

FACILITATING DOCUMENT ANNOTATION USING CONTENT AND QUERYING VALUE

C. Nancy * and Sha Sha Ali

Student of M.Tech, Bharath College Of Engineering And Technology For Women, Andhra Pradesh, India

Department of CSE, Bharath College Of Engineering And Technology For Women, Andhra Pradesh, India

Abstract— In the present computing world, computer based information technologies have been extensively used to help many organizations, private companies, academic and education institutions to manage their processes and information systems. Information systems are used to manage data. A general information management system that is capable of managing several kinds of data, stored in the database is known as Database Management System (DBMS). Databases are comprehensive element in private and public information systems which are essential in number of application areas. Databases are built with the objective of facilitating the activities of data storage, processing, and retrieval associated with data management in information systems. Due to the progress and in-deep applications of computer technologies, the widespread applications of web technology in several areas to be accurate, databases have become the repositories of huge volumes of data. In relational databases, to retrieve information from a database, one needs to formulate a query in such way that the computer will understand and produce the desired output.

Index Terms—Natural Language, User Language Words

Manuscript C. Nancy, Student of M.Tech, Bharath College of Engineering and Technology for Women, Kadapa, Andhra Pradesh, India. Email: nancymrln47@gmail.com

Sha Sha Ali, Assistant Professor Department of CSE, Bharath College of Engineering and Technology for Women, Kadapa, Andhra Pradesh, India.

I. INTRODUCTION

The Structured Query Language (SQL) norms are been pursued in almost all languages for relational database systems. Structured Query Language (SQL) is an ANSI standard for accessing and manipulating the information stored in relational databases. It is comprehensively employed in industry and is supported by major database management systems (DBMS). Most of the languages used for manipulating relational database systems are based on the norms of SQL. The SQL norms are based on a Boolean interpretation of the queries. But some user requirements may not be answered explicitly by a classic querying system. It is due to the fact that the requirements' characteristics cannot be expressed by regular query languages. Many novel-generation database applications stipulate intelligent information management necessitating efficient interactions between the users and database. In recent times, there is a rising demands for non-expert users to query relational databases in a more natural language encompassing linguistic variables and terms, instead of operating on the values of the attributes. Therefore the idea of using natural language instead of SQL has prompted the development of new type of processing method called Natural Language Interface to Relational Database systems (NLIRDB). NLIRDB is a step towards the development of interfaces to database systems to

enhance the users in performing flexible querying in databases.

II. LITERATURE SURVEY

Natural Language Processing (NLP) is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things. NLP researchers aim to gather knowledge on how human beings understand and use language so that appropriate tools and techniques can be developed to make computer systems understand and manipulate natural languages to perform the desired tasks. The foundations of NLP lie in a number of disciplines, viz. computer and information sciences, linguistics, mathematics, electrical and electronic engineering, artificial intelligence and robotics, psychology, etc.

NLP Tools and Techniques

A number of researchers have attempted to come up with improved technology for performing various activities that form important parts of NLP works. These works may be categorized as follows:

- Lexical and morphological analysis, noun phrase generation, word segmentation, etc. (Bangalore & Joshi, 1999; Barker & Cornacchia, 2000; Chen & Chang, 1998; Dogru & Slagle, 1999; Kam-Fai et al., 1998; Kazakov et al., 1999; Lovis et al., 1998; Tolle & Chen, 2000; Zweigenbaum & Grabar, 1999).
- Semantic and discourse analysis, word meaning and knowledge representation (Kehler, 1997; Mihalcea & Moldovan, 1999; Meyer & Dale, 1999; Pedersen & Bruce, 1998; Poesio & Vieira, 1998; Tsuda & Nakamura, 1999).
- Knowledge-based approaches and tools for NLP (Argamon et al., 1998; Fernandez & Garcia-Serrano, 2000; Martinez et al., 2000, 1998).

2.2 Already Developed NLIDB Systems

Work for developing Natural Language Interface to Database has started in early seventies. Since then many systems have been developed. Early systems have many flaws then some systems were developed to overcome these flaws.

NLP Applications

Applications of NLP include a number of fields of studies, such as machine translation, natural language text processing and summarization, user interfaces, multilingual and cross language information retrieval (CLIR), speech recognition, artificial intelligence and expert systems, and so on.

The applications utilizing NLP include the following:

Machine Translation:

Automatically translate text from one human language to another. This is one of the most difficult problems, and is a member of a class of problems colloquially termed "AI-complete", i.e. requiring all of the different types of knowledge that humans possess (grammar, semantics, facts about the real world, etc.) in order to solve properly.

Speech Recognition:

Given a sound clip of a person or people speaking, determine the textual representation of the speech. This is the opposite of text to speech and is one of the extremely difficult problems colloquially termed "AI-complete". In natural speech there are hardly any pauses between successive words, and thus speech segmentation is a necessary subtask of speech recognition. Note also that in most spoken languages, the sounds representing successive letters blend into each other in a process termed coarticulation, so the conversion of the analogue signal to discrete characters can be a very difficult process.

