International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 9 No. 4 Dec. 2014, pp. 1746-1754

http://www.ijias.issr-journals.org/

A Natural Language Retrieval System, Natural Language Interface to Data Warehouse (NLI to DWH)

Junaid Haseeb, Irfan Majeed, Fiaz Majeed, and Umair Shafique

Department of Information Technology,
University of Gujrat,
Gujrat, Pakistan

ABSTRACT: The Data Warehouse is the advanced form of the database which is used for decision-making by the executives. There are several front-end tools for data warehouse available to support decision making. These tools come under the categories such as OLAP, Data Mining and Enterprise Information System etc. The common thing in using these tools is to have knowledge about the schema. The users of decision-making systems are top management or executives who are normally non-technical having less knowledge of the data warehouse schema and about writing database technical queries. Therefore, Natural Language (NL) interface can facilitate the executives in their decision-making process. The users prefer to have an easy querying tool that free them from technicalities of back-end processes and let them focus on desired results.
This motivated us to develop a natural language retrieval system (Natural Language Interface to Data Warehouse) that supports users especially in the ad-hoc query development. Using this system, non-technical users can easily write any ad-hoc information need in their native language. Users without having the knowledge about back-end query processing and schema can retrieve any information they want that is available in Data Warehouse. As a result, the complexity and time is reduced as well as dependency is removed.

KEYWORDS: Business Intelligence, Decision Making, Retrieval, Query, Schema.

1 INTRODUCTION

A data warehouse is a subject oriented, non volatile, integrated and time variant collection of data used in support of management's decision making process and business intelligence [1]. It is the collection of different heterogeneous data bases. A multidimensional model will be very helpful for Executives, Business Analysts and Managers to know all dimensions for good and effective decision making. Top management may not be technically expert in order to assist top management for effective and better decision making there is a need of natural language interface to data warehouse. There are several front-end tools for data warehouse available to support decision making. The common issue in using all front-end tools is that user must have knowledge about schema and technical queries. Where the process of decision making is performed by managers, business analyst or executives and all these have little technical knowledge. The users prefer to have an easy querying tool that free them from technicalities of back-end processes and let them focus on desired results.

Considering the above scenario our intention in this work is to build a natural language interface to support non-technical users unlike other Business Intelligence (BI) tools that mostly focus on expert users so that we develop and design a system name as Natural Language Interface to Data Warehouse (NLI to DWH) which has some goals to achieve. The main goal is to assist Top Management by providing a natural language interface for data warehouse for easy and efficient querying as well as effective decision making. There are some other goals of developing the system and these are

- A system which can be used with minor domain knowledge
- Support in ad-hoc query with simplicity
- Provide broad scope of analysis
- Reduce dependence of I.T staff

Corresponding Author: Umair Shafique

2 LITERATURE REVIEW

Data Warehouse provides an effective way for analysis and statistic to the mass data, and helps to do the decision-making. Data Warehouse integrates several types of operational or external sources to provide multidimensional analysis of data. For decision purposes, multidimensional model plays a key role in data warehousing. Better understanding of multidimensional data of an organization could be very beneficial for a businessman or executives. It is also good for an organization that their executives, business analyst and managers to know about all dimensions of enterprise to take effective, good and fast decisions.

The Natural Language Interfaces to Databases (NLIDB) [2] facilitate users to write their request in natural language rather than technical database query language. Several NLIDBs have been developed. There are very few Natural Language Interfaces has been develop till know. One of them was developed with XML database [3]. Some interfaces are question answer based [4], [5] and some of the Natural Language Interfaces are domain specific that are made especially for that domain [6], [7], [8]. Multidimensional models from an implemented data warehouse can be generated by reverse engineering process [9]. The experiments proved that the approach has been successfully reverse engineered the data warehouse.

3 DEMONSTRATION OF THE SYSTEM (NLI TO DWH)

The system is design and develop provide a Natural Language Interface to Data Warehouse. This interface helps the users to query in natural language. A user query in natural language and the results display after searching for relevant context from data warehouse. When user query at run time it will break the query into parts and each part is checked against schema objects.

To develop Natural language Interface to Data Warehouse system the C#.Net and Crystal Reports used with SQL Server 2008 database. General architecture of the system is given below. System architecture is shown in Figure 1.

