

Voice Based Natural Language Query Processing

Puja Munde¹, Sayali Tambe², Afreen Shaikh³, Pratiksha Sawant⁴, Prof. Deepa Mahajan⁵

¹B.E. Students, Dept. of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Maharashtra, Pune

⁵Professor, Dept. of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Maharashtra, Pune

Abstract - Now-a-days there is a need to store and retrieve data from the database. To use database user need to learn SQL so non-expert users' needs a system which can interact with database using natural language such as English. The main aim of NLP is to ensure communication between human and computer to retrieve data. A voice based user interface is used to get desired output. Speech recognition technique is used to design this system. User speech is given as input and then converted into text and then text to SQL query and result is displayed in tabular format. Technique such as analyzer, parser, matching and dictionary mapping are used to get result in tabular format. NLP technique helps in solving simple as well as complex query. The most important applications of natural language processing include information retrieval and information organization, machine translation.

Key Words: Natural Language Processing, Structure Query Language, Speech recognition

1. INTRODUCTION:

The main objective of Natural language processing is to enable communication between human and computer. This system will help users who do not have knowledge about the query languages such as SQL (Structured Query Language). It mainly useful for Placement cell officers who work on student database can use the technique to extract data. NLP is a technique that makes the computer understands the languages used by humans. While natural language may it easy for people to learn and use, it has been proved to be hard for a computer to master. A NLQ Interface to Database system is an application that accepts a natural language query, creates a SQL query from it and executes it to retrieve the data from relational database. The result is retrieved from the database is a stream of elements. Speech recognition is the ability of a machine or program to recognize words and phrases which are either spoken or written in text form to a system-understandable format. The software accepts the Natural Speech and convert into text. Then text is converted into SQL query to get desired results. SQL is better than Excel sheet to store and retrieve data. There are various operations which can be efficiently performed by SQL than Excel. There will be three options for the user-one where they can view all the table .They can enter a query in SQL and Enter Query in natural language. Then query will get converted to SQL query. Many challenges like Ambiguity and mapping are faced during the process as mapping of attributes with table are done.

2. LITERATURE SURVEY:

This section reviews the different works concerned with the project. Some papers were studies and summarized as follows:

Paper[1] Anum Iftikhar, Erum Iftikhar, Muhammad Khalid Mehmood proposed a system for "Domain Specific Query Generation from Natural Language Text". It generally involves generation of SQL Queries using Stanford Parser. The paper revolves various ambiguity problems in NLP. Automated queries of NoSQL can be used as an application for the idea presented in the paper. It can also be used to design NL business etiquettes.

Paper[2] Hanane Bais, Mustapha Machkour, Lahcen Koutti proposed a system for "An independent domain Natural Language Interface for Relational Database :Case Arabic Language". The aim of this system is to allow communication between database and human user using Arabic language. They propose a Generic Arabic Natural Language Interface to database. This interface allow user to access data in database by answering question in Arabic language.

Paper[3] Shravankumar Hiregoudar, Karibasappa K G, Manjunath Gonal proposed a system "Speech to SQL Generator –A Voice Based Approach".A user interact with system through a voice based user interface to fetch the desired output. The system show that voice based user interface is an effective method of fetching and querying data from database.

Paper[4] Ruslan Posevkin ,Igor Bessmertny proposed a system "Translation of Natural language queries to structured data sources". User interface interact to database that contain information about existent program libraries and framework. Computational linguistics NLP methods are discussed and technique used.

Paper[5] Hangu Yeo proposed a system "A Natural Language Question and Answering System for Healthcare Data Search using Complex Queries". The proposed auxiliary system is machine learning based and extends existing NLIDB system to help it answer the complex queries.

Paper[6] MEINA SONG¹, (member, IEEE), ZECHENG ZHAN¹, AND HAIHONG E.¹proposed a system "Hierarchical Schema Representation for Text-to-SQL Parsing with Decomposing Decoding". A schema-aware neural network with decomposing architecture, namely HSRNet, which aims to address the complex a Text-to-SQL generation task.

2.1 DIFFERENCE BETWEEN SQL AND EXCEL:

SQL	Excel
1. It is used for large, heavy tasks.	1. It is used for small tasks
2. Works on large, infinite rows.	2. Works on limited rows
3. Separate analysis of data from database is done.	3. Separate analysis of data from database is not done
4. With the help of natural language processing there is no need to remember any formula to perform any tasks.	4. Difficult to remember the formulas to perform tasks
5. SQL flexibility doesn't create any risk.	5. Excel is flexible but with flexibility comes risk
6. SQL can easily replicate an old data on new data.	6. Hard to replicate an old data on new data
7. Works on multiple tables at a time.	7. Works on only one table at a time
8. Used for solving complex queries.	8. It is used for solving simple queries
9. Data integrity is maintained.	9. Data integrity is not maintained
10. It's secured.	10. It's not secured

3. RELEVANT MATHEMATICAL MODEL:

1. Accept the input from the user in the form of speech and convert it to text.
2. The conversion of speech to text is done by Speech Recognition Technique.
3. Split the input query and store it in a list, i.e. tokenize the input sentence.
4. Find all the attributes of all the tables which are required.
5. Examine the query and find the table presents in the query and the attributes present in the query.
6. Find the attributes which belong to table present in the query.
7. Find the attribute which do not belong to the table in the query and separate them.

8. Now find the tables which will contain the pair of attribute which do not belong to the table in the query, other attributes present in the table in the query.
9. Select any one table. Thus we will obtain the tables required for natural join.
10. For a natural join query, find out the common attribute of the two tables and form the inner query. Then form the outer query according to the different conditions. Merge both of them and generate the final query.
11. For a simple query, generate the final query by checking the different conditions accordingly.
12. If there are 2 tables, then perform a natural join on the 2 tables with appropriate attributes of the tables.
13. Obtain the conditions of the where clause, aggregate function and the relational operators between the conditions from the list of attributes. Add these to the final query.
14. Display the result on GUI.

