**Mahavir Education Trust's**

# SHAH & ANCHOR KUTCHHI ENGINEERING

**COLLEGE Chembur, Mumbai - 400 088**

## UG Program in Computer Engineering

Mini Project Synopsis Report on

**Generating SQL query from Natural Language using NLP**

Submitted in partial fulfillment of the requirements of the degree of Bachelor in Engineering

by

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Class/Roll No.** | **Contact No.** | **Email Id** |
| Rakesh Choudhary | TE3-08 | 8369422384 | [Rakesh.choudhary@sakec.ac.in](mailto:Rakesh.choudhary@sakec.ac.in) |
| Satyam Gupta | TE3-14 | 9702866020 | [Satyam.gupta@sakec.ac.in](mailto:Satyam.gupta@sakec.ac.in) |
| Yogesh Hole | TE3-15 | 8657093183 | [Yogesh.hole@sakec.ac.in](mailto:Yogesh.hole@sakec.ac.in) |

## Under the Guidance of

Prof. Deepshikha Chaturvedi

DEPARTMENT OF COMPUTER ENGINEERING

**SHAH AND ANCHOR KUTCHHI ENGINEERING**

**COLLEGE** CHEMBUR, MUMBAI-400088

2020-2021

**Abstract**

This paper describes a method for effectively automating the conversion of Natural Language Query to SQL Query. SQL is a potent tool for managing data in a relational databases. To retrieve or process data, the user must enter the appropriate SQL Query. However, users who are unfamiliar with SQL are unable to retrieve relevant data. To address this, we suggested a model in Natural Language Processing for converting Natural Language Query to SQL query. This allows a naive user to obtain the required data without having to know any complex SQL details. This system is also capable of handling complex queries. This system is intended for users who work with student databases but are unfamiliar with SQL.

The system will accept text input. This natural language query will be converted into a SQL query. The query will be executed by the system, and the user will receive the result

**Table of Contents**

[**Abstract**](#_Toc72246250) ii

**1.** [**Introduction**](#_Toc72246254) **1**

**2.** [**Literature Review**](#_Toc72246255) **2**

**3. Problem Definition &Objectives 4**

**5. Proposed Methodology 5**

**6. Summary 6**

**Chapter 1**

**Introduction**

Natural Language Processing is a branch of Artificial Intelligence that is used to create intelligent computers that can interact with humans in the same way that humans do. It connects the human-machine gap. The primary goal of Natural Language Query Processing is to have a computer interpret English Sentence structures. Despite these problems, it is commonly used for research purposes. Natural Language Processing is used to access the database by asking Natural Language questions and receiving the necessary results. Asking questions via natural language to databases is a very convenient and easy method of data access, particularly among users who are unfamiliar with complex database query languages such as SQL (Structured Query Language). This system is intended for users who deal with student databases but lack SQL expertise. The system suggests an algorithm for handling the English Query fired by the user in order to obtain a SQL query using feedback as text or expression. Tokenization, syntactic, and semantic analysis, as well as the use of dictionaries and grammars used for such analysis, can be performed using the natural language to SQL query conversion tool.

A query can be entered in natural language by the user. When the user enters the query in English, it is translated to a SQL query. There are many difficulties in converting natural language queries to SQL queries, such as complexity, which implies that a single term may have several meanings. In this case, a single word may be mapped to several meanings. Another difficulty is the development of complex SQL queries.