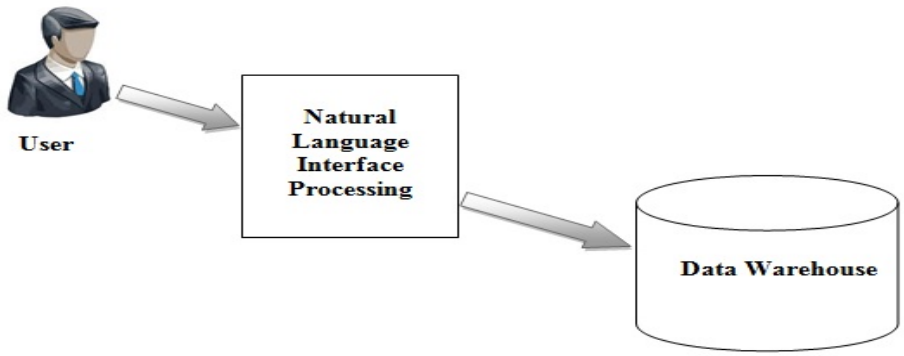

Fig. 1. System Architecture

The system provides ability to choose from two search techniques (search online or search offline) to its users. When a user has selected an option the progress bar shows the current state of system. Synonyms generated for all objects which will help for mapping the query [10], because user is free to query in natural language and can use any synonyms of any word. User can generate the synonyms of objects in both searching techniques (search offline and search online). Front end of Natural Language Interface to Data Warehouse is shown in Figure 2.

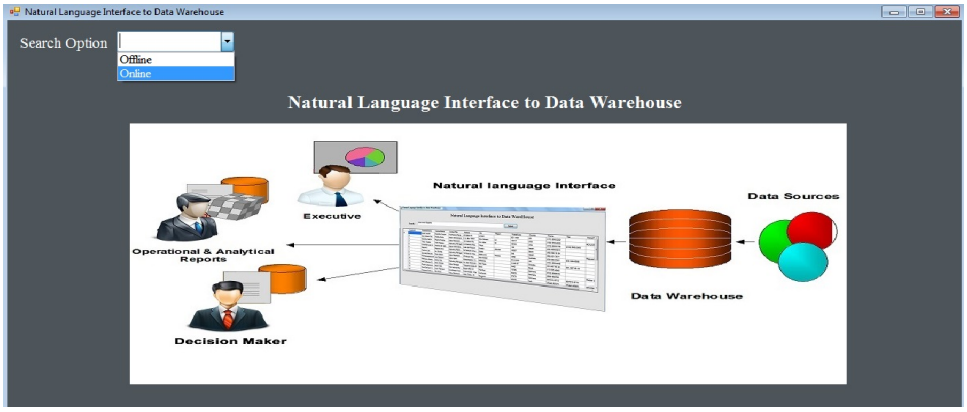

Fig. 2. Natural Language Interface to Data Warehouse

3.1 SEARCH OPTION

This system provides the ability to the users to choose any search option from online search or offline search. After choosing any of the above option users is provided an easy interface to make selection of searching.

3.1.1 OFFLINE SEARCH OPTION

The offline search option provides the ability to search results from XML file. In Offline Search system will map user's natural language query with the XML file objects and find out the results. User can generate the XML file of a database on clicking "Generate XML file" button in offline search. A message will be shown to user when XML file generated successfully. Offline Search Option is shown in Figure 3.

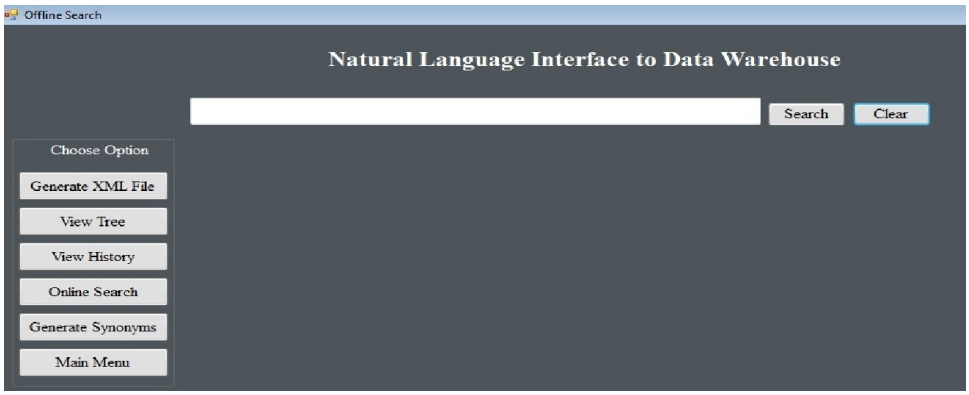

Fig. 3. Offline Search Option

Search offline will have the above dash board for user with features of

- Generate XML file
- View Tree
- View History
- Redirect to Online Search
- Generate Synonyms
- Redirect to main View

