

Лінгвістичні технології інформаційної діяльності

Analysis of a user query in natural language

Iana Potochniak¹, Alexei Kungurtsev²,

Department of System Software, Odessa national polytechnic university,
boulevard Shevchenko, 1, Odessa, UKRAINE,
E-mail: 1. yana_onpu@mail.ru, 2. abkun@te.net.ua

Abstract – The article considers questions of creation DB model on natural language and analysis of the user's requests in natural language for communication end-user with information system.

Keywords – database model, object, attribute, user request., natural language.

I. Introduction

Since the relational database (RDB) retain their dominant position, as a mechanism for data storage and processing in information system (IS), drafting requests for information from IS remains the lot of professionals. Thus, the problem of communication is not qualified users IS is becoming increasingly important [1].

II. Creation DB model

User query in natural language (NL) enters the query analyzer, which parses sentences and allocates their grammatical characteristics. Database model (DBM) has verbal description of the database structure and allows defining availability of the necessary information in the database to generate user response [2]. If user request can be performed, the queries generator produces a corresponding SQL-query. For each request (made or contain errors) response synthesizer generates a response in NL containing the result of the analysis of the query.

Are invited to submit a DBM as a set of objects:

$$M = \langle \{O_i\} \rangle, \quad (1)$$

Where each object is described by a tuple:

$$O_i = \langle Nt_i, TNo_i, \{A_{ij}\}, \{R_{ij}\} \rangle, \quad (2)$$

Where Nt_i – table name; TNo_i – object description; A_{ij} – object attribute; R_{ij} – link from object O_i to object O_j .

Each attribute is described by a tuple:

$$A_{ij} = \langle Nc_{ij}, TNa_{ij} \rangle, \quad (3)$$

Where Nc_{ij} – table column name; TNa_{ij} – description of attribute.

Proposed description of the object represented as a tuple:

$$TNo_i = \langle term_o_i, Syn_i, prop_i \rangle, \quad (4)$$

Where $term_o_i$ – object term (a word or phrase that defines the essence of the object); Syn_i – synonymously object which can be represented as, $Syn_i = \{synonym\}$ (a word or phrase that is close in meaning to the term in the subject area); $prop_i$ – some characteristic of the object.

Describing of attribute invited to submit as a tuple:

$$TNa_{ij} = \langle term_a_{ij}, Syn_{ij}, prop_{ij}, \{range_{ij}\}, \{agg_{ij}\}, \{ac_{ij}\} \rangle, \quad (5)$$

Where $term_a_{ij}$ – attribute term; Syn_{ij} – attribute synonym, which may be represented as: $Syn_{ij} = \{synonym\}$; $prop_{ij}$ – attributes characteristic; $range_{ij}$ – limitation to the values that attribute can take (less, more, equals); agg_{ij} – aggregation function, applicable to the attribute (maximum, minimum, average, ...); ac_{ij} – action allowed on attribute (find value, order, ...).

III. Analysis of the user's query

Based on a comparison of the object description from DBM (2) and representation of the user request (6), proposed to formulate a query to select data from the DBM and the user response in NL. If result of analysis does not allow to generate a request, the user receives a response indicating the reason in NL. If all the information found in the DBM, the response is generated in NL.

It is proposed the request Q of the user represented as:

$$Q = \langle \{to_i\}, \{ta_{ij}\}, \{Z_{Aij}\}, \{r_{ij}\}, agg_{ij}, ac_{ij} \rangle, \quad (6)$$

Where to_i – objects term of the query; ta_{ij} – term attribute for the term of the request object Q ; Z_{Aij} – attribute value; r_{ij} – range of attribute values ta_{ij} ; agg_{ij} – aggregation function applied to the attribute ta_{ij} ; ac_{ij} – action on the attribute ta_{ij} in the request.

Conclusion

Proposed DBM displays information stored in RDB as natural language constructions. It is proposed analyze user queries in NL and generate a response in NL using DBM.

References

1. M. Bates, "Models of natural language understanding". Proceedings of the National Academy of Sciences of the United States of America, vol. 92, No. 22, pp. 9977–9982, Oct. 24, 1995.
2. О.Б. Кунгурцев, Я. В. Поточняк, Создание модели РБД для анализа сообщений на естественном языке в информационных системах, Матеріали IV міжнародної конференції студентів і молодих науковців «Сучасні Інформаційні Технології 2014»: тези допов. – Одеса, 22-26 квітня 2014 р.