

УДК 004.415 (045)

Л.М. Бадьоріна, асп.

## АНАЛІТИЧНА МОДЕЛЬ СИНОНІМІЧНОЇ РЕЛЕВАНТНОСТІ ТЕКСТОВИХ ВІДПОВІДЕЙ У КОМП'ЮТЕРНИХ ТЕСТОВИХ СИСТЕМАХ

НАУ, кафедра комп'ютеризованих систем управління  
E-mail: vada@ukr.net

*Запропоновано аналітичну модель обчислення показника релевантності значущих термів еталонного визначення і відповіді. Наведено приклад розрахунку функції числових значень показника синонімії значущих термів.*

*The article describes analytical method by comparison analysis of the textual answer to the ideal text and identifying a concordance.*

### Постановка проблеми

Інтеграційні процеси впровадження засобів телекомунікації, комп'ютеризації людської діяльності становить безліч нових завдань у науковій галузі, яка знаходиться між напрямками комп'ютерних технологій і лінгвістикою.

Розвиток сучасних інформаційних технологій в освітній сфері створює необхідність автоматизованого контролю знань студентів. Велике значення для автоматизованих систем освітнього призначення мають моделі оцінки відповідей не у вигляді обраних варіантів, а у вигляді вільного тексту довільної довжини з урахуванням слів-синонімів. Актуальне завдання розробки моделі аналізу відповідей на завдання відкритого типу, що потребує ввести з клавіатури визначення формулювання того або іншого терміна предметної області.

З погляду на той факт, що загальна кількість слів і словоформ у будь-якій природній мові обмежена через загальні об'єктивні принципи побудови самої мови, можна стверджувати, що з математичного погляду кожна така множина, так само як і його підмножина, є скінченною множиною, тобто має скінченне число складових його елементів.

У той самий час лінгвістичне значення речення являє собою відношення деяких семантичних компонентів, відвернених одночасно від значення структурної схеми речення і лексичних значень тих слів, що зайняли позиції її компонентів і поширили собою речення [1].

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

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.

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

Сучасні наукові дослідження лінгвістів надають нам ряд підходів, що дозволяють формалізувати подання кожної лінгвістичної одиниці в зручному для машинної обробки вигляді [2]. Під час автоматизованого контролю знань термінології предметної області ставиться завдання порівняння двох визначень одного терміна: визначення, даного викладачем (еталонного визначення), і визначення, даного тим, кого навчають (відповіді). Оцінка правильності текстової відповіді заснована на методі абсолютного збігу відповіді з одним з еталонів, що зберігаються в базі даних тестової системи. Оскільки визначення терміна формується через систему термів (базових понять), кожне з яких має своє визначення, пропонується для обчислення показника релевантності відповідей на завдання відкритого типу використовувати кількісні показники синонімії термів предметної області.

Аналіз автоматизованих систем контролю знань показав, що широко застосовані методи обробки 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 [3].

### Обчислення показника релевантності

Для розрахунку показника релевантності використовується модель обчислення значення показника відповідності значущих термів еталонного визначення і відповіді [4].

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

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

де  $\omega$  – функція числових значень показника синонімії значущих термів в еталонному і фактичному визначеннях;  $\eta$  – функція кількості значущих термів в еталонному визначенні і відповіді. Параметр  $\omega$  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  $A$ . Кількість термів у відповіді дорівнює множині  $B$  до обробки:

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

Параметр  $\omega$  має такі властивості, якщо  $N = \text{const}$ :

- якщо  $K_{A,B}$  збільшується, значення  $\omega$  зростає;
- якщо кількість термів відповіді збільшується, значення  $\omega$  зменшується;

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

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

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

На підставі перерахованих властивостей отримана формула обчислення коефіцієнта  $\omega$ :

$$\frac{\sum_{e \in B} k_e}{M} = \omega. \tag{2}$$

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

- якщо  $M = N$ , то  $\eta = \eta_{\max}$ ;
- функція  $\eta$  симетрична щодо свого максимального значення;
- якщо  $|A - B| \rightarrow \pm\infty$ , то  $\eta \rightarrow 0$ .

На підставі перерахованих властивостей отримана формула обчислення коефіцієнта  $\eta$ :

$$\varphi(\Delta) = \frac{h}{\sqrt{\pi}} * e^{-h^2 \Delta^2}, \quad h > 0. \tag{3}$$

Параметр  $h$  і діапазон значень параметра  $\Delta$  підбираються експериментально і можуть варіюватися.

Використані параметри  $\Delta = 0,1 * \sqrt{M - N}$ ;  $h=2,050$

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

Підставивши формули (2) і (3) у вираз (1), одержимо

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

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

**Показники синонімії термів  $k_e$  і відповіді**

Фактичне визначення	Данн	Управл	Компонент	Систем	Обробк	Інформац	Цілий	Реалізац	Визначен	Алгоритм
Сукупн				0,8						
Дані	1		0,5			0,7				
Послідов				0,2						
Операц			0,6							
Обробк		0,3			1			0,5		
Інформац	0,7					1				
Реалізац					0,3			1		
Визначен									1	
Алгоритм				0,6						1

На основі таблиці отримано сукупність множин, що становлять інтерес для подальшого розгляду: {данн, данн}, {данн, інформац}, {управл, обробк}, {компонент, данн}, {систем, сукупн}, {систем, послідовн}, {систем, алгоритм}, {обробк, обробк}, {обробк, реаліз}, {інформац, данн}, {інформац, інформац}, {реаліз, обробк}, {реаліз, реаліз}, {визначен, визначен}, {алгоритм, алгоритм}.