3.1.2 ONLINE SEARCH OPTION

The online search option provides the ability to search desired results from Database. In Online Search system will map user's natural language query with the Database schema objects and find out the results. User can generate the schema tree

of a database on clicking “View Tree” button in online search. Schema tree will be shown to user. Online Search Option is shown in Figure 4.

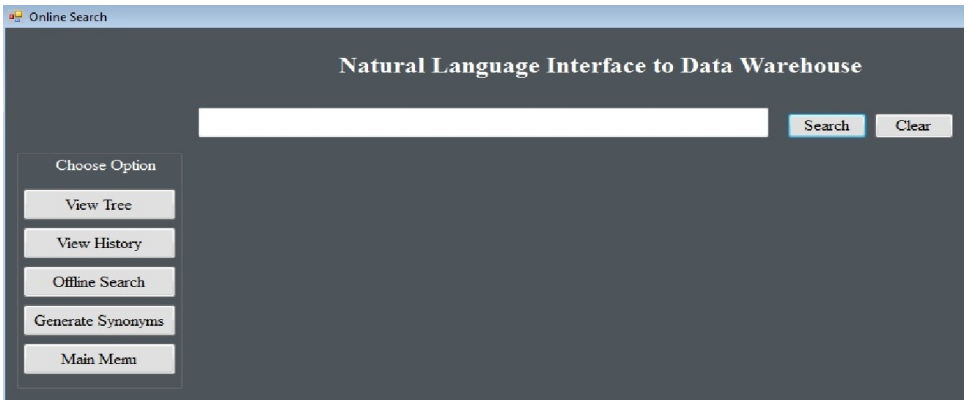

Fig. 4. Online Search Option

Search online will have the above dash board for user with features of

- View Tree
- View History
- Redirect to Offline Search
- Generate Synonyms
- Redirect to main View

3.2 NATURAL LANGUAGE INTERFACE FOR SEARCHING

User will write any query in natural language (without any syntax) and system will show the desired results e.g. “show me details about categories”.

User can also write any query in native language whether user wants to search selected results or any other results. User can query anything from available data according to need e.g. “Show me details about categoryid=1 from products”. NLI searching window is shown in Figure 5.

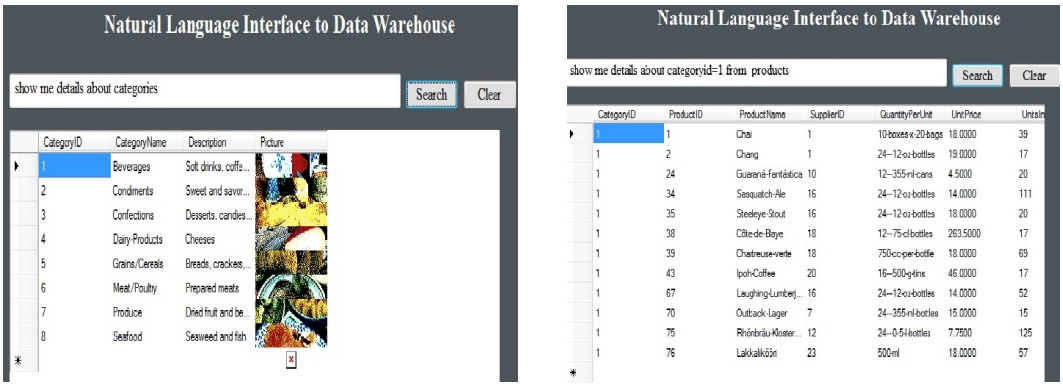

Fig. 5. NLI for Searching

3.3 GENERATE SYNONYMS OF DATABASE OBJECTS

Synonyms generated for all database objects which will help for mapping the query to Database, because user is free to query in natural language and can use any synonyms. Synonyms will be generated and stored in dictionary. Mapping of query to Database objects. User can generate the synonyms of objects in both searching techniques (search offline and search online). Synonyms generated for all objects which will help for mapping the query, because user is free to query in natural language and can use any synonyms of any word. Generate Synonyms window is shown in Figure 6.

