



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 9      Issue: V      Month of publication: May 2021**

**DOI: <https://doi.org/10.22214/ijraset.2021.34145>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# The Natural Language Query using NLP by Generation of SQL Query

Prajakta Devgade<sup>1</sup>, Prof. Jiwan Dehankar<sup>2</sup>, Prof. Priyanka Bhende<sup>3</sup>

<sup>1, 2, 3</sup>Department of CSE, T.G.P.C.E.T., Nagpur

**Abstract:** Database is a primary source of information. This information is accessed by a common man who isn't ready to write SQL query. Now-a-days most of the information is assembled in the form of a database. A user who wants to access a database but having limited or lack of knowledge of database languages, they find it difficult and challenging to access the database. Retrieving needful information from such repositories requires an extensive knowledge of database languages like SQL. However, nobody is in a position to write down the complex SQL queries to extract the useful information. Hence, there is a necessity for a system that enables the users to access the information from the database. A system using NLP by giving structured natural language as an input and receiving SQL query as an output, to access the related information from the database with ease.

**Keywords:** Natural Language Processing (NLP), Query Generation, SQL, Regular Expression, Syntax semantic parsing, parsing.

## I. INTRODUCTION

In the digital era, databases are the well-liked medium for data storage. Databases have applications almost in all information systems such as human resource management system, transport information system, financial information system, etc. It has become more important for humans to interact with computers to provide an assistance in many fields. Retrieval of the required information from the database is a monotonous process. In order to retrieve information from the database, requires user to know details of database such as relations, entities, etc that is the basic knowledge of database management system (DBMS), which is a software developed to store and handle the data in a database. Therefore, a naïve user faces difficulty in retrieving the data. Natural Language Processing (NLP) techniques are used as solution for such a problem and to make easier human interaction with computers.

The main problem is that the user who want to get information from the database, does not know formal languages like SQL. To provide a solution to this we propose a system that uses English language to query the database. Our system accept query in English language and convert it into SQL statement.

## II. OBJECTIVE

Now a days, everyone is having their own device who connects with the internet. Every user is trying to gather the information that they require from internet source. Most of the information is assembled in the form of a database. If a user who wants to access a database but having limited or lack of data about database languages it's difficult and challenging situation to access the database. Hence, there is a necessity for a system that enable the users to access the information from the database. Objective is to convert a natural language query into a SQL query to simplify data extraction.

## III. SYSTEM ARCHITECTURE

The proposed system is meant to attenuate the communication gap between a person's and computer. A system is meant which contains an intelligent layer that accepts common user's sentences in natural language as input, converts these sentences into standard SQL queries and executes them to retrieve data from relational databases. Our proposed system consists of several modules that are used to extract key words alone and leave out the redundant data. This is critical because presence of redundant data will certainly decrease the overall performance of the system. Input data initially goes through an NLP phase followed by a mapping phase.

### A. Graphical User Interface (GUI)

The system has a user-friendly GUI which consists of a query editor through which the user can provide the English language query to the system and a query generator that provides the resultant Query of the user's request.

### B. Natural Language Processor

The input provided by the user must be processed to identify the constraints and predicates required to formulate the desired SQL query. The NLP makes use of techniques such as tokenization, part of speech tagging, chunking and entity recognition. The NLP provides chunked tags which are processed to analyses the attributes and predicates such as table name, attribute name or column names and selection conditions or criteria for the input user’s query.

### C. Query Generator

The query generator makes use of a query translation. The algorithm helps in identification of predicates based on the set of rules. A prediction model is, used when the translation algorithm fails to generate sufficient information to formulate the query, also included in the query generator to predict the query. The formulation of query is done with the help of pre-defined structure and attributes and predicates identified using NLP.

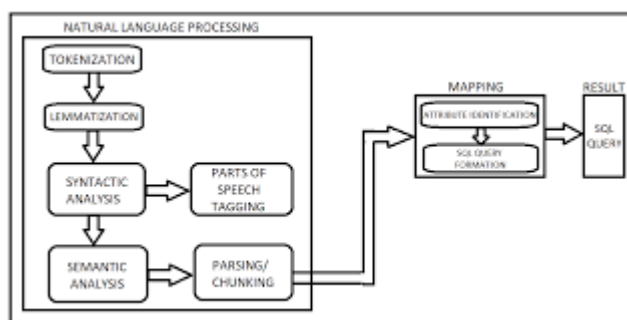


Fig. 1 Flow diagram of NLP to SQL

In the pre-processing phase, the input sentence is split into tokens. From these tokens, root is extracted by getting obviolate inflectional morphemes. This is done with the help of lexicon and suffix replacement rules. Second phase of database component is database management system (DBMS) which execute the SQL query over the database.

