ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 8.4
IJAR 2023; 9(9): 17-20
www.allresearchjournal.com
Received: 14-07-2023
Accepted: 18-08-2023

Anu
Department of Computer Science Engineering, Desh Bhagat University, Mandi Gobindgarh, Punjab, India

Khushboo Bansal
Assistant Professor, Department of Computer Science Engineering, Desh Bhagat University, Mandi Gobindgarh, Punjab, India

Corresponding Author:
Anu
Department of Computer Science Engineering, Desh Bhagat University, Mandi Gobindgarh, Punjab, India

A rule based approach to align natural language with query language

Anu and Khushboo Bansal

Abstract
In the modern world, accurate, quick, and exact data access is essential for analysis and commercial decision-making. Users can access data from a database using the NLIDB (Natural Language Interface to Database) system by inputting requests that are expressed in natural language. In this research, a novel method for creating NLIDBs for natural language interface to databases is presented. It employs a two-stage processing method that includes syntactic processing first, then semantic processing. The processing becomes more robust and general thanks to syntactic processing. These relations' proximity facilitates and improves the accuracy of semantic processing. The systems become more portable as a result. In this study, we utilised the Hindi language interface for the patient query database to conduct Hindi-language database queries.

Keywords: NLP, Hindi, NLIDB, SQL, Hospital

1. Introduction
One of the main sources of information is databases [1]. Information is kept in databases using the relational model due to the rise in big data and the requirement for quick, easy retrieval of data, and it can then be obtained by making a query in the database using Structured Query Language (SQL). As a result, SQL serves as a hub for database information retrieval. However, this is annoying for customers who are unaware of how these intricate SQL queries are created. Linguistics and artificial intelligence (AI) can be coupled to create software that can assist in understanding and producing data in natural language. These information retrieval systems can be used to manage and retrieve information in a variety of educational and informational institutions as well as businesses. The NLIDB approach employs natural language for the information seeker to submit the query and obtain the results from databases, saving the user from having to learn the database language/schemas to make the queries. To transfer the unprocessed Hindi text to the database query, NLIDB employs a number of procedures. After that, the database receives this query for effective information retrieval.
This interface enables the average user to enter a query in the system using a language other than English, such as Telugu, Punjabi, Hindi, or English, and to view the results in that language. Natural Language Processing (NLP) includes the discipline of NLIDB. In today's cutting-edge technology, particularly in artificial intelligence components, the computer understands the query in natural language and provides the answer rather to having the user type a command to search for the answer. NLP is often dependent on converting natural language into machine learning these days. Fig. 1 illustrates the relationship among deep learning, machine learning and natural language processing.

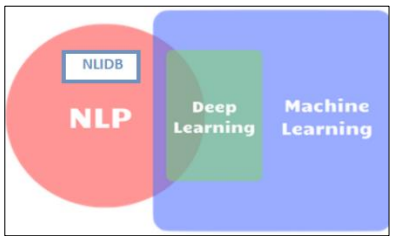

Fig 1: Relationship between NLP, Deep learning and Machine learning

A. Machine Learning

Machine Learning (ML) [2] is the most potent technology available today. It is an artificial intelligence (AI) application that concentrates on programme creation and can access and utilise data to learn new things. It gives the system the ability to learn automatically and get better based on experience without having to be explicitly coded. It offers a variety of applications, including a self-driving car, efficient web search, useful speech recognition, a photo tagging tool, and a spam detector. Machine learning is now being used by businesses to enhance decision-making, identify diseases, recognise faces, boost production, predict the weather, and do many other things. The main goal of ML is to provide the system the ability to learn automatically, without human supervision, and operate on its own.

B. Deep Learning

Big data can be utilised for knowledge application, knowledge discovery, and knowledge-based prediction using deep learning, a subset of machine learning (ML) in artificial intelligence. It has a learning network that can process unstructured or unsupervised data. Deep neural learning or deep neural networks are other names for it. Deep Learning is utilised in a number of algorithms and to address fundamental AI issues. According to economic data, is the new oil in the modern era. If data were analogous to crude oil, data warehouses and databases would serve as the drilling rings that fracture and generate the data on the internet, while deep learning would serve as the oil refinery that transforms crude oil into a usable end product. There is a lot of crude oil beneath, and there are many pumps and drills on the market, but we couldn't obtain anything useful without a refinery. Because of this, both established businesses and new start-ups rely on the idea of deep learning.

