

УДК 004.415 (045)

Л.М. Бадьоріна, к.т.н.

**СИНОНІМІЧНА РЕЛЕВАНТНІСТЬ ТЕКСТОВИХ ВИЗНАЧЕНЬ
В ІНТЕЛЕКТУАЛЬНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ**Національний авіаційний університет
E-mail: vada@ukr.net

Запропоновано методику оброблення відповідей того, кого навчають, на завдання відкритого типу, яка дозволяє зв'язувати терміни предметної галузі.

автоматизований контроль знань, інформаційні технології, синонім**Постановка проблеми**

Упровадження інтелектуальних інформаційних технологій у педагогічну практику потребує додаткового використання нетрадиційних для педагогіки методів досліджень, моделювання і кваліметрії. Вони забезпечують отримання кількісних оцінок параметрів процесів автоматизованого навчання.

Аналіз світових тенденцій розвитку вищої освіти, а також суперечностей, що стримують його розвиток, дозволяють сформулювати наукову проблему, що полягає в розв'язанні суперечності між можливостями традиційних методів навчання і тенденціями розвитку сучасного суспільства, що ґрунтується на знаннях і інтелектуальних інформаційних технологіях, та потребує системного підходу до автоматизації навчання фахівців.

Мета дослідження – підвищення ефективності підготовки фахівців шляхом використання інтелектуальних інформаційних технологій у процесі автоматизованого навчання студентів.

Сучасні інформаційні технології в освітній сфері створюють необхідність автоматизованого контролю знань студентів.

Велике значення для автоматизованих систем освітнього призначення мають моделі оцінювання відповідей не через обрані варіанти, а через текст довільного обсягу з урахуванням синонімічно близьких понять [1].

Актуальне завдання розроблення моделі аналізу відповідей на завдання відкритого типу, що вимагає введення з клавіатури визначення формулювання того або іншого терміна предметної галузі.

Оскільки загальна кількість слів і словоформ у будь-якій природній мові обмежена через загальні об'єктивні принципи побудови самої природної мови, можна стверджувати, що з математичного погляду кожна така множина, так само, як і його підмножина, є кінцевою множиною, тобто має кінцеву кількість складових його елементів.

Лінгвістичне значення речення є відношенням деяких семантичних компонентів, відвернених одночасно від значення структурної схеми речення і від лексичних значень тих слів, що зайняли позиції її компонентів і поширили собою речення [2].

У реченні формується нове, більш конкретне, ніж семантика його зразка, але теж абстрактне значення, що називається семантичною структурою.

Семантична структура речення – це його абстрактне значення з урахуванням мовних одиниць, що являють собою відношення семантичних компонентів, сформованих взаємною дією граматичних і лексичних значень членів речення [3].

Кількісна (сміслова) оцінка довільного речення враховує морфемний аналіз, який визначає, з яких морфем складається слово. Тобто визначено всі значущі частини слова (конкатенація окремих фрагментів: основ слів, префіксів, суфіксів).

Сучасні наукові дослідження лінгвістів [4] надають ряд підходів, що дозволяють формалізувати подання кожної лінгвістичної одиниці в зручному для машинного оброблення вигляді.

У разі автоматизованого контролю знань термінології предметної галузі ставиться завдання порівняння двох визначень одного терміна:

– визначення, пропонуваного викладачем (еталонного визначення);

– визначення тим, кого навчають (відповіді).

Оцінювання правильності текстової відповіді ґрунтується на методі абсолютної збіжності відповіді з одним з еталонів, що зберігаються в базі даних тестової системи. Оскільки визначення терміна формується через систему базових понять (термів), кожне з яких має своє визначення, пропонується для обчислення показника релевантності відповідей на завдання відкритого типу використати кількісні показники синонімії термів предметної галузі.

Аналіз автоматизованих систем контролю знань показав, що широко застосовані методи оброблення текстової відповіді ускладнюють можливість поелементного аналізу тексту відповіді й

еталонної відповіді, що зберігається в базі даних тестової системи, і накладають обмеження на формат завдання відкритого типу [5]. Потрібно побудувати аналітичну модель розрахунку показника релевантності.

Релевантність текстових визначень

Для текстових визначень використовують модель обчислення значення показника відповідності значущих термів еталонного визначення і відповіді [1].

Нехай $W_{A,B}$ – показник релевантності еталонного визначення і відповіді тих, кого навчають. Його значення обчислюється за формулою

$$W_{A,B} = f(\omega, \eta), \quad (1)$$

де ω – функція числових значень показника синонімії значущих термів в еталонному і фактичному визначеннях;

η – функція кількості значущих термів в еталонному визначенні та відповіді.

Параметр ω є функцією трьох змінних:

– коефіцієнта синонімії термів $K_{A,B}$ еталонного і фактичного визначень;

– кількості значущих термів в еталонному визначенні N і кількості значущих термів у фактичному визначенні M :

$$\omega = \theta(K_{A,B}, N, M).$$

Кількість термів в еталонному визначенні дорівнює множині A , а кількість термів у відповіді – множині B до оброблення:

$$N = \|A\|, M = \|B\|.$$

У разі $N = \text{const}$ параметр ω має такі властивості:

– зі збільшенням $K_{A,B}$ значення ω зростає.

