CORE
Metadata, citation and similar papers at core.ac.uk
Provided by MAT Journals

MAT JOURNALS
Journal of Data Mining and Management
eISSN: 2456-9437
Volume 3 Issue 3

Mining Based Natural Language to Database Interface

[1]Miss. Pooja M Rajannavar, [2]Prof. Gambhir Halase
[1]M.Tech Student, [2]Professor
Department of CSE, KLE Dr M S Sheshgiri College of Engineering & Technology,Udyambag, Belagavi, India
Email: pooja5.rjn@gmail.com, g_halse@rediffmail.com
DOI: http://doi.org/10.5281/zenodo.1745408

Abstract

Data/Information plays an important role in our daily life. This data can be generated from many sources like Hospitals, Organizations, and Educational Institutions etc. These data need to be managed and stored in a database. The database is the main source of information. To access, store and manipulate the data stored in a database is a critical task. This requires the knowledge of high level database languages like SQL, where the user writes the high level query to retrieve data from database. But this creates a complex problem for normal users, who are not aware of database languages. To minimize this complexity, NLQP (Natural Language Query System) is designed. This system provides an interface for the end users to write the query in natural language such as English and obtains the result back in Natural Language. The query written in natural language will be converted to SQL like queries by the NLQP system and the required results will be fetched from the database for the user. The main goal of NLQP system is to provide user friendly communication between the end user and the computer from where the data is to be fetched.

Keywords: *Natural language processing, Parser, Database interface, Semantic analysis.*

INTRODUCTION

In the Current days massive amount of data/facts is spawned and conserved by many birthplaces, for example: Organizations, Hospitals, Company, Universities, etc. We, at all the times have to deal with the data/facts which plays a vital starring role in our regular life.

Database is the principle wellspring of data. The exercises, for example, information recovery, information stockpiling and information preparing are the principle goal of database. To get to/recover the information from the database requires the learning of database dialect, for example, Structured Query Language (SQL). Just the people who have the ability to manage these inquiry dialects can utilize the information put away in the database specifically. Yet, be that as it may, everybody can't have the capacity to compose SQL inquiries to recover the data from database which causes the multifaceted nature of utilizing the information from database. Likewise, a few programming's does not be able to control complex SQL questions.

To streamline this complication and to handle the data kept in the catalog for public people, the novel scheme has been developed where Non-Professional/Non-expert consumer can request the questions/queries in ordinary language and achieve the preferred result from the database. This novel scheme is called as Natural Language Query Processing (NLQP) system.

In Natural Language Query Processing (NLQP) system, the regular user can create question/query in his/her natural language as a replacement of query language (SQL) to bring the records from the database. The NLQP system is developed and aimed in such a way that it agree to take the queries in normal

Page 1-9 © MAT Journals 2018. All Rights Reserved

language like English and interpret the English language into SQL queries and produce the looked-for result from the database in normal/natural language that can be comprehensible by human creatures. Here the query request is implemented in the DBMS and the response will be provided back in natural language. This makes an easy method and shrinks the complexity of normal users to deal with the database dealing out system.

Aim and Objective

The leading objective of NLQP system is to offer a manageable atmosphere and to build an easy dealing with the database using regular language like English, in the sense that NLQP system won't have need of any assistances of high level language to attain information from catalog. The indication here is to use natural language as a substitute of query language like SQL, which is difficult for public users to contract with facts warehoused in database.

The goal line of NLQP scheme is to offer communiqué between the usual users and the computers (Electronic Device) deprived of calling for any complex measures and approaches to obtain from the database- the data. It equates that NLQP structure works as a boundary in the middle of computer and user. It makes Natural language like English to be understandable by the computer.

