Natural Language Query to SQL Query using NLP and LSTM (Deep Learning)

Prof. M.D. Sale[1], Shraddha Shekhar[2], Samiksha Shirgave[3], Shreeteeja Salunke[4,] Harshada Jadhav[5]

[1-4]B.E. Student, Dept of Computer Engineering, Sinhgad College of Engineering, Vadgaon, Pune-411041, Maharashtra, India

---***---

Abstract - *Data is produced at a massive amount every day and this trait is going upward, creating tons of data that is continuous in nature and cannot be stopped. Managing such data needs a lot of time investment. The main part of database management is to write or construct precise queries for fetching accurate records. Accurate query writing is quite difficult as the query should be syntactically as well as logically correct. It should also contain valid attributes of the table from which data is to be extracted. Here comes the need for a system that can generate SQL queries by simply understanding natural language queries. Such a system accepts natural language queries written in simple English and converts it into appropriate equivalent SQL queries.*

***Key Words*: LSTM Model, NLP**

1. INTRODUCTION

Database Management System (DBMS) is a software system used to store, retrieve, define and manage the data in the databases. It offers an efficient way to handle large amounts of data and multiple types of data. Generating SQL queries require solid syntactic and logical knowledge in order to retrieve data accurately. This involves investment of time and capital which can be unaffordable in some situations. The system focuses on reducing these assets and utilizing them in some other beneficial way. It will save time for training newbies with SQL and will use simple instead. Such a system will even help non-technical people to generate proper SQL queries. Using NLP processes and a Deep Learning model like LSTM it can be possible to develop such a system.
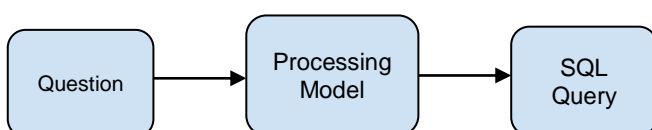

Fig. 1.1 NL to SQL system

1.1 Input Question

The Input is the question written in the English language that will be converted to an SQL query.

1.2 Processing Model

The Processing model comprises NLP, LSTM, and basic AI. The question entered is tokenized and respective POS tags are added to every token. These tokens are segregated and passed to the LSTM model and basic AI model respectively. After processing the output generated is integrated and formulated into the SQL syntax.
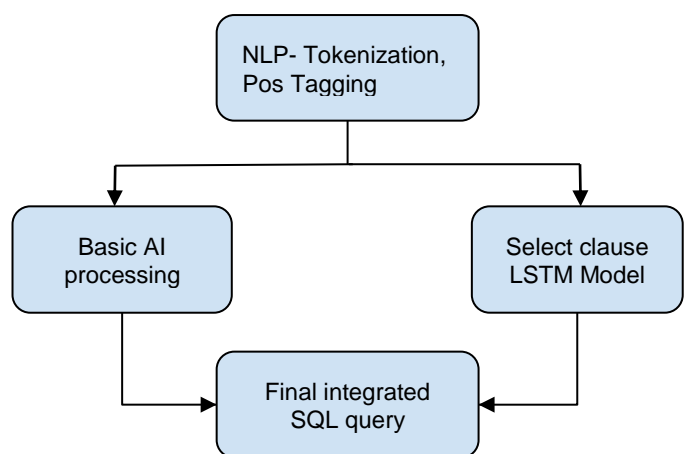

Fig. 1.2 Processing Model

1.3 SQL Query Output

The final output is the generated SQL query.

2. LITERATURE SURVEY

The first paper presents an approach on how one can automate the conversion of Natural Language Query to Structured Query Language. This system is prepared for Training & Placement Cell Officers so that they can retrieve proper records from a student's database without involving an expert in it. It also accepts input in speech format. The paper tells us how one can generate SQL queries from NL Query in 7 steps. The main levels/steps involved in the conversion are Tokenization, Lexical Analysis, and Syntactic Analysis & Semantic Analysis.

