A Rule-Based User Interface for Accessing Relational Databases via the Use of Natural Language Processing

Abstract: The data has been stored and retrieved using database management systems. Databases are notoriously difficult to work with due to the inflexibility of their user interfaces. A database is essential to almost every e-governance program. Agricultural, banking, weather forecasting, railway, and other services are all part of what they provide. These programs are necessary for those who are comfortable with the Hindi language since they can take a Hindi phrase as a query, process it, and then display the response to the user in Hindi. When improving the Hindi language for the average person, natural language processing (NLP) is a crucial tool. Here, we have created a rule-based system that will meet the needs of the user; it will only take queries written in Hindi and provide output in Hindi.

Introduction

1. As a subfield of AI, natural language processing encompasses tasks such as language analysis, information retrieval, and machine translation. The goal of implementing a natural language processor into database access is to provide simpler dataset access for the general public. Although various groups of individuals may find natural language to be the simplest or most difficult. To this day, computers still struggle to fully grasp and implement this structure. In order to access the database, a user has to be familiar with DBMS and Structured Query Language (SQL). Users with expertise in SQL and other database-related languages are severely restricted in their ability to access data and information. The data can only be accessed via a rule-based graphical user interface, which requires a basic understanding of how to utilize the system. Many Indians are familiar with the English language, yet they lack the proficiency to ask complex questions in the language. Natural languages like English, Hindi, Marathi, etc., may be used to query the system using this rule-based graphical user interface, and the response can be shown in the same language.

2. Related Work

These days, a lot of public and commercial businesses, as well as schools and universities, rely on computer-based information technology to assist them manage their data and operations. Managing data is the job of information systems. Database Management System (DBMS) [1] is an information management system that can handle any kind of data contained in a database. Databases are an integral part of both public and private information systems, and they have many uses. [2]. Databases are very much important for large variety of application areas employing private and public information systems. Retrieval of a large amount of the same type of datais very efficient in relational databases [3], but still the user has to master the DB schema completely to formulate the queries.
(SQL) Structured Query Language is an ANSI standard for manipulating and accessing the information stored in databases. It is comprehensively employed in industry and issupported by major DBMS. Most of the languages used for manipulating relational database systems are based on the norms of SQL. In the past few years many advances have been made in the field of databases and many other fields of critical relevance to information technology. An intelligent database is an emerging database technology that has dramatic impact on the way we think and work [4].
In recent times, there has been a rising demand for non- expert users to query relational databases in a more natural language encompassing linguistic variables and terms,instead of operating on the values of the attributes. Talking to a computer in a natural language such as plain English is always a dream that drives the progress of human-computer interaction work [5, 6].
For the last thirty years, numerous attempts have been made to build useful natural language interface. It has turned outto be much more difficult than what was originally expected.There have been large numbers of research works introducing the theories and implementations of NLIDBs. There are mainly four kinds of four NLIDBs framework [7]. The first type of framework is based on pattern matching. A typical application of this type of framework is SAVVY [8]. In this system, various patterns are written in some different kind of queries and these patterns are

executed after the complete queries are entered. The main advantage of pattern matching approach is that no elaborate parsing and modules of interpretation are required and the systems are very easy to implement. Some pattern matching systems were able to perform impressively well in certain applications. One of the NLP system that is based on pattern-matching approach is (), SUM () and AVG (). The user will type the query in Hindi language and that natural language has been processed and will give the output in Hindi language only. Time difference has been calculated, system will give translation time and execution time in milliseconds as well as in nanoseconds.

2.1 System Overview

This system has been developed for the people who are not having any knowledge of query languages. Asking query in Hindi language and getting the result in same language is the biggest advantage for the person who knows Hindi language. Objectives that are covered are given as:-

Design the GUI through which user can easily give input query in Hindi language also he will get output in the same language. We have design the Hindi language interface to relational database. We have developed a system that can handle Hindi query for the extraction of single or multiple columns from tables stored in databases. This system can perform all the operations such as select, insert, update, delete on relational databases as well as it will perform aggregate functions such as sum (), max (), min (), avg () on the same relational database. We have also worked on semantic behavior of the independent query. Translation time and execution time is also calculated whenever the query fired.

2.2 Methodology

To achieve the above objective methodology used is given as- we are going to use the rule based system which will follow and execute each and every query as per the rules made for it. First it will identify the nature of the query i.e. select, update, delete, create, insert and also it will identify that the query is with aggregation functions or not. We are using the relational database so it is very much flexible we can easily store all Hindi as well as English values in it and also we can easily retrieve it. Randomize automatic record generation technique is also there so that we can easily generate maximum number of records in very less time. Appropriate mapping of tokens with database values should be done by extracting table, columns information from input Hindi sentences. With the help of stored values of databases generate SQL query by mapping input query. Finally we will execute the Hindi query and also get the output in Hindi language itself.

