



REVIEW: STUDY OF NATURAL LANGUAGE QUERY INTERFACE IN VARIOUS APPLICATION DOMAINS.

Mrs.Rupali Narkhede,Dr.Varsha Pathak

Assistant Professor,
Computer Department
KCES's I.M.R., Jalgaon, India

Abstract: This paper is the study of various Natural Language Query Interfaces in various application domains such as healthcare, web mining, tax, data warehouse, industrial applications, spatial domains, Academic domain and business etc. This paper focuses on the various application domain areas of NLP in which Natural Language Query Interface is used to interact with users and various databases required in many applications domains. To retrieve information from database, we need Natural Language query interface so that any local user with his native language can query the database through interface.

Keywords: *NLQI, NLP, NLIDB,SQL,NLI, ML,RDBMS*

1. Introduction

At present, Natural Language Query Interface is implemented in many application domains like clinical databases, military databases, bank databases, biological databases, spatial databases, regional databases, domain-independent databases and various businesses etc. It is the need of the hour to access or to collect various information in different domain areas and get the desired result. The idea of using natural language instead of using Structured Query Language (SQL) has promoted to a new type of processing method called Natural Language Interface to Database systems (NLIDB) which is a one step next to the development of intelligent database systems which enhance the users for easy querying. The importance of querying and retrieving information with the help of Natural language Query Interface(NLQI) is increased nowadays.

2. Goal of Natural language Interface

The main goal of Natural Language Interface (NLI) is to provide proper responses to user requests These requests may ask the interface for information or they may ask for actions to be performed. The interface may issue several responses to a single request. It may return the number of lines of information as a response. Another goal is domain coverage. In general, the recent craze is the creation of an interface that provides communication with respect to a large number of datasets, domain areas, and possible actions. Another aspect of NLI is easy to use. An Interface should accept unlimited natural language input and it should display interactivity for clarity and for other purposes. One important aspect of NLI is to handle complex queries.

3. Purpose & Need of Natural Language Interface to Databases

Many of these implementations follow the approaches like Dialogue, Corpus, Menu based and lexical analysis and other some sort of parsers. These approaches could be classified into four categories:

- **Pattern Matching:** Many of the system's trust is on pattern matching to directly map the user input to the database. Pattern matching systems are easy to implement but have critical limitations. The main advantage of this system is its simplicity. Example ELIZA system was programmed by Dr. Joseph Weizenbaum in nearly from 1964 to 1966
- **Syntax based:** In this system, the users queries are parsed. The resulting parsed tree is directly mapped to an expression in some database query language. The Syntax-based system uses Grammars and converts parsed trees into SQL queries.
- **Semantic Grammars:** Semantic Grammar uses syntactic and semantic processing. This approach provides a way to assign a name to a certain node in the tree which results in less ambiguity.
- **Intermediate Representation Language (IRL) Systems:** Many difficulties are found in previous approaches of translating parsed trees into structured database query languages. Instead of this we can break the problem into simple steps. For

example user's query is parsed syntactically that is resulted in parsed tree. It is then translated using semantic interpreter into intermediate representation such as logical query representation.

People store huge volumes of information in repositories or databases and they need to manage those databases for this purpose. DBMS is implemented but to access these large databases, we need an interface that can be used by local people who do not have knowledge of SQL queries can enter their query in natural language through Natural language Query Interface (NLQI).

4. Methodology:

We have downloaded many articles of this research area. From these articles, we found 50 articles interesting for the topic under study Viz. "NLQI for databases of different application domains". From these articles, 24 relevant articles are selected for further study. The study has focused to categorize these articles based on their application domains. The Natural Language Processing (NLP) approach, Machine Learning (ML) approach and specific native language from which it is to be translated to SQL.

5. Literature Review:

1. Darshil Shah, D Vanusha [7] have demonstrated Natural Language Query Interface for relational databases. Their system translates natural language input to SQL. Using this system, naïve users can able to accomplish logically defined tasks.

2. Yunyao Li/ et. al. [8] has demonstrated the DaNalix system demonstrated how to take advantage of domain knowledge to improve its usability and accuracy of query translation.

3. Meriana Kobeissi et. al. [9] has proposed an intent based Natural Language Interface (NLI) for querying process execution data.

The interface facilitates the querying activity by understanding the intent of the user from a natural language question, constructing automatically the corresponding Cypher Query to be executed over the process data stored in a graph database, and returns the answer. Their main focus for future work is to generalize results by creating a benchmark of questions collected from experts in the domain and that covers different domains.

