

A Constructive Review for Natural Query Language

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- Data has been stored in the database and the databases are the major sources of information. Information is playing an important role in daily life. This database technology has the major effect on the growing use of computer and internet. Database management system has been used for accessing, storing and retrieving data. However, database system is not understandable to each and every user because they are hard to use and understand. People with no knowledge of database language may find it difficult to access database. Therefore, there is need to find out the new technique and methods to access the database with the use of Natural Language Processing. Therefore this idea of using natural language instead of SQL triggered the development of a different type of processing method called Natural Language Interface to Database (NLIDB). Natural language has been the holy grail of query interface designers, but has generally been considered too hard to work with, except in limited specific circumstances. Where user does not have any need to learn the formal language, they can give query in their native language. For the people who are comfortable with the Punjabi and English language need this application to accept Punjabi and English sentence as a query, process it and after execution provide result to the user in the same language which is nothing but the Punjabi and English Language Interface to Database Management System.

Keywords – Natural Language, Database, SQL, Punjabi, English.

I. INTRODUCTION

As the requirement of information is essential part of our life [1]. In the today's computing world computer based technologies are extensively used in managing information in almost all the sort off organizations e.g educational, private, government etc. Information [15] management systems are used to manage data. Information can be accessed from anywhere and by anyone. Largest source of information in computer world is database. Database: A database is collection of related the data, organized in a systematic way, in such a way that computer can quickly select a particular data from huge data. Databases 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.

Language Processing [14] (NLP) is the development of a natural language interface to database systems (NLIDB). In the last few decades many NLIDB systems have been developed. Through these systems, users can interact with database in a more convenient and flexible way. Because of this, this application of NLP is still very and widely used today. Natural Language Interface has been a very interesting area of research since past times. The aim of Natural language Interface to Database is to provide an interface where user can interact with database more easily using their natural language and access or retrieve their information using the same. We can also say that NLIDB is a system that converts the query in native language into SQL and vice-versa.

Large number of e-governance applications use database. For the easy and frequent use such database applications, people who are more comfortable with Punjabi language and English, require these applications to accept a simple sentence in Punjabi or English, and process it to generate a SQL query, which is further executed on

the database to produce the results. Therefore, any interface in Punjabi or English language will be an asset to these people. With the help of Punjabi or English interface, they will be able to make use of database applications very well.

II. LITERATURE REVIEW

Saravjeet Kaur et al. [2] proposed a method of querying with the databases by means of a natural language interface. For non technical users, this issue is to give a high level interface in the area of database management. Normal users are not aware with the formal language like SQL. Then the problem is how they interact with the database system. To deal with the database system, a normal user may find him/her self as handicapped. This paper provides an interface module that translates user's given query from natural language to a corresponding SQL command. For normal users who are unable to understand the database query language such as SQL, so asking questions to databases in natural language like English is a very simple and easy method of accessing data from database system. This paper proposed the translation for English Query into SQL.

Amol Jadhav et al. [16] discussed that 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. The idea of using Natural Language instead of SQL has prompted the development of new type of processing called Natural language Interface to Database. NLIDB is a step towards the development of intelligent database systems (IDBS) to enhance the users in performing flexible querying in databases. This system prepares an "expert system" implemented in prolog which it can identify synonymous words in any language. It first parses the input sentences, and then the natural language expressions are transformed to SQL language. This paper is an introduction to Intelligent Database System and Natural Language Interface to Databases.

