ON AN APPROACH TO STATEMENT OF SQL QUESTIONS IN THE NATURAL LANGUAGE FOR CYBER2 KNOWLEDGE DEMONSTRATION AND ASSESSMENT SYSTEM

Tea Munjishvili, Zurab Munjishvili

Abstract*: The paper focuses on the midterm and final exams in 180 disciplines based on "Cyber1" system developed by the authors. For four years, the exams were held at Sh. Rustaveli State University in the sea-port town of Batumi, Georgia. On the grounds of analyzing the results thereof the authors developed "Cyber2". In the publication they speak about the part the test result analysis plays, its pre-conditions, tasks, means and solutions. They emphasize significance of visualization, generalization and statistical approach to the test results. They also describe method of semantic analysis and algorithms of the questions in the natural language put to a certain section of the database. The programs are in VB. NET.*

Keywords: Analysis of the test results, Semantic analysis

Introduction

Development and use of the knowledge demonstration and assessment systems put forward several tasks, such as making out the optimal timetable of the tests or classes, distribution of disciplines to professors, an automatic understanding of a student's essay, analysis of the test results and the relevant conclusions and recommendations.

As is well known, tests provide important information on the knowledge and demonstration thereof. A test is indispensable for verifying significance of the material imparted to the students, professor's approach to instruction etc.

As a result of analysis of "Cyber1" [Tea Munjishvili, Zurab Munjishvili, 2014] knowledge demonstration and assessment system operated first at Shota Rustaveli State University in the town of Batumi (BSU, Georgia) for four years and then for five years, until 2014/2015 academic year, applied in teaching accounting at the Faculty of Economics and Business at the Tbilisi State University, we developed "Cyber2" [Thea Munjishvili; Zurab Munjishvili , 2015] – the knowledge obtainment, demonstration and assessment software.

Our experience at BSU makes it possible to define conditions **relevant** [Thea Munjishvili; Zurab Munjishvili , 2015] to the functionality and application of “Cyber2” or the other knowledge demonstration and assessment computer systems for that matter:

1. Processing closed and open tests;
2. Setting a task by way of textual or graphic, video or textual and graphic or textual and video information;
3. A prompt to a subject or a test by way of textual or graphic or video information or a combination thereof;
4. Availability of max. three correct answers out of the seven implied ones in a closed test;
5. Giving answers only upon marking the right number of the correct ones and giving “answer” order;
6. Availability of a number answers in open tests;
7. Clicking $q_i \in Q$ mark relevant to $\forall n_i, n_i \in N$ task answer. The mark may be an integer or a decimal positive number;
8. Using words, numbers, sentences or a combination thereof and, also, an abbreviation in an open test;
9. Understanding a statement used in answers in case of desynchronization or insertion of words;
10. In a designational statement and generally answers, writing words in any case and using the wrong variants thereof;
11. Comparison of actual answers to the tasks related to certain subjects, topics or subtopics with the reference value in which case desynchronization, writing words in any case, their insertion or omission will be unacceptable;
12. Formulation the test task by subjects, topics, subtopics and professors;
13. Introducing complex topics or subtopics in the task;
14. Holding examinations by student groups;
15. At the beginning of an examination, arrangement and selection of tests by students in terms of probability;
16. Making out reports after an examination (e.g. a protocol reflecting the course of an examination, examination sheet etc.);
17. Obtainment of analytical information on the results of exams upon completion thereof or at any time;
18. Selection from the database upon putting a question in the natural language;
19. Displaying diagnostic messages, such as using an unknown word, omission of words, numbers etc. during an examination;

20. In case of interruption of an examination for technical etc. reasons, its continuation from the breakpoint;
21. Training: changing the training direction according to the answers given during learning, diagnosing the mistakes and pointing out the ways of prevention thereof.

The necessity of meeting the conditions is detailed in [Thea Munjishvili; Zurab Munjishvili , 2015]. We'll try to prove how important it is to meet the conditions related to obtainment of information on the test result analysis and maintenance of a dialog with the database in the natural language.

Problem Statement

Relevance of reports is unquestionable. For instance: it is necessary to print and familiarize oneself with the test result sheet, test protocol etc. reports, the content of which will have to be predefined.