**Chapter 2**

**Literature Review**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Author | Title | Methodology | Objective | Limitation |
| M. Uma, V. Sneha, G. Sneha,  J. Bhuvana and B. Bharathi, 2019 | Formation of SQL from Natural Language Query using NLP | NLP Phase: -Tokenization -Lemmatization -Syntactical Analysis -Semantic Analysis Mapping Phase: - Attribute Identification -Sql Query Formation | Formation of SQL query using Natural Language Processing on train seats and fare datasets. | They have taken input in text form and they have make use of only one table. |
| Alaka Das, Rakesh Chandra Balabantaray, 2019 | MyNLIDB: A Natural Language Interface to Database | Input🡪Preprocessing 🡪POStagger🡪 🡪Node generator🡪 SQL generator🡪 Database🡪 Result | To make Natural Language Interface to Database.(i.e. Interface through which we retrieve data from database) | It is made only for simple queries |
| A. kate, S. Kamble, A. Bodhke, M. Joshi, 2018 | Conversion of Natural Language Query to SQL Query. | Tokenization , Lexical Analysis , Syntactic Analysis , Semantic Analysis. | To help T&P officer to easily retrieve and manage student data from database. | This system does not provide high accuracy in results. |
| Tanzim Mahmud, K. M. Azharul Hasan,  Mahtab Ahmed,  Thwoi Hla Ching Chak, 2015 | A Rule Based Approach for NLP Based Query Processing | Design context free grammars  word check,removing excess words, tokenization and mapping to CFG rules. | Develop NLID for alumni database using CFG base system | Accuracy is very low i.e 47% |
| N. Sangeeth ,  R. Rejimoan , 2015 | An Intelligent System For Information Extraction From Relational Database Using HMM | Two modules used: Linguistic module, database module. | develop NLIDB system based on HMM using GEOQUERY database. | Accuracy and Performance can be increased, audio input is not taken |
| Vishal Wudaru, Aruneswara Reddy, Radhika Mamid,  2019 | Question Answering on Structured Data using NLIDB Approach | Intermediate query approach Syntactic approach 1.Tokenizer & POS Tagging 2.Stop word remover 3.Dependency parsing. | Proposed system able to handle most queries related to movie domain through database | Focused on limited number of database attributes |

**Chapter 3**

**Problem Definition & Objectives**

* **Problem Definition:**

To create a system which converts Natural Language To SQL query by taking input as text format, process it and generate SQL query and execute it to provide output.

* **Objectives:**

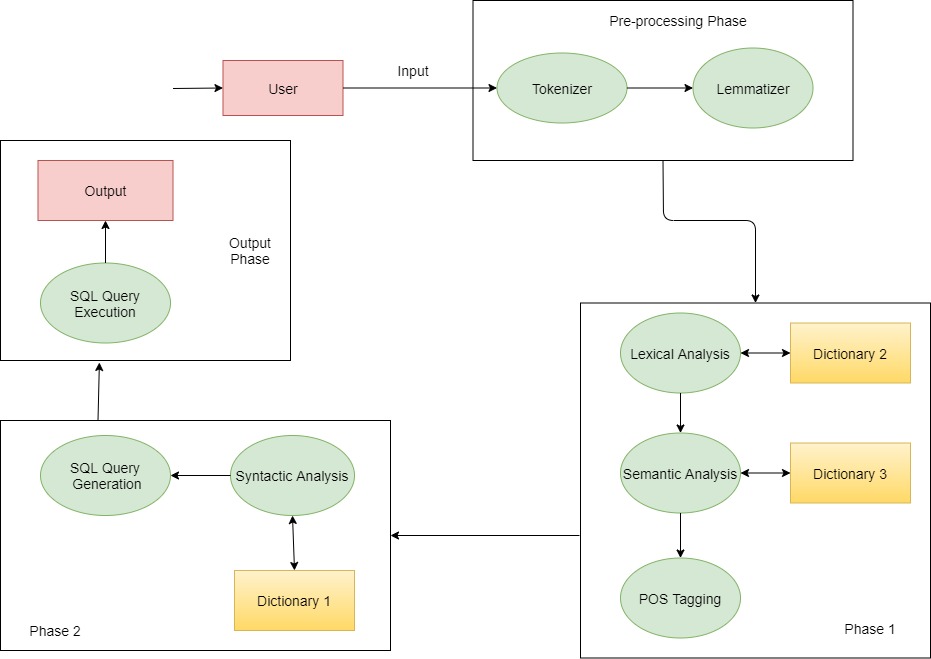
To create a system which converts Natural Language To SQL query and also achieve more accuracy for various queries.

* **Scope:**

This system can be used for many sectors from pharmaceutical to any government sector. Those who don’t know SQL can access database easily without having much trouble.

**Chapter 4**

**Proposed Methodology**

****

**Chapter 5**

**Summary**

Those who don’t know anything about Database Management Systems (DBMS) they can’t use database for their need as they don’t have any knowledge regarding any database query language. To overcome that the system has been developed based on NLP to assist the non-technical user to convert natural language given by naïve users to SQL and then this query will be run in database and it will provide output.