Users can view the output query as well as execute it on the database. This paper also discusses the difficulties faced during conversion like ambiguity, formations of complex queries & Discourse knowledge in which immediately preceding sentence affects the interpretation of the next sentence for example-If the user enters a SELECT and INSERT query at the same time, this creates an anomaly in the conversion process.[1]

The proposed system in the second paper helps to convert English Query into an SQL Query using semantic grammar. The main purpose of this system is to make a user-friendly DBMS handling system just by using the English Language. This is the working flow of the proposed system is as follows: English query -> Tokenization -> lexical analysis -> Syntactic, semantic analysis -> Mapping -> SQL query **->** execution -> result. A finite data dictionary is used accomodating limited words associated with the table names. Here, the data dictionary updates regularly & the system model makes use of a static framework or the database.[2]

The third paper proposes a system that effectively converts natural language queries to corresponding SQL queries. The approach used here is by understanding the natural language statement, its POS (Parts of speech) tags, and constructing the SQL query. NLP processes like Tokenization, Lemmatization, POS tagging are first carried out. The next step divides the query according to the SELECT clause and CONDITION clause. Finally, all the segments formed are integrated together to formulate a SQL query. Further, there is a provision to implement JOIN Queries. The system faces limitations when it comes to complex queries containing table interlinking. The result obtained is not accurate and satisfactory.[3]

The fourth paper gives fundamental but essential details of the Long Short-Term Model (LSTM) Model. Many traditional text classification techniques are being used in Natural Language Processing. However, the accuracy of text classification is not satisfactory. The paper describes the introduction of the Deep Learning Algorithm in the text classification module. How accuracy of text classification can be increased by using LSTM Networks and Convolutional Neural Network(CNN) is discussed. The paper gives a basic yet useful overview of the LSTM model.[4]

Table -1: Comparison Table

Paper Title	Year	Seed Idea
Conversion of Natural Language Query to SQL Query	2018	- Traditional Approach for nlp - Syntactic analysis - Semantic analysis
SQL Query Formation Using Natural Language Processing (NLP)	2016	- NLP - DBMS handling - Lexical analysis
Automated SQL Query Generation by Understanding A Natural Language Statement	2018	- Grammar - POS tags - Mapping - Table interlinking
Research on Text Classification Based on CNN and LSTM	2019	- LSTM model and its working - Concepts related to LSTM

3. CONCLUSION

This paper gives an overview of various techniques used for translating natural language query into SQL query and specifies a combined methodology using NLP and Deep learning (LSTM), further gives an idea of how this can be implemented.

REFERENCES

[1] Abhilasha Kate, Satish Kamble, Aishwarya Bodkhe, Mrunal Joshi, "Conversion of Natural Language Query to SQL Query", 2nd International Conference on Electronics, Communication, and Aerospace Technology (ICECA 2018) IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-1.

[2] Prof. Debarati Ghosal, Tejas Waghmare, Vivek Satam, Chinmay Hajirnis, "SQL Query Formation Using Natural Language Processing (NLP)" , International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016.

[3] Amit Pagrut, Ishant Pakmode, Shambhoo Kariya, Vibhavari Kamble and Yashodhara Haribhakta, "Automated SQL Query Generator By Understanding A

Natural Language Statement", International Journal on Natural Language Computing (IJNLC) Vol.7, No.3, June 2018.

[4] Yuandong Luan, Shafu Lin, "Research on Text Classification Based on CNN and LSTM", 2019 IEEE International Conference on Artificial Intelligence and Computer Applications, March 2019.

[5] Sachin Kumar, Ashish Kumar, Dr. Pinaki Mitra, Girish Sundaram, "System and Methods for Converting Speech to SQL", International Conference on "Emerging Research in Computing, Information, Communication, and Applications", ERCICA 2013 pp: 291-298, Published by Elsevier Ltd ISBN:9789351071020.

[6]Tanzim Mahmud, K. M. Azharul Hasan, Mahtab Ahmed, Thwoi Hla Ching Chak, "A Rule Based Approach for NLP

Based Query Processing" , International Conference on Electrical Information and Communication Technology (EICT 2015).

[7]Aditya Narhe, Chaitanya Mohite, Rushikesh Kashid, Pratik Tade, Santosh Waghmode," SQL Query Formation for Database System using NLP" , International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 8 Issue 12, December-2019.

[8]https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/

[9]https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21

[10]https://www.datasciencecentral.com/profiles/blogs/top-nlp-algorithms-amp-concepts

[11]http://www.iaees.org/publications/journals/selforganizology/articles/2016-3(3)/algorithm-to-transform-natural-language-into-SQL-queries.pdf