2.3 Implementation & Architecture of the system

Architecture of Hindi language interface to relational database using NLP is given and explained below from fig1. This architecture is known as HLIDBMS i.e. Hindi Language Interface to Database management System. There are important phases i.e. Tokenizer, query type rule, query table rule, basic queries and its sub rules, query generator engine DBMS & database server. In tokenize phase Hindi sentence is split into tokens. This is done with fact that all the tokens are separated by a space gap from each other. All the tokens which we get in this phase are stored in an array. Tokens are words of Hindi language. Token may be a table name, column name, condition, any value, command name, operation name or any non-useful word. To understand this; let the user query is as:
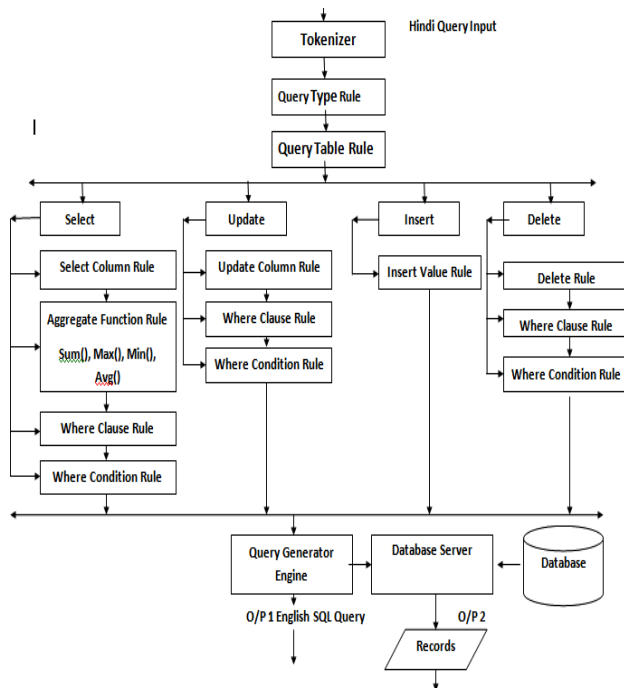
Figure 1: Architecture

सभी विद्यार्थी का नाम,अंक बताओ. This Hindi sentence has7 tokens. First token is सभी which is the starting of sentence.Now सभी means it is reflecting like select all i.e. in SQL we say „Select *‟‟, another token is विद्यार्थी it is reflecting the name of the database table i.e. „student table‟‟ Some tokens

may be fields name as in the above query नाम and अंक are

the field names. There is conjunctions also like का as well aswe also included the comas (,) in the list of tokens & finally last thing is बताओ which is reflecting as the „select query‟‟

Therefore after this step we have all the tokens from which the sentence is composed of.

 After that we will apply the query type rule. Query type ruleis a rule which will identify which type of query it is whether it is select, insert, update, delete type of query. We are given with the query properties through which we can easily identify the associated Hindi word which is given in the sentence within a query and is given below in figure 2.

Later it will identify the table name with the help of query table rules. It will just see whether the given table is present there or not. These both the things have been possible because of the tokenizer and its tokens which we are matching under each rule. Once the query rules and table rules has been applied then we will proceed with the further tokens and we will apply the sub rules of the selected query.

If the query in Hindi will be the select query then it will lookfor the rules like column rules, aggregate function rule, where clause and where condition rule.
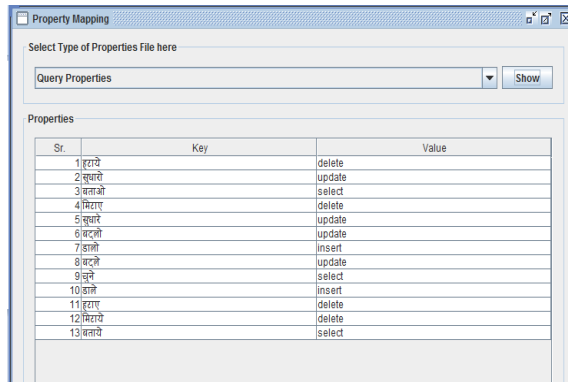
Figure 2: Query Properties

It will work with the help of tokens only. It is like column rules it will select the number of columns given in the Hindi query.

Aggregate function will identify whether it is min (), max (), sum (), avg () query or not . Other rules like where clause forthat we are given with the properties i.e. It will identify all the associated Hindi English words which are given belowin the fig 3.
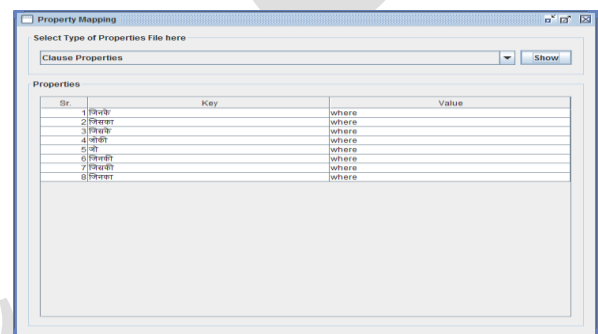

Figure 3: Where clause property

Similarly where condition is also there it will work like sameas given above it is consisting of all the conditional part and its associated Hindi words including <,>,=,logical and ,or not etc. Similarly for update query it is having update column rule ,where clause rule and where condition rule and its working is same as explained above. The same way insertand delete also work. At last there is query generator which will generate query from Hindi sentence .that query generated will be fired to database and all the selected records selected rows has been displayed in Hindi Language.SQL is generated in this phase according to Hindi sentence. Execute query and display result to user the above SQL query is executed and result of which in Hindi language is displayed to user.