Apart from that, information search and selection by various parameters is also significant.. Such questions are actually probabilistic and they should be formulated by a management specialist, not a programming professional. The task is related to communication with the databases in the natural language.

Knowledge description methods, such as frames, semantic networks, generating regulations etc. have become widespread in the natural language interactive systems. In most cases a question is put for the purpose of obtainment of the desirable information from the database.

Upon displaying the semantics of a question in the natural language, our task is translation thereof into the one of the database, mainly SQL and then gathering the desirable information.

Along with the uniform algorithmic methods of understanding, the semantics of a question in the natural language, the so-called beyond-the system engineering methods [Thea Munjishvili; Zurab Munjishvili , 2015] relying on the axiom: "***Rules of understanding the semantics can be detected in any problematic area and a logical-semantic model be developed on the grounds thereof***" - are also widespread.

For this purpose, within the knowledge demonstration and assessment "Cyberr2" system, we developed an engineering approach to the semantic analysis of a sentence. It is the so-called "Productive Grammar" method, which is somewhat universal and free from the shortcomings of the tabular suitability principle. The method is discussed in [Z.Munjishvili, 1990], while algorithmic-programming representation of understanding a sentence written in the natural language is detailed in [Thea Munjishvili; Zurab Munjishvili , 2015].

Semantic Analysis Method and Algorithm of Questions to the Database

The objective of our research is understanding a question in the natural language put to the database and generally, realization of semantics of a sentence and its presentation as SQL instruction. We are examining a problematic area, namely, obtainment of the analytical information on the examination results by way of figures and tables. Representing the process as a figure is required by a set of values selected according to a certain criterion(a) marked on X and Y axes.

The content of questions depends on the structure of a certain database but that of any question is typical to SQL instruction. Any question may be considered as a designational sentence made up of more than one word describing searchable items, conditions of selection and classification.

We aim at understanding a sentence at the system entry and translating it into the SQL instruction. There may be omissions or insertion of words or desynchronization in a sentence entered into the system. The words in various cases, synonyms and homonyms may also be used.

The analysis of the structure and content of the questions leads us to the conclusion that in most cases, namely, in the selected problematic area and the task, the logical-semantic model is the basis of presenting examination results as a figure. The model has the following structure: name of the items to be placed on <X axis ; name of those to be placed on a><Y , < selection conditions>.

The selection structure is as follows =:<object><temporal parameters: a semester, an academic year><instruction status><instruction stage>.

Statements are acceptable in terms of questions put for the sake of obtainment the examination results

$$G = \bigcup g_j,\ \mathrm{j} = 1....\mathrm{n}.$$

Let us formulate the G set conditions as follows:

G is a pre-known definite set, with the sentences in the natural language or the order of words being its elements.

1. g_i- marks a sentence with “i” as its conditional number, while $g_{i,\lambda}$ – a word thereof with λ as its conditional number. Then the $g_{i,\lambda}$ used words make up dictionaryL. Composition of G depends on the structure of the system, namely the Cyber 2 database, the knowledge demonstration and assessment by subjects and the appraisal system.
2. Any two elements of G differ at least by a single word;

$$\forall(g_i, g_{i_0} \in G) \Rightarrow (g_{i_0} \setminus (g_{i_0} \cap g_i) \neq \{\} \wedge g_i \setminus (g_{i_0} \cap g_i) \neq \{\})$$

3. Any pair may contain the words similar in content.
4. In Cyber 2 knowledge demonstration and assessment system, a certain selection condition (action) corresponds to a question or a phrase - ψ_i , i.e. according to $g_i \rightarrow \psi_i$. logical – semantic model, the selection condition may contain one or more words. ψ_i may be a word, a sequence thereof, a set of symbols etc. Consequently, the G set $g_i \in G$ elements are reflected in G^* set of operations $\psi_i \in G^*, f : G \rightarrow G^*$, $f(g_i) = \psi_i$.