4. Galina Datskovsky Moerdler [10], has confirmed NLI to Expert Systems in which user is permitted to ask queries about the information contained in the rules but not usually provided by expert systems.

5. Fadi H. Hazboun et.al. [11], has brought forward a system in which focus is on the use of plain English language and NLP will help users to extract and manage data from databases. The user do not require to learn SQL and user interfaces that can be combined with relational databases built through Online Analytical Processing (OLAP) which makes user to cope with databases using this technology.

6. Hasan M Jamil [12] has introduced NLI plug-in for cooperative query response for biological database. The goal of this paper was to discuss the outline of a smarter and cooperative natural language interface for the biological database where the semantics of a query usually have many interpretations.

7. Hanane Bais et.al. [13] has presented a model of a generic natural language query interface for querying the database. This model is based on a machine learning approach which allows the interface to automatically improve its knowledgebase through experience.

8. Fei Li and H.V.Jagadish [14] have described an Interactive Natural Language Interface for relational databases. With the help of this system, the first natural language query is translated to SQL and then evaluated against the Relational Database Management System (RDBMS).

9. GARY G. HENDRIX et.al. [16] has described the LIFER and LADDER systems. The LIFER parser interprets natural language inputs and translates them into appropriate interactions with application software. The LADDER system provides natural language access to distributed databases.

10. Jingjing Li et.al. [17] has proposed the spatial comprehension model which is able to recognize the meaning of spatial entities based on the semantics of the context. Their SpatialNLI is able to capture semantic structure of the question and translate it into the corresponding syntax of an executable query.

11. Anthony Colas et.al. [18] has proposed a conversational data collection system that accelerates the deployment of conversational NLI applications which utilize structured data.

12. ABHIJEET R. SONTAKKE, AMIT PIMPALKAR [19] proposed rule-based GUI for SQL queries of relational database. This system receives Hindi sentences as a query. The responses of queries are executed and translated back to Hindi Language.

13. Yunyao Li et.al. [20] has proposed NLQI for a database. A large class of natural language queries can be translated into XQuery expressions, these expressions can be evaluated against an XML database. The system is practically implemented and actual user experiences are gathered.

14. Sudarshan Awale, S.J.Karale [21] focused on the accomplishment of a question-answering system that works for small domain specific ontology only.

15. Hongyu Xiong and Ruixiao Sun [22] developed a Structured Query Inference Network (SQIN) to separate schema-related information from the NL inputs which enhances the learning of sequence-based models on general NL knowledge from source domains.

16. Sharath Chander P et.al. [23] proposed an approach that performs concept identification by using event-related concepts available in WordNet to find candidate events from natural language constraints.

17. Jadhav Sneha et.al. [24] introduced Natural Language to database interface where information is extracted from the database by entering query in Natural Language. This framework is most suitable for performing extraction on text corpus which is in Natural language. This extraction approach saves much more time.

18. Dr. M. Humera Khanam and S. Venkata Subbareddy [25] have implemented QA system which follows keyword based matching approach. The word which contains semantic information will be found in knowledgebase.

19. Mona Gharib Mohamed Reda Zahraa E. Mohamed [26] have proposed a new approach for generating SQL query using a multidimensional database interface for non expert user by posting a question which is translated into SQL query.

20. Neelu Nihalani et.al. [27] has proposed intelligent interface for database querying. This system accepts flexible acceptable user queries and converts them into standard SQL queries.

21. Jigha Kapoor and Harpreet Kaur [28] have proposed Punjabi natural language interface to database to retrieve data of patient. In this system, user enters input query about patient or doctor in Punjabi Language and gets the output in the same language.

22. Mr. Ashish Kumar [29] Proposed Hindi Language Interface to Database in which natural language query in Hindi is given as input to the interface. It is tokenized and semantic mapping is performed on it. Finally, SQL query is generated which is executed to display output to user.

23. A. B. Patil, S. C. Pawar [30] have proposed a keyword-based Marathi Language interface to Database systems which accepts a query in Marathi and it is converted to SQL at last by using Natural Language Processing.

24. Sree Harsha Ramesh et.al. [31] has proposed a novel system for building NLI to real-world databases. They applied syntactic parse trees and token mapping to SQL.