Motivation

The present system of processing queries requires the high level programming languages like SQL. SQL is also knows as Structured Query Language which is designed for defining, modifying and manipulating large amount of data in relational data base management system (RDBMS) generated from many organizations. More complex queries can be handled by SQL like languages. SELECT statement is one of the most commonly used operations in SQL, Which allows users to retrieve selected data from the database. Some of the administrative and management functions in SQL are - insert, select, update and delete. These functions allow the user to deal with the data processing.

It will be difficult task for the common user to deal with complex SQL queries, as it requires to remember different commands to perform different operations to retrieve the information from the database. The knowledge of writing the queries in high level syntax such as SQL is essential. To overcome this problem Natural Language to Query Processing System (NLQP) has been introduced. An easy and simple way of using natural language to access data from database has prompted the development of this new system known as NLQP. NLQP is an intelligent database system to deal with SQL query processing.

LITERATURE SURVEY

[1] In 2012, Luis Tari at al[1] describes about how the information can be extracted using database queries which enables generic extraction and reduces the repeated processing of information by performing extraction incrementally which will identify the affected part of data by changing the components. It uses automated query generation system that allows normal users need not have the knowledge of query language to extract the information. The extraction system consists of two phases:

1. Primary Phase**:** In this phase, one-time parsing, tagging and entity recognition is performed to identify individual entries and the resulting parse tree is stored in Parse Tree Database (PTDB)
2. Mining (Extraction) Phase: This phase is then performed by issuing queries of the database to PTDB. The Parse Tree Query Language (PTQL) is designed to express the patterns which are extracted. Queries are issued to

Page 1-9 © MAT Journals 2018. All Rights Reserved

identify the newly recognized patterns. Then the extraction can be performed.

Incremental Information Extraction system however is an active research, But has several disadvantages, that is accuracy of the system has to be improved and the computational cost of extracting information can be high. These systems do not provide the ability to manage intermediate data which is processed such as semantic information and parse tree.

[2] Mrs. Neelu Nihalani in 2011 proposed a scheme about Natural Language Interface to Databases (NLIDB). These are the systems which will accept the query request in natural language such as English or other and translate the request into a database query. Using this system any of the users can collect the information from database. NLIDB systems provide flexible approach, which leads to maximize the usage of database. It consists of two sub components: 1) Verbal Component– which will translate natural language request into query. **2)** Catalog Component – Which is responsible for performing Database Management Functions.
Advantages of NLIDB systems includes 1) Artificial Languages are not used 2) Simple and Easy to use 3) Better response for some questions 4) Fault tolerance 5) Multiple database tables can be used easily.
Disadvantages includes: 1) The cause of system failure will not provide any explanation 2) Only the natural language subset can be handled by the system.

[3] In 2013, Rukshan Alexander at al [3] designed the Natural Language Web Interface for Database (NLWIDB) system to support the search of a user query by using natural language. Retrieval of information using traditional method is combined with research of questions and answers. Suppose the user is required to fetch data stored in the table from database, technically the user has to write the query for the data of his interest in high level language like SQL. NLWIDB system eliminates this complexity of writing query in SQL, and allows the user to fetch tables from database in his/her preferable language. NLWIDB system uses robust language of processing techniques, which makes it convenient and reliable for query access. But it cannot handle complex queries

[4] The author NikulaeStratica& Bipin C. Desai in 2005 proposed semantic templates [4] that is mainly used for converting the queries written in English into the SQL like queries. This system is generally developed consisting of pre-processor module and run time module which is based on semantic templates. The system also includes a dictionary for different type of languages, which contains API (Application Programming Interface) called the WordNet. This WordNet allows to use synonyms, antonym and hyponym of keywords in writing SQL queries.

[5] In 2007, Mahboob A K at all[5] recommended a method, that behaves exactly like a human being while asking questions in natural language query and returns the result back in natural language. In this system machine learning technique is used for questions and answers. This system works in two phases: 1) A classifier is used for the question to retrieve the particular column from the table. 2) After the first phase the required answer is retrieved from the database. Unlike the NLIDB systems, this system does not require any of the information related to semantics.