Word segmentation:

Separate a chunk of continuous text into separate words. For a language like English, this is fairly trivial, since

words are usually separated by spaces. However, some written languages like Chinese, Japanese and Thai do not mark word boundaries in such a fashion, and in those languages text segmentation is a significant task requiring knowledge of the vocabulary and morphology of words in the language.

Natural Language Interfaces to Databases:

A Natural Language Interface to a Database (NLIDB) is a system that allows the user to access information stored in a database by typing requests expressed in some natural language.

III. SYSTEM ANALYSIS

Databases are gaining prime importance in a huge variety of application areas employing private and public information systems. Databases are built with the objective of facilitating the activities of data storage, processing, and retrieval associated with data management in information systems. Due to the progress and in-deep applications of computer technologies, the widespread applications of web technology in several areas to be accurate, databases have become the repositories of huge volumes of data. In relational databases, to retrieve information from a database, one needs to formulate a query in such way that the computer will understand and produce the desired output. The Structured Query Language (SQL) norms are been pursued in almost all languages for relational database systems. The SQL norms are based on a Boolean interpretation of the queries. But some user requirements may not be answered explicitly by a classic querying system. It is due to the fact that the requirements' characteristics cannot be expressed by regular query languages. Many novel-generation database applications stipulate intelligent information management necessitating efficient interactions between the users and database. In recent times, there is a rising demands for non-expert users to query relational

databases in a more natural language encompassing linguistic variables and terms, instead of operating on the values of the attributes.

Related Work

As databases are concerned with different types of User Interfaces, some of them were discussed in this section.

User interface - The features of a computer system which allows the user to interact with it.

A user interface, also sometimes called a human-computer interface, comprises both hardware and software components. It handles the interaction between the user and the system. There are different ways of interacting with computer systems which have evolved over the years. There are four main types of user interface:

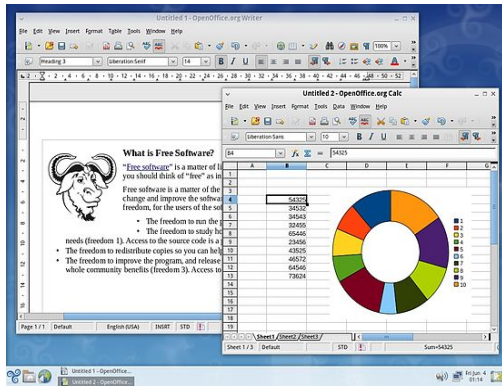
- Command line
- Graphical user interface (GUI)
- Menu driven
- Form based

Command Line Interface:

A command line interface is the oldest of the interfaces as shown in the fig 3.1. It involves the computer responding to commands typed by the operator. This type of interface has the drawback that it requires the operator to remember a range of different commands and is not ideal for novice users.

```
Enter today's date (m-d-y): 08-04-81
The IBM Personal Computer DOS
Version 1.00 (C)Copyright IBM Corp 1981
A>dir *.com
IBMDIO COM      1920 07-23-81
IBMBIOS COM      6400 08-13-81
COMMAND COM     3231 08-04-81
FORMAT COM      2560 08-04-81
CHKDSK COM      1395 08-04-81
SYS COM          896 08-04-81
DISKCOPY COM    1216 08-04-81
DISKCOMP COM    1124 08-04-81
COMP COM        1620 08-04-81
DATE COM         252 08-04-81
TIME COM         250 08-04-81
MODE COM         860 08-04-81
EDLIN COM       2392 08-04-81
DEBUG COM       6049 08-04-81
BASIC COM      10800 08-04-81
BASICA COM      16256 08-04-81
A>_
```

Command Line Interface



Graphical User Interface

Graphical User Interface:

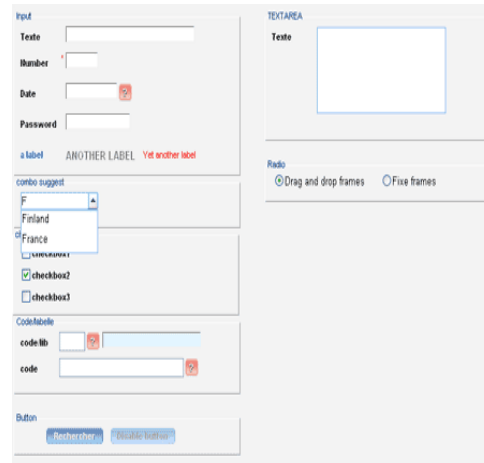
Graphical user interfaces (GUI) are sometimes also referred to as WIMP because they use Windows, Icons, Menus and Pointers. Operators use a pointing device (such as a mouse, touchpad or trackball) control a pointer on the screen which then interacts with other on-screen elements as shown in figure.

