International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 3 | ISSN : 2456-3307

Generation of Natural Language Queries from ER-Diagram

Srinivasa K, Kavitha M

Department of Computer Science and Engineering, Siddaganga Institute of Technology, Tumakuru, Karnataka, India

ABSTRACT

The concept of object-oriented modelling makes the real world represented in some kinds of logical forms.ER-diagram is one of those logical forms, which is used in database design. Many tools are available to generate ER-diagram from natural language database requirement specifications and SQL queries from natural language queries (NLQs). Here we address the problem of generating natural language queries from ER diagram so that user efforts in typing NLQs can be reduced. The proposed method first converts the given ER-diagram to understandable format that is intermediate text or sentence form and then the NLQs are generated from the intermediate form using Natural Language Processing (NLP) techniques.

Keywords: ER-diagram**,** NLP, NLQ, SQL, NLU.

I. INTRODUCTION

Database applications play an important role in today's commercial systems. Most of the businesses need these types of applications in their native language i,e. natural language. Processing of natural language is becoming one of the most active research area which is used in Information Retrieval, Machine Translation and Language Analysis. The main goal of NLP is to enable communication between human and computer more interactive like human-to-human communication. NLP is the ability of a computer program to understand human language. Many challenges in natural language processing involve natural language understanding (NLU) that is enabling computer to derive meaning from human or natural language input and other involve natural language generation, which is the natural language processing task of generating natural language from a machine representation such as a knowledge base or a logical form. Our project lies in natural language generation.

Here the focus is on generation of natural language queries from a conceptual model i.e. ER diagram by extracting text from ER-diagram and generation of sentences from text. Finally, natural language queries are generated from the sentences.

The organization of this document is as follows. In Section 2 we provided few works related to the proposed work. Section 3 dicusses the proposed work and result. Section 4 and 5 include conclusion and future work. Finally, section 6 gives the references used in the paper.

II. RELATED WORK

In [1] Pranali P. Chaudhari discussed generation of SQL queries from the questions, which is in human understanding language so that non-experts can easily generate SQL queries to access the database. In [2] XavierTannier et al proposed Natural language queries for Information retrieval in structured documents. In [3] Eman S. Btoush et al discussed about the generation of ER diagram from natural language requirement specifications. In [4] Saravjeet Kaur et al discussed about SQL generation and execution from natural language processing. SQ-HAL [5] is the powerful system that can translates different types of select queries, which include retrieving of data from multiple tables with or without conditions.

There are many tools available to generate ER diagram from natural language database requirements and SQL queries from natural language queries. As for

our knowledge is concerned, there are no tools to convert ER diagram to natural language queries.

III. PROPOSED WORK

Here we developed a tool that takes ER diagram as input and generates all possible natural language queries, which can be given as input for any tool that converts NLQs to SQL for accessing the database. The following sections discuss the system architecture and the results.

A. SYSTEM ARCHTECTURE

Figure 1 gives the overall architecture of the proposed method.

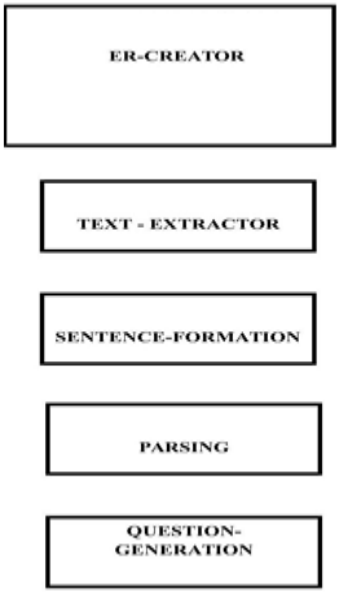

Figure 1. System Architecture

Here we developed a tool to draw an ER-diagram using ER-creator, which considers shapes like rectangle to represent entity, oval to represent attributes and rhombus to indicate relationship.

Text-extractor is a extractor which extracts texts or words which are present in the ER-diagram. All the text present in each entity, attributes and relationships are extracted using text extractor. After extraction of text these words are converted into sentence form.

After extraction of text using text-extractor, sentences are formed by the concatenation method.

Parsing is done for obtained sentences. For the purpose of parsing we took standford parser to parse

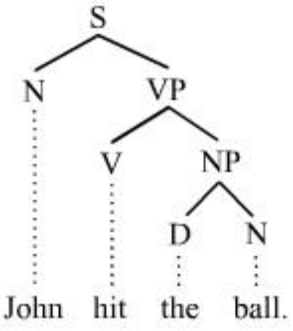

Figure 2. Example of a Parse Tree

the required sentences which provides the parts of speech of a given sentence that helps us in generating NLQs. A sample parse tree is as shown in Figure 2.