[6] Dr. Jyoti Pareek et al, in 2013 explains about NLKBIDB system which is the combination of two methodology NLIDB (Natural Language Interface for Database) & KBIDB (Keyword Based Interface for Database). The NLKBIDB system accepts

Page 1-9 © MAT Journals 2018. All Rights Reserved

the query input in natural language and provides the output in table format only if the produced SQL query is a valid query. This system implements Executor of SQL, Agent which is based on Keyword, Lexical Analyzer and SQL Generator for the purpose of KBIDB systems. The interface is provided by NLKBIDB system which includes the area for the user to enter the query in preferred language such as English and also it includes an area for displaying result which consists of converted SQL query. Using this system, incorrect query which is written in natural language as input will be solved using KBI.

DESIGN

System Architecture

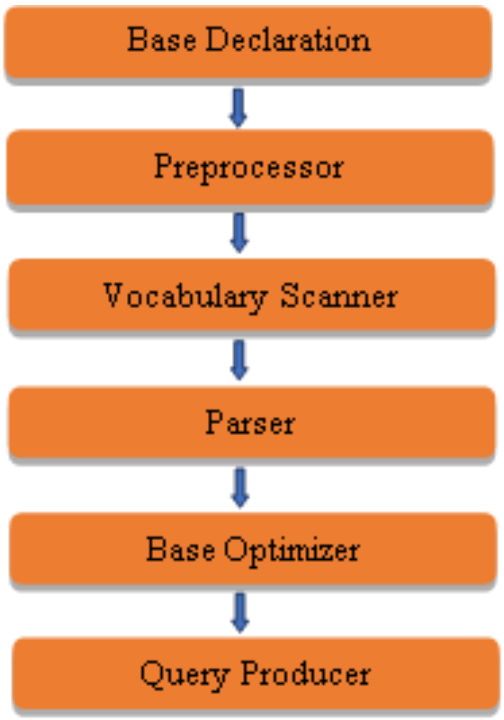

Fig1: *System Architecture*

The Above figure illustrates the architecture of NLQP system. Initially, the system accepts the query or question from user in a preferred language like English. Then it will be converted into SQL query internally which will be in the form of high level language. The query is then processed and the required result will be returned back. The stages included in the system are as follows:

a) Base Declaration:

In this phase, it includes the user written question/statement/declaration that is nothing but the input to the NLQP system. This input statement may consist of different types of questions related to the information in which the user is interested and the input will be in the human understandable format.

b) Preprocessor:

In the pre-processing, the given user input is further processed and mapped to related synonyms which will be used to access desired data from the particular table of the database.

c) Vocabulary Scanner (Lexical Scanner):

Here the user provided input will be divided into tokens, each token is compared with the attributes to fetch the required result. For ex: if the user input is: “display salary of franklin”. The input is divided into 4 tokens like- display, salary, of, franklin. Then each token is compared with attributes from selected database to get the result.

d) Parser:

In this phase, Parsing of data is carried out. As soon as the informations is separated into Chunks/Tokens by succeeding some fixed rules, these tokens/chunks need to be interpreted, managed, or conveyed properly. This comprises corresponding synonym description for the given attributes by removing /, - , periods (.), extra white spaces, eliminating noisy words written by the user etc. Throughout the parse platform it drafts for the declaration syntax and semantics validity.

e) Base Optimizer:

The succeeding phase takes in spawning SQL query. In Base optimizer, subsequently parsing of facts/data have been passed out, the user input is more optimized to spawn the query in SQL language.

Page 1-9 © MAT Journals 2018. All Rights Reserved

f) Query Producer:
Finally, SQL query is produced which is the translation of user written query. At this phase user can view the query for the given base declaration in the form of both SQL language and natural language.

