



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: V Month of publication: May 2020

DOI: <http://doi.org/10.22214/ijraset.2020.5166>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Natural Language Processing with some Abbreviation to SQL

Chandrakala Kombade¹, Monika More², Shweta Patil³, Anjalidevi Pujari⁴

^{1, 2, 3, 4}Computer Engineering, JSMP'S Jayawantrao Sawant College of Engineering, Pune, Maharashtra.

Abstract: This paper presents an approach to convert the Natural language processing (NLP) abbreviation to SQL. SQL is domain specific language which is used for storing, manipulating and retrieving data stored in relational database, but any ordinary person who don't have any knowledge about SQL are unable to retrieve the data from database. To overcome this problem we proposed a model conversion of NLP to SQL but in this system user can also use an abbreviations to retrieve the data from the database. Those who are expert in SQL language can access information from database but non-technical user cannot retrieve data from database. This system can be used everywhere because database is everywhere. If we want some data from database this system is very much useful. In this system user can also enter the query using speech.

Keywords: NLP (Natural Language Processing), NLTK (Natural Language Tool Kit), AI (Artificial Intelligence), Speech to Text.

I. INTRODUCTION

Natural language processing is a field of Artificial intelligence to build the intelligence systems. The main purpose of Natural Language Processing is for an English sentence to be interpreted by the computer and appropriate action taken. NLP can be used to access the database by asking question in natural language with using some abbreviations and getting the required results. This system can be used in many organizations such as education, medical, etc. Because database is everywhere and if we want some data from database this system is very much useful. This system has a ability to accept natural speech. Speech recognition is voice recognition of human being. This system is containing supporting feature GUI. When user will enter a query in natural language it will be converted to SQL directly. If user enters query in speech, system will convert speech to text.

II. LITERATURE SURVEY

According to the traditional Systems and prior research, the methodologies and technologies used for conversion of natural language processing to sql are giving below in the table.

Sr.no	Paper Title	Published Year	Technique used
1.	IQS- Intelligent Querying System Using Natural Language Processing	2017	Natural Language Interface to Database Systems (NLIDB)
2.	Impact of Intellisense on the accuracy of Natural Language Interface to Database	2015	Natural Language Interface to Database Systems (NLIDB), GUI
3.	A Survey of Natural Language Query Builder Interface to Database	2015	Natural Language Interface to Database.

III. TECHNOLOGY USED

A. Python

Python is a high-level, interpreted, general-purpose programming language.

B. Speech Recognition

Speech recognition is a voice recognition of human being. Speech recognition used in home automation, artificial intelligence, etc.

C. NLTK

NLTK3 is library of python used for input stemming. This library serves as a toolkit for computational linguistics.

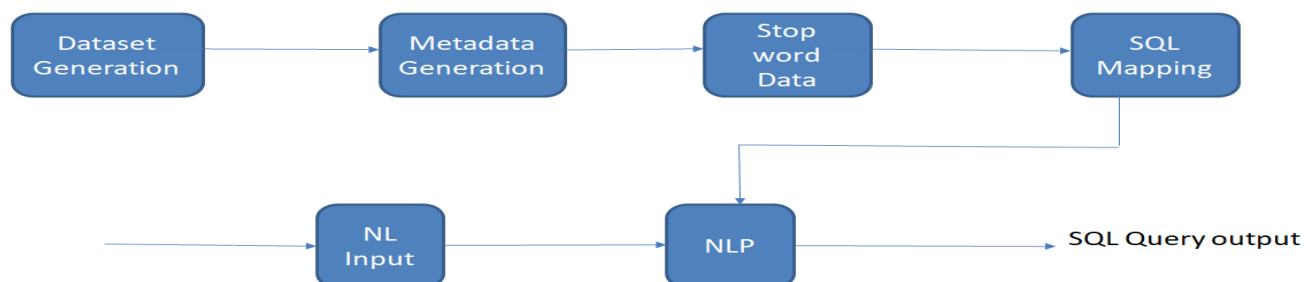
D. Database

Mysql is a relational database management system. The records are stored inside the tables in mysql database system.

IV. OBJECTIVE

- A. To generate valid SQL queries after parsing natural language using open source tools and libraries.
- B. To also analyze the abbreviations provided in NLP to get the required records.
- C. To execute the query and fetch the data from database.