Following are the rules for generating questions:

1. If the sentence has noun (by considering subject and object).
<question> -> <List all> + <object> + <of> + <subject> .
Example : Doctor has name.
List all the names of doctor.
<question> -> <Find all> + <object> + <of> + <subject> .
Example: Customer has phone number
Find all customer's phone number.
2. If the sentence consists of noun (by considering object).
<question> -> <Sort all> + <object> + <in> + <ascending order> .
Example: Doctor has name.
Sort all names in ascending order.
<question> -> <Sort all> + <object> + <in> + <descending order> .
Example: Student has Id.
Sort all Ids in descending order.

3. If the sentence consists of compound noun (by considering subject).
<question> -> <Print all information of> + <subject> .
Example: teachers, student belongs to school.
Print all information of teachers, students.

4. If sentence consists of noun in plural form (by considering object).
<question> -> <Count total number of> + <object> .
Example: College has many students.
Count total number of students in college.

IV. RESULTS

User can give ER-creator tool to draw ER-diagram by selecting the shapes in the menu, as shown in Fig 3.

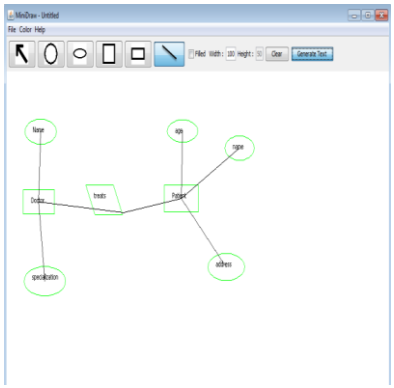

Figure 3. ER Creator

When an ER-diagram is drawn, all the text present in it is extracted in backend and sentences are formed by concatenating the text. Figure 4 shows the output for parsing of sentences and question generation phase.

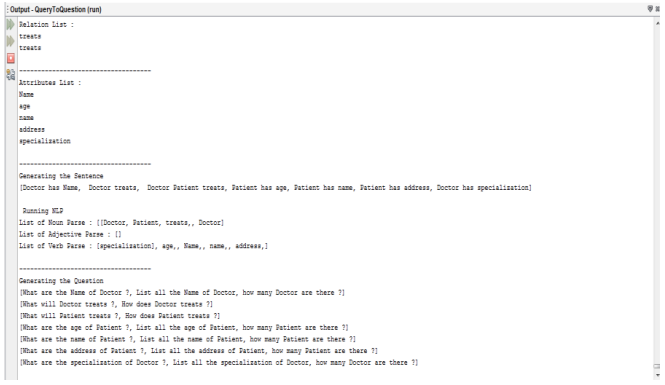

Figure 4. Parsing and Question Generation

V. CONCLUSION

This system is currently capable of handling ER-diagram as an input, extracts all the text in it, and generates all possible natural language queries. It helps the database users to access the data or information easily from the database without wasting time for understanding and analysing ER-diagram. Hence, this tool may serve database users in a better way by reducing the time and complexity.

VI. FUTURE WORK

We have presented only one among many other possible conceptual models, i.e. ER-diagram. More work need to be done for understanding and implementing other conceptual models. However, the system presented in this paper is currently capable of generating some type of questions. In future, we need to develop a system that can generate all types of questions.

VII. REFERENCES

[1]. Natural Language Statement to SQL Query Translator .Pranali P. Chaudhari Asst. Professor, I.T. Department MIT Academy of Engineering Alandi (D), Pune.

[2]. Natural language queries for Information Retrieval in Structured Documents. XavierTannier, Jean-Jacques Girardot and Mihaela Mathieun...cole Nationale SupÈrieure des Mines de Saint-Etienne 158 Cours F auriel F-42023 Saint ...tienne cedex2.

[3]. Generating ER Diagrams from Requirement Specifications Based On Natural Language Processing. Eman S. Btoush and Mustafa M. Hammad Department Of Computer Science, Mutah University, Jordan.

[4]. SQL generation and execution from natural language processing. Saravjeet Kaur, Rashmeet Singh Bali, Student, MMU, Mullana.

[5]. SQ-HAL: Natural language to SQL Translator, http://www.csse.monash.edu.au/hons/projects/2000/Supun.Ruwanpura/