Third phase of database component is response generator. It takes the output of DBMS as its input and converts it into the English language.

## IV. CONCLUSIONS

In this paper, we present an approach to convert an English sentence into SQL query. The intelligent interface developed uses semantic matching technique which translates natural language query to SQL. It also uses set of production rules and data dictionary which consists of semantics sets for relations and attributes. A series of steps like small letter conversion, tokenization, speech tagging, database element extraction, SQL element extraction and ambiguity removal is employed to convert Natural Language Query to SQL Query.

## V. ACKNOWLEDGMENT

I would like to express my thanks of gratitude to my guide “Prof. Jiwan Dehankar” for their able guidance and support in completing my project. I would like to extend my gratitude to the HOD “Prof. Dr. N. V. Chaudhari Sir” and all professors for providing me with all the facility that was required.

## REFERENCES

- [1] Uma, M., Sneha, V., Sneha, G., Bhuvana, J., & Bharathi, B. (2019). Formation of SQL from Natural Language Query using NLP. 2019 International Conference on Computational Intelligence in Data Science (ICCIDS).
- [2] Das, A., & Balabantaray, R. C. (2019). MyNLIDB: A Natural Language Interface to Database. 2019 International Conference on Information Technology (ICIT).
- [3] Kate, A., Kamble, S., Bodkhe, A., & Joshi, M. (2018). Conversion of Natural Language Query to SQL Query. 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA).
- [4] Reshma, E. U., & Remya, P. C. (2017). A review of different approaches in natural language interfaces to databases. 2017 International Conference on Intelligent Sustainable Systems (ICISS).
- [5] Li, Q., Li, L., Li, Q., & Zhong, J. (2020). A Comprehensive Exploration on Spider with Fuzzy Decision Text-to-SQL Model. IEEE Transactions on Industrial Informatics, 16(4), 2542–2550.
- [6] Mahmud, T., Azharul Hasan, K. M., Ahmed, M., & Thwoi Hla Ching Chak. (2015). A rule based approach for NLP based query processing. 2015 2nd International Conference on Electrical Information and Communication Technologies (EICT).



- [7] Sangeeth, N., & Rejimoan, R. (2015). An intelligent system for information extraction from relational database using HMM. 2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI).
- [8] Kumar, R., & Dua, M. (2014). Translating controlled natural language query into SQL query using pattern matching technique. International Conference for Convergence for Technology-2014.
- [9] Agrawal, R., Chakkarwar, A., Choudhary, P., Jogalekar, U. A., & Kulkarni, D. H. (2014). DBIQS — An intelligent system for querying and mining databases using NLP. 2014 International Conference on Information Systems and Computer Networks (ISCON).
- [10] Shah, A., Pareek, J., Patel, H., & Panchal, N. (2013). NLKBIDB - Natural language and keyword based interface to database. 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI).
- [11] Siasar djahantighi, F., Norouzifard, M., Davarpanah, S. H., & Shenassa, M. H. (2008). Using natural language processing in order to create SQL queries. 2008 International Conference on Computer and Communication Engineering.
- [12] Adwait Ratnaparkhi, "Learning to Parse Natural Language with Maximum Entropy Models," *Machine Learning*, vol. 34, issue 1-3, Feb. 1999, pp. 151-175.0
- [13] I. Androutsopoulos, G. D. Ritchie and P. Thanisch, "Natural Language Interfaces to Databases – An Introduction," in *Natural Language Engineering*, vol. 1, part 1, 1995, pp. 29-81
- [14] George A. Miller (1995), "WordNet: A Lexical Database for English," in *Communications of the ACM* Vol. 38, No. 11: 39-41.
- [15] Ana-Maria Popescu, Oren Etzioni and Henry Kautz, "Toward a Theory of Natural Language Interfaces to Databases," in *IUI Proceedings* 2003.



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)