

VOICE BASED SQL QUERY USING NATURAL LANGUAGE PROCESSING

Mr.M Jaiganesh Asst.prof

*Computer science and Engineering
SRMTRP Engineering college, Trichy, Tamil Nadu, India*

J M Akshaya Mohan, T Monisha, V subasri, R Yalini

*Computer science and Engineering
SRMTRP Engineering college, Trichy, Tamil Nadu, India*

Abstract- Natural Language is one of the easiest systems for people to Learn and use [1]. The goal of NLP provides communication between people and computers without resorting to memorization. Implement the framework for translating English queries into SQL queries using semantic grammar. The system will take the user's query in natural language as input in the form of voice and convert it into a query [2].

Keywords:

NLP, SQL, Natural language, Query

I.INTRODUCTION

One of the major fields in computer science is natural language which is mainly concerned with the interaction occurring between the computer and the human language .There it contains more attractive areas in the interaction between human and the computer. These include spoken language[3] that incorporate speech and natural language. There should be a foregoing knowledge about DBMS(Database Management System) in the direction of extracting information from the database. The DBMS is a software developed for the purpose of storing and manipulating the data in a database. Here an unspecialized user finds trouble in extracting the data. To discover a solution for such a problem and ease human interaction with computers[4], Natural language processing techniques are worned. Natural language processing has applications in varied sectors like tourism, where a tourist can get information about the famous tourist spots in a particular city, the hotels available, best places nearby and so on. our work spotlights on extracting the correct query by providing the input in the form of voice. The conversion of a natural language into a query to ease the data extraction by granting the input natural human language in the form of voice is the main intension of the paper.

II.LITERATURE SURVEY

For the utilization of Natural Language Processing various methodology have been discovered in the beginning. At first, [5]In 2010,Gauri Rao found a technique called Natural language processing using semantic grammar for the semantic rules construction but it result in the drawback of Manually labeled the complex questions. Later, In 2014,[6]Akshay G proposed a technique called a proposed Natural Language Query Processing System for word pair mining which gives the better process than anything else by providing more accuracy but it leads to the disadvantage of may need to handle some bugs in the process. Later in 2015[7], Javubar sathik proposed a system for Morphological analysis which provide the way to extract the sources from various web using the natural language and this process have been successful for some particular content of SQL. This proposed system has been named as Natural Language to SQL generation in extracting semantic knowledge from social web sources. But it leads to difficulty to extract the semantic content of the query.After this ,In 2016[8],Garima singh proposed a technique called an algorithm to transform natural language into SQL queries for relation database from the Three tier architecture of NLTSQLC which mostly result in the drawback of Computational of several data.

III.METHODOLOGY

English is the most preferred one in Natural Language Processing i.e., the input dictum have to be specified in the language of English. Input query can be taken from user in the form of voice. And the processing of input query is used by Query formation. i.e., the process of breaking the query of the user into tokens. These tokens are the individual words in the query. Further extract the keywords to execute the query.

IV.EXISTINGSYSTEM

The existing model of the Natural language processing is that it converts the Query from Natural language[9] which is given as text i.e., Input is in the form of text. At first the Natural language processing step is done by the following steps.

1. TOKENIZATION:[10]The activity of partitioning the string sequence into keywords, words, phrases and symbols called tokens.

2. LEMMATIZATION:It is a process of grouping together the inflected forms words so they can be analyzed as a single item, identified by the word's lemma or dictionary form

3. SYNTACTIC ANALYSIS: It is a procedure of examining a string of symbols, either in natural language computer language or data structure for the formal grammars rules

4. SEMANTIC ANALYSIS: From the level of phrases, clauses, sentences and paragraph a relating process of syntactic structure is occurred to the level of the writing as a whole, to their language independent meanings is called as semantic analysis.

5. MAPPING:

Which consists of two modules

ATTRIBUTE IDENTIFICATION:

In the table of relational database, the table has rows and column which stores the attributes which is identified in the process of natural language. An attribute is a column in a database tables. The natural language consists of several words which is given as input which after that Natural language processing the output got as in the level of phrases, clauses, sentence and paragraphs then into the attribute identification. This module identifies the useful attributes in the column of the table to create the query by following a rule-based paradigm.

