УДК 004.415(045)

L.N.Badorina, post-graduate

METHOD OF THE RELEVANCE DEGREE ESTIMATION OF THE TEXT ANSWER IN COMPUTER TRAINING SYSTEMS.

НАУ, кафедра комп'ютеризованих систем управління, vada@ukr.net

The article proposes the method of a trainee's text answer analysis on the task of the open type allowing establishing conformity with terms of a subject domain.

Предложен метод оценки степени релевантности текстового ответа в компьютерных обучающих системах.

Introduction

Introduction of progressive forms of training and development of modern information technologies creates the necessity of automated assessment of the student's knowledge. Great value for automated educational systems have models of assessing answers not in the form of chosen variants, but in the form of a free text of any length with synonym concept estimation. A special urgency gets the problem of development of answers' analysis model on the task of the open type, demanding to enter from the keyboard the certain formulation of this or that term of a subject domain. There is an objective necessity of transition to computer testing of students' knowledge. Thus, on the foreground rises the problem of automatic assessment of students' knowledge. This problem is simple enough, if the student is offered to choose one or more right answers from a set of variants, but it becomes considerably difficult, if the procedure of testing provides a developed answer in any form, that is with his or her own words in natural language. In the latter case it is possible to appreciate the student's knowledge only by comparative text analysis of the answer with the set standard reference text and to assess their relevance. Thus, all word-forms, terms of the subject domain and grammatical structures of the statement should be considered and assessed with the use of all possible synonyms. [1].

This approach is based on the assumption, that knowledge of a subject domain is defined by its thematic dictionary, by the skill to correctly place the words, that is to give a correct formulation of concepts of a subject domain which has semantic concurrence or semantically close value to the interpretation of the given concept of the thesaurus of a subject domain. It is considered, that the actual formulation of the term of a subject domain can be not in the form of unequivocal sequence of the subject area language (firm formulation), but using close synonym concepts and terms (in the trainee's mind).

At the automated knowledge control of the terminology of a subject domain the try to solve a task to compare two definitions of one term: the definition given by the teacher (reference definition), and the definition given by a trainee (answer). The result of the comparison should be conformity with these definitions. The development of this method of standard reference definition and answer analysis is the purpose of the given article. [3].

Statement of the problem

To estimate the relevance degree of standard reference definition and the answer of a trainee it is necessary:

To establish mutual monosymantic synonymic conformity with terms of standard reference definition and answer;

To calculate the value of a relevance parameter of standard reference definition and answer.

Statements are considered as a set of terms. Thus, the standard reference definition should be considered as a set of base terms, and the answer should be considered as a set of terms *t,* for each of them it is necessary to find a corresponding base term *e.* The search of conformity of a base term and an answer term proposes the definition of function $e = \varphi(t)$ and the calculation of the size of synonymic conformity $k = \theta(e,t)$. Thus, the pair $< e,k >$ will allow characterizing a term t in relation to a term-standard e. It means conformity of answer terms with base terms. [2].

Let *A* be a set of standard-term definition, *B*- a set of answer terms.

Then the description of standard definition and answer is as follows:

$\grave{A} = \{\mathring{a}_1, \mathring{a}_2, ..., \mathring{a}_i, 1 \le i \le N\}$;

$B = \{t_1, t_2, ..., t_i, 1 \le i \le M\}$;

N – quantity of terms of standard definition;

M – quantity of answer-terms.

To calculate the conformity with terms of standard definition and answer it is necessary to characterize terms t according to terms-standards e. We are going to define synonymic conformity of answer-terms with standard definition.

Comparative analysis of terms

The purpose of synonymic conformity is to bring into line the terms containing in the answer to terms to the standard definition. In the result we will have the following ratio between sets A and B:

1. $A = B$ – the answer of a trainee completely coincides with the standard definition.

2. $A \subset B$ – the answer of a trainee contains all terms from the standard definition and superfluous terms.

3. $B \subset A$ – the answer of a trainee partially correspond the standard definition, but the answer lacks some base terms.

4. $A \cap B \neq \varnothing$ – the answer of a trainee and the standard definition have identical terms.

5. $A \cap B = \varnothing$ – the answer of a trainee does not correspond to the standard definition.

Interaction of a trainee with the monitoring system of knowledge assumes that there is a question in the system about the definition of some term E of a considered subject domain. In the answer the trainee should generate the definition being analogous to the formation of standard definition of this term.

To assess the knowledge of a trainee it is necessary to establish conformity with terms of the standard definition and answer-terms and on the basis of this conformity to calculate a parameter of relevance of standard definition and answer, being based on the concept of synonymic conformity of terms.

Comparing separate terms of standard definition and answer there can be following situations which should be solved. [4].

I. One term of standard definition corresponds to only one base term of the answer.

It can be presented as a biactive display between sets A and B:

$$\alpha : B \to A, a_i = \alpha(b_j), a_i \in A, b_j \in B.$$

In this case between terms of standard definition and answer there is a mutual monosymantic tie. All relations between terms of standard definition and answer should be brought to a similar kind.

