International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Natural Language Interface to Databases: A Survey

Manika Tyagi

PG Scholar, CSE, Galgotias University, Greater Noida, UP, India

Abstract: ***We have to deal with information in our day-to-day life. Databases are the main source of information. In order to retrieve information from database, one has to understand the structure of database languages like SQL. However, general users are not able to write SQL queries since they may not have knowledge of databases. So, this led to developing such systems where common users write their questions in natural language and obtain results in form of table. This paper introduces natural language processing to databases.***

Keywords: SQL, Information, Natural language processing, table, database.

1. Introduction

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. Therefore everybody is not able to write SQL queries. Many researchers have turned out to use Natural Language (NL) i.e. English, French, Tamil, Arabic 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). It is a type of computer human interface. It is a system where user access information stored in databases by typing request in natural language. It makes accessing of data from database very easy for a person who has no knowledge of query language.

1.1 NLIDB

Persons with no knowledge of database language may find it difficult to access database. In recent time there is a rising demand for non-expert users to query relational database in a more natural language encompassing linguistic variables and terms. Therefore the idea of using natural language instead of SQL triggered the development of new type of processing method Natural Language Interface to Database.

1.2 Motivation

Research has been very active in developing interfaces for accessing structured knowledge, from faceted search, where knowledge is grouped and represented through taxonomies, These interfaces still require that the user is familiar with the queried knowledge structure. However, casual users need to be able to access the data despite their queries not matching exactly the queried data structures. According to the interface evaluation systems developed to support Natural Language Interfaces are perceived as the most acceptable by end-users. This conclusion is drawn from usability study, which compared four types of query language interfaces to knowledge bases and involved 52 users of general background. Also, much data on the Web is accessible through the use of applications based on relational databases.

1.3 Advantages of NLIDB

Like other systems, NLIDB also have some merits as well as demerits. This section discusses NLIDB's advantages over formal query language and form based interfaces [1]. Advantages are following;

- No need to learn artificial language
- No need To know Physical structure of data
- Easy to use

1.4 Disadvantages of NLIDB

Disadvantages of NLIDB have been discussed below;

- Deals with limited set of natural language
- Difficult to decide failure of query
- Ambiguity
- Wrong assumption by users

1.5 Applications of Hindi Language Interface to Database Management System

Hindi is spoken in mostly northern and central India, Pakistan, Fiji, Mauritius and Suriname. Approximately Seven hundred Million people speak Hindi as either the first or second language. Large numbers of applications uses database. For these people a system should be developed where they can access the database with Hindi language. We identified some areas where interface to database using Hindi language can be applied.

- **Agriculture**
  Government has developed many systems to help farmers solving their queries. Data related to their queries is stored in databases as we said above farmers are not much literate so we can't expect a SQL query from them so there should be a system which is friendly to farmers. Using Hindi as a query language will be a good idea for this. Therefore interface to database with Hindi language can be applied to achieve this goal.
- **Students**
  STUDENT is the cheapest and largely used by public for transport. STUDENT has a database regarding the arrival and departure time, frequencies of trains, journey fares, reservation and cancelation of reservation etc. in India most of the people speak Hindi so a system is required for STUDENT database where passenger give query in Hindi

language and result is also shown to them in Hindi. This will help them a lot. This is where Hindi language interface to database management system is used.

- **Weather Forecasting**
  To know the weather conditions may be required for some persons or organizations. Farmers are one of them because for all the activities from reaping to harvesting of crops depends heavily on weather. So to provide queries for weather in Hindi is very important. This is done with interface to database with Hindi language. Presently there is a vast amount of NLP-based research carried out for the development of such systems. One modern natural language system is Jupiter.
- **Sports**
  Sports are played in nook and corner of India. Specially, if talk about cricket. In cricket huge data is stored in form of tables. One who want to know anything about the game, any match result or scorecard they just need to type what they want in Hindi. The interface to database will provide them their result. Therefore here also interface to database is helpful.

2. Literature Review

Since the end of 1960 there have been a large number of research works introducing the theories and implementations of NLIDBs. Asking question to databases in natural language is very convenient and easy method of data access especially for casually users who do not understand complex database query language.

2.1 Existing NLIDB systems

Prototype for NLIDB had appeared in late sixties and early seventies. Since then a number of systems have developed. Here we discuss some of them.

2.1.1 LUNAR

The system was introduced in 1971.It answer the questions about samples of rocks brought back from the moon [4]. The meaning of the system name is in relation to the moon. LUNAR system has two databases, one for chemical analysis and other for literature references. LUNAR system uses an augmented transition network (ATN) parser and procedural semantics [2]. The performance of LUNAR was very impressive; it managed to handle 90% of requests without any error [5].

2.1.2 LADDER

The LADDER (Language Access to Distributed Data with Error Recovery)[6] was designed as a natural language interface to database of information about US Navy ships. It takes queries in English language. The system was developed as a management aid to navy decision makers. LADDERS applies all the necessary information concerning the vocabulary and syntax of question, The name of specific fields, how they are formulated and even where the fields are physically located to provide the answer. LADDER's first component is INALAND (Informal Natural Language Access to Navy Data). The third component of the LADDER system is for FAM (File Access Manager) [7].The work of FAM is to find the location of the generic files and manage the access to them in distributed database. The system LADDER was implemented in LISP. At the time of creation of LADDER, system was able to process a database that is equivalent to relational database with 14 tables and 100 attributes.

