/SQL Query: - `SELLECT Employee_Mst.*, Employee_dtl.* FROM Employee_Mst INNER JOIN Employee_dtl ON Employee_Mst.nID = Employee_dtl.nEmployeeID;`

The next step of the research is to solve more complex queries.

IX. CONCLUSION

In this work we try to achieve that system is able to execute both PL/SQL queries, input by the user in his/her natural language (English). Natural Language Processing can bring powerful enhancement to virtually any computer program interface. An SQL server Database is used to store the data. Execution of the query is done by using semantic grammar technique. The aim of our project is to accommodate more and more complex queries.

ACKNOWLEDGMENT

Our special thanks to Prof. P.A. Kale for continuous inspiration and valuable guidance in throughout our dissertation work.

REFERENCE

- [1] Gauri Rao et al, "NATURAL LANGUAGE QUERY PROCESSING USING SEMANTIC GRAMMAR", International Journal on Computer Science and Engineering, Vol. 02, 219-223, 2010.
- [2] Saravjeet Kaur, Rashmeet Singh Bali, Student, MMU, Mullana, "SQL GENERATION AND EXECUTION FROM NATURAL LANGUAGE PROCESSING", International Journal of Computing Business Research ISSN (Online): 22296166
- [3] Arati K. Deshpande and Prakash. R. Devale Student and Professor & Head, Department of Information Technology, Bharati Vidyapeeth Deemed University, Pune, India , "NATURAL LANGUAGE QUERY PROCESSING USING PROBABILISTIC CONTEXT FREE GRAMMAR", International Journal of Advances in Engineering & Technology, May 2012. ISSN: 2231-1963
- [4] Anil M. Bhadgale, Sanhita R. Gavas, Meghana M. Patil & Pinki R. Goyal Pvg" S COET, Pune, Maharashtra, India, "NATURAL LANGUAGE TO SQL CONVERSION SYSTEM", International Journal of Computer Science Engineering and Information Technology Research (IJCEITR) ISSN 2249-6831 Vol. 3, Issue 2, Jun 2013, 161-166.
- [5] Ashish Tamrakar, Deepty Dubey Dept. of CSE, Chhatrapati Shivaji Institute of Technology, CG, India, "Query Optimization using Natural Language Processing", IJCST Vol. 3, Issue 1, Jan. - March 2012.
- [6] Anuradha Mohite, Varunakshi Bhojane, "CHALLENGES AND IMPLEMENTATION STEPS OF NATURAL LANGUAGE INTERFACE FOR INFORMATION EXTRACTION FROM DATABASE" , International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-3, Issue-1, March 2014
- [7] Mrs. Neelu Nihalani, Dr. Sanjay Silakari, Dr. Mahesh Motwani, "NATURAL LANGUAGE INTERFACE FOR DATABASE: A BRIEF REVIEW", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011 ISSN (Online): 1694-0814
- [8] Jasmeen Kaur, Bhawna chauhan , Jatinder Kaur Korepal, "Implementation of Query Processor Using Automata and Natural Language Processing", International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013 ISSN 2250-3153