Wiqas Ghai et al.[4] discussed that Punjabi, Marathi, Gujarati, Hindi, Bengali, Nepali, Sinhala, Oriya, Sindhi, Assamese, Urdu are important members of the group of Indo-Aryan languages. These languages are mainly spoken in India, Bangladesh, Pakistan, Maldives Islands, Nepal and Sri Lanka. All these languages contain huge diversity of phonetic content. In the last two periods, some researchers have worked for the evolution of Automatic Speech Recognition Systems for most of these all languages in such a way that progress of this technology can attain at point with the research work where it has been done and it is being done for the multiple languages in the world. Punjabi language is the 10th most widely spoken, for which no possible work has been done in the region of automatic speech recognition. Being a member of Indo-Aryan languages family and a language rich in literature, Punjabi language deserves attention in this highly growing field of Automatic speech recognition. In this paper, the efforts made by various researchers to develop automatic speech recognition systems for most of the Indo-Aryan languages, have been investigated and then their appropriate to Punjabi that has been discussed so a real work can be commenced for Punjabi language.

Simarjeet Kaur et al.[5] discussed that the accurate and timely advice for the productive insect pest management is an prominent constituent component in controlling the pest on vegetables crops. This web based insect pest management system aims to transfer the pest management practices in different vegetable crops recommended by Punjab Agricultural University to the farmers for their advice to take fast and appropriate work for pest management. This system has been evaluate by using PHP, JavaScript, CSS, HTML and Ajax and database has designed using My SQL. The proposed system is advantageous as it is simple to use, effectual and well organized to control the insect pests by giving appropriate and accurate information at affordable cost.

D.Ramesh et al. [6] discussed that the lavish information accessible on internet produce the need to store data in well organized manner so that searching, redeem and maintaining of data becomes simple. A database is a technology that stores the data in an logical and organized manner. To systematic operate these databases, knowledge of SQL becomes necessary. But the usage of SQL limits the access to databases from users who don't have the idea about it. Even for non-expert users this interface comes into the picture to allow the access of these databases. This paper narrate the design to improve Telugu language interface to databases. Bank database is used as a case study to develop telugu language interface. The performance of the system has shown to be satisfactory

Jaspreet Kaur et al. [7] discussed that as there is increasingly high advancement in technology, so Question Answering is becoming major area of research for the researchers. Different queries are provided by the user in aim of getting accurate answers in Question Answering Systems. Question Answering provides perfect solution to retrieve valid and accurate answers to user question asked in natural language instead of query. Hindi, Telugu, Bengali etc are popular languages that are spoken in India. Currently these languages are taken into consideration by the researchers and a lot of work is being done in these and other Indian languages. In this paper we compare Question Answering Systems performance for different Indian languages. We discuss the best features of Question Answering systems built in different Indian languages and compare their performances.

Manu Bansal et al. [8] discussed that the term data mining has been the oldest yet one of the interesting buzzwords. Many organizations often underutilize their already existing databases. There is a need to mine information and interesting patterns from these databases. The focus of the current research is to apply data mining on a library management system. Data mining is usually done on a data warehouse or a data mart. It incurs various cost factors like software, hardware, maintenance and experts. The objective here is to study how the real-time data stored in database can be turned informative without setting up a separate data warehouse. The main emphasize is on understanding the problem perspective, competing objectives and constraints and generating a model for information extraction from the real-time library database using ARM (Association Rule Mining) mining technique. As SQL (Structured Query Language) can also be used for mining data instead of using specialized data mining algorithm, the study also compares SQL based mining with ARM. The results shows that association rule mining performs better than SQL based mining as type of pattern to be extracted can be controlled much effectively in ARM as compared to SQL because of the parameters (support and count) used in the data mining algorithm. Algorithms are implemented using SQL and MATLAB (Matrix Laboratory) Tool - ARMADA.

Ravinder Kumar et al. [9] discussed that we require information in our daily life. One of the major sources of information is database. Almost all applications need to retrieve information from database that requires knowledge of database languages like SQL. To write SQL query one need to have knowledge of formal query language. Therefore everybody is not able to write SQL queries. To override the complexity many research have turned out to use Natural Language (NL) i.e. English, French, Tamil, Arabic, Hindi, Punjabi etc. instead of SQL. The idea of using NL has prompted the development of new type of processing method called Natural Language Interface to Database systems (NLIDB). Lots of works have been already done in natural language processing but in Hindi and other Indian languages have scope of research and improvement. So for this intention we are trying to design a tool and database that helps the farmers to solve out the different queries requested in their native language (Hindi) related to agriculture. To design the tool that fulfills the objectives of our work we will use Java swings as front end and for storing the data we will use MySQL 5.0 as backend.