SQL QUERY FORMATION:

In this module, After identifying the attribute in the column of the table, the SQL query have been formed, by using the SQLmap() function as the algorithm. The algorithm has the formation of query for the train database. The algorithm has the query SELECT which has the value , train name, train no, source station ,destination station . The query have written as Query(SELECT) with the syntax as, " FROM railways.train WHERE". Finally the resultant query formed after done the tokenization, lemmatization, POS tagging, stop words removal.

6. RESULT:

In the process of formation of SQL query from natural language processing, the steps include Tokenization, Lemmatization, Syntactic and Semantic analysis[11] , POS tagging, parsing and chunking and mapping include attribute identification and SQL query formation ,the result is formed.

The output for the process is taken as SQL query from the Natural language(human language).

V.PROPOSED SYSTEM

5.1 SPEECH SYNTHESIZER:

A Synthesizer is normally an instrument which is operated through the keyboard, is used for combining and generating different frequencies of voice signal. A Speech synthesizer is used for generating the speech from the voice. It is the step from which the process of converting the speech to text is initiated[12]. The "Hidden Markov Model" is a algorithm used to convert the query.



FIGURE 1 : This figure shows the speech in English from the user is taken as input converted into text using Hidden Markov Model and Speech Recognition Library. The audio can be provided through a microphone or similar device. The input is provided by the user in the above user interface which can be accessed by both admin and user.

5.2HMM ALGORITHM

There are many applications in continuous speech recognition (CSR) it contains command and control, transcription of recorded speech, interactive spoken dialogs, and so on. The feature extraction stage seeks to give a compact representation of the speech waveform. This form should reduce the loss of information between the words and convert the words from the acoustic models[13].

THE HMM ALGORITHM for the speech to text conversion is as follows,

GOAL:

Given acoustic data $A= a_1, a_2, \dots, a_k$ Find word sequence $W=w_1, w_2, \dots, w_k$ BAYES RULE:

$$P(W|A) = \frac{P(A|W) \cdot P(W)}{P(A)}$$

P(A) is a constant for a complete sentence.

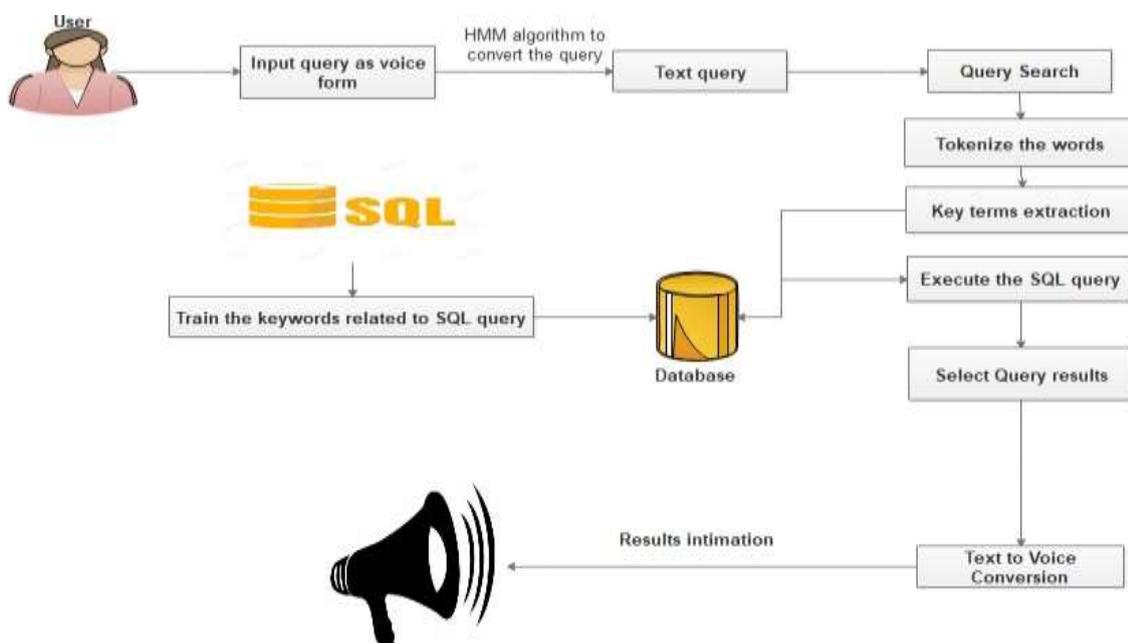


FIGURE 2: This figure shows the user is providing input as voice query converted into text query. Next step is splitting of sentences into minimal meaningful units is referred to as tokenization. Then extract the results from database and execute the SQL query to selected query results. Result can be obtained in the form of voice which is obtained from the database.

