

# Implementation of Natural Query Processing in English and Punjabi using Keywords Approach

Kamalpreet Singh

*Department of Computer Science Engineering  
ASRA College of Engineering & Technology, Bhawanigarh, India*

Rajnish Kansal

*Faculty of Computer Science Engineering Department  
ASRA College of Engineering & Technology, Bhawanigarh, India*

**Abstract-** Information is playing an important role in our lives. One of the major sources of information is databases. Databases and database technology are having major impact on the growing use of computers. Almost all IT applications are storing and retrieving information from databases. Retrieving information database requires knowledge of database languages like SQL. This idea of using NL has induced the development of new sort of processing method in database systems. This new system can be named as Natural Language Interface to Database Systems (NLIDBs). It is a type of communication channel between the user and the computer. The user no longer needs to learn any SQL queries. Without the knowledge of any programming language, a user can act as a programmer. No such hectic queries are required from the user, by the system. It becomes very easy for a person to access data from database who has no knowledge of formal query language. In this paper we developed an algorithm that eliminates the problem of normal user to interact with database with rigid language SQL. The users are able to access information's by issuing query in simple Punjabi or English language.

**Keywords –** NLP, SQL, English, Punjabi, Database.

## I. INTRODUCTION

Databases [4] are comprehensive and important element in organizations. No organization or industry is possible without the use of database. In order to extract information from database one needs to formulate a query that the computer can understand and produce the desired data. Plenty of database tools are available in the market for retrieving information from database. To retrieve information from database, applications require knowledge of database language such as Structured Query Language (SQL). One has to learn language SQL to access the information stored in relational databases. Hence it becomes difficult for normal user to retrieve information without SQL. The most common way for people to obtain information is by asking questions in their natural language. To find a alternate to this many researches turned out to the use of Natural Language like Hindi, English, Punjabi, Tamil, French etc. for retrieving information. This approach is helpful for naïve users in accessing the databases. Each and every computer based application need to access information from database that requires knowledge of formal query language like SQL. But it is not possible for everyone to learn or write SQL queries. To overcome this problem many researchers for nontechnical users have brought out Natural Language (NL) i.e. Punjabi, Arabic, English, Bengali etc. in place of formal query language which will be very much useful for layman users. This idea of using NL has induced the development of new sort of processing method in database systems. This new system[3] can be named as Natural Language Interface to Database Systems (NLIDBs). It is a type of communication channel between the user and the computer. The user no longer needs to learn any SQL queries. Without the knowledge of any programming language, a user can act as a programmer. No such hectic queries are required from the user, by the system. It becomes very easy for a person to access data from database who has no knowledge of formal query language. In this thesis we will interface the Natural language (Punjabi and English) to the Database systems.

## II. ARCHITECTURE OF NLIDB

There are four phases in the Architecture of NLIDB (Natural Language Interface to Database Systems) system as shown in figure 1. The four phase are- Token filters, Parser, Query creator and DBMS [1]. Output from the upper phase becomes input for the lower phase. Descriptions of all phases are given below:

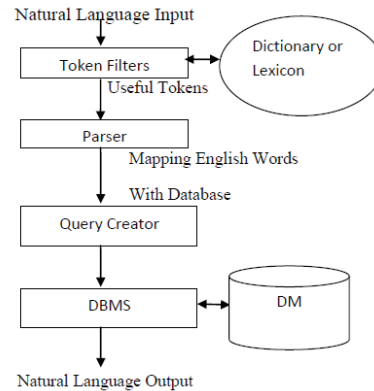


Figure 1. Architecture

### 1. Token Filter

This phase takes input from the user in Natural language. It then splits the sentence of Natural Language into tokens. All the tokens must be separated by comma, space etc. from each other. These tokens are then stored in an array. Tokens may represent name of a table, column, row, command, operator or it may be any value or any non-useful word. These all the tokens are output of this phase. Some tokens may be conditions or values etc. This phase's output is given to the next phase.

### 2. Parser

Parsing [2] is the most important phase. It takes inflow from the Token Filter phase. All tokens and their English word and token type are stored in database dictionary. The output by the upper phase is then matched with lexicon words one by one. If the match found then its corresponding English word with its type are saved. The tokens which are found useless are discarded in this phase. Only the useful tokens are kept. Now, we have with name of the table, attribute, conditions, and functions etc. that are used further to frame SQL query. The output of Token Filter phase is its input and it converts the tokens (that it takes from the above phase) into SQL query. Only those tokens are stored in lexicon that is enough to create an SQL query. Only the tokens that are required to generate a query are stored in lexicon. For every token of natural language its corresponding English word and its type is also stored. The token stored may be a table name, column, function, condition etc.