4. SYSTEM ARCHITECTURE:

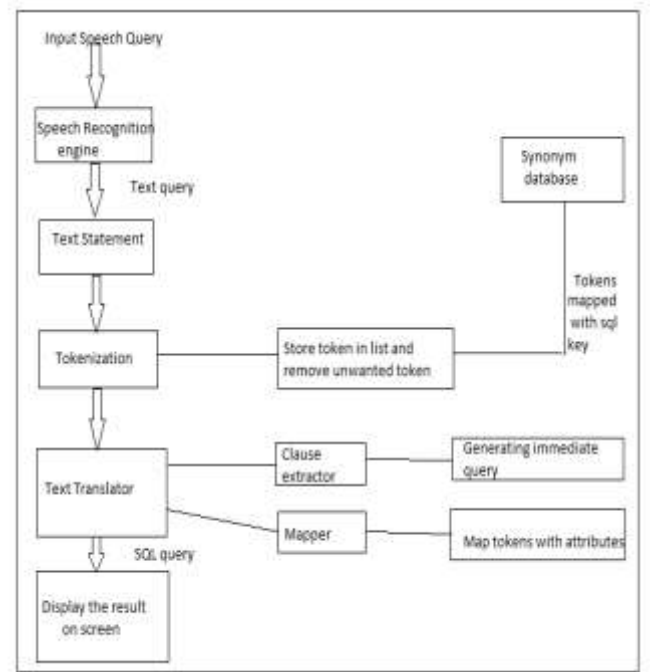


Fig.1. System architecture for NLP

As given in the architecture, Fig.1, the systems takes input as spoken query language and sent it to speech recognition engine. The output will be the input query text extracted from the speech. The accurate input query is extracted and sent to tokenization.

A. Tokenization

System will perform tokenization on the entered query by separating it into single words. Each word represents a token. Then these words will be stored in a separate list and passed to Lexical Analyzer.

B. Lexical Analysis

The tokenized list will be mapped with the dictionary. These words will get replaced by the database words from the dictionary and passed to syntactic analysis.

C. Syntactic Analysis

In this step dictionary of table names, attributes and keywords are maintained. Each tokenized word gets mapped with attributes in the dictionary. It is passed to Semantic Analysis for further processing.

D. Semantic Analysis

System will find words which represent conditions or symbols and that word will get mapped with the dictionary. (For Example: If there is "less than or equal to" in the query, it will get mapped with the symbol "<=").

5. ADVANTAGES:

1. User can easily access the system to convert natural language into SQL.
2. System provides options such as viewing the table, Enter SQL statement, Enter natural language.
3. System will prompt user about the error while entering the query and provide auto-correction feature.
4. System is eco-friendly to use.
5. Handles complex queries.

6. RESULTS :



Fig. 2.View Sections

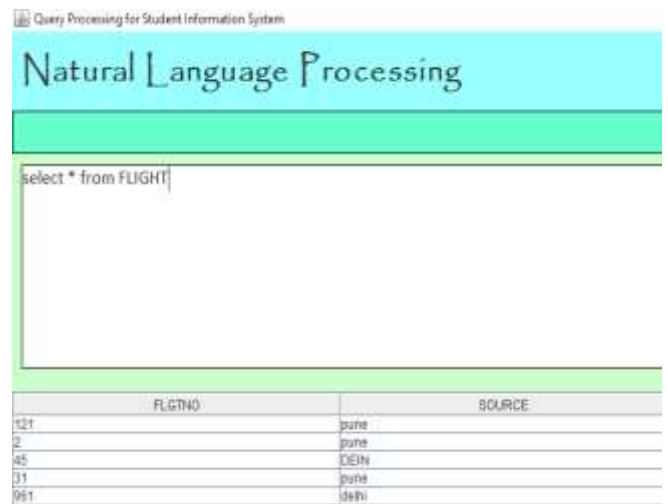


Fig.3. SQL Converter

Above figures shows sample results generated by the system. First student need to register and enter the details required for placement. While Admin has access to view and search the data from database.

7. CONCLUSION

The main goal of this system is to allow communication between database and its human users using natural language. Use of Natural Language brings ease for any human being. This system will help T&P officer to easily retrieve and manage data from student database using simple English language. There is no need for the user to learn complex query syntax to retrieve data. The facility to accept the input in speech format makes the system user-friendly.

8. REFERENCES

[1] Anum Iftikhar, Erum Iftikhar, Muhammad Khalid Mehmood "Domain Specific Query Generation from Natural Language Text", The Sixth Generation Conference on Innovative Computing Technology,2016 IEEE.

[2] Hanane Bais, Mustapha Machkour, Lahcen Koutti "An independent domain Natural Language Interface for Relational Database: Case Arabic Language",2016 IEEE.

[3] Shrivankumar Hiregoudar, Karibasappa K G, Manjunath Gonal, "Speech to SQL Generator –A Voice Based Approach". International Journal of research in signal processing, computing and communication system design,Volume 4,Issue 2,2018.

[4] Ruslan Posevkin, Igor Bessmertny, "Translation of Natural language queries to structured data sources",2016

[5] Hangu Yeo, "A Machine Learning Based Natural Language Question and Answering System for Healthcare Data Search using Complex Queries",2018 IEEE International conference of Big Data.



[6] MEINA SONG¹, (member, IEEE), ZECHENG ZHAN¹, AND HAIHONG E.1 ,“Hierarchical Schema Representation for Text-to-SQL Parsing with Decomposing Decoding”, Volume 4,2016 IEEE Access.