2.1.3 CHAT-80

The system CHAT-80[7] came in eighties. It was implemented in prolog. According to [8] the database of CHAT-80 consist of facts about 150 countries of the world and a small set of English vocabulary that are enough for querying the database. The system translates the English language question by the creation of logical form as process of 3 serial and complementary functions. First word is represented by logical constraints. In second function words, nouns and adjectives with their associated preposition are represented by predicates. Third function is representation of phrase by conjunction of predicates. The functions are following;

- Parsing
- Interpretation
- Scoping

The parsing module function determines the grammatical structure of a sentence and interpretation and scoping consist of various translation rules, expresses directly as prolog clauses. The basic strategy followed by CHAT-80 is to append some extra control information to the logical form of a query in order to make it an efficient piece of prolog program, which can be executed directly to produce the answer. Similarly, many other system were developed which can be summarized as NLIDB systems.

3. Component of NLIDB

Computing scientists have divided the problem of natural language access to a database into two sub-components [15].

3.1 Linguistic component: It handle the natural language input, translate it into formal query and generate a natural language output from the result which come after execution of query.

3.2 Database Component: It performs traditional database management functions. A lexicon consists of a number of tables that store natural language words and their corresponding mapping to formal objects that will be used to create a formal query. These tables can have entries of relations name; attribute names, verbs, adverbs etc. questions entered in natural language translated into a statement with the help of parser which tokenize the input. Then by mapping these tokens into lexicon tables a formal query is formed. Which is executed and the result in natural language is given to user.

4. Architecture of NLIDB Systems

Researchers have adopted a no of architectures in different existed systems. These are given below [12].

- Pattern Matching systems
- Syntax-Based Systems

- Semantic grammar system
- Intermediate Representation Languages.

5. Conclusion

NLP is relatively recent area of research and application. We have described a theoretical as well as practical approach to the problem producing a reliable NLI to database. Now-a-days such interfaces are increasingly important. In our novel framework we store an intermediate processed data introducing new knowledge can be issued using simple SQL insert statements on the top of the processed data. Our framework is most suitable for performing extraction written in natural sentences. This approach saves much more time.

References

[1] Ana-Maria Popescu, Oren Etzioni, Henry Kautz, "Towards a Theory of Natural Language Interfaces to Database", University of Washington, Computer Science Seattle, WA 98195, USA
[2] B.W. Ballard, J.C. Lusth and N.L. Tinkham, "LDC-1: A Transportable, Knowledge based Natural Language Processor for Office Environments", ACM Transactions on Office Information Systems, 2(1): (January 1984), pp. 1–25.
[3] M. Dua, S. Kumar, "Hindi Language Graphical User Interface to Database Management System", 12th International conference on Machine learning and Applications, Miami, USA 2013.
[4] A. Shingala, P. Virparia, "Enhancing the Relevance of Information Retrieval by Querying the Database in Natural form", International Conference on Intelligent Systems and Signal Processing (ISSP), 2013
[5] N. Nihalani, S. Silakari, M. Motwani"Natural Language Interface for Database: A Brief Review", International Journal of Computer Science Issues, vol. 8, Issue 2, March 2011.
[6] A. Shingala, P. Virparia ," Enriching Document Features for Effective Information Retrieval using Natural Language Query Interface", International Journal of IT, Engineering and Applied Science Research, ISSN: 2319-4413, 2012.
[7] M. E. Saleh, "Semantic Based Query in Relational Database Using Ontology", Canadian Journal on Data, Information and Knowledge Engineering, vol.2, 2011.
[8] T. Amble, "BusTUC - A Natural Language Bus Route Oracle." 6th Applied Natural Language Processing Conference, Seattle, Washington, USA, 2000.
[9] B.J. Grosz, "TEAM: A Transportable Natural-Language Interface System", In Proceedings of the 1st Conference on Applied Natural Language Processing, California, (1983), pp. 39–45.
[10] A. Shingala, R. Chavda, P. Virparia, "Natural Language Interface for Student Information System", Journal of Pure and Applied Sciences, Vol. 19:41-44, ISSN: 0975 – 2595, 2011
[11] R.J.H., Scha., "Philips Question Answering System PHILIQA1", In SIGART Newsletter, no.61. ACM, New York,(February 1977).
[12] P. Resnik, "Access to Multiple Underlying Systems in JANUS", BBN report 7142, Bolt Beranek and Newman Inc., Cambridge, Massachusetts, (September 1989).
[13] A. Popescu, A. Armanasu, O. Etzioni, David Ko, and Alexander Yates, "Modern Natural Language Interfaces to Databases: Composing Statistical Parsing with Semantic Tractability", COLING (2004).
[14] B. Sujatha, S. Viswanadha Raju and Humera Shaziya "A Survey of Natural Language Interface to Database Management System" International Journal of Science and Advance Technology", vol.2, no.6, June 2012.
[15] Abrahams P. W. et al. "The LISP 2 Programming Language and System", in proceeding of FJCC, No. 29, USA, 1996, pp. 661- 676.
[16] H. Jain, P. Bhatia, "Hindi language interface to database," 2011.

Author Profile

Manika Tyagi received Bachelor Degree in Computer Science and Engineering from Greater Noida Institute of Technology, Greater Noida. She is currently pursuing M. Tech in CSE from Galgotias University, Greater Noida, UP, India. Her area of interest is "Natural Language Interface to Databases".