Figure 4: GUI & timing results

include whether the query has been successfully executed or not if it is failed it will show the unsuccessful message as shown in the fig4. It will also give the translation time in milliseconds as well as nanoseconds to notice the minute difference during conversion and same in the case of execution time also, it will show the time required to executethe query.

3. Conclusion

Rule based graphical user interface to relational database is presented in this paper. The system will accepts Hindi sentence as a query and gives output in Hindi itself. It is very much useful for the people who do not have any prior knowledge of database and SQL queries languages. We are using different rule along with the NLP to perform operationsuch as insert, update, delete, select as well as the aggregate functions such as min (), max (), sum (), avg () etc. This system can be enhanced by making it more generic. We can also implement it for very complex queries like join operations & order by operations (queries). To make the system more friendly the dialogue based system can be used in which user will provide the input Hindi query through speech interface.

References

[1.] Zongmin Ma's 2007 book "Intelligent Databases: Technologies and Applications" (320 pages) published by IGI.
[2] Contributors include Donald K. Burlesan, Joe Celko, John Paul Cook, and Peter Gulutzan. that year 2003. Manuel for Database Programmers with Advanced SQL Skills. DBAZine and BMC Software.
[3] "Applications of SQL for Informetric Data Processing" by Dietmar Wolfram was published in 2005 in the Proceedings of the 33rd meeting of the Canadian Association for Information Science.
retrieved from "http://searchsqlserver.techtarget.com/sDefinition/0,,sid87_gci1124415,00.html" [4] "IntelligentDatabase" Chapter 5 of Readings in Artificial Intelligence and Databases, edited by M. Brodie and published by Morgan Kaufman in 1988, discusses the integration of AI and database technologies in future intelligent information systems.
[6] "The Intelligent Database Interface: Integrating AI and Database systems" (pp. 677--684, 1990) in Proceedings of the 1990 National Conference on Artificial Intelligence, written by Donald P. McKay and Timothy L. Finin.
The article "Natural language interfaces to databases—an introduction" by Androutsopoulos, G. Ritchie, and P. Thanisch was published in the Journal of Natural Language Engineering in 1995 and can be found on pages 29–81.
[8] Natural Language Computing: The business Applications, by Johnson T. (1985). This is Ovum Limited in London.
In 1966, Weizenbaum published an article titled "ELIZA— A computer program for the study of natural language communication between man and machine" in the Communications of the ACM, volume 9, issue 1, pages 36–45.
[10] Bolt Beranek and Newman Inc., Cambridge, Massachusetts, 1972. The Lunar Sciences Natural Language Information System: Final Report, BBN eport 2378. Contributed by W.A. Woods, R.M. Kaplan, and B.N. Webber.
In the Proceedings of the Sixth International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, held in Edinburgh in 1993, Androutsopoulos, G. Ritchie, and P. Thanisch presented MASQUE/SQL, an efficient and portable natural language query interface for relational databases. The paper can be found on pages 327–330.
[12]The authors of the paper "A Domain Independent Natural Language Interface to Databases Capable of Processing Complex Queries" are Rodolfo Rangel, Joaquín Pérez, Juan Javier González, Gelbukh, Grigori Sidorov, and Rodríguez. The paper was presented at the MICAI 2005 conference on advances in artificial intelligence. Pages 833–842 of LNCS 3789.
[13] "Issues in Translating from Natural Language to SQL in a Domain-Independent Natural Language Interface to Databases" (LNCS 4293, MICAI: Advances in Artificial Intelligence, 2006, pp. 922-931), by B. Juan J. González, Rodolfo A. Pazos Rangel, I. Cristina Cruz C., H. Héctor J. Fraire, and L. de Santos Aguilar, et al.
In their 2012 article "A Servey of Natural Language Interface to Database Management System," published in the International Journal of Science and Advance Technology, B. Sujata, S. Viswanadha Raju, and Humera Shaziya discuss this topic.
[15]"Hindi Language Graphical User interface to Database Management System" (Mohit Dua, Sandeep Kumar, Zorawar Singh Virak) Machine Learning and Its Applications: The 12th International Conference, 2013.
"Hindi Language Interface to Database using Semantic Matching" [16] Ashish Kumar Volume 6, issue 2, pages 133–140, September 2013, issue 0974–6471, Oriental Journal of Computer Science and Technology.
[17]Prof. Amit Pimpalkar, Abhijeet R. Sontakke "An Evaluation of the Hindi-Language GUI for Relational Databases Utilizing Natural Language Processing" Article published in October 2014 in the International Journal of Advanced Research in Computer Engineering and Technology (IJARCET) volume 3, issue 10.