C. Natural Language Processing

Natural Language Processing (NLP) is a field of study that enables computers to comprehend human speech. Artificial intelligence's field of NLP focuses on developing systems and using mathematical and computational techniques to automate many parts of human (natural) language. It makes considerable use of compiler phases like lexical and syntactic analysis.

Natural Language Processing = Text Processing + Machine Learning

One way to describe a language is as a system of symbols or rules. Rules dominate the symbols in the language, which are combined and used to communicate information. When a user wishes to use natural language to interact with a machine, NLP is created. NLP assists users who lack the time to learn new machine specific language and become proficient in it because not every user can be trained in machine specific language. NLP is helpful in teaching computers how to carry out difficult tasks in natural language, such as conversation generation, machine translation, text summarization, sentiment analysis, and Natural Language Interface to Database (NLIDB). The primary advantage of NLP is that it eliminates the need to learn machine-specific languages like ML or database languages like SQL in order to operate with the NLP system. Numerous fields, including computer and information science, electrical and electronics engineering, linguistics, robotics, psychology, mathematics, artificial intelligence (AI), agriculture, and weather forecasting, constitute the basis of NLP.

D. NLIDB (Natural Language Interface to Databases)

A system known as a natural language interface to a database (NLIDB) [3] enables users to access data from databases by entering queries that are stated in one or more natural languages (for example, Hindi) or subsets of natural languages. It could be challenging for anyone without any database language understanding to access databases. There has been an increase in recent years in the need for non-expert users to query relational databases in a more natural language that includes linguistic concepts and variables. Thus, the notion of using natural language rather than SQL led to the creation of a new category of processing technology known as Natural Language Interface to Database. NLIDB is still an open research problem even though the early studies began in the late 1960s. We will gain from having a comprehensive NLIDB system in numerous ways. By using such methods, anyone can extract information from the database. Additionally, it could alter how we view the data in a database. People are accustomed to working with forms; hence, their expectations are highly reliant on the capabilities of the form. In this way, a database will be utilized to its full potential.

Architecture of NLIDB

In Fig. 2, the NLIDB architecture is depicted. The system is broken down into five primary parts, which are as follows:

1. User Interface

This element has direct communication with the user. The user can enter an inquiry in Hindi here. Here, you may also see the outcome of SQL execution or any error messages.

2. Question Analyzer

The analysis of natural language queries will take place in this component. This component will employ a parser to assist in breaking down the user-provided sentence, and the system will further process the dependencies to simplify the translation procedure.

3. Query Generator

Based on a set of criteria, this component accepts some dependencies from Question Analyzer and converts them into a SQL query. This component will attempt to build a piece of a SQL query while checking each dependency's label and contents (head and dependent). The portions created will be integrated at the end of the procedure into a complete SQL query that is prepared for execution.

4. Keywords Mapping

This module maps the database keywords with tokens to generate the query language.

5. Database Interface

This element communicates with the relational database that it is built upon. A mapping must be created before the system can convert a natural language query into a SQL query. This component facilitates keyword mapping by offering the selected database's metadata. A SQL query is

afterwards provided to this component for execution after being generated by Query Generator. The user interface will receive the execution result and present it to them.