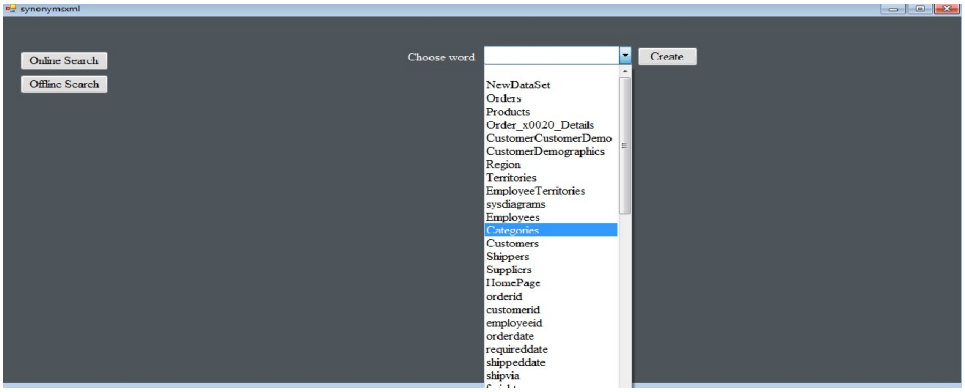

Fig. 6. Generate Synonyms

3.4 INTELLISENSE OPTION (SUGGESTIONS)

When user writes any word into query box system will show all the suggestions matching with that word existing in the form of objects in available data so that user can easily and accurately write the query and also helps in saving time. System's IntelliSense menu (suggestions pop-up) also focuses on retrieving the desired results for users with providing the time saving facility. IntelliSense Option (suggestion) window is shown in Figure 7.

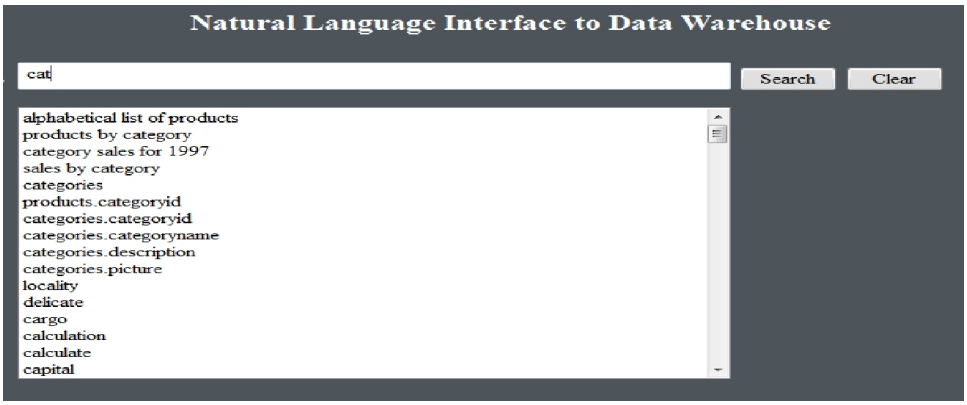

Fig. 7. Suggestions

3.5 RANKING MULTIPLE INTERPRETATIONS

When user searches for the data by writing a natural language query that exist in more than one tables system will automatically map all the existing results for user query and allowing the user to choose their desired results e.g. User searches "show me details of productid=1" if this information exists in 2 different tables system will show all the available options that contain this data so that user can view desired results. Ranking Multiple Interpretations shown in Figure 8.

