International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.9 May - June – 2025
DOI:10.46647/ijetms.2025.v09i03.009 ISSN: 2581-4621

Text-to-SQL Conversion by using Deep Learning/Machine Learning: Integrating Natural Language with Database Queries.

Amar Kaygude[1], Onkar Rajguru[2], Sandesh Karad[3], G.T.Avhad[4]
[1]*UG - Computer Engineering, Vishwabharati Academy's College of Engineering, Ahmednagar, Maharashtra*
[2]*UG – Computer Engineering, Vishwabharati Academy's College of Engineering, Ahmednagar, Maharashtra*
[3]*UG - Computer Engineering, Vishwabharati Academy's College of Engineering, Ahmednagar, Maharashtra*
[4]*Associate Professor, Computer Engineering, Vishwabharati Academy's College of Engineering, Ahmednagar, Maharashtra*

Abstract
Accessing and extracting insights from databases can be challenging for non-technical users unfamiliar with
Structured Query Language (SQL). This project presents a novel solution that converts plain language into SQL queries using advanced deep learning (DL) and machine learning (ML) techniques. The system enables users, including academics, business professionals, and students, to interact with databases without needing to understand complex SQL syntax. Through a user-friendly interface, users input queries in natural language and receive accurate, real-time database responses. Transformer architectures (BERT, GPT, and T5) are employed to understand language structure, context, and syntax, generating accurate SQL queries for complex operations involving joins, nested statements, and conditions. This approach democratizes data access, reduces reliance on technical staff, increases productivity, and lowers training costs. The technology fosters a data driven culture by enabling direct interaction with databases, benefiting industries such as banking, healthcare, education, and customer service. Despite challenges in maintaining high accuracy, handling ambiguous queries, and managing diverse database schemas, extensive training on real-world datasets and robust error management are essential. The project's potential to revolutionize data access and utilization underscores its significance in promoting informed decision-making and innovation across various sectors.
Keywords: Natural Language Processing, Deep Learning, Machine Learning, SQL Query Generation, Transformer Models, Data Access Automation, Database Interaction etc.

1. INTRODUCTION

This project Text-to-SQL systems represent a significant leap in simplifying the interaction between humans and databases. By enabling users to frame their queries in natural language, these systems eliminate the need for in-depth knowledge of SQL syntax or database schemas. This innovation democratizes access to structured data, making it accessible to a broader audience, including those without technical expertise. However, these systems are not designed to perform logical reasoning or render judgments. Instead, they focus on generating
SQL queries to retrieve relevant data for users to analyse further. By bridging the gap between natural language and database querying, Text-to-SQL systems are transforming how data is

DOI:10.46647/ijetms.2025.v09i03.009 ISSN: 2581-4621

accessed and utilized, paving the way for improved decision making and insights across various industries.

SQL :

Relational databases are now widely used to store vast amounts of structured data from the Internet as a result of its growth. For structural data, the relational database offers easy query features together with reliable storage. While relational databases may be effectively accessed by proficient programmers using structured query languages (SQL), individuals without an understanding of SQL can access the databases with the help of natural language interfaces to databases (NLDB). As a result, the academic and industrial groups have taken an interest in text-to-SQL, which attempts to convert natural language (NL) descriptions and queries into SQL. As we cover in Chapter 3.1, the majority of text-to-SQL techniques now use neural networks, which are mostly based on the seq2seq model structure.

1.1 Text-to-SQL Assumption:

Text-to-SQL systems are specifically designed to convert natural language (NL) descriptions into SQL queries that retrieve structured data from databases. These systems assume a prior understanding of the underlying database schema, requiring users to phrase queries in a way that aligns with SQL syntax and the database's fields. Unlike traditional Question Answering (QA) models, Text-to-SQL systems do not perform logical comparisons or provide direct answers to questions requiring reasoning or judgment. Instead, they retrieve raw data for users to analyse on their own. For instance, to find someone's age, users might ask, "How old is Lily?" rather than phrasing the question as "Is Lily older than 18?" While the former retrieves Lily's age, the latter would require logical interpretation, which the system does not perform. Similarly, indirect information can be inferred by framing queries creatively, such as asking, "What is her nationality if she is older than 18?" Overall, these systems streamline access to structured data, but the responsibility of interpreting and reasoning with the retrieved data remains with the user. This approach highlights their role as data access facilitators rather than tools for quality control or direct reasoning tasks.

Table 1.

Name	Age	Gender	Nationality	Phone Number
Ram	25	male	Indian	7894561235
Kartik	36	male	Indian	9856231233
Rani	59	female	Indian	8854613254
Nami	26	female	India	9658432025

Aspect **Description**

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.9 May - June – 2025
DOI:10.46647/ijetms.2025.v09i03.009 ISSN: 2581-4621

Purpose Converts natural language (NL) descriptions into SQL queries to retrieve data from databases.

Focuses on creating SQL queries **Functionality** rather than direct Question-

Aspect **Description**
Answering (QA) or logical reasoning.

Query Phrasing Requires users to phrase queries in ways that match the database structure (e.g., "How old is Lily?" rather than "Is Lily older than 18?").

Limitations Does not directly perform logical comparisons or render judgments; provides data for users to interpret.

Facilitates structured access to data
Data Access but leaves quality control or interpretation to users.