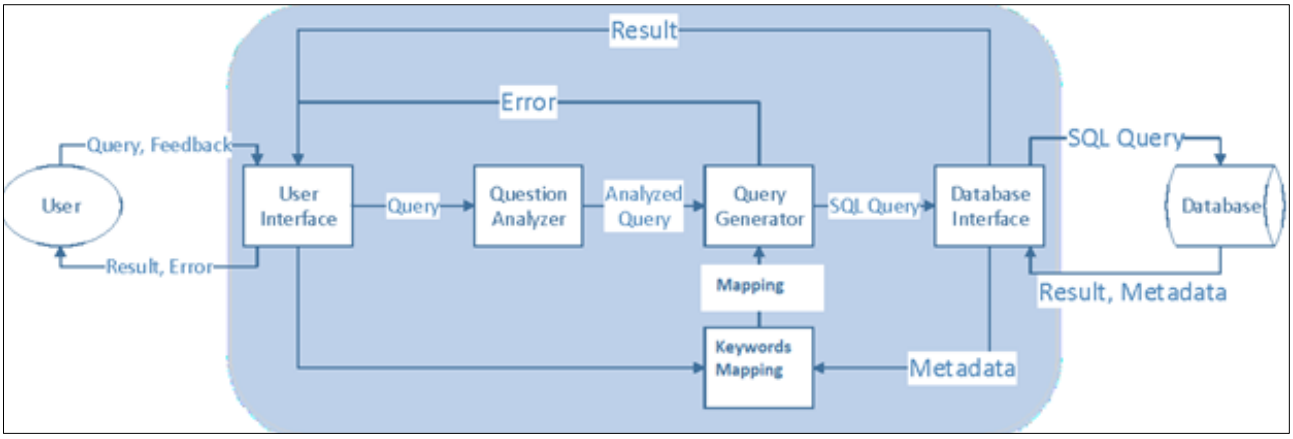

Fig 2: NLIDB System Architecture

2. Related work

According to Pertik Garg *et al.* [2], the need for information is a vital component of our lives. Although there are several sources of the information, a database is the primary one. We can access, save, and retrieve information with the use of a database. No business or sector can survive in the modern technological age without using databases. Every computer-based programme needs to retrieve data from a database, which necessitates familiarity with database languages like SQL. However, not everyone can receive SQL training. Researchers offer the Natural Language Interface Database (NLIDB) as a remedy to solve this issue. Natural language can be such as Bengali, Urdu, Punjabi, Hindi, and English. An ideal user interface between a layperson and a computer application can take the place of SQL. The many NLIDB parameters will be covered in this essay, along with a description of the hospital enquiry system's organisational structure. In this research, we will create a medical inquiry application that will accept queries in Punjabi, translate them into formal query language, or SQL, and then output the results back in Punjabi. The outcomes of the application's development were also examined.

According to D. Ramesh *et al.* [5], the abundance of data on the internet necessitates the need for organised information storage in order to make finding, retrieving, and maintaining information easier. An innovation that saves information in a standardised and coordinated manner is an information base. Information about the structures question language (SQL) becomes fundamental to efficiently working with these data bases. However, the usage of SQL restricts access to data bases from clients that lack the necessary information. To enable the entry of these information bases even to non-master clients, an interface need enters the picture. This article outlines the strategy for developing a Telugu-language information base interface.

Data plays a vital role in our life when it comes to registration, according to Ashish Kumar *et al.* [6]. Information bases are one of the main sources of data. Innovation in information bases and databases has a big impact on how PC usage is growing. Almost all IT applications store data or information and then retrieve it from an information base. Information storage and retrieval have often been handled via database management systems (DBMS). Information bases, however, are frequently challenging to utilise because of their interface's rigid approach to client collaboration. Data set language knowledge, such as SQL, is needed for storing and retrieving data from information bases. ANSI's Organised Query Language (SQL) standard is used to access and manage data stored in information bases. But since everyone might not be familiar with the punctuation and syntax of SQL and the information base separately, nobody will be able to write the SQL query. Hindi is most people's default language when they are in India. Many e-administration programmes also leverage databases.

In their study, Jaspreet Kaur *et al.* [7] found that as technological advancement increases, Question Answering becomes a more important area of study for experts. Clients ask a variety of questions in an effort to get precise answers from Question Answering Systems. When a client asks a question in everyday language as opposed to an inquiry, question answering is the best response to obtain detailed and accurate answers. Famous Indian dialects include Hindi, Telugu, Bengali, and others. Analysts are now debating these dialects, and a great deal of work is being done in these and other Indian dialects. In this paper, they examined how Question Answering Systems work across different Indian dialects.

Data mining has been the most established at this point among the fascinating trendy phrases, according to Manu Bansal *et al.* [8]. Many organisations routinely underuse their generally available information sources. These data stores must be mined for knowledge and emphasis is on comprehending the fascinating examples. The use of information mining on a library the board framework is the main focus of the flow research. Information stores or distribution centres are typically the targets of information mining. It raises costs for a variety of things, including programming, hardware, support, and specialists. The goal is to concentrate on how the ongoing data stored in the data set may be made usable without creating a new information distribution hub.

According to Mohit Dua *et al.* [9], DBMS have been utilised for data retrieval and archiving. However, regular users find it difficult to access the database's data. Many e-government applications, including those for banking, agriculture, railroads, and weather forecasting, employ databases to retrieve and store data. For those who are comfortable using Hindi, these applications should accept a Hindi sentence as a

query, process it, and then return responses in Hindi. The author of this paper explains the system's architecture for mapping Hindi sentences into SQL queries. The established technology allows users to enter queries in Hindi and receive results in that language, making data retrieval simple for the average person. To process the database, users do not need to understand any database or structure languages like SQL.