Flowchart



V. ALGORITHM DETAILS

A. Multinomial Logistic Regression

Logistic Regression is a Machine Learning algorithm which is used for the classification problems, it is a predictive analysis algorithm and based on the concept of probability. Logistic Regression a Linear Regression model but the Logistic Regression uses a more complex cost function, this cost function can be defined as the Sigmoid function or also known as the logistic function instead of a linear function. The hypothesis of logistic regression tends it to limit the cost function between 0 and 1. Therefore linear functions fail to represent it as it can have a value greater than 1 or less than 0 which is not possible as per the hypothesis of logistic regression.

- 1) In this project machine learning approach is used to predict type of query.
- 2) Predicting type of query come under classification problem therefore multinomial Logistic Regression is used for predicting the type of query.
- 3) On the Basis of tagged tokens , the noun map and verb list is prepared through one iteration over the tokens. then the tokens are given a unique identifier using Label encoder and the data is trained on logistic regression algorithm.

VI. RESULT ANALYSIS

In this project machine learning approach is used to predict type of query. Predicting type of query come under classification problem therefore multinomial Logistic Regression is used for predicting the type of query. On the Basis of tagged tokens , the noun map and verb list is prepared through one iteration over the tokens. then the tokens are given a unique identifier using Label encoder and the data is trained on logistic regression algorithm. then model will predict whether the natural language statement represents a data retrieval query (SELECT) or a DML query (ALTER, DELETE) is taken at this stage with the help of certain data arrays for denoting type of query. For example, when words like select and its certain synonyms appear in the input, the type of query is predicted as select as per the trained model and so on. The model predicts the query type accurately with accuracy of 98.65. The confusion matrix is shown in following diagram.

VII. CONCLUSION

Use of Natural Language helps user to easily retrieve data. This system will help many organizations such as education, medical, etc. to easily retrieve and manage data from database using simple English language. There is no need for the user to learn complex query language like SQL. The facility to accept the input in speech as well as in text format makes the system user-friendly. Our system will convert natural language query into SQL language query and provides required information from database to organizations.

VIII. ADVANTAGES

- A. User dont need to learn sql for query creation.
- B. Any user can fetch data from database without prior knowledge of sql.
- C. User friendly system.
- D. Require less time for query generation.

IX. APPLICATION

- A. Banking SystemDatabase Management.
- B. College Database Management.

REFERENCES

- [1] Prof. Debarati Ghosal, Tejas Waghmare, Vivek Satam, Chinmay Hajirnis "SQL query formation using natural language processing", International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016.
- [2] Anum Iftikhar, Erum Iftikhar, Muhammad Khalid Mehmood "Domain Specific Query Generation from Natural Language", Sixth International Conference on Innovative Computing Technology, IEEE 2016.
- [3] Prasun Kanti Ghosh, Sagarja Dey, Subhabrata Sengupta "Automatic SQL Query Formation from Natural Language Query", International Journal of Computer Applications (0975 – 8887), International Conference on Microelectronics, Circuits and Systems (MICRO-2014).
- [4] Niket Choudhary, Prof. Sonal Gore "Impact of IntelliSense on the accuracy of Natural Language Interface to Database", IEEE 2015.
- [5] Rajender Kumar, Mohit Dua "Translating Controlled Natural Language Query into SQL Query using Pattern Matching Technique", IEEE 2014.
- [6] Prof. Wei Chen "Parameterized Spatial SQL Translation for Geographic Question Answering", IEEE International Conference on Semantic Computing 2014.
- [7] Natural Language To SQL Conversion System, International Journal of Computer Science Engineering and Information Technology Research (IJCEITR) Vol. 3 Issue 2, June 2013, 161-
- [8] Deepshikha, Kiran Devi "A Survey of Natural Language Query Builder Interface to Database" International Journal of Advanced Research in Computer Science and Software Engineering 2015.
- [9] Prashant Gupta, Aman Goswami, Sahil Koul, Kashinath Sartape "IQS- Intelligent Querying System using Natural Language Processing" 978-1-5090-5686-6/17/\$31.00 ©2017 IEEE
- [10] Rajender Kumar, Mohit Dua "Translating Controlled Natural Language Query into SQL Query using Pattern Matching Technique", IEEE 2014.
- [11] Anum Iftikhar, Erum Iftikhar, Muhammad Khalid Mehmood "Domain Specific Query Generation from Natural Language", Sixth International Conference on Innovative Computing Technology, IEEE 2016.
- [12] Rajender Kumar, Mohit Dua "Translating Controlled Natural Language Query into SQL Query using Pattern Matching Technique", IEEE 2014.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)