5.3 TEXTMINING

Text mining is also referred to as text data mining roughly equivalent to text analysis [14]. once it completed the conversion of speech to text, the next process steps include: Query search, Tokenization, Key terms extraction, Execution of query with train the keywords related to SQL query and Query results. The basic steps in text mining are,

1. Choosing the scope of document
2. Tokenization
3. TokenNormalization
4. Stop words removal
5. Stemming the word



FIGURE 3: This figure shows given voice query sentence is splitting into minimal meaningful units is referred to as tokenization. Then removal of usual words like a, but, of, the etc.is referred to as the stop words which are also removed. Removal of prefix and suffix words is referred to as stemming words which are also removed in the process. After these steps are done the result can be obtained from the database.

5.4 IMPLEMENTATION

The process initiated by creating the home page. The home page has Admin login and User login and search. Admin login has SQL database connection which has the access to the admin of the web. The admin has the following access.

1. CREATE atable.
2. ALTER atable.
3. ADD attributes to the table.
4. RETRIVE data from a table.
5. CREATE a userlogin.

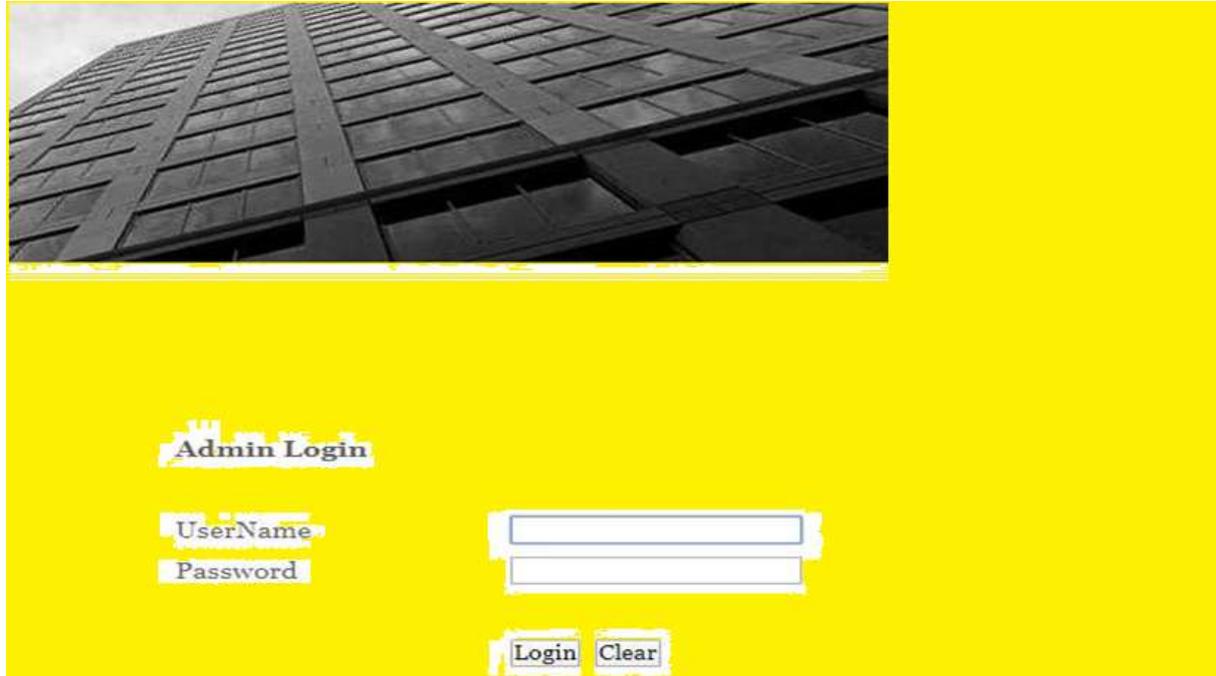


FIGURE 4: This figure shows the Admin login page. The admin page has three modules such as Homepage, Adding details to the database and the user login. The Admin have the access to creating new table, add attributes to the table, add details to the table and giving access to the table. Connecting user and admin done in this module.

