AN APPROACH FOR TRANSFORMING HINDI REGIONAL LANGUAGE INTO DATABASE QUERY

Amit Kumar[1], Dr. Pertik Garg [1]

Abstract- **In our life, information is really vital. Databases are one of the main sources of information. Databases and database technologies are significantly influencing the rise in computer use. Almost all IT applications need databases to store and retrieve data. SQL and other database languages are needed for information retrieval. Nearly all languages used in relational database systems adhere to the Structured Query Language (SQL) standards. Nevertheless, not everyone has the ability to create SQL queries since they might not be familiar with the database's structure. As a result, the Intelligent Database System (IDBS) has been developed. Instead of dealing with the values of the attributes, non-expert users must be able to query relational databases in their native tongue. As a result, a variety of intelligent natural language database interfaces have been created, offering numerous choices for modifying searches. Natural language Interface to Database is a novel processing method that was created in response to the notion of using Natural Language rather than SQL.**

In order to improve users' ability to conduct flexible database searching, NLIDB represents a step towards the development of intelligent database systems (IDBS). The Intelligent Database System and Natural Language Interface to Databases are introduced in this essay. The subject then shifts to NLIDB designs and various methods for developing NLIDB systems after providing a quick overview of NLIDB subcomponents. In this paper, we used the Hindi language interface for patient query database to query the database in Hindi language. The system is proposed in .NET.

Keywords - Natural Language Processing, Hindi, Natural language interface, Database, Hospital.

I. INTRODUCTION

Language is the essential methods for correspondence utilized by people. It is the instrument we use to communicate most of our thoughts and feelings. Learning new ideas and communicating thoughts through them is normal to such an extent that we barely acknowledge how we measure characteristic language. Natural Language Processing (NLP) is worried about the advancement of Computational models of parts of human language preparing. There are two primary reasons of such turn of events:

- To create computerized apparatuses for language handling.
- To acquire a superior comprehension of human correspondence.

Building computational models with human language handling capacities requires Knowledge of how people procure store and cycle language. It likewise requires the information on world and of language. Organizations have related the issue of extricating information from a Data Base Management System (DBMS) by utilizing the apparatuses like MS Access, Oracle and others. An individual with no information on Structured Query Language (SQL) may get oneself debilitated in relating with these instruments. Accordingly it makes the need to create items for individuals to connect with the information base in their own local language. In India, Hindi is utilized as a wellspring of language for correspondence. We have proposed a model for Hindi language interface to empower the individuals of Punjab to utilize the information bases in their own local language.

[1]*Department of Computer Science Engineering, SVIET, Banur, Punjab, India.*

Information base applications assume a significant part in the present business framework [1]. The majority of the organizations need these kinds of uses by u sing the SQL language. Regular language handling (NLP) is getting perhaps the most dynamic strategies utilized in Human-PC Interaction. It is a part of AI which is utilized for Information Retrieval, Machine Translation and Language Analysis. The primary objective of NLP is to empower correspondence among human and PCs without remembrance of complex Commands and methodology. At the end of the day, NLP is the procedures that can make the PC to comprehend the normal dialects utilized by people. The present prerequisite of business framework is to extricating information from a Database Management System, for example, MS Access, Oracle and others. An individual without information on SQL may get himself/herself disabled in relating with the data set. Hence in this work the improvement of framework for individuals to interface with the database in natural English language is executed. This empowers a client to enter their inquiries in natural English and find the solution in same language. This is known as a Natural Language Interface to a Database (NLIDB) [9].

1.1 TYPES OF INTERFACES
A user interface is the perspective on an information base interface that is seen by the client.
• Form based Interfaces
• Text based Interfaces
• GIS Interface

1.1.1 Form based Interfaces:
This interface comprises of structures which are adjusted to the client [2].User can fill in the entirety of the fields and make new sections to the information base. They are anything but difficult to utilize and have the preferred position that the client needn't bother with exceptional information about information base dialects like SQL. Eg: SBB Travel Online site. Be that as it may, a few activities may be confined by the application.

1.1.2 Text based Interfaces:
Utilizing this interface, the clients speak with the DBMS straightforwardly in the inquiry language through an information/yield window. Text-based interfaces are extremely useful assets and permit an extensive association with a DBMS. Nonetheless, the utilization of those depends on dynamic information on the individual information base language. For Eg: SQL.