NLQP Block diagram

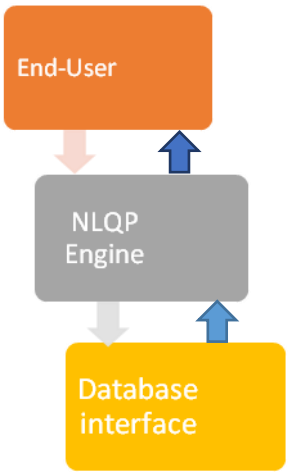

***Fig. 2** Block diagram of NLQP system*

The NLQP system consists of three main components:

1. End-User:
End-User is the normal person who wishes to fetch data from the database as per his interest. The user provides input to NLQP engine in his own language like English.

2. NLQP Engine:
This will parse the user given input question, collect the synonyms related to that and the input will be converted to SQL statement. This SQL statement is given to database interface to which it is connected to retrieve the required information.

3. Database interface:
This will handle connection establishment with database, access and retrieve the data from the particular table. Result will be sent back to user in natural language through NLQP engine.

Detailed Workflow of the system
This section describes the flow of NLQP system. The workflow mainly consist of 4 modules, each module performs the specific operations as shown in figure 3 below.

1. Analyzer/Requirement collection Module:
In this module, the End-User appears. This Module gets the client input or the inquiry required by the client in characteristic dialect which the end client can undoubtedly get it.

2. Verbal Scanner Module:
This module performs parsing of client contribution by taking out unique administrators like - , /, additional blank areas, period image (.), unfortunate words and so on. At that point it separates the contribution to little tokens for the further handling on input.

3. SQL Query Constructor Module:
As the name itself shows, this module builds an inquiry as SQL dialects with the goal that it can be utilized for getting to information from database.

4. Database Handler Module :
This module is in charge of tolerating the SQL question and playing out an association foundation with the required database and afterward getting the information/result from the database. The outcome is then shown to the client in normal dialect as it were.

Page 1-9 © MAT Journals 2018. All Rights Reserved

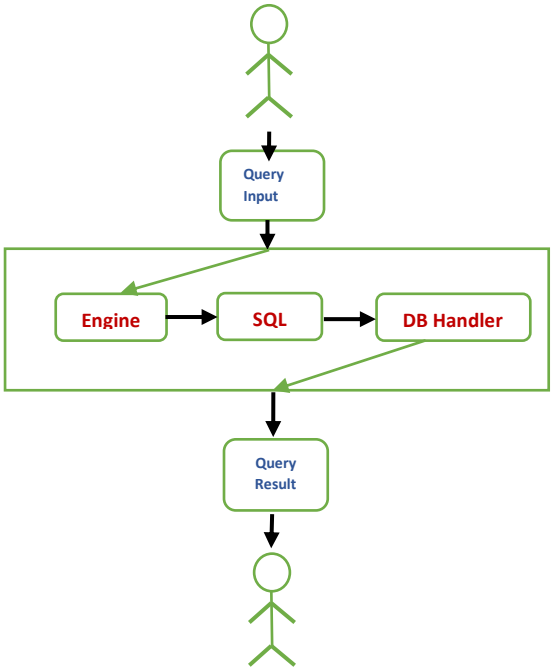

Fig. 3 *Detailed Workflow of the system*

RESULTS

There are two ways of providing user query in this project:

1) User can select any one of the query from "Examples" which will be displayed by default once the database is selected by the user.
2) User can also write query in natural language like English.