The user login has the following access:

1. CREATE a new user login.
2. LOGIN using the credentials.
3. SEARCH using the search page.
4. INPUT is given in the form of voice.
5. OUTPUT is obtained as an SQL query and
6. RETRIEVE the data as in the query.

Search page done the following.

Enabling the voice symbol we can record the speech at first.

The recorded speech is converted into text by using the speech synthesizer (HMM Algorithm).

The page has query information, result and the label.

Although several technologies are already employed in the process of converting the NLP into SQL query, this process mainly employed in making more interactive framework. This process include get the input from the user in the voice format and then convert it into query without using any tools but all with the code(C#.NET).

This process include HMM algorithm and text mining process which recognize text and convert it into speech and then into subsequent steps to form a query and the retrieval of query in text format.

VI.CONCLUSION

Already there are more techniques are available but those techniques used more steps which are lemmatization, tokenization with other steps like syntactic and semantic analysis for the conversion. We developed the framework for the conversion includes the input in the form of voice and which is converted into text by using the HMM algorithm. After that the Text mining process was used. The final output got from the framework by extracting data which is already trained in the database. Finally the output got in the form of both voice and text. Then the text mining process is followed. The output of the process is then converted by retrieving the trained data from the table and the query also taken as the output. Finally the output is get in the format of SQL query and retrieved data also get in the output.

REFERENCE:

- [1]T.Patten,P.Jacobs(1994),”Natural-language processing”, IEEE Expert,volume:9 issue:1.
- [2]Abhilasha Kate,Satish Kamble,Aishwarya Bodkhe,Mrunal Joshi(2018),”Conversion of Natural Language Query to SQL Query”,Second International Conference on Electronics,Communication and AerospaceTechnology(ICECA).
- [3]Haizhou li,Bin Ma,Kong Aik Lee(2013),”Spoken Language Recognition: From Fundamentals topractice”,Proceedings of the IEEE,Volume 101issue:5.
- [4]Gaurav Sinha,Rahul Shahi,Mni Shankar(2010),”Human Computer Interaction”,3rd International Conferenceon Emerging Trends in Engineering andTechnology.
- [5]Rao, Gauri, Agarwal, Chanchal, Chaudhry, Snehal, et al., (2010) “Natural language query processing using semantic grammar,” International journal on computer science and engineering, vol. 2, Issue 2, pp. 219–223.
- [6]Satav, Akshay G, Ausekar, Archana B and Bihani, et al., (2014) “A Proposed Natural Language Query Processing System,” International Journal of Science and Applied Information Technology, vol. 3, Issue 2, pp.219–223.
- [7]Sathick, K Javubar and Jaya, A, (2015) “Natural language to SQL generation for semantic knowledge extraction in social web sources,” Indian Journal of Science and Technology, vol. 8, Issue 1, pp.1–10.
- [8]Singh, Garima and Solanki, Arun, (2016) “An algorithm to transform natural language into SQL queriesfor relational databases,” Selforganizology, DirectoryofOpen Access Journals, vol. 3, Issue 3, pp. 100–116.
- [9] uma M,sneha V,sneha G,bhuvana J,bharathi B (2019),”Formation of SQL from natural language query using NLP,”Second International conference on computational Intelligence in Data Science(ICCIDS-2019).
- [10]xuefeng li, xiaochuan wu ,xin pei, zhujun yao(2019) “Tokenization:open asset protocol on blockchain,” 2nd International Conference on Information and Computer Technologies(ICICT).
- [11]N.Rishe,K.Naboulsi,O.Wolfson,B.Ehlmann(1999)”19th IEEE International Conference on Distributed Computing System Workshops on Electronic Commerce and Web-based Applications.Middleware”.
- [12]Sneha lukose,Savitha s.upadhya(2017),”Text to speech synthesizer-formant synthesis”,International Conference on Nascent Technologies inEngineering(ICNTE).
- [13]Soon Suck Jarng(2011),”HMM Voice Recognition Algorithm Coding”International Conference on Information Science andApplications.
- [14]YuZhang,Mengdong Chen,Lianzhong Liu(2015),”A Review on Text mining”,6th International Conference on Software Engineering and ServiceScience(ICSESS)