NLP TO SQL

Siddhika Anil Kapure, Priyanka Vitthal Ankush, Priyanka Bhimdev Yadav

[1] *Government college of engineering and research, Awasari(kh), Pune-412405*
[2] *Government college of engineering and research, Awasari(kh), Pune-412405*
[3] *Government college of engineering and research, Awasari(kh), Pune-412405*

ABSTRACT

Now days data is increasing rapidly. There are so many new database tools and technologies are rowing, therefore we can store large data, but the problem is that the technology or an interface which can process data and display the data as per the user request is not familiarized with many of the people.
It means many people don't have proper knowledge of handling database.

Actually, NLP to SQL conversion is part of machine learning. The main idea of NLP to SQL conversion is to generate SQL query from natural language. This technique is useful for accessing data from database without having prior knowledge of SQL. This technique can be used by many common people.
In this system input is simple English text and query is generated by using POS tagger in python. In this project we are going to implement a system which will generate SQL query from natural language.

Keyword: **-** *SQL (Structured Query Language), Database, NLP (Natural Language Processing)*

1. INTRODUCTION

Storage of data is a crucial task in today's commercial system specially social media, database size is increased and accessing data from database become more crucial part in the recent research world. So many new database tools and technologies are growing, therefore we can store large data, but the problem is that the technology or an interface which can process data and display the data as per the request is not familiarized with many of the people. Most of the businesses and social sites need these types of applications by using the SQL language.

Natural language processing (NLP) is becoming most active techniques to process on human language. In case of social media, the query conversion is very crucial task in terms of getting exact data which is requested by the users.

The query or request can be of simple English language statement such as blog, comment, tweets etc., these statements must be converted into proper sql statement so that exact data can be fetch from database. so, these factors are acting as a precious evidence for implementing the proposed work through this article. The objective of NLP is to facilitate communication among human and computers without multifaceted instructions and procedures. In other words, NLP is the technique that can used the natural languages used by users. An end user can be easily processing their query without knowledge of SQL.

Therefore, in this work the development of system for people to interact with the database in simple English language is implemented and analyzed for the accuracy. This enables a user to input their queries in simple English and get the answer in same language which is referred as Natural Language Interface to a Database (NLIDB) The knowledge extraction is enabled with the successful implementation of SQL generation from the natural language statement**.**

2. LITERATURE SURVEY

There have been a large amount of research works introducing the theories and applications of NLIDBs. Asking question to databases in natural language is very appropriate and easy method of information access especially for informally users who do not comprehend complex database query language. In fact, database NLP may be one of the most significant successes in NLP since it began. Asking questions to databases in natural language instead of the data base complex queries.

3.PROPOSED METHOD

1.User Interface: The user interacts with the system via Graphical User Interface and types his/her Natural Language Query for the further output.

2.Lowercase Conversion: The Natural Language Query is then translated into lowercase and passed to the tokenization.

3.Tokenization: The query after lowercase conversion is then transformed into stream of tokens and a token id is providing to each word of NLQ.

4.Escape word removal: The extra/stop words are removed which are not needed in the analysis of query.

5.Part of Speech Tagger: The tokens are then classified into nouns, pronouns, verb and string/integer variables.

6.Relations-AttributesClauses Identifier: Now the system classifies the tokens into relations, attributes and clauses on the basis of tagged elements and also separates the Integer and String values to form clauses.

7.Ambiguity Removal: It removes all the ambiguous characteristics that exists in multiple relation with the same attribute name and maps it with the correct relation.

8.Query Formation: After the relations, attributes and clauses are extracted, the final query is built.

9.Query Execution and Data Fetching: The query is then executed and data is got from the database.