– зі збільшенням кількості термів відповіді значення ω зменшується;

– якщо $\frac{\sum_{e \in B} k_e}{M} = 1$, то $\omega = \omega_{\max}$.

Загальний параметр $K_{A,B}$ дорівнює сумі максимальних значень коефіцієнтів відповідності термів еталонного і фактичного визначень:

$$K_{A,B} = \sum_{e \in B} k_e$$

На підставі наведених властивостей отримано формулу обчислення коефіцієнта ω :

$$\frac{\sum_{e \in B} k_e}{M} = \omega, \quad (2)$$

Функція η задає залежність показника релевантності $W_{A,B}$ від кількості термів в еталонному визначенні й відповіді та має такі властивості:

– якщо $M = N$, то $\eta = \eta_{\max}$.

– функція η симетрична відносно свого максимального значення.

– якщо $|A - B| \rightarrow \pm\infty$, то $\eta \rightarrow 0$.

На підставі цих властивостей отримано формулу обчислення коефіцієнта η :

$$\varphi(\Delta) = \frac{h}{\sqrt{\pi}} * e^{-h^2 \Delta^2},$$

$$h > 0. \quad (3)$$

Параметр h і діапазон значень параметра Δ підбирають експериментально, і їх можна варіювати.

Використано такі параметри:

$$\Delta = 0,1 * (M - N); \quad h = 2,050.$$

Такі значення параметрів забезпечують значення максимуму функції в точці (0; 1) і збіжність до коефіцієнта $\varphi = 0,01$ за $M - N = 12$.

Підставивши формули (2), (3) у рівняння(1), дістаємо

$$W_{A,B} = \omega \eta = \frac{\sum_{e \in B} k_e}{M} \frac{h}{\sqrt{\pi}} e^{-h^2 \Delta^2}. \quad (4)$$

Обчислимо за формулою (4) значення $W_{A,B}$ для наведеного прикладу, взявши значення k_e з таблиці.

Показники синонімії термів еталонного визначення і відповіді

Фактичне визначення	Еталонне визначення									
	Дан	Управл	Компонент	Систем	Обробк	Інформац	Цілий	Реаліз	Визначен	Алгоритм
Сукупн				0,8						
Дан	1		0,5			0,7				
Послідов				0,2						
Операц			0,6							
Обробк		0,3			1			0,5		
Інформац	0,7					1				
Реаліз					0,3			1		
Визначен									1	
Алгоритм				0,6						1

На основі таблиці отримано сукупність множин. У таблиці наведено тільки ті з них, що становлять інтерес для подальшого розгляду:

{дан,дан};
 {дан,інформац};
 {управл,обробк};
 {компонент,дан};
 {систем,сукупн};
 {систем,послідовн};
 {систем,алгоритм};
 {обробк,обробк};
 {обробк,реаліз};
 {інформац,дан};
 {інформац,інформац};
 {реаліз,обробк};
 {реаліз,реаліз};
 {визначен, визначен};
 {алгоритм, алгоритм}.

Одержимо сукупність функцій $k_n = \theta(a_i, b_j)$, що позначають числові показники синонімії термів еталонного і фактичного визначень:

$k_1 = \theta(a_1, b_2) = \text{синонім}\{\text{дан, дан}\}=1;$
 $k_2 = \theta(a_1, b_6) = \text{синонім}\{\text{дан,інформац}\}=0,7;$
 $k_3 = \theta(a_2, b_4) = \text{синонім}\{\text{управл,операц}\}=0,2;$
 $k_4 = \theta(a_1, b_5) = \text{синонім}\{\text{управл,обробк}\}=0,8;$
 $k_5 = \theta(a_3, b_2) = \text{синонім}\{\text{компонент,дан}\}=0,5;$
 $k_6 = \theta(a_3, b_2) = \text{синонім}\{\text{компонент,операц}\}=0,6;$
 $k_7 = \theta(a_4, b_1) = \text{синонім}\{\text{систем, сукупн}\}=0,8;$
 $k_8 = \theta(a_4, b_3) = \text{синонім}\{\text{систем, послідовн}\}=0,2;$
 $k_9 = \theta(a_4, b_9) = \text{синонім}\{\text{систем, алгоритм}\}=0,6;$
 $k_{10} = \theta(a_5, b_5) = \text{синонім}\{\text{обробк, обробк}\}=1;$
 $k_{11} = \theta(a_5, b_7) = \text{синонім}\{\text{обробк, реаліз}\}=0,3;$
 $k_{12} = \theta(a_6, b_2) = \text{синонім}\{\text{інформац, дан}\}=0,7;$
 $k_{13} = \theta(a_6, b_6) = \text{синонім}\{\text{інформац, інформац}\}=1;$

$k_{14} = \theta(a_8, b_5) = \text{синонім}\{\text{реаліз, обробк}\}=0,5;$

$k_{15} = \theta(a_8, b_7) = \text{синонім}\{\text{реаліз, реаліз}\}=1;$

$k_{16} = \theta(a_9, b_8) = \text{синонім}\{\text{визначен, визначен}\}=1;$

$k_{17} = \theta(a_{10}, b_9) = \text{синонім}\{\text{алгоритм, алгоритм}\}=1;$

З наведених даних бачимо, що деякі множини $\{a_i, b_j\}$ перетинаються в точці a_i або b_j .