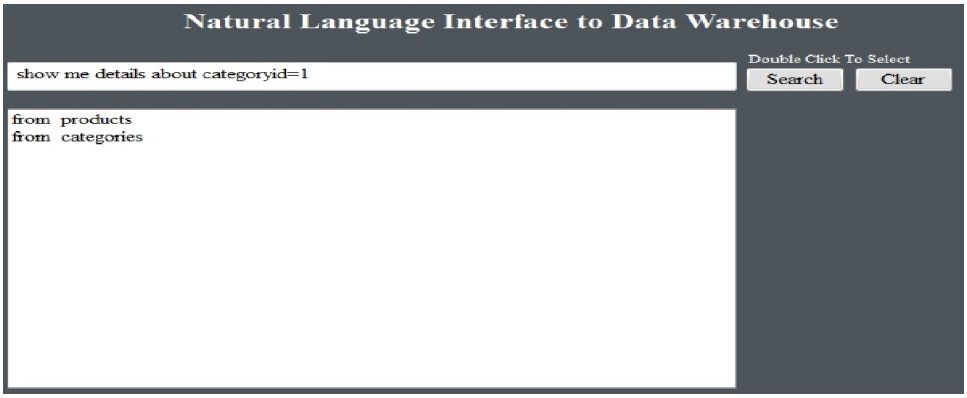

Fig. 8. Ranking Multiple Interpretations

3.6 GROUP BY FUNCTION

System facilitates the user by helping them in setting parameters of query for more efficient search e.g. user viewing the results of total units in stock system will give them facility of viewing this result against all parameters of (productid, product name etc). The system also shows results to user for the query in a well formatted structure that is easily understandable to user. Group by Function is shown in Figure 9.

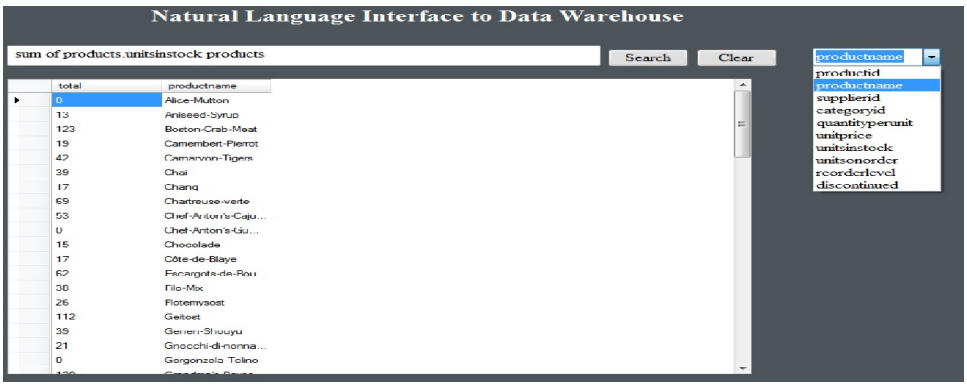

Fig. 9. Group by Function

3.7 DECISION SUPPORT SYSTEM

System is a dully decision support system e.g. user in a situation wants to see the details of products who are not available in stock user can simply write “show details about products having unitsinstock=0”. Decision Support System window is shown in Figure 10.

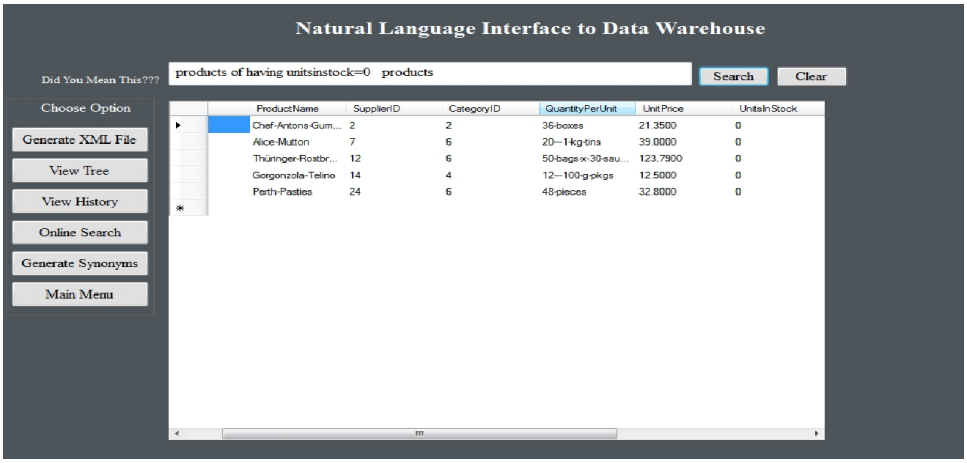

Fig. 10. Decision Support System

4 LIMITATIONS

The system has some limitations and these limitations can be addressed in future to enhance our Natural Language Interface to Data warehouse system. The limitations are