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

$$k_1 = \theta(a_1, b_2) = \text{синонім}\{\text{данн, данн}\} = 1;$$

$$k_2 = \theta(a_1, b_6) = \text{синонім}\{\text{данн, інформац}\} = 0,7;$$

$$k_3 = \theta(a_2, b_4) = \text{синонім}\{\text{управлен, операц}\} = 0,2;$$

$k_4 = \theta(a_1, b_5) = \text{синонім} \{ \text{управлен, обробк} \} = 0,8;$   
 $k_5 = \theta(a_3, b_2) = \text{синонім} \{ \text{компонент, данн} \} = 0,5;$   
 $k_6 = \theta(a_3, b_2) = \text{синонім} \{ \text{компонент, операц} \} = 0,6;$   
 $k_7 = \theta(a_4, b_1) = \text{синонім} \{ \text{систем, сукупн} \} = 0,8;$   
 $k_8 = \theta(a_4, b_3) = \text{синонім} \{ \text{систем, послідовн} \} = 0,2;$   
 $k_9 = \theta(a_4, b_9) = \text{синонім} \{ \text{систем, алгоритм} \} = 0,6;$   
 $k_{10} = \theta(a_5, b_5) = \text{синонім} \{ \text{обробк, обробк} \} = 1;$   
 $k_{11} = \theta(a_5, b_7) = \text{синонім} \{ \text{обробк, релізац} \} = 0,3;$   
 $k_{12} = \theta(a_6, b_2) = \text{синонім} \{ \text{інформац, данн} \} = 0,7;$   
 $k_{13} = \theta(a_6, b_6) = \text{синонім} \{ \text{інформац, інформац} \} = 1;$   
 $k_{14} = \theta(a_8, b_5) = \text{синонім} \{ \text{реалізац, обробк} \} = 0,5;$   
 $k_{15} = \theta(a_8, b_7) = \text{синонім} \{ \text{реалізац, реалізац} \} = 1;$   
 $k_{16} = \theta(a_9, b_8) = \text{синонім} \{ \text{визначен, визначен} \} = 1;$   
 $k_{17} = \theta(a_{10}, b_9) = \text{синонім} \{ \text{алгоритм, алгоритм} \} = 1.$

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

Наприклад, множина  $\{a_1, b_2\} = \{ \text{данн, данн} \}$  перетинається з множиною  $\{a_1, b_6\} = \{ \text{данн, інформац} \}$  у точці  $a_1 = \text{данн}$ :  $\{a_1, b_2\} \cap \{a_1, b_6\} = \{a_1\}$ .

Це означає, що терму «данн» еталонного визначення відповідають два терма фактичні визначення «данн» і «інформац» [5].

За умови потрібно вибрати один із двох термів. Для цього порівнюємо їхні функції  $k_i = \theta(a_i, b_j)$ :

$k_1 = \theta(a_1, b_2) = 1;$   
 $k_2 = \theta(a_1, b_6) = 0,7;$   
 $\theta(a_1, b_2) > \theta(a_1, b_6).$

Множина  $\{a_1, b_6\}$  з подальшого розгляду виключається і не бере участь у процесі формування оцінки знань [1].

Взявши значення  $k_e$  з таблиці, підставимо у формулу обчислення параметра  $\omega$ , використовуючи при цьому множину термів відповіді і кількість значущих термів у фактичному визначенні:

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

$$R(C, D) = \frac{2,050}{\sqrt{\pi}} * e^{-(2,050)^2 * 0,1 * 1} * \frac{7,4}{9} = 0,912 \approx 0,9.$$

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  $R_{A,B} = 1$ , прийнятно для практичного застосування.

### Висновки

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

### Література

1. Цаленко М.Ш. Моделирование семантики в базах данных. – М.: Наука, 1989. – 288 с.
2. Kinshuk & Patel A. A conceptual framework for Internet based intelligent tutoring systems. Vol. II. Knowledge transfer / Ed. A. Behrooz. – London. – P. 117–124, URL: <http://ifets.ieee.org>.
3. Авенсов В.С. Научные основы тестового контроля знаний. – М.: Исслед. центр, 1994. – 135 с.
4. Абросов В.М., Хрусталева Е.Ю. Классификация критериев смыслового соответствия // Науч.-техн. информ: Сб. Сер. 2, № 4. – М.: ВИНТИ, 1979. – С. 15–24.
5. Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities // Proc. Natl. Acad. – 1982. – Sci. 7 9. – P. 2554–2558.

Стаття надійшла до редакції 30.11.06.