1.1.Figures

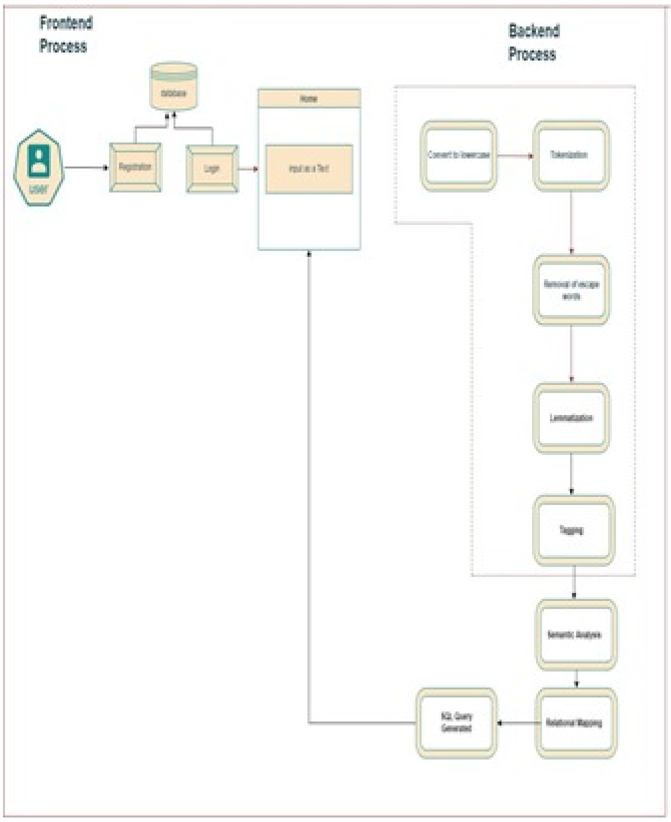

FIGURE 1 : **Proposed System**.

DOI:10.46647/ijetms.2025.v09i03.009 ISSN: 2581-4621

1.2 Results

Finally, the obtained data is given to the user in an understandable way. This might include tables, charts, or other visual aids that make the material more accessible and easier to understand. This step concludes the workflow by giving users with the information they sought, hence improving decision-making processes.

1.3 Discussion

After receiving input, the system uses Natural Language Processing (NLP) methods to evaluate and enhance the query. This includes techniques like tokenization, which divides the text into individual words or phrases; stop word removal, which removes popular terms that may not contribute value; and lemmatization, which reduces words to their simplest forms. This phase ensures that the query is clear and accurate, which improves the correctness of future actions.

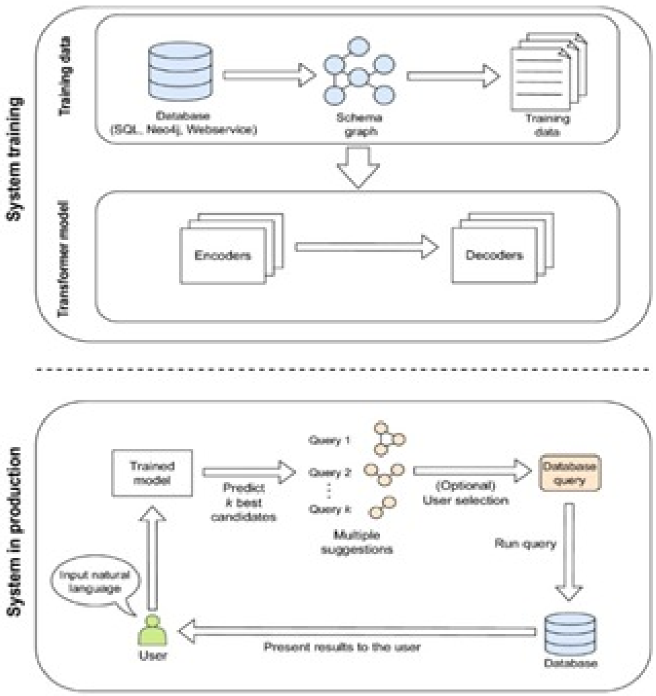

FIGURE 2 : **Existing System.**

CONCLUSION

By integrating cutting-edge advancements in deep learning and machine learning, Text-to-SQL conversion systems can evolve into powerful tools for democratizing data access. Future developments will focus on improving adaptability, understanding complex queries, and enhancing user experience. These systems are poised to play a crucial role in various industries, from healthcare to finance, by empowering users to interact effortlessly with their data and make informed decisions.

ACKNOWLEDGEMENTS :

I would like to express my sincere gratitude to all those who contributed to the successful completion of my project titled "Text-to-SQL Conversion using Deep Learning/Machine Learning: Integrating
Natural Language with Database Queries."

DOI:10.46647/ijetms.2025.v09i03.009 ISSN: 2581-4621

I extend my deepest appreciation to my project guide, for their constant support, guidance, and valuable insights throughout the course of this project. Their encouragement played a vital role in shaping my understanding and approach to this topic.

REFERENCES :

1. Stinivasan, 1, Lei, J, Tau, Y., & Smith, E. (2023) butraction Tuning of Large Pre-trained Language Models. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL), 1098-1107. https://doi.org/10.18653/v1/2023.acl- main. 109
2. C. Zhu, Y. Xu, X. Ren, B. Lin, M. Jiang, and W.
Yu, "Knowledge augmentedmethodsfornaturallanguageprocessing,"in Proc.49thAnnu. Meeting Assoc. Comput.
Linguistics, 2022.
3. R. Cao, L. Chen, Z. Chen, Y. Zhao, S. Zhu, and K. Yu, "LGESQL: Line graph enhanced text-to-SQL model with mixed local and non-local relations," in Proc.49thAnnu.MeetingAssoc.Comput.Linguistics,2 021
4. MeCann, B., Raffel, C., & Socher, R. (2020), decaNLP: A Natural Language Processing Decathlon. Transactions of the Association for Computational Linguistics, 8, 422-438 https://doi.org/10.1162/tacl_a_00352
5. Chen, S, Xic, P. & Xing, E. P. (2021). Multi-Task Learning in Natural Language Processing: An Overview. ACM Computing Surveys (CSUR), 54(8), 1-28 https://doi.org/10.1145/3427714