### 3. Query Creator

This phase [5] takes input from Parser phase. It takes name of a table, column, conditions, command and functions. SQL query is created by this phase with the help of all these tokens.

### 4. Database Management System

SQL query is executed in this phase. This component then converts the result into user's own Natural Language.

## III. RESULTS

### 1. Train Enquiry Form

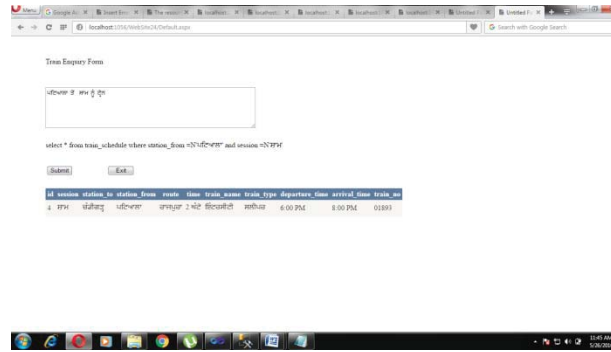


Figure 2. Train Enquiry form for searching station

User can search the trains going from Patiala in evening in Punjabi language. The entered Punjabi text is converted into sql query and fetches the results from the database as shown in figure 2.

2. Train Enquiry Form



Figure 3. Train enquiry form to chandigarh station

User can search the entire information like stoppage no., Train type, and train name etc for the trains going to Chandigarh by entering in Punjabi language. The entered Punjabi text is converted into sql query and fetches the results from the database as shown in figure 3.

3. Train Enquiry Form

User can search the entire information like train type, train name, departure time, arrival time and route etc for the trains going from Sangrur to Delhi by entering in Punjabi language. The entered Punjabi text is converted into SQL query and fetches the results from the database as shown in figure 4.



Figure 4. Train enquiry form for Sangrur to Delhi station

4. Train Enquiry Form

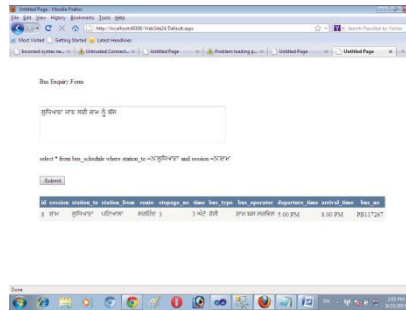


Figure 5. Train enquiry form for different sessions

User can search the entire information like train type, train name, departure time and arrival time etc. for the trains going to Ludhiana for the session in evening by entering in Punjabi language as shown in figure 5. The entered Punjabi text is converted into sql query and fetches the results from the database.

5. Train Enquiry Form

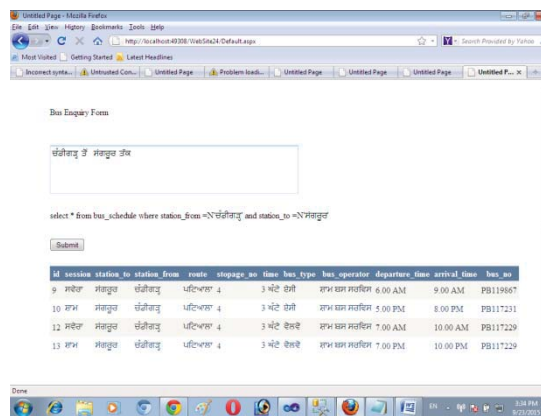


Figure 6 Train enquiry form for particular source to destination i.e from Chandigarh to Sangrur

User can search the entire information for the trains going from particular source to destination like search all the trains going from Chandigarh to Sangrur as shown in figure 6. The entered Punjabi text is converted into sql query and fetches the results from the database.

6. Train Enquiry Form



Figure 7. Train enquiry form for particular source to destination i.e from Sangrur to Delhi with Train Type

User can search the entire information for the trains going from particular source to destination like search all the trains going from Sangrur to Delhi with train type like mentioned as Passenger as shown in figure 7. The entered Punjabi text is converted into SQL query and fetches the results from the database.

7. Train Enquiry Form



Figure 8 Train enquiry form for particular source to destination i.e from Luckhnow to Chandigarh

User can search the entire information for the trains going from particular source to destination like search all the trains going from Chandigarh to Sangrur as shown in figure 8. The entered Punjabi text is converted into sql query and fetches the results from the database.