1.1.3 GIS Interface:
A GIS UI regularly incorporates highlights of an information base interface. The information base cooperation happens through the blend of various interfaces:
• Graphical communication through a choice on the guide
• Combination of structure based and text-based communication (for example exceptional Query-Wizards for the simpler making of information base inquiries)

1.1.4 NLIDBs favorable circumstances over formal query language and form based interfaces
• No compelling reason to learn artificial language
The client isn't needed to gain proficiency with a fake correspondence language. It is hard to learn Formal question dialects and expert in any event by non-PC trained professionals. In NLIDB client utilize their local language to question the information base so there would be no requirement for the clients to invest energy learning.
• No need to know Physical structure of data
To question in proper language one ought to know about area of information where it is store. However, there is no necessity of that information in NLIDB.
• Simple to utilize
To recover information NLIDB framework require a solitary info while a structure based may contain different info relying on the ability of structure. In the event of inquiry language an inquiry may should be expressed utilizing numerous explanations which may contain at least one sub inquiries for certain joint activities.
• Discourse
Another preferred position of NLIDB innovation in concern normal language interface that help anaphoric and circular articulation. NLIDB of this sort permit the utilization of brief underspecified questions where the significance of each question is joined by the talk setting.

1.1.5 NLIDB's disadvantages
• Semantic inclusion isn't self-evident
As of now all NLIDB [3] frameworks can just deal for certain subsets of a characteristic language and it is difficult to characterize these subsets. Indeed, even some NLIDB frameworks can't respond to specific inquiries have a place with their own subsets. This isn't the situation in a conventional language. The conventional language inclusion is evident and any articulations that adhere to the offered rules are ensured to give the relating response.

- Phonetic versus reasonable disappointments

On account of NLIDB framework disappointments, it is regularly the situation that the framework doesn't give any clarification of what makes the framework fizzle. A few clients may attempt to reword the inquiry or simply leave the inquiry unanswered. More often than not, it is dependent upon the clients to decide of the causes the mistakes.

- Bogus assumptions

Individuals can be misdirected by a NLIDB framework's capacity to handle a characteristic language: they may expect that the framework is keen .Therefore instead of posing exact inquiries from an information base, they might be enticed to pose inquiries that include complex thoughts, certain decisions, thinking abilities, and so forth, which a NLIDB framework can't be depended upon.

II.RELATED WORK

D. Ramesh et al. [4] examined that the bountiful data accessible on web creates the need to store information in a coordinated way so that looking, recovering and keeping up of information gets simpler. An information base is an innovation that stores the information in a consistent and coordinated way. To productively work these information bases, information on structures question language (SQL) gets basic. Yet, the use of SQL confines the admittance to information bases from the clients who don't have the information on them. A requirement for interface comes into the image to empower the entrance of these information bases even to non-master clients. This paper depicts the plan to create Telugu language interface to information bases. Bank information base is utilized as a contextual investigation to create Telugu language interface. The exhibition of the framework has demonstrated to be palatable.

Ashish Kumar et al. [5] talked about that in the realm of registering, data assumes a significant job in our lives. One of the significant wellsprings of data is information base. Information base and Database innovation are having significant effect on the developing utilization of PCs. Practically all IT applications are putting away and recovering the data or information from the information base. Information base Management Systems (DBMS) have been generally utilized for putting away and recovering information. Be that as it may, information bases are regularly difficult to use since their interface is very unbending in co-working with clients. For putting away and recovering the data from information base requires the information on data set language like SQL. Organized Query Language (SQL) is an ANSI standard for getting to and controlling the data put away in information base. Nonetheless, everybody will most likely be unable to compose the SQL inquiry as they may not know about the punctuation and structure of SQL and information base individually. In India, the characteristic language of individuals is primarily Hindi. Likewise huge number of e-administration applications use information base. Thus, to utilize such information base applications effortlessly, individuals who are more alright with Hindi language, requires these applications to acknowledges a straightforward sentence in Hindi, and cycle it to create a SQL inquiry. The SQL inquiry is additionally executed on the information base to create the outcomes. Subsequently, any interface in Hindi language will be a resource for these individuals. This paper examines the design of planning the Hindi language inquiry entered by the client into SQL question.

Jaspreet Kaur et al. [6] examined that as there is progressively high headway in innovation, so Question Answering is turning out to be significant territory of examination for the specialists. Various inquiries are given by the client in point of finding exact solutions in Question Answering Systems. Question Answering gives ideal answer for recover substantial and exact responses to client question asked in regular language rather than inquiry. Hindi, Telugu, Bengali and so forth are famous dialects that are spoken in India. At present these dialects are mulled over by the analysts and a ton of work is being done in these and other Indian dialects. In this paper we analyze Question Answering Systems execution for various Indian dialects. We examine the best highlights of Question Answering frameworks worked in various Indian dialects and analyze their exhibitions.