Priyanka Arora et al. [1] discussed that database Management Systems have been used extensively for accessing, storing and retrieving data. However, database systems are not understandable to every user because they are hard to use. A plethora of e-governance applications railways, billings, agriculture, banks etc. use databases. Some users face difficulty in using these database systems because they do not have knowledge of the languages used in this system. So, they want a system that accepts a Hindi sentence as a query and after processing it, execute it and provides the output in the same language. Then the users have no need to learn any low-level languages those are hard to learn, use in databases such as SQL.

Rohini B. Kokare et al. [10] discussed that natural language query builder interface retrieves the required data from database when query is given in natural language. To retrieve the correct data from database, the user should have sufficient technical knowledge of Structured Query Language (SQL) statements. Natural Language Query Builder Interface (NLQBI) will solve this problem. In natural language parsing, getting highly accurate syntactic analysis is a crucial step. Parsing of natural languages can be seen as the process of mapping an input string or a sentence to its syntactic representation. One of the parsing techniques is dependency parsing. Dependency parsing focuses on relations between words which resolve ambiguity. Most of the recent efficient algorithms for dependency parsing work by factoring the dependency trees. Graph based dependency parsing models are prevalent in dependency parsing because of their state-of-art accuracy and efficiency. This paper covers some recent developments in NLQBI systems and survey on dependency parsing techniques.

Preeti Verma et al.[3] discussed that unlike most user-computer interfaces, a natural language interface allows users to communicate fluently with a computer system with very little preparation. Databases are often hard to use in cooperating with the users because of their rigid interface. A good NLIDB allows a user to enter commands and ask questions in native language and then after interpreting respond to the user in native language. For a large number of applications requiring interaction between humans and the computer systems, it would be convenient to provide the end-user friendly interface. Punjabi language interface to database would prove fruitful to native people of Punjab, as it provides ease to them to use various e-governance applications like Punjab Sewa, Suwidha, Online Public Utility Forms, Online Grievance Cell, Land Records Management System, legacy matters, e-District, agriculture, etc. Punjabi is the mother tongue of more than 110 million people all around the world. According to available information, Punjabi ranks 10th from top out of a total of 6,900 languages recognized internationally by the United Nations. This paper covers a brief overview of the Natural language interface to database, its different components, its advantages, disadvantages, approaches and techniques used. The paper ends with the work done on Punjabi language interface to database and future enhancements that can be done.

Jyothsna Cherapanamjeri et al. [11] discussed that natural language interface to database system is one of the areas of Natural language processing. A recent development in Natural Language Processing is Question Answering

System. Question Answering accepts queries formulated in natural language and respond to the user with a specific answer in Natural Language Sentence. Keyword based question answering system in NLIDB for crop statistic information has been described. Different components had been developed in this system for processing the query and translating it into formal SQL query. Based on the results from the SQL query, a natural language answer is generated.

Jaina Patel et al.[12] discussed that information plays an important role in our everyday life and databases are widely used for storing and retrieving information. Database technology is having major impact in the world of computing. To access the information from database one need to have knowledge of database query language such as SQL. Because the naïve user may not be aware of the syntax of SQL and structure of database, s/he may not be able to write the SQL queries. Non -technical users may query relational databases in their natural language (i.e. English) instead of using SQL. This idea of using a Natural Language instead of SQL has lead to an approach of building Natural Language Interface to Relational Database. This paper is an introduction to the natural language interface to databases (NLIDB).