8. Train Enquiry Form



Figure 9. Train enquiry form for source to destination by mentioning train type and session

User can search the entire information for trains going from particular source to destination like search all the trains going from Sangrur to Delhi as shown in figure 9 by mentioning the train type and session. The entered Punjabi text is converted into SQL query and fetches the results from the database.

9. Train Enquiry Form in English

Figure 10 to figure 13 shows that the entered English text is converted into SQL query and fetches the results from the database.

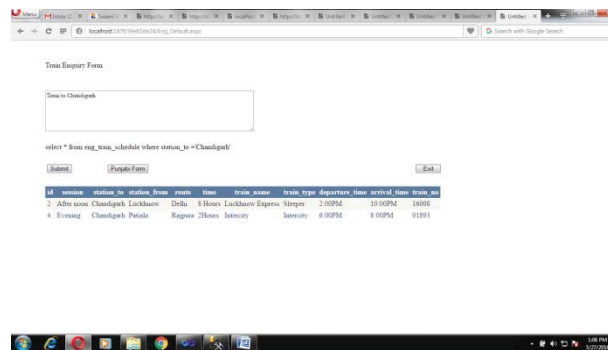


Figure 10. Train enquiry form for source to particular destination

Figure 10 shows the Train enquiry form for source to particular destination.

10. Train Enquiry Form

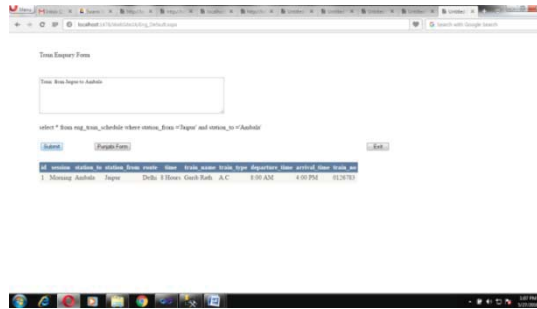


Figure 11. Train enquiry form for particular station to particular destination

Figure 11 shows the enquiry form for particular station to particular destination

11. Train Enquiry Form

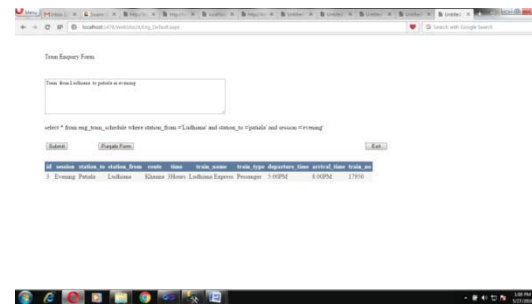


Figure 12. Train enquiry form for particular station to particular destination in the particular session

Figure 12 shows the Train enquiry form for particular station to particular destination in the particular session.

12. Train Enquiry Form



Figure 13. Train enquiry form for particular station to particular destination in the particular session for the particular type of train

Figure 13 shows the enquiry form for particular station to particular destination in the particular session for the particular type of train.

IV.CONCLUSION

In this thesis we developed an algorithm that eliminates the problem of normal user to interact with database with rigid language SQL. The users are able to access information's by issuing query in simple Punjabi or English language. The Proposed system is developed in .NET. The application accept enquiry in natural language, convert it into technical SQL query and display the results in the same language. The features of this application are that it is online application, admin control panel, easy to operate and export to excel options. User can operate their query in both English and Punjabi language. The benefit of two languages for enquiry is the major feature of the research.

REFERENCES

[1] Priyanka Arora, Punet Goswami, "An Efficient Hindi Language Interface using Relational Databases", International Journal of Research in Computer and Communication Technology, Vol 3, Issue 6, June – 2014.  
 [2] Abhijeet R. Sontakke and Amit Pimpalkar(2014), "A Review Paper on Hindi Language Graphical User Interface to Relational Database

- using NLP”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 10.
- [3] Saravjeet Kaur, Rashmeet Singh Bali, “SQL Generation and Execution from Natural Language Processing”, International Journal of Computing & Business Research, Proceedings of ‘I-Society 2012’ at GKU, Talwandi Sabo Bathinda (Punjab).
  - [4] Joginder Singh, Pratiba Verma and Navneet Kaur(2016), “Natural Language Interface to Database-An Introduction”, Imperial Journal of Interdisciplinary Research (IJIR) Vol.2, Issue-1.
  - [5] Ashish Kumar and Kunwar Singh Vaisla, “Hindi Language Interface to Database using Semantic Matching”, Oriental Journal of Computer Science & Technology, June 2013, vol. 6, no. (2):Pgs.133-140.