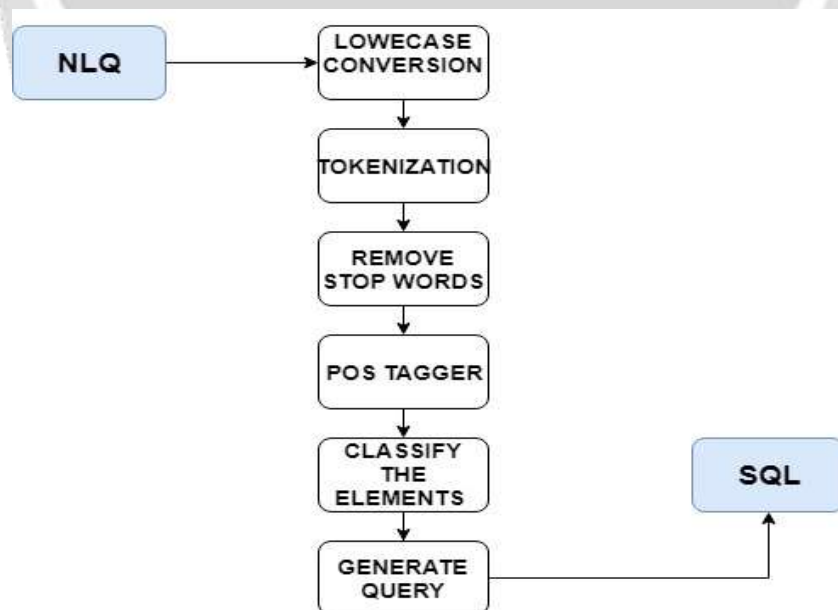**10.Results**: The final query result is displayed to the user on the Graphical User Interface.

Fig 1.0: NLP to SQL Flow Chart

4. IMPLEMENTATION

In Proposed system is natural language text which is "Show me details of students who are in division B." First of all, system will convert the natural language query into lowercase. After lowercase conversion system will convert it into stream of tokens and it is stored in python list. And then stop words/extra words are removed.

Fig 1. starting window

The tokens are tagged as nouns, verbs, adjectives, etc. this result is obtained by Part of Speech tagger. The system will classify the tokens into relations on the basis of tags.

If the user types the Natural Language Query in the input field then the SQL Query will be generated thorough the processing.
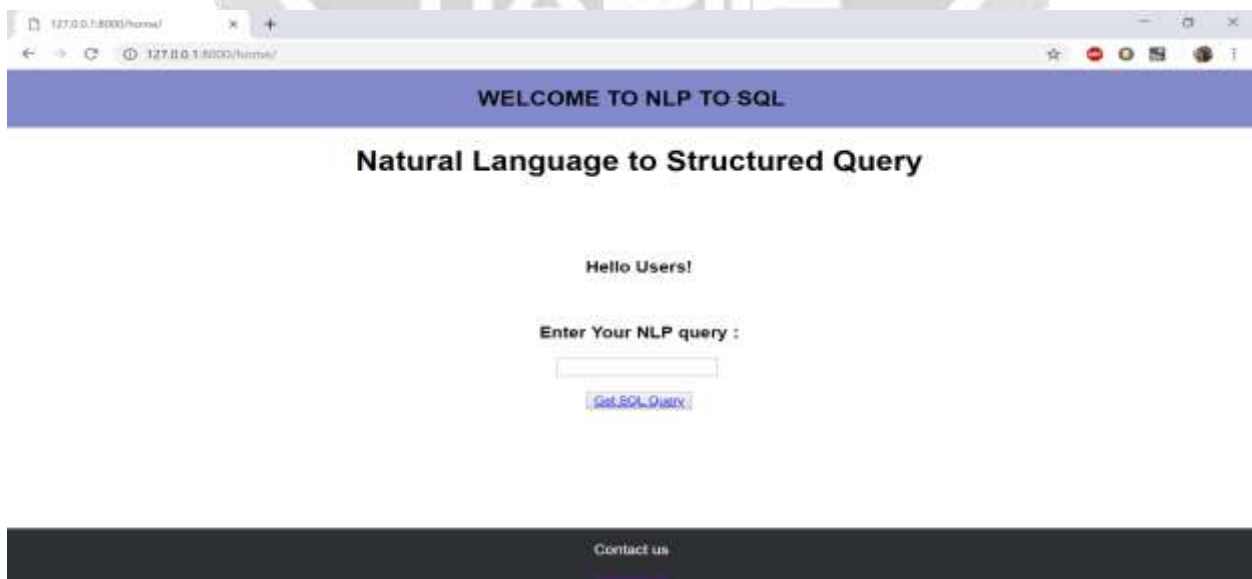