The desired results are obtained if $\forall g_i \in G$, then g_i may be regarded as the product, the $a_k \in L$ words included in g_i as conditions and the names of objects to be placed on X and Y axes, while ψ_i as selection conditions corresponding to g_2 as operation. In this sense, “the analysis” of a question – statement in the selected problematic area may be brought down to the formation of a productive system, the dictionary arrangement and the SQL instruction relevant to the incoming facts.

As exemplified by Cyber 2, method of the semantic analysis and the algorithm is based on those of understanding the answers to the open tests. The method and algorithms are detailed in [Tea Munjishvili; Zurab Munjishvili , 2015].

In Cyber 2, two tables reflecting the examination results, such as tbmain (information on a student’s activities, the examinations) and the tbmosasmeni (list of the students to do a repeated course of the same subject) form the basis of presenting them by way of a figure and understanding questions

By scores, the tbmain table contains the result –related information detailing a student’s activities from the very beginning of an ‘i’’ course (an activity, a midterm exam, the final, the repeated exam) and the order of appraisals.

The principles presented by way of the productive ones and formulated according to the logical-semantic model of the questions form the algorithmic basis of understanding them In the case in question, the structure of the productiveprinciples is as follows:

Conditions: name of the object to be placed on X axis, name of the object to be placed on Y axis

action: selection condition(s)

In this case, the following principles were formulated:

P1: Subjects, appraisals ⇨ <a student> AND <optional selection conditions >

P2: studens, appraisals,⇨ <subject> AND <<optional selection conditions

P3: subjects, quantity ⇨<appraisal> AND <<<optional selection conditions >

P4:groups, quantity ⇨<appraisal> AND subject AND <optional selection conditions>

In this case, the optional selection conditions are: appraisal type, number, semester, academic year, instruction stage and status.

A dictionary made up of three ones (table) $L = L_1 \bigcup L_2 \bigcup L_3$.

L_1 is the structure of an entry

<L_1 a dictionary entry>:=<word $a_k \in g_{i,\lambda}$><$a_k \in g_{i,\lambda}$ a word or its wrong variant or a synonym>

L_1 the words used in the productive principles (textual data) are entered into the dictionary

L_2 the structure of an entry:

<L_2 < a dictionary entry>:=the $a_k \in g_i$ morphological root of the right version of words in the dictionary>< name of the word by the base table>

L_2 dictionary is compiled along with L_1. The system enters an L_1 word unchanged The administrator edits it and creats its morphological root. There are no wrong variants in the dictionary.

L_3 The dictionary structure is as follows

<L_3 dictionary entry >:=<naming the selection parameter by the base table ><a sentence made up of the $a_k \in g_i$ >words in L_1 dictionary.

L_3 dictionary is compiled by the administrator

Understanding a question and its presentation by way of SQL instruction: /fig. 1/

Understandingly, for the discussed problematic area, there is no need to formulate a question in the natural language and apply the specified complex pattern in order to understand its semantics. In the case in question, the task may be solved by a simple selection from the lists or another method. The objective of the article is to highlight “a forgotten” problem: a dialog with databases in the natural language.