Manu Bansal et al. [7] examined that the term information mining has been the most seasoned at this point one of the fascinating trendy expressions. Numerous associations regularly underutilize their generally existing information bases. There is a need to mine data and intriguing examples from these information bases. The focal point of the flow research is to apply information mining on a library the board framework. Information mining is normally done on an information distribution center or an information store. It causes different cost factors like programming, equipment, support and specialists. The target here is to concentrate how the continuous information put away in data set can be turned useful without setting up a different information distribution center. The primary stress is on understanding the difficult viewpoint, contending goals and imperatives and producing a model for data extraction from the constant library information base utilizing ARM (Association Rule Mining) mining method. As SQL (Structured Query Language) can likewise be utilized for mining information as opposed to utilizing specific information mining calculation, the examination additionally contrasts SQL based mining and ARM. The outcomes shows that affiliation rule mining performs in a way that is better than SQL based mining as sort of example to be extricated can be controlled a lot of viably in ARM when contrasted with SQL due to the boundaries (backing and check) utilized in the

information mining calculation. Calculations are actualized utilizing SQL and MATLAB (Matrix Laboratory) Tool - ARMADA.

Garg et al. [8] discussed that as the requirement of information is the essential part of our life. There are numerous resources for the information, but the main provider is a database. The database helps us to access, store and retrieve information. In technical era, no organization or industry is surviving without the use of databases. Each and every computer based application need to retrieve information from a database that requires knowledge of database language like Structured Query Language (SQL). But it is not possible for everyone to be trained in SQL. To overturn this problem researcher giver a solution of Natural Language Interface Database (NLIDB). Natural language, i.e. Hindi, English, Punjabi, Urdu, Bengali etc. In place of SQL can be a perfect interface between a layman user and an application of computer. In this paper, we will discuss the various parameters of NLIDB and also explain the structure of hospital enquiry system. In our research, we will develop an application of hospital enquiry which will accept query in Punjabi language, convert this into formal query language i.e. SQL and produce the result again in Punjabi language. We will also analysis the results of the developed application.

III. METHODOLOGY

Hindi Language Interface to database system gives an interface to the client which causes him/her to question the data set in his/her characteristic language. In this work just Hindi language is utilized as a mean for giving information sources. In this framework we consider an information base SQL Server and client made tables are utilized. A framework is built up that takes out the issue of ordinary client to associate with information base with inflexible language SQL. The clients can get to data's by giving inquiry in basic Hindi language.

The principle issue is to plan a Natural Language Interface of Database for Hindi language utilizing tolerant inquiry database. The question is asked in the Hindi language for recovering the pertinent data from the information base. The organization of the inquiries asked by the client should be basic, not intricate. Consequently, there is need of building up a Hindi language interface for understanding question information base to inquiry the information base in Hindi language. The framework is proposed in .NET. The procedure embraced for setting up this product as appeared in figure 3.1 can be characterized through following steps:

3.1 MORPHOLOGICAL ANALYSIS

The dictionary of a language is its jargon that incorporates its words and articulations. Morphology portrays dissecting, recognizing and depiction of structure of words.

Singular words are examined into their parts and non-word tokens, for example, accentuation are isolated from the words.

3.1.1 Syntactic Analysis

Direct groupings of words are changed into structures that show how the words identify with one another. The goal of syntactic investigation is to locate the syntactic structure of the sentence. This progression separates the sentence into less complex components that are called tokens. Token Analyzing capacity is utilized to part the info string into a succession of crude units considered tokens that is treated as a solitary consistent unit. At that point Spelling Checker work ensures that every token is in the frameworks word reference (dictionary) and in the event that this isn't the situation, at that point the spelling revision is performed or new words are added to the systems‟ jargon. Equivocalness decrease work lessens the uncertainty in a sentence and rearranges the undertaking of the parser.

3.1.2 Semantic Analysis

In this structures made by the syntactic analyzer are doled out implications. Semantics is related with the significance of language. Semantic investigation is worried about making portrayals for significance of etymological sources of info. In this emphasis is on legitimate words, and no consideration is paid to non-sensible words. It manages how the significance of sentence is resolved from the implications of its parts. Furthermore, accordingly, it creates a coherent question which is taken care of as contribution to the information base inquiry generator.

3.1.3 Discourse integration

The importance of an individual sentence may rely upon the sentences that go before it and may impact the implications of the sentences that follow it. The substances that structure those sentences might be identified with elements that additionally were past or might be presented unequivocally. The general talk should be cognizant.

3.1.4 Pragmatic Analysis

The structure speaking to information disclosed is rethought to figure out what was really implied. To decipher the importance of a sentence it is vital for the program to know the setting of objectives and plans in which the sentences are given. This is that are utilized to speak to the projects translation with accuracy. The approach will be utilized for taking care of the above issue.

