VOICE CONTROLLED DATABASE ANALYSIS

Prof. Minal Nerkar , Aditya Salvi , Mohammad Sameer , Ashutosh Kshirsagar , Hemendra Patil

Computer Department , All India Shri Shivaji Memorial Society's Institute of Information and Technology , Pune , Maharashtra , India

Abstract : Information is an important aspect in today's growing technology. Data in an organised manner can be stored in database and Structured Query Language is used for accessing, managing and processing of the database.

The technical terms and specifications used in SQL can be challenging to the non-technical users. It may be difficult even for the experts in database programming language to know the exact schema or database structure, attributes in the different tables, the roles of various entities in the database query and the exact join operations. This may arise the need for the accessibility of the database for the non-technical users using natural language (NL).

Natural Language Interface for Database (NLIDB) is an intelligent interface designed for the access of the database. This interface allows users to provide input by typing or giving voice as an input in their natural language and gain access to the database. Intelligent Query System (IQS) designed in this interface, enables the non-technical user to fire queries in their natural language (English). The first step in this system is a conversion of voice input from the user into the text format. Then the parsing of input sentences is done by the system and from the natural language expressions SQL queries are generated. The final output can be obtained by mapping these SQL queries with the database. This system makes the accessing of the information susceptible, reliable and effective.

***IndexTerms* – Natural Language Interface for Database(NLIDB),Intelligent Query System(IQS),Structured Query Language(SQL)**

I. INTRODUCTION

As we know the application of database is very important in our life. The application of database is used almost everywhere example: it is used in information system like transport, human resource, school and college, commercial business, banking and finance.

Since the database is very massive, the queries which are generated are very difficult to comprehend .Because of this the relation between the entities is very complex. The users who are unfamiliar to the concept of database and the queries related with the database, they find it difficult to write these queries. It is very important to know the details of the database.

The biggest issue is that the user who wants to retrieve information from the database does not have any knowledge about the SQL language. To overcome this challenge we have natural language toolkit which utilises natural language processing. This helps users to interact with the system and generate queries. The biggest challenge in computer technology has been when it comes to designing a model for automatically mapping natural language semantics into programming language. For example: Normal users does not have the knowledge of structured query language and machine readable instructions. In ideal condition the users should only ask questions in natural language without any technical knowledge related to database scheme. The output is in the form of formal query language that is translated from the questions entered in natural language. Once the query is formed, the DBMS processes the query to retrieve the required data. The average user is not well versed in SQL query because they don't have knowledge about the database and their aspects like table, fields, types, primary keys etc.

In order to provide a solution to this problem the system has been created that takes the input as natural language through speech recognition and this is further converted to SQL query and the desired result is retrieved and displayed from the database.

II. RELATED WORK

[1] "System and methods for converting speech to SQL" by Sachin Kumar, Ashish Kumar, Dr. Pinaki Mitra, Girish Sundaram, Department Of Computer Science And Engineering, IIT Guwahati, Guwahati, India

To translate voice input into SQL query, the system uses speech recognition models which go hand in hand with classical rule based technique and semantic knowledge. The database schema is used by the system which is converted into the graph structure for joining of the tables. Single tables and multiple tables are checked in the system. The system checks for the syntactic consistency with the syntactic rule of the input query to give the correct results. The system need not to be manually configured for different database. This shows database independency of the system.

[2]"Constructing An Interactive Natural Language Interface For Relational Database" by Fei Li, H.V. Jagadish, Univ. of Michigan, Ann Arbor.

In this paper, an interactive natural language query interface for relational database is used. First, a natural language query is translated into a SQL statement and then against Relational Database Management System evaluation of statement is done. The entire processing of the query is explained to the user resulting a highly reliable system. The system allows the user to choose between different interpretations when ambiguity is arise. These interpretations is provided by the system itself.

[3]"Ruled based Domain Specific Semantic Analysis for Natural Language Interface for Database" by Probin Anand, Zuber Farooqui, M. Tech Scholar, Assistant Professor, All Saints' College Of Technology, Bhopal

This paper defines derivation involving the ambiguous term and domain specific rules and with this approach this paper makes a NLIDB system portable and generic or smaller as well as large number of applications. This paper only focuses on context based interaction along with SELECT, FROM, WHERE and JOIN classes of SQL query and also handles complex query that results from the

ambiguous Natural Language query. Both DDL and DMM can be executed in the system. This system is developed in Java Programming Language and various tools of Java are used to build a system. Oracle is the database.

[4]" An Algorithm for Solving Natural Language Query Execution Problems on Relational Databases" by Enikuomehin A.O., Okwufulueze D.O. Dept. of Computer Science, Lagos State University, Lagos, Nigeria.

User friendly non-expert search processing is done by modelled algorithm. The modularity of SQL conversion is also shown. The users request is efficiently processed in a human understandable format by this model. The limitations are domain independent limited on query domain.

[5]" A Rule Based Approach for NLP Based Query Processing" by Tanzim Mahmud, K. M. Azharul Hasan, Mahtab Ahmed, Thwoi Hla Ching Chak, Department of Computer Science and Engineering Khulna University of Engineering & Technology (KUET), Bangladesh.