II. One term of standard definition corresponds to some various answer-terms.

In this case there exist the intersected sets $\{a_i, b_j\} \cap \{a_i, b_b\} \neq \varnothing$. Each of these sets is characterized by the function θ designating a synonym parameter of the terms

$$a_i, b_j : k_m = \theta(a_i, b_j),$$

To achieve the aim it is necessary to remove from consideration one of the intersected sets by the following rule:

1. If $\theta(a_i, b_j) \rhd \theta(a_i, b_b)$, it is possible to remove the set $\{a_i, b_b\}$ as the term b_j is the closest synonym to the term a_i, and to use for further processing a parameter k_h, which describes numerical value of synonymic term c_i and b_j.

2. If $\theta(a_a, b_j) = \theta(a_i, b_j)$, it is possible to remove any of sets $\{a_a, b_j\}$ or $\{a_i, b_b\}$, as terms b_j and b_b are equally close synonyms to the term a_i and to use for further processing the parameter k, describing numerical value of synonymic term c, and the remained term of actual definition.

3. If $\theta(a_i, b_j) \prec \theta(a_i, b_b)$, it is possible to remove the set $\{a_i, b_j\}$ as the term b_b is the closest synonym to term a_i, and to use parameter k_a, for the further processing describing numerical value of synonymic term c_i, b_b.

III. Several various terms of standard definition correspond to the same term of actual definition.

In this case there exist the intersected sets $\{a_a, b_j\} \cap \{a_i, b_j\} \neq \varnothing$. To achieve the stated aim it is necessary to remove from consideration one of the intersected sets according to the following rule:

1. If $\theta(a_a, b_j) \lhd \theta(a_i, b_j)$, it is necessary to remove set $\{a_i, b_j\}$ and k_h. To use the parameter k for further processing which describes the numerical value of synonymic term c_a and b_j.

2. If $\theta(a_a, b_j) = \theta(a_i, b_j)$, it is possible to remove any of sets $\{a_a, b_j\}$ or $\{a_i, b_j\}$ and to use parameter k, for the further processing which describes the numerical value of synonymic term c_i and the remained term of actual definition.

3. If $\theta(a_a, b_j) \lhd \theta(a_i, b_j)$, it is possible to remove the set $\{a_a, b_j\}$ and to use parameter k, for further

processing which describes the numerical value of synonymic term c_i and b_j.

In this case it is impossible to use proportionally to calculate both numerical parameters because of the peculiarities of the result in the system of assessment. There is a methodological aspect of the chosen decision: if semantics of the sentence is as follows, so it is necessary to use some terms, but the trainee used instead of them only one, and to his opinion, the generalizing term, hence, in according to his consciousness these terms are poorly distinguished, and it is necessary to make special methodical job.

IV. Some various terms of standard definition have some common synonyms.

In this case to achieve the aim it is necessary:

1. To choose the set $\{a_i, b_j\}$, characterized by the highest numerical parameter $k_h = \theta\{a_i, b_j\}$.

2. To remove from the further consideration all other sets in which there are the chosen elements a_i and b_j.

3. Among other sets to continue choosing and removing the sets with the maximal parameter $k_a = \theta\{a_b, b_a\}$according to the same rule until all intersected sets are not found.

In case if there are simultaneously several sets with identical maximal numerical parameter k, it is necessary to choose only one of them and again to make the analysis.

As a result all intersected sets are removed. It means that mutual monosymantic conformity with significant terms of standard and actual definitions is established.

Thus, mutually monosymantic conformity with terms of standard definition and answer is established. The result of synonymic conformity of terms, which is considered at the estimation of knowledge assessment, is the ratio between sets A and B. We are choosing the no intersected sets, which participate in the estimation of knowledge. The set of no intersected sets are considered, which have the biggest value of parameter k = 1 or k = 0,8.

Conclusions

The developed method of a trainee's answer analysis on the task of the open type allows to establish conformity with the terms of a subject area, used in standard definition and answer of the trainee. The result of conformity is the ratio between sets. The analysis of the results of synonymic conformity of terms of standard definition and answer give the possibility to account the numerical parameter of relevance.

References

1. *Лантух Д. А.* Математичний апарат системи контролю коректності текстів // Технології Microsoft у теорії і практиці програмування: Тези доп. конкурсів-конф. – СПб: С-ПбГпу, 2004. –122 с.
2. *Цаленко М. Ш.* Моделирование семантики в базах данных. – М.: Наука, 1989. – 288 с.
3. *Hopfield J. J.* Neural networks and physical systems with emergent collective computational abilities // Proc. Natl. Acad. Sci.79, 1982. –P. 2554–2558.
4. *Бадьоріна Л. М.* Метод оцінювання довільних відповідей у комп'ютерних системах тестування знань // Математичні машини і системи.–2006. – №4. – С.138–144.

Стаття надійшла до редакції 04.02.07.