3. Results

Processing the database could be challenging for anyone without any background in structured language or databases. There has been an increase in recent years in the need for non-technical users to query databases in a more natural language that includes linguistic concepts and variables. As a result, a Hindi language interface for hospital inquiry is required in order to map with a Hindi language database. In this study, a Hindi typeface called Unicode is being employed. The system was created using SQL server and ASP.NET. The developed system's code is written in C#.

The user enters data into the database system using a Hindi language interface. The system will break the sentence into tokens after receiving the user's input. After that, these tokens will be mapped in order to determine the names of the columns and tables, as well as the conditions, instructions, and values. The tokens are then translated into English words. Unneeded tokens are thrown away. Then, using the useful tokens, a SQL query is created. Only a few queries that will display the results will be used to test the system. Any queries' results are displayed, for instance:

- Example-1

राजेश नाम का मरीज कौनसे कमरे मे है।

The SQL statement for the natural language query is:

select * from patient_info where patient_name = N' राजेश %'

Following query execution, the result is shown in Fig. 3.

id	name	father_name	address	admit_date	room_no
5	राजेश	शाम	हरमन नगर	3 मई 2022	कमरा नं 6

Fig 3: Result of Example-1

- Example-2

हरजीत की चमडी के डाकटर के साथ सोमवार को मुलाकात पक्की करो

The SQL statement for the natural language query is:

select * from doctor_schedule where department like N' चमडी %'

Following query execution, the result is shown in Fig. 4:

id	patient_name	appointment_day	doctor_name
1031	हरजीत	सोमवार	गोबिंद

Fig 4: Result of Example-2

It has been designed to use NLIDB for hospital queries, which accepts Hindi input and returns results in Hindi after processing a SQL query that was originally written in natural language. Finding the correct table to deliver the results is the major task in the development.

4. Conclusion

Through this technology, a user can access database data without being familiar with formal query languages like SQL. Users who are accustomed to using Hindi in daily life are the target audience for this system. An example of a human-computer interface is the Hindi Language Interface to Database. In this paper, the user enters a Hindi-language input inquiry pertaining to the patient and doctor to receive a Hindi-language response.

5. References

1. Pooja Malhotra, Yash Kapadia, Krishna Saboo, Ankita Sarda. Question-Answering System using Natural Language Processing with NLIDB Approach, International Journal of Current Research. 2017;9(09):57575-57577.
2. Pertik Garg, Rupinder Singh, Manju Bala Goel. A Novel Approach to Implement Natural Language Interface for Database in Punjabi Language. Journal of Xidian University. 2020;14(4):1230-1238.
3. Shripad Pramod Rane, Akanksha Kulkamil. Natural Language Interfaces to Database (NLIDB). International Journal of Innovative Science and Research Technology. 2022;7(11):616-618.
4. Reinaldha F, Widagdo TE. Natural Language Interfaces to Database (NLIDB): Question handling and unit conversion. International Conference on Data and Software Engineering (ICODSE); c2014. Doi: 10.1109/icodse.2014.7062663
5. Ramesh D, Suresh Kumar Sanampudi. Telugu Language Interface to Databases. International Journal of Advanced Research in Computer and Communication Engineering. 2013 July;2(7).
6. Ashish Kumar, Kunwar Singh Vaisla. Hindi Language Interface to Database using Semantic Matching, Oriental Journal of Computer Science & Technology. 2013;6(2):133-140.
7. Jaspreet Kaur, Vishal Gupta. Comparative Analysis of Question Answering System in Indian Languages, 128X International Journal of Advanced Research in Computer Science and Software Engineering. 2013;3(7).
8. Manu Bansal, Mandeep Kaur. Analysis and Comparison of Data Mining Tools Using Case Study of Library Management System. International Journal of Information and Electronics Engineering. 2013;3(5).
9. Dua M, Kumar S, Virk ZS. Hindi Language Graphical User Interface to Database Management System, in Proc. of ICMLA'13. 2013;2:555-559.
10. Pertik Garg, Anu Marwaha, Manju Bala Goel. Identification and Classification of Reduplication Words in Punjabi Language. International Journal of Scientific & Technology Research. 2020;9(6):532-537.