Author has proposed a system which is convenient for the end user. The querying access is very reliable which bridges the gap between system and end user. Searching of the terminals is done by context free grammar. The data dictionary of the system is updated regularly with new words which increases the performance of the system.

[6]"Natural Language to SQL Generation for Semantic Knowledge Extraction in social web sources." by K. Javubar Sathick and A. Jaya, Department Of Computer Application, B.S. Abdur Rahman University, India

The author has proposed a system which is developed to compute both DDL and DML queries. The limited dictionary is referred where all possible words are related to a specific system are included. Equivocation among the words handled while natural language processing. Oracle database is used for the system. Java Programming Language and its various tools are used to develop this system.

III. METHODOLOGY

Step 1:
.The input from the user is supposed to be in the form of speech. After the input is given it is passed to Speech to Text Convertor and Communicator and is further converted to text form. The user have the utility to analyse and the update the text if required.

Step 2:
With the use of tokenizer the natural language query is converted to a stream of tokens and the further id is provided to every single of the NLQ.

Step 3:
Again with the help of token ids, a parsed tree is generated by the parser and a set of words is identified. A collection of identified words will be the output of this process.

Step 4:
With the use of MR generator the collection of identified words is converted into some meaningful representation.

Step 5 & 6:
Relevant attributes are extracted from the database after the semantic builder accepts the output generated by MR generator.

Step 7:
The relation between attributes and the structure of the words are extracted from the database. These are then identified in the relation identifier and lexicon builder.

Step 8:
After identifying the relation between the attributes and the word structure and semantic maps, a new relation is generated.

Step 9:
The semantic map takes the input from the query generator and the SQL query is generated.

Step 10:
The generated SQL query is then executed on the database.

Step 11:
After the SQL query is executed the output is displayed to the user.

The query generation for natural language is achieved by this algorithm.

IV. ALGORITHM

Step 1: Extricate meaningful keywords

In this step meaningful keywords are extracted from input sequence.
To achieve this the tree tagger tool is used to filter empty words with regards to their grammatical classes and also perform stemming.

Step 2: Architecture of database recovery

In order to achieve this result two methods have been implemented. The collection of information by implementing SQL queries from the database queries like "SHOW TABLES, SHOW COLUMNS, DESCRIBE, etc." To analyse the backup or database creation file is the second method. In this method we don't need a pre-processing connection to the database, but what we need is a universal SQL schema.

Step 3: Constructing a thesaurus.

In thesaurus file, we have synonyms of almost all words. The query should work for "student" as well it should work for "students". For that, this pairing is must.

Age: generation, era, period.

First name: Nickname, forename.

Step 4: Natural Query Division.

In the first step of division of natural query, a cut out of the input sentence is performed according to the tagged keywords. "Table" and "Column" found in the sentence, in order to know which segment of the sentence corresponds to which part of the request to build. The presence of SELECT and FROM segment is compulsory in the sentence.

Step 5: Query Structure Determination.

DDA and DMA are the two types of queries from the given natural input to decide which SQL clause to use. This is decided in the steps explained in the below table

Keywords	Operations
"what is the number", "How much is there", …	count
"Do not […]", "do not […]" …	negation
"bigger than", "greater than", …	superiority
"smaller than", "less than", …	inferiority
"what is the sum", "adds", …	aggregate
"what is the average", …	average

Step 6: Query Generation.

This is the last step of the algorithm that consist of the final query generation.

V. CONCLUSION

- Intelligent Query system using NLP is a system which is used for retrieving data from database easier and more efficient.
- The system which we are using is closing the technological differences between a user who is not well versed with the computer technicalities .This issue is solved by this system.
- The human speech is converted into database query by the system and is fed to the database that gives the desired output.

VI. FUTURE WORK

- In this system the process of creating queries from natural language is independent of one another. The process of searching is not achieved in a single step.
- The user may have questions regarding the output. It is essential to provide a system that can perform follow up queries.

VII. REFERENCES

[1] Sachin Kumar, Ashish Kumar, Dr. Pinaki Mitra, Girish Sundaram, "System and Methods for Converting Speech to SQL" Appeared in proceedings of International Conference on "Emerging Research in Computing, Information, Communication and Applications" ERCICA 2013.

[2] Fei Li, H.V. Jagadish, "Constructing an interactive natural language interface for relational database" Journal proceedings of VLDB endowment, vol. 8, Issue 01, Sept. 2014.

[3] Probin Anand, Zuber Farooqui, "Rule based Domain Specific Semantic Analysis for Natural Language Interface for Database" International Journal of Computer Applications (0975 – 8887) Volume 164 – No 11, April 2017.

[4] K. Javubar Sathick, A. Jaya, "Natural Language to SQL Generation for Semantic Knowledge Extraction in Social Web Sources." Middle-East Journal of Scientific Research 22 (3): 375-384, 2014.

[5] Tanzim Mahmud, K. M. Azharul Hasan, Mahtab Ahmed, "A Rule Based Approach for NLP Based Query Processing" Proceedings of International Conference on Electrical Information and Communication Technology (EICT 2015).

[6] Enikuomehin A.O., Okwufulueze D.O, "An Algorithm for Solving Natural Language Query Execution Problems on Relational Databases" *(IJACSA)* International Journal of Advanced Computer Science and Applications, Vol. 3, No. 10, 2012.