M.C. Nikose et al.[13] discussed that Query optimization is the refining process in database administration and it helps to bring down speed of execution. Some object-oriented languages allow to express queries explicitly in the code, which are optimized using the query optimization techniques from the database domain. With respect to this, a formalized object query language (OQL) has been developed that performs optimization of queries at compile time. Object Oriented Data Base (OODB) has all the features, functionality of a relational database system and also offers an Object Oriented Programming language interface. We follow the stack based approach to query languages, which is responsible for naming-scoping-binding principle. In this paper we proposed one of the methods of query optimization depending on rewriting. Optimization by rewriting concerns queries containing so called independent subqueries. It consists in detecting them and then factoring outside the loops implied by query operators.

IV.CONCLUSION

Natural Language Interface is the prominent research area in the field of artificial intelligence. From the last few decades many systems were developed on Natural Language Interface to Databases System which provide considerable improvement over another in terms of usability and accuracy. Now-a-days such interfaces are increasingly important on websites, particularly as access information from cell phones and PDA's and other small screen devices where GUI is less appealing. Different architectures are followed by different systems which distinguish their approach from one another. Also, with the revolutionary advancement in hardware technology and their processing speed, NLIDB system becomes more efficient, accurate and user-friendly.

Research is done from the last few decades on Natural Language Interfaces. With the advancement in hardware processing power, many NLIDBs mentioned in historical background got promising results. Though several NLIDB systems have also been developed so far for commercial use but the use of NLIDB systems is not wide-spread and it is not a standard option for interfacing to a database. This lack of acceptance is mainly due to the large number of deficiencies in the NLIDB system in order to understand a natural language.

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] 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).
- [3] Preeti Verma, Suket Arora, Kamaljit Batra, "Punjabi Language Interface to Database: A Brief Review".
- [4] Wiqas Ghai, Navdeep Singh, "Analysis of Automatic Speech Recognition Systems for Indo-Aryan Languages: Punjabi A Case Study", International Journal of Soft Computing and Engineering (IJSCE),ISSN: 2231-2307, Volume-2, Issue-1, March 2012.
- [5] Simarjeet Kaur, B K Sawhney, Sandeep Kaur, "Design and Development of Software for Insect Pest Management of Vegetable Crops using Web Technology", International Journal of Emerging Science and Engineering (IJESE) ISSN: 2319–6378, Volume-1, Issue-11, September 2013.
- [6] 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.
- [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, Volume 3, Issue 7, July 2013.
- [8] 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.
- [9] Ravinder Kumar, "A Convenient Hindi Language Interface to Database for Agriculture Based Queries", International Indexed & Referred Research Journal, ISSN-2250-2556, Nov-2014.
- [10] Rohini B. Kokare & Kirti H. Wanjale, "A Survey of Natural Language Query Builder Interface for Structured Databases using Dependency Parsing", International Journal of Computer Applications (0975 – 8887),Volume 107 – No 5, December 2014.
- [11] Jyothsna Cherapanamjeri, Lavanya Lingareddy and Himabindu. K, "Keyword based Question Answering System in Natural Language Interface to Database",International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)", Volume 3 Issue 12, December 2014.

- [12] Jaina Patel and Jay Dave(2015), "A Survey: Natural Language Interface to Databases", National Conference on Recent Research in Engineering and Technology (NCRRET -2015, International Journal of Advance Engineering and Research Development (IJAERD), e-ISSN: 2348 - 4470 , print-ISSN:2348-640.
- [13] M.C. Nikose, S.S. Dhande and G. R. Bamnote(2012), "Query Optimization in Object Oriented Databases through Detecting Independent Subqueries", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 2.
- [14] Ashish Kumar and Kunwar Singh Vaisla(2013), "Natural Language Interface to Databases: Development Techniques", Elixir Comp. Sci. & Engg , 58(4).
- [15] 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.
- [16] Amol Jadhav, Sanket Bhuskat, Sushant Raut and Umesh Sutar(2015), "Interface to Database By Using English Language", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 3, Issue V.