Menu Driven:

A menu driven interface is commonly used on cash machines (also known as automated teller machines, or ATMs), ticket machines and information kiosks (for example in a museum) as shown in figure 3.3. They provide a simple and easy to use interface comprised of a series of menus and sub-menus which the user accesses by pressing buttons, often on a touch-screen device



ATM-Menu Driven



Form Based Interface

Form Based:

A form-based interface uses text-boxes, drop-down menus, text areas, check boxes, radio boxes and buttons to create an electronic form which a user completes in order to enter data into a system as shown in figure 3.4. This is commonly used on websites to gather data from a user, or in call centers to allow operators to quickly enter information gathered over the phone.

IV. EXISTING & PROPOSED METHODS

Existing:

In the existing computing world, computer based information technologies have been extensively used to help many organizations, private companies, academic and education institutions to manage their processes and information systems. Information systems are used to manage data. A general information management system that is capable of managing several kinds of data, stored in the database is known as Database Management System (DBMS). Databases are comprehensive element in private and public information systems which are essential in number of application areas. Databases are built with the objective of facilitating the activities of data storage, processing, and retrieval associated with data management in information systems. Due to the progress and in-deep applications of computer

technologies, the widespread applications of web technology in several areas to be accurate, databases have become the repositories of huge volumes of data. In relational databases, to retrieve information from a database, one needs to formulate a query in such way that the computer will understand and produce the desired output.

Proposed:

We propose a general architecture for a database interface and also a real implementation of such a system which can be connected to any database. One of the main characteristics of this interface is domain-independence, which means that this interface can be used with any database. Another characteristic of this system is ease of configuration. The intelligent interface employs semantic matching technique to convert natural language query to SQL using dictionary and set of production rules. The dictionary consists of semantics sets for tables and columns. The shaped query is executed and the results are presented to the user.

NLP is concerned with the development of computational models of aspects of human language processing. There are two main reasons of such development:

- To develop automated tools for language processing.
- To gain a better understanding of human communication.

Building computational models with human language processing abilities requires knowledge of how humans acquire store and process language. It also requires knowledge of the world and of language.

So, the developed natural language interface can easily understand the user defined query in natural language, executes it and return the results respectively

V. CONCLUSION AND FUTURE

ENHANCEMENTS

One of the goals shared by artificial intelligence and database research is developing natural language interfaces that allow a user to query a database for information freely in his native language. Using natural language interface to RDBMS, the end user requests for desired information in natural language (i.e., in English) and can get the desired information. No training is required for retrieving the information. The NLIRDB has been tested successfully for a limited set of sentences. The efficiency of the natural language interface has been demonstrated with the aid of experimental results.

Future work is directed towards general formation of grammar which accepts more number of queries.

REFERENCES

- [1]. Harjit Singh, "Providing Inferential Capability to Natural Language Database Interface", IJECSE, Vol.1, NO.3, pp. 1634 - 1639, 2012.
- [2]. Mrs. Neelu Nihalani, Dr. Sanjay Silakari, Dr. Mahesh Motwani, "Natural language Interface for Database: A Brief review", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, pp. 600 - 608, March 2011.
- [3]. Mrs. Neelu Nihalani, Dr. Sanjay Silakari, Dr. Mahesh Motwani, "An Intelligent Interface for relational databases", IJSSST, Vol. 11, No.1. pp 29 - 34, Jan 2010.
- [4]. Lappoon R. Tang, "Using machine learning approach for natural language interfaces for databases: Application of advanced techniques in inductive logic programming", Journal of Computer Science, Informatics and Electrical Engineering, vol 2, Issue 1, 2008.
- [5]. D.L. Waltz., "An English Language Question Answering System for a Large Relational Database",

Communications of the ACM, Vol.21, No.7, pp. 526 - 539, July 1978.

[6]. Donald P. Mckay and Timothy W. Finin, "The Intelligent Database Interface: Integrating AI and Database systems", In Proceedings of the 1990 National Conference on Artificial Intelligence, pp. 677- 684, 1990.

[7]. Reis, P., Matias, J., Mamede, N. Edite, "A Natural Language Interface to Databases: a New Dimension for an Approach", In Proceedings of 4th International Conference on Information and Communication Technology in Tourism, Edinburgh, Scotland, 1997.

[8]. I. Androutsopoulos, G.D. Ritchie and P. Thanisch, "Natural language interfaces to databases – an introduction", Journal of Language Engineering, Vol.1, No.1, pp 29 - 81, 1995.

[9]. D. Jurafsky, and J. H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition", Prentice Hall, Upper Saddle River, NJ, 2000.

[10]. Bertino, B. Catania, G.P. Zarri, "Intelligent database systems", Reading, Addison Wesley Professional, 2001.

[11]. Alfred V. Aho, Monica S. Lam, Ravi Sethi, and Jeffrey D. Ullman, "Compilers: Principles, Techniques, and Tools", Pearson Education, 2006.

[12]. Stuart Russell, Peter Norvig, "Artificial Intelligence: A Modern Approach", Pearson, 2nd Edition, 2003.