ISSN 2395-1621



Intelligent Query System using Natural Language Processing

ATHARVA SARODE, KEYA PATEL, MIMOH KULKARNI, SAMRUDDHI PATIL

DEPARTMENT OF COMPUTER ENGINEERING

STES'S SMT. KASHIBAI NAVALE COLLEGE OF ENGINEERING VADGAON BK, OFF SINHGAD ROAD, PUNE 411041.

ABSTRACT

In today's fast growing world of computing, information plays an important role. Information is stored in an organized manner in database and SQL allows user to access, manage, process the data on the database. SQL is too difficult for users who do not have any technical knowledge of databases. Even for users who are experts in database programming languages, it can be challenging because it requires that users know the exact schema or structure of the database, different tables in database along with its attributes, the roles of various entities in a query, and the correct join operations to be followed. This results in need for non-sophisticated users to access the database using natural language.

A Natural Language Interface for Database (NLIDB) acts as an intelligent interface for database, which allows users to access the database by typing or giving voice, input in their natural language. In this system we propose a methodology for building an Intelligent Querying System (IQS) by which a layman or a person without any technical knowledge of databases can fire queries in natural language (English). The system first accepts voice input from the user and converts it into text format. The system then parses the input sentences and then generates SQL queries from the natural language expressions of the input. These queries are in turn mapped with the database to generate the required output. Hence, it makes the information retrieval process simple, effective and reliable.

Keywords: NLP, IQS, NLIDB, SQL

I. INTRODUCTION

Application of databases is increasing day by day. Databases have application in almost all information systems such as transport information system, financial information system, human resource management system etc.

Complicated SQL queries are generated due to continuous increase in the size of database and complexity in the relation among the entities. These complex queries are very difficult to write for a layman or users who do not have knowledge of. The main problem is that the users who want to get information from the database, does not know formal languages like SQL. It is required to know all the details of the database such as relations, entities etc. Natural language interface to database presents an interface for non-expert users to interact with the system and database. To design models for automatically mapping natural language semantics into programming languages has always been a major and interesting challenge in Computer Science. For

ARTICLE INFO

Article History

Received: 25th April 2022 Received in revised form : 25th April 2022 Accepted: 28th April 2022 **Published online :** 29th April 2022

example, accessing a database requires the knowledge of structured query language (SQL) and machine readable instructions that common users are not supposed to know. Ideally, they should only ask questions in natural language without knowing either the underlying database schema or any complex machine language. Questions entered in natural language form are translated into a statement in a formal query language. Once the statement is formed, the query is processed by the DBMS in order to produce the required data. Databases are the common entities that are processed by experts and with different levels of knowledge. Databases respond only to standard SQL queries which are based on the relational algebra. It is nearly impossible for a layman to be well versed in SQL querying as they may be unaware of the structure of the database namely tables, their corresponding fields and types, primary keys and so on.

Providing a solution to this problem, this system has been proposed that uses natural language speech through voice

recognition, converted to SQL query and displaying the results from the database.

PROBLEM DEFINITION:

To convert user's speech generated pseudo code to software specific MYSQL query with the help of Natural Language Processing (NLP). Natural language Processing is used to ease database operation and make database accessible to layman.

The proposed system will work in the following problem domains,

- It will accept user's natural language voice input and convert it into text format using Google's Speech-to-Text API.
- The natural language input will then be converted to relevant MySQL queries.
- MySQL query obtained is then fired and data is retrieved from the database.

II. LITERATURE SURVEY

Literature survey is another term that carries similar meaning to "literature review". It targets to review the papers. It is an activity a researcher will perform during research by reviewing all the relevant literature which include journal articles, books, documented artifacts /scripts etc. It shows various analysis and research made in the field of interest and the result already published, considering the various parameters of the project and the extent of the project. Surveying the papers related to the proposed system helped to gather information like limitations and advantages, which gives the scope of improvement for the proposed system. For the proposed system "Intelligent Querying System using Natural Language Processing" various papers have been reviewed whose survey report is given below.