Наприклад, множина

$\{a_1, b_2\} = \{\text{дан,дан}\}$

перетинається з множиною

$\{a_1, b_6\} = \{\text{дан, інформац}\}$

у точці

$a_1 = \text{дан}: \{a_1, b_2\} \cap \{a_1, b_6\} = \{a_1\}$

Це означає, що терму «дан» еталонного визначення відповідають два терми фактичні визначення «дан» і «інформац» [3].

За умовою потрібно вибрати один із двох термів. Для цього порівнюємо їхні функції:

$k_n = \theta(a_i, b_j);$

$k_1 = \theta(a_1, b_2) = 1;$

$k_2 = \theta(a_1, b_6) = 0,7;$

$\theta(a_1, b_2) > \theta(a_1, b_6).$

Множина $\{a_1, b_6\}$ з подальшого розгляду виключається і не бере участі у процесі формування оцінювання знань [6].

Узявши значення k_e з таблиці, підставимо у формулу обчислення параметра ω , використовуючи під час цього множину термів відповіді і кількість значущих термів у фактичному визначенні:

$$\omega = \frac{\sum_{e \in B} k_e}{M}; \quad \varepsilon = \frac{7,4}{9};$$

$$R(C, D) = \frac{2,050}{\sqrt{\pi}} e^{-(2,050)^2 \cdot 0,11} \frac{7,4}{9} = 0,912 \approx 0,9.$$

Розрахувавши за формулою (4) значення $K_{A,B}$ для еталонного визначення, одержимо $R_{A,B} = 1$.

Доведено, що значення показника відповідності значущих термів еталонного і фактичного визначень $R_{A,B} = 1$ прийнятне для практичного застосування.

Висновки

Розроблена аналітична модель оброблення відповідей тих, кого навчають, на завдання відкритого типу дозволяє зв'язувати еталонні визначення і відповіді в просторі термінів предметної галузі, що дає можливість аналізувати відповідь тестуючого, даний у вигляді вільного тексту довільної довжини, при цьому враховуючи значення показника релевантності, який бере участь у формуванні оцінювання знань.

Література

1. *Абросов В.М.* Классификация критериев смысловой ответственности / В.М. Абросов, Е.Ю. Хрусталев // Наук.-техн. инфор. Сер. 2. – М.: ВИНТИ, 1979. – № 4.

2. *Цаленко М. Ш.* Моделирование семантики в базах данных / М.Ш. Цаленко. – М.: Наука, 1989. – 288 с.

3. *Hopfield J.J.* Neural networks and physical systems with emergent collective computational abilities / J.J. Hopfield // Proc. Nat. Acad. Sci. 1982. – N 79. – P. 2554–2558.

4. *Kinshuk A* conceptual framework for Internet based intelligent tutoring systems / Kinshuk, A. Patel // Knowledge transfer (ed. A.Behrooz), рAce. – London. Vol. II. – P. 117–124.

5. *Авенсов В.С.* Научные основы тестового контроля знаний / В.С. Авенсов. – М.: Исслед. центр 1994. – 135 с.

6. *Бадьорина Л.М.* Метод оцінювання довільних відповідей у комп'ютерних системах тестування знань / Л.М. Бадьорина // Математичні машини і системи. – 2006. – № 4. – С. 138–144.

Стаття надійшла до редакції 21.01.10.

Л.Н. Бадерина

СИНОНИМИЧЕСКАЯ РЕЛЕВАНТОСТЬ ТЕКСТОВЫХ ОПРЕДЕЛЕНИЙ В ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Национальный авиационный университет

автоматизированный контроль знаний, информационные технологии, синоним

Введение прогрессивных форм обучения и развитие современных информационных технологий создает потребность в автоматизированной оценке знания студента. Автоматизированные образовательные системы представляют собой модели оценки ответов не в форме избранных вариантов, а в форме текста любой длины с учетом синонима. Анализ ответов моделируют на задаче открытого типа. Студенту предлагается выбрать правильный ответ из набора вариантов. Формы слова, термины, утверждения ответов необходимо оценивать с применением всех возможных синонимов. Автоматизированный контроль знаний терминологии предметной области должен позволить сравнить определение, данное преподавателем, с определением, данным стажером. Результат сравнения должен соответствовать этим определениям. В статье рассмотрен метод стандартного справочного определения и анализа ответа.

Lyubov M. Badorina

THE SYNONYMIC RELEVANCE DEGREE ESTIMATION OF THE TEXT ANSWER IN MODERN INFORMATION SYSTEMS

National Aviation University

automated educational control, information technologies, synonym

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. 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. 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.