Execution Snapshots

Fig. 4 *Administrator Login page*

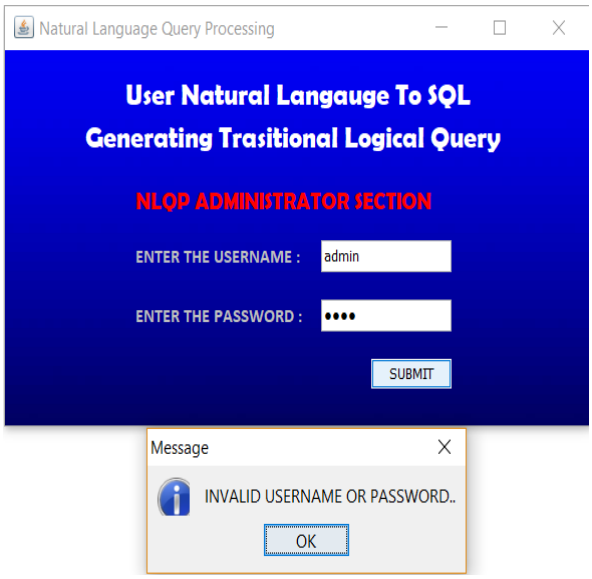

Fig 5: *Failure of Administrator Log-in*

Page 1-9 © MAT Journals 2018. All Rights Reserved

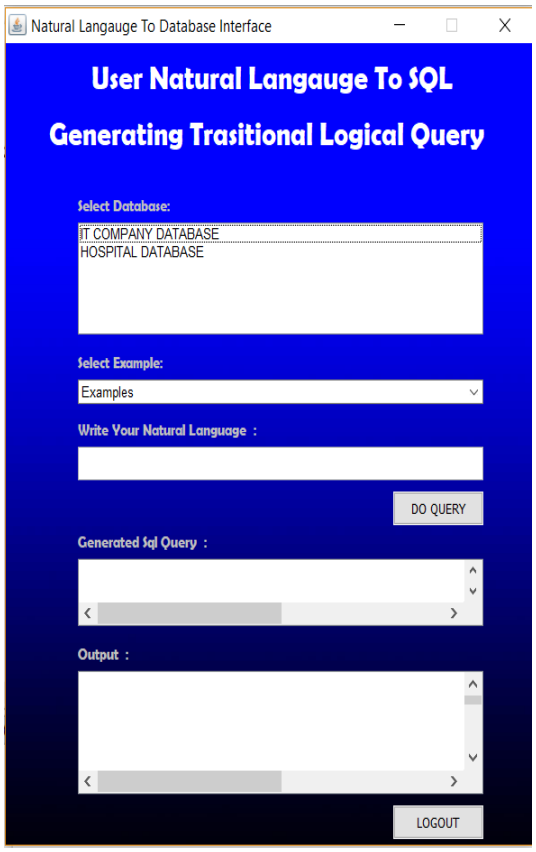

Fig. 6 *Main User Interface*

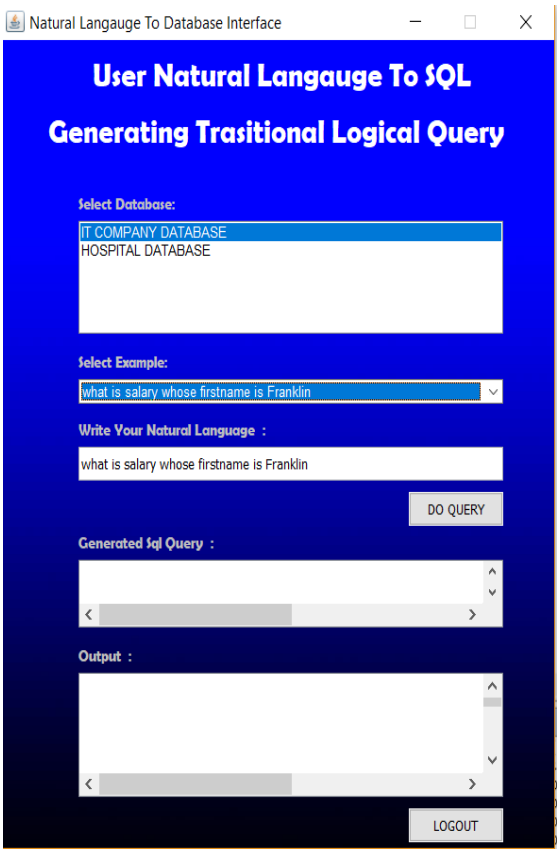

Fig. 8 *Selection of Query from the Examples List*

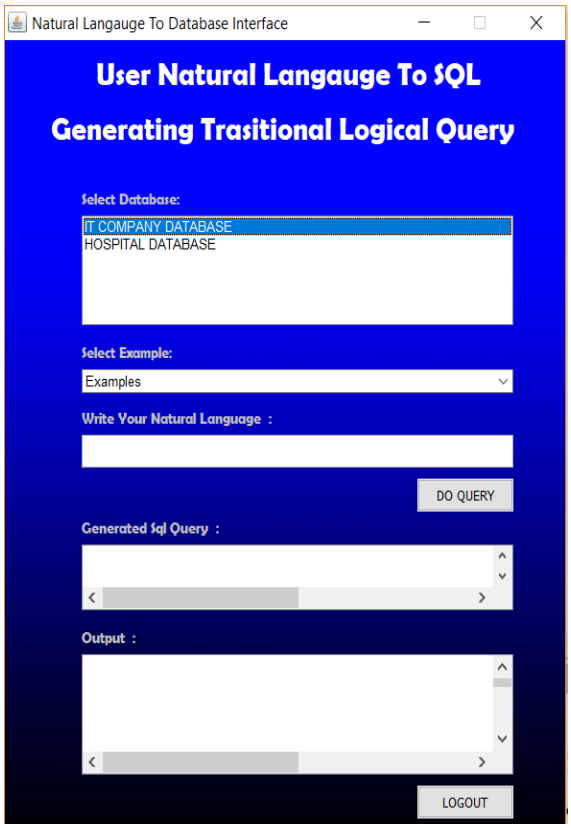

Fig. 7 *Database selection by the user*

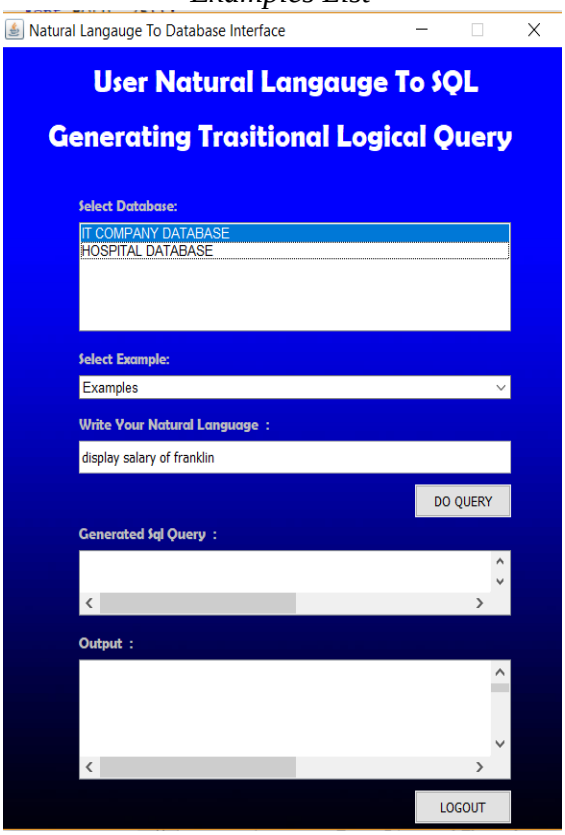

Fig. 9 *Query in natural language by user*

Page 1-9 © MAT Journals 2018. All Rights Reserved

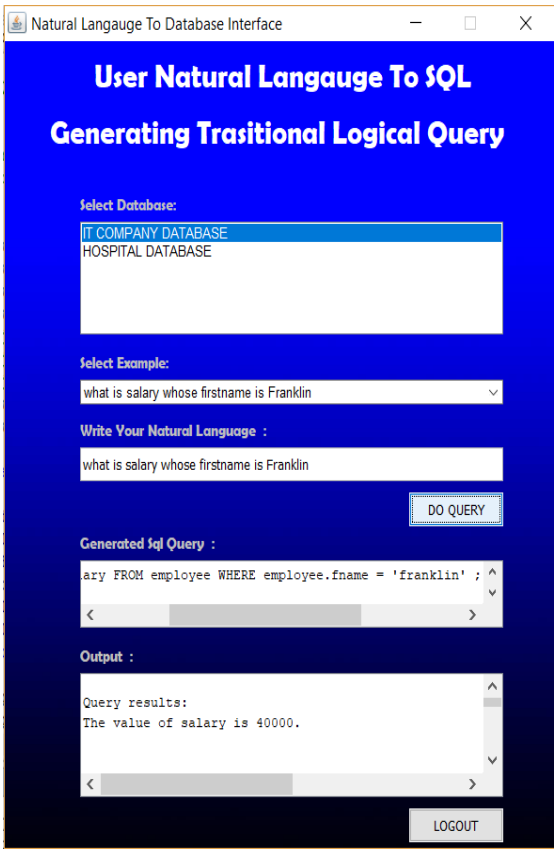

***Fig. 10** Result of the query*

FUTURE SCOPE

The work accepted in our proposed system will be enhanced more with some additional effects which includes the following points:

- We can outspread the project by counting more compound queries of data management (Insert, Update and Delete), GROUP, ALTER etc.
- The organism can be more realized for all natural languages rather that only English.
- The system can also comprise the article of voice appreciation for accepting the queries from the manipulator.

CONCLUSION