6. Review Analysis:

Article No. & Reference Paper	Domain Name	NLP Approach	ML Approach	Language (From and to)
Article 1 [7]	Wiki SQL dataset provided by Salesforce	Recurrent Neural Network(RNN) + Semantic Parsing	Reinforcement Learning(Policy Gradient Algorithm)	English to SQL
Article 2 [8]	XML data domain	Parse Tree Tokenization, Validation and Translation		English to XQuery
Article 3 [9]	Loan application processes	Named Entity Recognition(NER)	Deep learning	English-Cypher Query-SQL
Article 4 [10]	Tax domain	ATN Parser + Semantic Processing		English to SQL
Article 5 [11]	Data warehouse	Tokenization, Lemmatization, Lexical and Semantic module, POS tagging, Syntactic module		English to SQL
Article 6 [12]	Biological Databases	Term Analyzer and Semantic Graph Matcher		English to SQL
Article 7 [13]	Organizational Database	Morphological analysis, Syntactic and semantic analysis	Classification	English -IXLQ-SQL
Article 8 [14]	Microsoft Academic Search(MAS) dataset	Dependency Parser, Parse Tree Node Mapper, Parse tree structure Adjustor		English to SQL
Article 9 [15]	SQL for Industrial Applications	Semantic Parsing and GAN based augmentation Technique		English to SQL
Article 10 [16]	Data Analysis	Lexicon, Semantic Map Generator(SMG), Meaningful Representation Generation(MRG)		NL-Semantic Map-SQL
Article 11 [17]	Spatial domain	Semantic injection, Seq2seq model		Spatial Query in NL to Executable Query

Article 12 [19]	Sample Student database	Tokenization, Query type Rule, Query Table Rule		Hindi-SQL-Hindi
Article 13 [25]	Regional database in Telugu	Syntax and Semantic approach, Tokenization		Telugu to SQL
Article 14 [26]	Multidimensional databases	Semantic and Lexical Analysis		English to SQL executable query
Article 15 [27]	Domain independent	Semantic matching technique		English to SQL
Article 16 [28]	Patient database	Token Filter, Parser, Query Creator		Punjabi-SQL-Punjabi
Article 17 [29]	Hindi Language database	Tokenization, Semantic Mapping		Hindi to SQL
Article 18 [30]	Student database	String Tokenizer, Keyword Mapper, Query Generator		Marathi-English-SQL
Article 19 [31]	Sales, Human Resources , Product domain	Syntactic Parser, NER, Entity Constraint Recognizer		English to SQL

Table 1.1

7. Results and discussion

The SQL generation from Natural language is still a challenging field. In this study, 24 relevant articles are reviewed and the observations are noted as above in tabular form 1.1. Our study shows that most of the work is done using NLP approach for implementing NLQI in various domains. In 6-7 articles, Tokenization technique is used. In 12-13 articles, Semantic approach is applied to implement the system. 6-7 articles have used Parsing techniques. In 2-3 articles, hybrid approach of NLP and ML is found to be used. In these articles, a rare combination of NER and deep learning is found as well as in another RNN and Reinforcement learning approach is implemented. From this study, we come to know that very few work is done by using ML approach and much of the systems are implemented using NLP approach.

8. Conclusion

In this paper, we found important points for the study of literature related to NLQI. In future this study will be extended further to understand chronological progress in last 10 years. This study will be mainly focused to extract important issues and challenges handled by the researchers in this vibrant research area. We studied approaches and methodologies applied by them and their experimental work performed by them.

References:

- 1) "Natural language Interface for Database: A Brief review", Mrs. Neelu Nihalani, Dr. Sanjay Silakari, Dr. Mahesh Motwani, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011 ISSN (Online): 1694-0814 www.IJCSI.org
- 2) "Efficient Natural Language Query interface to databases", B.SUJATHA, Dr. S.Vishwanatha Raju, S. Nagaprasad, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 9, September 2014
- 3) [https://www.google.com/imgres?imgurl=https%3A%2F%2Fstatic.packt-cdn.com%2Fproducts%2F9781838550295%2Fgraphics%2FC13783_01_04.jpg &imgrefurl= https%3A%2F%2Fsubscription.packtpub.com%2Fbook%2Fdata%2F9781838550295%2F1%2Fch011v11sec04%2Fapplicationsofnaturallanguagprocessing&tbnid=401Sxtx81mcDM&vet=12ahUKEwiAtpColtD5AhVB9gFHQSACmcQMygKegUIARDWAQ..i&docid=C3LSIyCgb92XLM&w=738&h=489&q=applications%20of%20natural%20language%20processing%20with%20the%20help%20of%20diagram&ved=2ahUKEwiAtpColtD5AhVB9gFHQSACmcQMygKegUIARDWAQ](https://www.google.com/imgres?imgurl=https%3A%2F%2Fstatic.packt-cdn.com%2Fproducts%2F9781838550295%2Fgraphics%2FC13783_01_04.jpg&imgrefurl=https%3A%2F%2Fsubscription.packtpub.com%2Fbook%2Fdata%2F9781838550295%2F1%2Fch011v11sec04%2Fapplicationsofnaturallanguagprocessing&tbnid=401Sxtx81mcDM&vet=12ahUKEwiAtpColtD5AhVB9gFHQSACmcQMygKegUIARDWAQ..i&docid=C3LSIyCgb92XLM&w=738&h=489&q=applications%20of%20natural%20language%20processing%20with%20the%20help%20of%20diagram&ved=2ahUKEwiAtpColtD5AhVB9gFHQSACmcQMygKegUIARDWAQ)
- 4) <https://www.analyticsvidhya.com/blog/2020/07/top-10-applications-of-natural-language-processing-nlp/>
- 5) "An Analysis of the Applications of Natural Language Processing in Various Sectors", Priya B., Nandhini J.M and Gnanasekaran T, Smart Intelligent Computing and Communication Technology V.D. Ambeth Kumar et al. (Eds.) © 2021 The authors and IOS Press.
- 6) "A Natural Language Interface for Mobile Devices", Boris Katz, Gary Borchardt, Sue Felshin, and Federico Mora, Boris Katz, Gary Borchardt, Sue Felshin, and Federico Mora. A Natural Language Interface for Mobile Devices. In Kent L. Norman and Jurek Kirakowski (Eds.), The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition, John Wiley & Sons, 2018, pp. 539–559
- 7) "Optimizing Natural Language Interface for Relational Database", Darshil Shah, D Vanusha, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-4, April 2019
- 8) "DaNaLIX: a Domain-adaptive Natural Language Interface for Querying XML", Yunyao Li/, Ishan Chaudhuri, Huahai Yang, Satinder Singh, H. V. Jagadish, Copyright is held by the author/owner(s). SIGMOD '07, June 12-14, 2007, Beijing, China. ACM 978-1-59593-686-8/07/0006.
- 9) An Intent-Based Natural Language Interface for Querying Process Execution Data, Meriana Kobeissi, Nour Assy, Walid Gaaloul, Bruno Defude, Bassem Hai, © 2021 IEEE. Pre-print copy of the manuscript published in the Proceedings of the 2021 3rd International Conference on Process Mining (ICPM 2021), available at <https://doi.org/10.1109/ICPM53251.2021>