In this paper author has proposed that the system converts the queries from the Natural Language form to its equivalent Structured Query Language form. It starts with the Syntactic analysis performed by Stanford POS tagger. Then, the keyword extractor use the information from the POS tagger to extract the keywords that are used by the Named Entity Recognition tool. The Named entity Recognizer defines the related domain concepts like person or department. The identified keywords are handled by a SQL Generator class.[1]

The author has proposed an interactive natural language query interface for relational databases. Given a natural language query, the system first translates it to an SQL statement and then evaluates it against an RDBMS. To achieve high reliability, the system explains to the user how the query is actually processed. When ambiguities exist, for each ambiguity, the system generates multiple likely interpretations for the user to choose from, which resolves ambiguities interactively with the user. [2]

The Rule based domain specific semantic analysis Natural Language Interface for Database converts a wide range of text queries (English questions) into formal (SQL query) ones that can then be run against a database by employing generic and simpler processing techniques and methods. This paper defines the relation involving the ambiguous term and domain specific rules and with this approach this paper makes a NLIDB system portable and generic for smaller as well as large number of applications. This paper only focuses on context based interaction along with SELECT, FROM, WHERE and JOIN clauses of SQL query and also handles complex query that results from the ambiguous Natural Language query. [3]

In this paper, a system is developed that can execute both DDL and DML queries, input by the user in natural language. A limited Data dictionary is used where all possible words related to a particular system are included. Ambiguity among the words is taken care of while processing the natural language. The system is developed in java programming language and various tools of java are used to build the system. An oracle database is used to store the information. [4]

The author has proposed a system which provides a convenient as well as reliable means of querying access, hence, a realistic potential for bridging the gap between computer and the casual end users. The system employs CFG based system which makes it easy search the terminals. As the target terminals become separated to many non-terminals. To get the maximum performance, the data dictionary of the system will have to be regularly updated with words that are specific to the particular system. [5]

III. PROPOSED SYSTEM



Fig 1. System Architecture

The proposed system architecture is as follows:

Step 1 - The user is expected to give input in the form of speech. This input is then given to the speech to text converter and communicator, which converts it in the text form. The user can analyze the text and can update it manually if required.

Step 2 - This natural language query is then converted into a stream of tokens with the help of tokenizer and a token id is provided to each word of the NLQ.

www.ierjournal.org

Step 3 - The parse tree is generated through the parser with the help of the token id's and a set of words is identified. The output of this analysis will be a collection of identified words.

Step 4 - The set of identified words is then represented into a meaningful representation with the MR Generator. The identified words are transformed into structures that show how the words relate to each other.

Step 5 and 6 – The semantic builder takes the output generated by the MR generator and extracts the relevant attributes from the database.

Step 7 – The relations between the word structures and the attributes extracted from the database are identified in the lexicon builder and relation identifier.

Step 8 –The word structures and the attributes are mapped by identifying the relation between them and a semantic map is created.

Step 9 – The SQL query is constructed with the help of the semantic map input to the query generator.

Step 10 – This SQL query is then fired on the database.

Step 11 – The output after the execution of the SQL query is then displayed to the user.

IV. CONCLUSION

Intelligent Query System using Natural Language Processing is a system used for making data retrieval from database easier and more interactive.

Proposed system is bridging the gap between computer and casual user. Without any technical training handling databases is not possible for naïve user. This drawback is overcome by this system.

This system converts the human speech input i.e. natural language input to the SQL query after converting the natural language to SQL query the generated query is given to database which gives the desired output.

This report consists of various chapters which covers all the points step by step. This report gives functional nonfunctional requirements of the system along with hardware and software interfaces. Chapters covers the SDLC and the detailed System implementation plan.

V. REFERENCES

[1] Hessa Alawwad, Emdad Khan, "An Intelligent Database System Using Natural Language Processing" International Journal of Computers ,volume 1,2016.

[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.

[7] 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.