Agnel Charities'
Fr. C. Rodrigues Institute of Technology, Vashi, Navi Mumbai, India

IEEE International Journal of Engineering and Management Research November, 2021

AI Auto Query

Aditya Sawant, Rohit Raina, Anuja Patil

FR. Concecicao Rodrigues Institute of Technology, Vashi.

University of Mumbai

ABSTRACT

Data plays the most important role in the development of industries, small businesses. Even world leaders need the data to make analyses and make better policies for people. In almost every field where the work process is digitalized need to store data and then retrieve it. According to statistics most of the data is stored in the relational database and for manipulations of the data, Structured Query Language(SQL) is commonly used. So for handling databases a person need to have specialized knowledge regarding the queries and had to remember the syntax of many complex queries. So to enhance data manipulation using SQL and to efficiently get the required query, the paper proposes a method for the generation of SQL query from natural language input, spoken(audio input) by the user. The model is constructed on NLP (Natural Language Processing) and Neural Networks (Deep Learning) technologies. Long Short Term Memory(LSTM) Model is used for predicting queries and is trained on the dataset with natural language as input and returns outline skeletal structure of the query as output. Then the output will be processed and the final query will be displayed to the user. The project also aims to benefit the people who are suffering from Repetitive Stress Injury (RSI), causing pain in the finger joints, which has been attributed to work requiring a long period of typing and also to those who are not familiar with SQL queries. As this system will readily provide the required query.

CONTACT

Aditya Sawant
Email: adityaxt9990@gmail.com Contact No:- 8850258385

Rohit Raina
Email: rohitraina885@gmail.com Contact No:-9682527319

Anuja Patil
Email: anujaspatil99@gmail.com Contact No:- 9028068535

INTRODUCTION

In the era of technology, many jobs have become much simpler and efficient. Technologies like AI, ML, DL are assisting humans in many domains such as medical, space exploration, Bio-technology, Automobiles and many more. Technology has removed the barrier of having specialized knowledge to do certain work like anyone can create web pages by using pre-build templates and modify them by using drag and drop methodology, Anyone can make 3D models using 3D printers.
Retrieving the required data or information from relational databases is a tedious process. The person has to have specialized knowledge about Database Management systems, need to learn the syntax of the various complex queries. This makes it difficult for a non-technical person to retrieve data from databases. Advance technology can help to tackle this problem, where users don't have to remember the code snippets and have to do specialization in databases. Natural language processing can be used to interact with computers. And neural networks can be used to predict the semantic of user requirements and convey them to computers. This paper aims to combine this technology and create an environment where anybody can create a SQL query and work on relational databases.

OBJECTIVES

The objective of the project is to give user a platform where they can ask for queries through voice as input. The query in natural language will be processed to give output as correct syntax. This will avoid chances of syntactical errors.
Mainly coder faces an issue of repetitive stress injury in fingers due to constant movement of fingers while typing the code, So they have to go through a lot of pain while typing the code.
So the solution to above problem is code by voice which will help user to code efficiently.

LITERATURE SURVEY

The authors in [1] proposed a system which divided input in tokens and then generated the output after processing.
Authors in [2] described the model which uses lexical analysis, syntax analysis and semantic analysis for query generation.
The authors in [3] developed a system to generate valid SQL queries after parsing natural language using open source tools and libraries.
Authors of [4] proposed model which predicts the query based on given input and the ability to improve knowledge based on machine learning approach
The authors of [5] proposed NLP system using "levels of language" method. The process breaks the sentence for tokenizing using four levels called Morphology, lexical, Syntactic and Semantic

Methods/Diagram/Algorithm/Hardware Implementation

System Architecture :

1. **Creation of Original Dataset**
2. **Pre-Processing of Data**
3. **Creation of Input and output set**
4. **Training of model**

As the proposed system is a voice based application, there are chances the user may provide redundant data with his requirement, which will affect the performance of the model. It becomes necessary to pre-process the data before feeding it to the model.
So there are various steps taken to convert the data into the required format. Data pre processing steps are as shown in Fig. 1.
Flowchart for the implementation approach is as shown in Fig. 2.