- 10) "Building a Natural Language Interface to Expert Systems", Galina Datskovsky Moerdler, Technical Report #CUCS-306-87
- 11) "A Natural Language Interface to Relational Databases Using an Online Analytic Processing Hypercube", Fadi H. Hazboun, Majdi Owda and Amani Yousef Owda, Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>), AI 2021, 2, 720–737. <https://doi.org/10.3390/ai2040043>.
- 12) "A natural language interface plug-in for cooperative query answering in biological databases", Hasan M Jamil, *amil BMC Genomics* 2012, 13(Suppl 3):S4 <http://www.biomedcentral.com/1471-2164/13/S3/S4>.
- 13) "A Model of a Generic Natural Language Interface for Querying Database", Hanane Bais, Mustapha Machkour and Lahcen Koutti, *I.J. Intelligent Systems and Applications*, 2016, 2, 35-44 Published Online February 2016 in MECS (<http://www.mecspress.org/>) DOI: 10.5815/ijisa.2016.02.0.
- 14) "Constructing an Interactive Natural Language Interface for Relational Databases", Fei Li and H.V.Jagadish, *Proceedings of the VLDB Endowment*, Vol.8 No.1
- 15) "Transferable Natural Language Interface to Structured Queries aided by Adversarial Generation" Hongyu Xiong and Ruixiao Sun 8, No. 1 Copyright 2014 VLDB Endowment 2150-8097/14/09.
- 16) "Developing a Natural Language Interface to Complex Data", GARY G. HENDRIX, EARL D. SACERDOTI, DANIEL SAGALOWICZ AND JONATHAN SLOCUM, *ACM Transactions on Database Systems*, Vol. 3, No. 2, June 1978, Pages 105-147.
- 17) "SpatialNLI: A Spatial Domain Natural Language Interface to Databases Using Spatial Comprehension", Jingjing Li, Wenlu Wang, Wei-Shinn Ku, Yingtao Tian, Haixun Wang, arXiv:1908.10917v1 [cs.CL] 28 Aug 2019.
- 18) "Efficient Deployment of Conversational Natural Language Interfaces over Databases", Anthony Colas, Trung Bui, Franck Dernoncourt, Moumita Sinha, Doo Soon Kim, arXiv:2006.00591v2 [cs.CL] 4 Jun 2020.
- 19) "GRAPHICAL USER INTERFACE AND RELATIONAL DATABASE ACCESSING USING NLP", ABHIJEET R. SONTAKKE, AMIT PIMPALKAR, *International Journal of Advanced Computational Engineering and Networking*, ISSN: 2320-2106, Volume-3, Issue-12, Dec.-2015
- 20) "Constructing a Generic Natural Language Interface for an XML Database", Yunyao Li, Huahai Yang, H. V. Jagadish, Supported in part by NSF 0219513 and 0438909, and NIH 1-U54-DA021519-01A1
- 21) "INFORMATION RETRIEVAL FROM ONTOLOGY USING NATURAL LANGUAGE INTERFACE", Sudarshan Awale, S.J.Karale ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-2, ISSUE-2, 2015.
- 22) "Transferable Natural Language Interface to Structured Queries aided by Adversarial Generation", Hongyu Xiong and Ruixiao Sun, arXiv:1812.01245v1 [cs.CL] 4 Dec 2018.
- 23) "Data Analysis of Natural Language Querying Using NLP Interface", Sharath Chander P, J. Soundarya, R. Priyadharsini and B. Bharathi, *International Journal of Applied Engineering Research* ISSN 0973-4562 Volume 13, Number 8 (2018) pp. 5792-5795 © Research India Publications. <http://www.ripublication.com>.
- 24) "Natural Language to Database Interface", Jadhav Sneha, Raut Shubhangi, A.S.Zore, *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 3, Issue 2, February 2014, ISSN (Online) : 2278-1021 ISSN (Print) : 2319-5940.
- 25) "QUESTION ANSWERING SYSTEM WITH NATURAL LANGUAGE INTERFACE TO DATABASE", Dr. M. HumeraKhanam and S. Venkata Subbareddy, *International Journal of Management, IT & Engineering* Vol. 7 Issue 4, April 2017, ISSN: 2249-0558 Impact Factor: 7.119.
- 26) "INTELLIGENT MULTIDIMENSIONAL DATABASE INTERFACE", Mona Gharib Mohamed Reda Zahraa E. Mohamed, *International Journal of Scientific & Engineering Research*, Volume 4, Issue 11, November-2013 ISSN 2229-551.
- 27) "An Intelligent Interface for relational databases", Neelu Nihalani, Dr. Mahesh Motwani, Dr. Sanjay Silakari, *IJSSST*, Vol. 11, No. 1 ISSN: 147 29 3-804x online, 1473-8031 print
- 28) "Punjabi Natural Language Interface to Database", Jigha Kapoor and Harpreet Kaur, *Dogo Rangsang Research Journal UGC Care Group I Journal* ISSN : 2347-7180 Vol-10 Issue-08 No. 02 August 2020, Copyright @ 2020 Authors.
- 29) "DESIGN AND IMPLEMENTATION OF HINDI LANGUAGE INTERFACE TO DATABASE", Ashish Kumar, *INTERNATIONAL JOURNAL OF CURRENT ENGINEERING AND SCIENTIFIC RESEARCH (IJCESR)*, ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-4, ISSUE-10, 2017.
- 30) "Keyword based Marathi Interface to the Database using Natural Language Processing", A.B. Patil, S. C. Pawar, *International Journal of Scientific Research in Computer Science and Engineering* Vol.6, Issue.1, pp.50-55, Feb (2018), E-ISSN: 2320-7639.
- 31) "Towards Building A Domain Agnostic Natural Language Interface to Real-World Relational Databases", Sree Harsha Ramesh, Jayant Jain, Sarath K S, and Krishna R Sundaresan, D S Sharma, R Sangal and A K Singh. *Proc. of the 13th Intl. Conference on Natural Language Processing*, pages 305–314, Varanasi, India. December 2016. c 2016 NLP Association of India (NLP AI).
- 32) "A Natural Language Interface for Querying Graph Databases", Christina Sun, submitted to Department of Electrical Engineering and computer science on May 25, 2018.