Fig. 2: NLQ area window

Then final query will be generated. The expected output of the system will be like, Select * from Student where division='B'.
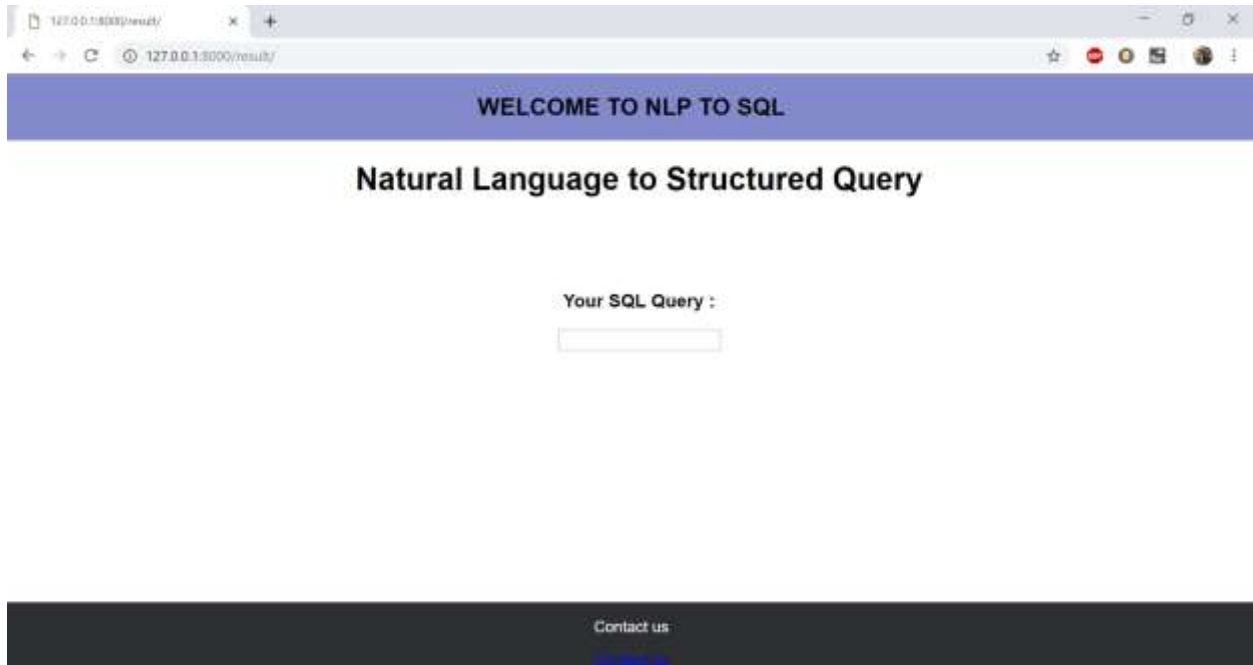Then the query will get executed and data is fetched from database.

Fig. 3: The result area

5.CONCLUSION

Natural Language Processing can bring commanding enhancements to virtually any computer program. Retrieving data from the database requires knowledge of technical languages like SQL.

In this project we consider a lightweight approach of translating English queries into equivalent SQL queries. In this approach we look at extracting certain keywords and indicators from an English query written using POS tagger method, and then using a system to generate the query based on the key.

6.FUTURE WORK

Retrieving information from the database requires knowledge of technical languages like SQL. In this project we consider a lightweight approach of translating English queries into equivalent SQL queries. In this approach we look at extracting certain keywords and indicators from an English query spoken or written using Hash Table technique, and then using a system to generate the query based on the keywords. Currently our project is built around the fixed database i.e. it can understand the queries about the static database only. But the project can be extended to refuge any database. Support for other languages should be provided for enhancing the project. The program should categorize the vendor of the selected database and should use proper drivers to connect them.

7. REFERENCES:

1. Natural Language To SQL Conversion System, International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) Vol. 3 Issue 2, June 2013, 161-166.

2. Automatic SQL Query Formation from Natural Language Query, International Journal of Computer Applications.

3. Huangi,GuiangZangi, Phillip C-Y Sheu ―A Natural Language database Interface based on probabilistic context free grammar, IEEE International workshop on Semantic Computing and Systems 2008.

4. A Survey of Natural Language Query Builder Interface to Database, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 5 Issue 4, 2015.

5. Gauri Rao(IJCSE) International Journal on Computer Science and Engineering