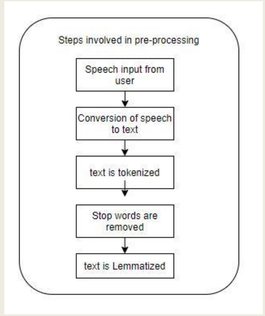

Fig 1. Preprocessing of data

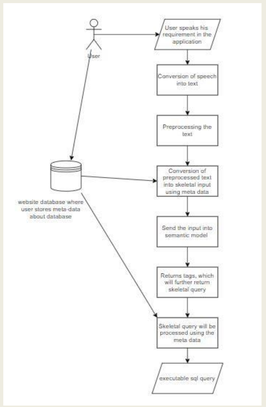

Fig 2. Flowchart for the implementation approach

The dataset created is then trained to take input as natural language and give required query with correct syntax. The accuracy of the model is as shown in Fig. 3. The input is processed and it gives the output as shown in Fig. 4 as the query which will give output from database.

```
Epoch 195/200
14/14 [==============================] - 0s 2ms/step - loss: 0.0499 - accuracy: 0.9851
Epoch 196/200
14/14 [==============================] - 0s 2ms/step - loss: 0.0291 - accuracy: 1.0000
Epoch 197/200
14/14 [==============================] - 0s 2ms/step - loss: 0.0200 - accuracy: 1.0000
Epoch 198/200
14/14 [==============================] - 0s 2ms/step - loss: 0.0589 - accuracy: 0.9701
Epoch 199/200
14/14 [==============================] - 0s 2ms/step - loss: 0.0621 - accuracy: 0.9701
Epoch 200/200
14/14 [==============================] - 0s 2ms/step - loss: 0.0528 - accuracy: 0.9701
```

Fig 3. Accuracy of model

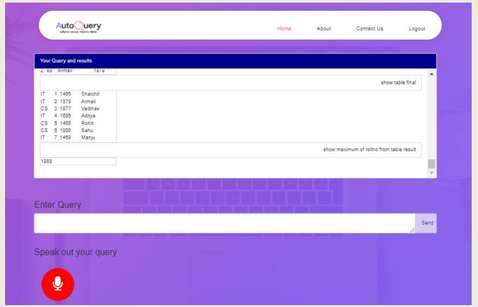

Fig 4. Test case for querying database

RESULTS/DISCUSSION

The model has an epoch of 200, which is sufficient for getting the best accuracy. To verify the accuracy of the model, it was run on several test cases and accordingly the accuracy was calculated. The accuracy the model gave was 94.54. It was also observed that the model takes the time to return the query in order of microseconds, this tells that the model is quite fast. The model when asked complex query gives sometimes wrong semantics. It is observed that the more the requirements are spelt in simple English more the accurate result will be returned. So the result also depends on the user, how efficiently and simply the user can specify its requirements. In a nutshell, the development of the system is an ideal for retrieving the data from a database.

FUTURE SCOPE

Currently in our project a web application is proposed which can convert natural language instructions for database queries via voice to corresponding SQL queries. It is currently capable of handling simple queries along with some complex queries. In future the support of other query languages as well as programming languages can be implemented using the same model proposed. The user will be able to switch between any programming language with the same instructions given. An extension can also be implemented using this AI for integrating into editors like VScode, Atom, Sublime etc. The user after installing the extension will be able code through voice with added functionalities like suggestions and auto complete. The feature of converting code from one programming language to another can also be implemented.

CONCLUSION

Artificial Intelligence and Natural Language Processing can bring significant enhancements to any computer program because human language is so natural and effortless for humans. Our model could be a big step forward in making database usability utilities accessible to people who have no knowledge of how a database query works. In our proposed system various processes like tokenization, syntactic, semantic, analysis are carried out to generate an equivalent SQL query for any instruction given as input in natural language. The use of Natural Language helps users to easily retrieve data. This system will help many organizations such as education, medical, etc. For the maximum performance of the system, the database has to be updated frequently with specific words to the particular system.

REFERENCES

1. Bais, Hanane, Mustapha Machkour and Lahcen Koutti. "A Model of a Generic Natural Language Interface for Querying Database." international journal of intelligent systems and applications. 8. 35-44. 10.5815/ijisa.2016.02.05.
2. Ghosh, Prasun Saltlake, Kolkata Kolkata, Saparja Dey, Kolkata Sengupta, Subhabrata Assistant, Kolkata Saltlake,. (2014). "Automatic SQL Query Formation from Natural Language Query", International Conference on Microelectronics, Circuits and Systems (MICRO-2014).
3. [3] Nagare, Indhe, Sabale, Thorat and Chaturvedi. "Automatic SQL Query Formation from Natural Language Query" International Research Journal of Engineering and Technology (IRJET) Mar -2017
4. Nandhini S, B.Viruthika, Almas Saba, Suman Sangeeta Das "Extracting Sql Query Using Natural Language Processing" International Journal of Engineering and Advanced Technology (IJEAT) April 2019
5. Chandrakala Kombade, Monika More, Shweta Patil, Anjalidevi Pujari "Natural Language Processing with some Abbreviation to SQL" International Journal for Research in Applied Science Engineering Technology (IJRASET) Dec 2019