



CONVERSION OF NLP TO SQL USING MACHINE LEARNING APPROACH

Ms. Sneha.A.Khaire¹,

Faisal Sayyed², Sudeep Mishra³, Kanchan Patel⁴

1 ASSISTANT PROFESSOR 2,3,4 UG SCHOLAR
INFORMATION TECHNOLOGY ENGINEERING,

SANDIP INSTITUTE OF TECHNOLOGY AND RESEARCH CENTRE, NASHIK, INDIA

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.

1.INTRODUCTION

Storage of data is a crucial task in today's commercial system especially 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.OBJECTIVES

- I. The objective of our project is to generate accurate and valid SQL queries after parsing natural language using open source tools and libraries.
- II. Users will be able to obtain SQL statement for the major 5 command words by passing in an English sentence or sentence fragment.
- III. We wish to do so in a way that progresses the current open source projects towards robustness and usability.
- IV. Objective of this system is to convert a natural language query into a SQL to simplify data extraction.

3.PROJECT IMPLEMENTATION

A module is a collection of source files and build settings that allow you to divide your project into discrete units of functionality. Your project can have one or many modules, and one module may use another module as a dependency. You can independently build, test, and debug each module.

- 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.

4. SYSTEM REQUIREMENTS

4.1 Database Requirements

- MySQLdb

4.2 Software Requirements

- Core i3 processor
- Windows 10
- Python Libraries: re, ScrolledText, Tkinter, nltk.

4.3 Hardware Requirements

- Desktop/Laptop
- 6 GB RAM, 1 TB Hard disk

5.PROJECT SCOPE

Retrieving the required information from a database is quite difficult for any common man and requires a lot of effort which needs the knowledge of the database structure. DBMS is incapable of dealing with queries framed in any other languages other than the standard database languages. So to make the retrieval more effortless and interactive for naive user, our proposed work provides a facility through which a user is free to pose a query in English, which will be processed by several modules to form an equivalent SQL query

6. RESULTS AND OUTPUTS

I.

```
NLP To SQL (QA)
NLP Chatbot database QA System
-----
Talk to the program by typing in plain English, using normal upper-
and lower-case letters and punctuation. Enter 'quit' when done.
-----

Chatbot >Please Enter your name using the format name {your name}. Example: name Yash
```



II.

```
NLP To SQL (QA)
NLP Chatbot database QA System
-----
Talk to the program by typing in plain English, using normal upper-
and lower-case letters and punctuation. Enter 'quit' when done.
-----

Chatbot >Please Enter your name using the format name {your name}. Example: name Yash
You > name Sam
Chatbot > Hello Sam! type your question.. or type help
```



III.

```
NLP To SQL (QA)
-----
Talk to the program by typing in plain English, using normal upper-
and lower-case letters and punctuation. Enter 'quit' when done.
-----

Chatbot >Please Enter your name using the format name {your name}. Example: name Yash
You > name Sam
Chatbot > Hello Sam! type your question.. or type help

You > help
Chatbot > You can ask me questions like:
* What is the average salary?
* What is the average salary in development departments?
* What is the minimum salary?
* How many employees in development department?
* Who is the highest paid employee?
* Who is the highest paid employee in development department?
* Which employee has last name Facello?
```



IV.

```
NLP To SQL (QA)
-----
Talk to the program by typing in plain English, using normal upper-
and lower-case letters and punctuation. Enter 'quit' when done.
-----

Chatbot >Please Enter your name using the format name {your name}. Example: name Yash
You > name Sam
Chatbot > Hello Sam! type your question.. or type help

You > help
Chatbot > You can ask me questions like:
* What is the average salary?
* What is the average salary in development departments?
* What is the minimum salary?
* How many employees in development department?
* Who is the highest paid employee?
* Who is the highest paid employee in development department?
* Which employee has last name Facello?

You > What is the average salary?
Chatbot > Ok Sam! Let me think ...
Cool! your sql statement is:

SELECT avg(salary) FROM salaries

avg(salary) --> 63810.7448
-----
```



V.

```

NLP Chatbot database QA System
-----
Talk to the program by typing in plain English, using normal upper-
and lower-case letters and punctuation. Enter 'quit' when done.
-----
Chatbot >Please Enter your name using the format name {your name}. Example: name Yash
You > name Sam
Chatbot > Hello Sam! type your question.. or type help

You > help
Chatbot > You can ask me questions like:
  * What is the average salary?
  * What is the average salary in development departments?
  * What is the minimum salary?
  * How many employees in development department?
  * Who is the highest paid employee?
  * Who is the highest paid employee in development department?
  * Which employee has last name Facello?

You > What is the average salary?
Chatbot > Ok Sam! Let me think ...
Cool! your sql statement is:

SELECT avg(salary) FROM salaries

avg(salary) --> 63810.7448
-----

You > What is the minimum salary?
Chatbot > Ok Sam! Let me think ...
Cool! your sql statement is:

select salary from salaries order by salary ASC limit 1

salary --> 38623
-----

```

7.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.

8.REFERENCES

- [1]Natural Language to SQL Conversion System, International Journal of Computer Science Engineering and Information Technology Research (IJCEITR) Vol. 3 Issue 2, June 2013, 161- 166.
- [2] Automatic SQL Query Formation from Natural Language Query, International Journal of Computer Applications.
- [3] Huang,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
- [6] TO THEXANDER R., R U KS HA N P. & MAHESAN S. (2013). Natural Language Web Interface for Database (NLWIDB).
- [7] Faraj A. El-Mouadib, Zakaria Suliman Zubi, Ahmd A. Almagrous, "Generic interactive natural language interface to databases (GINLIDB)", EC'09 Proceedings of the 10th WSEAS international conference on evolutionary computing, ISBN: 978-960-474-067-3
- [8] Rani Nelken, Nissim Francez, "Querying temporal databases using controlled natural language", Proceeding COLING '00 Proceedings of the 18th conference on Computational linguistics - Volume 2
- [9] Alessandra Giordani, Alessandro Moschitti, "Semantic mapping between natural language questions and SQL queries via syntactic pairing", Proceeding NLDB'09 Proceedings of the 14th international conference on Applications of Natural Language to Information Systems, ISBN:3- 642-12549-2 978-3-642-12549-2