Natural Language to Query Processing (NLQP) is one of the topical studies of user approachable interface to the databases. Interface to the database in today's scenario is a complex problem. Our proposed system has the ability to minimize such complexity. It permits to handle simple and complex queries related to the database. It helps common users to access and view the contents of database in user understandable form. The system is smart enough to process the user request from natural language to query language and providing the response back in natural language. A lot of user time will be saved in dealing with simple and easy way of writing natural language queries instead of writing complex SQL queries.

REFERENCES

1. Incremental Information Extraction Using Relational Databases [IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 1, JANUARY 2012] Luis Tari, Student Member, IEEE Computer Society, Phan HuyTu,Jo¨ rgHakenberg, Yi Chen, Member, IEEE Computer Society, Tran Cao Son, Graciela Gonzalez, and Chitta Baral
2. Natural language Interface for Database: A Brief review IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011, www.IJCSI.org
3. Natural Language Web Interface for Database Proceedings of the Third International Symposium, SEUSL: 6-7 July 2013, Olivia, Sri Lanka
4. Using semantic templates for a natural language interface to the CINDI virtual library - NiculaeStratica, Leila Kosseim, Bipin C. Desai [Department of Computer Science, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal H3G 1M8, Canada]
5. Machine Learning for Question Answering from Tabular Data - Mahboob Alam Khalid, Valentin Jijkoun, Maarten de Rijke [ISLA, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The

Page 1-9 © MAT Journals 2018. All Rights Reserved

Netherlands]
6. NLKBIDB - Natural Language and Keyword Based Interface to Database- Axita Shah, Dr. Jyoti Pareek, Hemal Patel, Namrata Panchal
7. Information Extraction Using Database Queries.(Luis Tari1, Phan Huy Tu1, Jorge Hakenberg1)Department of Computer Science and Engineering, Arizona State University Tempe, AZ 85287, USA
8. Efficient Information Extraction over Evolving Text Data, Jun Yang, Raghu Ramakrishnan Duke University, Yahoo! Research

Cite this article as: Miss. Pooja M Rajannavar, & Prof. Gambhir Halase. (2018). Mining Based Natural Language to Database Interface. Journal of Data Mining and Management, 3(3), 1–9. http://doi.org/10.5281/zenodo.1745408

Page 1-9 © MAT Journals 2018. All Rights Reserved