Initially, an information inquiry is passed in regular language like Hindi.

At that point the Morphological examination are recognized every world. Singular words are dissected into their segments, and it isolated thing and descriptor in the sentences.

A restricted information word reference is additionally used to store all connected words about the framework.

After this, Syntactic principles checks the linguistic errors of a sentence and Semantic examination should plan singular words into fitting articles in the information base or data set and the implications of the individual words join with one another and discover the importance of basic English question.

At that point the interpreter will change the above sentence with SQL question and with the assistance of SQL inquiry, we will ready to discover the outcomes.
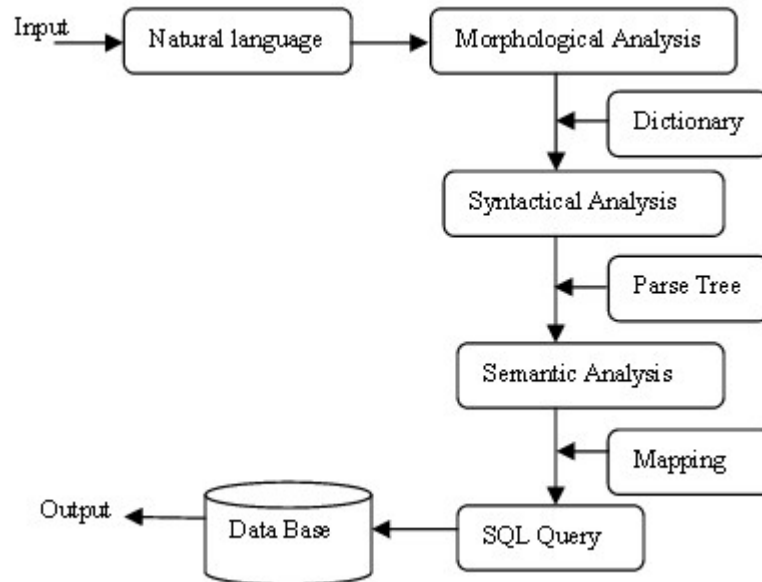

Figure 3.1 Structure of the System

3.2 PROPOSED METHODOLOGY

Algorithm for the proposed methodology is given beneath:

a.   Input the Hindi enquiry which is entered by client.

b.   Divide this enquiry into tokens and store it in the exhibit.

c.   Compare the token with the database assortment.

d.   Check whether the word is found in the database. On the off chance that it isn't discovered, at that point leap to the progression number "d" in any case play out the accompanying advances:

1.   Analyze the situation of that word from the database.

2.   Then make an interpretation of that word into English/Keyword.

3.   Adjust the word as per its situation in the pieces of SQL inquiry.

4.   Now check whether the following word is accessible in the exhibit. On the off chance that it is accessible, at that point brings it from the cluster and goes to the progression number "3" in any case join all the pieces of question and execute it in the SQL and speaks to the outcome to the client.

IV.CONCLUSION

In Hindi Language Interface to Database system the input is taken from the user in Hindi language. After obtaining the input from the user, the system will split the sentence into tokens. Then mapping of these tokens will be done and it is done to find out name of column and table, conditions, commands and values. After this, the tokens are converted into English words. Useless tokens are discarded. Then SQL query is generated from the useful tokens. The system will support operations like- selection, updating and deletion on database.

REFERENCES

[1] 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).

[2] Preeti Verma, Suket Arora, Kamaljit Batra, "Punjabi Language Interface to Database: A Brief Review".

[3] Neelu Nihalani, Dr. Sanjay Silakari and Dr. Mahesh Motwan(2011), "Natural language Interface for Database: A Brief review", International Journal of Computer Science Issues, Vol. 8, Issue 2.

[4] D.Ramesh, Suresh Kumar Sanampudi, "Telugu Language Interface to Databases", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 7, July 2013.

[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, pp.133-140.

[6] 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, Vol. 3, Issue 7, July 2013.

[7] Manu Bansal and Mandeep Kaur, "Analysis and Comparison of Data Mining Tools Using Case Study of Library Management System", International Journal of Information and Electronics Engineering, Vol. 3, No. 5, September 2013.

[8] Garg, P., Singh, R. & Goel, M. B. (2020). A Novel Approach to Implement Natural Language Interface for Database in Punjabi Language. Journal of Xidian University, Vol. 14, No. 4, pp. 1230-1238.

[9] Garg, P., Marwaha, A., Goel, M. B.(2020). Identification and Classification of Reduplication Words in Punjabi Language. International Journal of Scientific & Technology Research Vol. 9, No. 6, pp. 532-537.