- Search is pop-up based suggestion list
- System Interface shows resultant data only in tabular format
- Dashboard menu has limited facilities for end user
- To attach another data warehouse change in connection string is necessary and to remove the spaces between data value
- Results against the user searched query is populated in grid view only

5 CONCLUSION

The expected output from the system was providing a Natural Language Interface to Data Warehouse (NLI to DWH) so that users with little technical knowledge can easily retrieve desired results and the system achieved the expected output efficiently and conveniently. By using this system User will query to data warehouse through a natural language interface. User can retrieve any desired results he/she want to know about data. User can query to data warehouse whether he/she is not awarded of schema. Decision makers (non-technical users) can easily write any ad-hoc information need in their native language.

REFERENCES

[1] W.H. Inmon "Building the Data Warehouse. 1st Edition." Wiley and Sons, 1992.

[2] Umair Shafique and Haseeb Qaiser, "A Comprehensive Study on Natural Language Processing and Natural Language Interface to Databases," International Journal of Innovation and Scientific Research, vol. 9, no. 2, pp. 297–306, September 2014.

[3] Li, Y., Yang, H., Jagadish, H.V. "Nalix: an interactive natural language interface for querying xml", SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data, 2005.

[4] Woods, W., Kaplan, R. and Webber, B. "The Lunar Sciences Natural Language Information System" Bolt Beranek and Newman Inc., Cambridge, Massachusetts Final Report. B. B. N. Report No 2378, 1972.

[5] R.J.H., Scha, "Philips Question Answering System PHILIQA1", In SIGART Newsletter, no.61. ACM, New York, February 1977.

[6] Hendrix, G., Sacrdoti, E., Sagalowicz, D. and Slocum, J. "Developing a natural language interface to complex data". ACM Transactions on Database Systems, Volume 3, No. 2, USA, Pages 105 – 147, 1978.

[7] David L. Waltz, "An English Language Question Answering System for a Large Relational Database", Communications of The ACM, Volume 21, July 1978.

[8] Ana-maria Popescu , Alex Armanasu , Oren Etzioni , David Ko and Alexander Yates, "Modern natural language interfaces to databases: Composing statistical parsing with semantic tractability", COLING '04 Proceedings of the 20th international conference on Computational Linguistics Article No. 141, 2004.

[9] Hausi A. Müller, Jens H. Jahnke, Dennis B. Smith, Margaret-Anne Storey, Scott R. Tilley and Kenny Wong, "Reverse Engineering: A Roadmap" ICSE '00 Proceedings of the Conference on The Future of Software Engineering Pages 47-60 ACM New York, NY, USA, 2000.

[10] Renée J. Miller, Laura M. Haas and Mauricio A. Hernández "Schema Mapping as Query Discovery", VLDB '00 Proceedings of the 26th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, Pages 77-88, 2000.

AUTHOR'S BIOGRAPHY

JUNAID HASEEB- Received BS (Information Technology) degree from University of Gujrat, Gujrat, Pakistan in 2013. He is currently doing MS degree from COMSATS Institute of Information Technology (CIIT) Islamabad Pakistan. He is the members of this research project on the development of "A Natural Language Retrieval System, Natural Language Interface to Data Warehouse (NLI to DWH)".

IRFAN MAJEED- Received BS (Information Technology) degree from University of Gujrat, Gujrat, Pakistan in 2013. He is the members of this research project on the development of "A Natural Language Retrieval System, Natural Language Interface to Data Warehouse (NLI to DWH)".

FIAZ MAJEED- Received MS degree from COMSATS Institute of Information Technology (CIIT) Lahore Pakistan in 2009. He is currently PhD scholar in University of Engineering and Technology (UET) Lahore Pakistan. Further, he is Lecturer in University of Gujrat (UOG), Gujrat, Pakistan and working on couple of research projects. His research interests include data warehousing, data mining, data streams and information retrieval. This paper is part of the research project on the development "A Natural Language Retrieval System, Natural Language Interface to Data Warehouse (NLI to DWH)" and he is the supervisor of this project.

UMAIR SHAFIQUE- Received M.Sc degree in Information Technology, from University of Gujrat, Gujrat Pakistan in 2014. He is the part of this research project "A Natural Language Retrieval System, Natural Language Interface to Dat