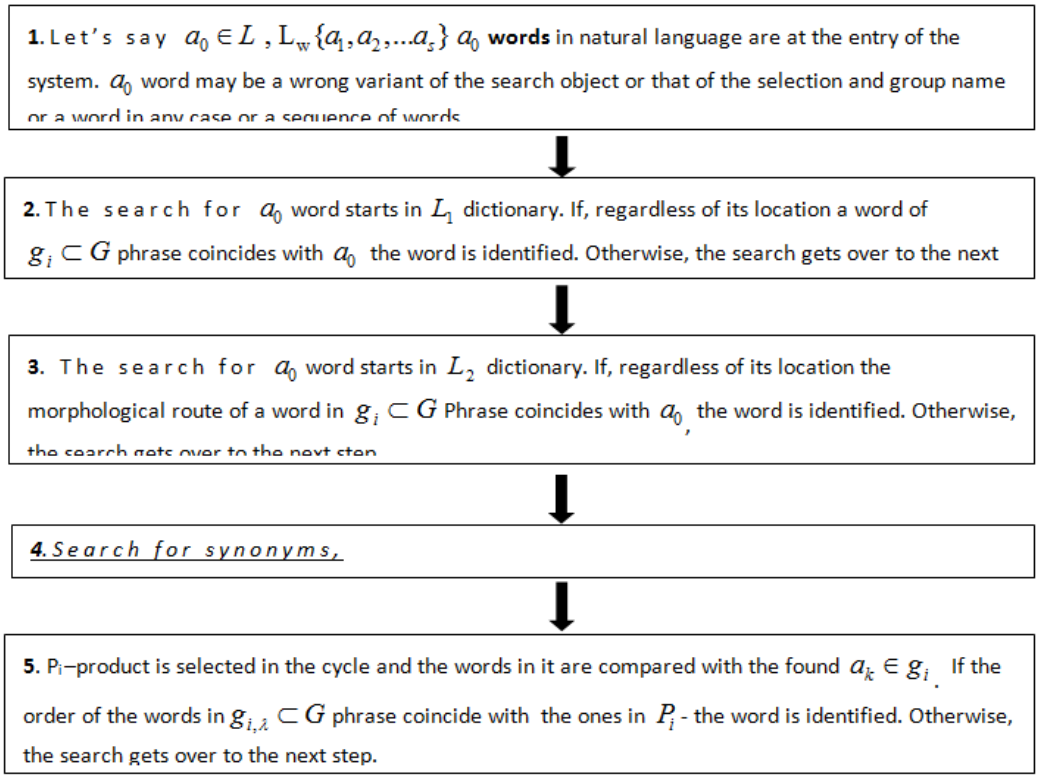

Fig. 1. Understanding and Presenting a Question by Way of SQL Instruction.

Conclusion

1. By means of generalization of theoretical issues and on the grounds of handson experience, we developed the conditions of functionality and application of the knowledge demonstration and appraisal computer systems, namely obtainment of the analhtical information on the test results by way of diagrams, to this end, formulation of questions in the natural language and understanding a question in case of desynchronization, insertion of words, putting them into any case or using a wrong variant of a word;
2. A question is understood by means of the logical-semantic model and application of the knowledge demonstration method by means of productive principles widespread in the artificial intellect;
3. After further research, the described approach to understanding questions to the database presented in the diagrams that reflect the analytical data regarding the examination results may be applied to the other problematic areas, as well.

Bibliography

[Tea Munjishvili, Zurab Munjishvili, 2014] Tea Munjishvili, Zurab Munjishvili. Knowledge Demonstration and Assessment System “Cyber1”, international Journal "Information Technologies & Knowledge" Volume 8, Number 3, 2014, pp. 271-279.

[Munjishvil T., Munjishvil Z., Nakashidze V, 2014] Munjishvil T., Munjishvil Z., Nakashidze V. System of knowledge revealing and rating – “Cyber 2”. 9th MIBES ANNUAL INTERNATIONAL CONFERENCE 2014 THESSALONIKI, GREECE, 30/5– 1/6 .CD ISBN# 978-960-93-6161-3. pp. 111-121.

[Tea Munjishvili; Zurab Munjishvili , 2015] Tea Munjishvili; Zurab Munjishvili. "The system of Discovery and Estimation of Knowledge “Cyber2””, Scholars' Press, Saarbrücken HRB 18918. Published on: 2015-01-15 Number of pages: 108. Book language: English. ISBN-13: 978-3-639-76094-1.

[Z.Munjishvili, 1990] Z.Munjishvili. Problem-oriented method of semantic analysis for sentence of natural language. Collection of Knowledge, dialogue, decision, Kiev, Ukraine, “Naukova Dumka”, 1990.

Authors' Information

Tea Munjishvili *– Iv.Javakhishvili Tbilisi State University, 0129-Tbilisi, Georgia. Associated Professor, e-mail: tmunjishvili@gmail.com*

Major Fields of Scientific Research: Practical information systems research, Logical-probability models while assessing risk by enterprises.

Zurab Munjishvili *– Interbusiness Acadeny, 8, Shalva Djaparidze str.*

Tbilisi, 0178, Georgia. Associated Professor, e-mail: zurab_ztm@ rambler.ru

Major Fields of Scientific Research: General Practical information systems research, Creator for knowledge E-Systems.