

УДК 004.415 (045)

ФОРМАЛЬНИЙ АНАЛІЗ МОВНИХ СТРУКТУР В АВТОМАТИЗОВАНИХ СИСТЕМАХ ОСВІТНЬОГО ПРИЗНАЧЕННЯ

Л.М. Бадьоріна

Національний авіаційний університет

e-mail: vada@ukr.net

Вступ

Інтеграційні процеси, впровадження засобів телекомунікації, комп'ютеризація людської діяльності становить безліч нових завдань в науковій області, 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 виді вільного тексту довільної довжини з урахуванням слів-синонімів [2]. Впровадження прогресивних форм навчання створює необхідність переходу комп'ютерного тестування знань студентів. Оцінювання знань студентів можливе лише шляхом порівняльного аналізу тексту відповіді з еталонним текстом та визначення їх релевантності. Теоретико-множинна модель на основі синонімії термінів предметної області дозволяє встановити відповідність між еталонним та фактичним визначенням, які представлені у виді тексту довільної довжини з використанням слів-синонімів [3]. Оцінка правильності текстової відповіді заснована на методі абсолютного збігу відповіді з одним з еталонів. 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 області.

Для оцінки релевантності еталонного визначення і відповіді необхідно:

- ✓ встановити відповідність між термами еталонного визначення і відповіді шляхом їхньої нормалізації;
- ✓ обчислити значення кількісного показника релевантності еталонного визначення і відповіді.

Постановка задачі:

Нехай S – множина усіх термів національної мови. Оскільки кількість термів у мові не є нескінченною, множина S є кінцевою і чисельною, а також неупорядкованою:

$$S = \{s_i : 1 \leq i \leq n\}, \text{ де } s_i \text{ – терм; } n \text{ – ціле число.}$$

Представимо сукупність термів будь-якої предметної області як множину T . Дана множина є частиною множини термів мови S і має всі її властивості:

$$T \subset S; T = \{t_i : 1 \leq i \leq m\}, \text{ де } m \text{ – ціле число; } m \leq n.$$

Запровадимо такі умовні позначення:

$p(T)$ – сукупність всіх можливих підмножин множини T ;

e – елемент підмножин сукупності $p(T)$, який найбільше повно відповідає описуваному об'єктові або ситуації.

Будемо вважати, що один терм не може бути використаний для опису декількох об'єктів або ситуацій, тобто $e_i \neq e_j$.

Із сукупності $p(T)$ можна виділити такі підмножини D_e , елементи яких характеризуються семантичною близькістю, що відповідає синонімічному зв'язку слів один з одним:

$$D_e \subseteq p(T).$$

Елемент e , щодо якого утворюється множина (синонімічний ряд) D_e , назвемо базовим термом, інші елементи множини D_e (слова, об'єднані синонімічним зв'язком з базовим термом) назвемо залежними термами. В рамках кожної множини D_e введемо відношення між його елементами, що позначає наявність синонімічного зв'язку між базовим термом і залежними термами. Позначимо цей зв'язок функцією $\theta(e, t)$. На кожній підмножині D_e вона задає частковий порядок, такий що:

$$t \leq s \leftrightarrow \theta(e, t) \leq \theta(e, s); \quad \forall t \in D_e; \quad \forall s \in D_e.$$

Цій частковий порядок може бути заданий у числовому виді відображенням пари термів на множину дійсних чисел:

$$\theta: T \times T \rightarrow R,$$

Тоді для двох довільних термів e і t можна вказати кількісний показник їхньої синонімії:

$$k = \theta(e, t); \quad e, t \in D_e,$$

значення якого знаходиться у інтервалі $[-1, 1]$.

Властивості та обмеження функції $\theta(e, t)$ такі:

- 1) якщо $k = \theta(e, t) = 1$, то терм t є повним синонімом терму e ;
- 2) якщо $\theta(e, t) \geq \theta(e, s)$, то $t < s$; $e, t, s \in D_e$;
- 3) якщо $\theta(e, t) - \theta(t, s) > 0$, то $t < s$; $e, t, s \in D_e$;
- 4) якщо $\theta(e, t) > 0$ і $\theta(t, s) = 1$, то $\theta(e, s) = \theta(e, t)$;
- 5) якщо $\theta(t, s) = 0$, то $\theta(t, s) = 0$.

Деякі залежні терми з множини D_{e_i} можуть бути базовими й у множинах D_{e_j} . Це означає наявність непрямого синонімічного зв'язку між базовим термом множини D_{e_i} і залежними термами множини D_{e_j} . Встановивши відповідність між термами еталонного і фактичного визначення можливо на основі синонімічної відповідності термів обчислити показник релевантності.

Використовуючи метод аналізу еталонного визначення і відповіді можливо провести нормалізацію цих визначень.

Література:

1. Цаленко М.Ш. Моделирование семантики в базах данных. - М.: Наука, 1989.- 288 с.
2. Широков В.А. Інформаційна теорія лексикографічних систем, -К.: Довіра, 1998, 331 с.
3. Бадьоріна Л.М. Кількісна оцінка змістової близькості довільних та еталонних текстових визначень в комп'ютерних навчальних системах // Наук.-практ. конф. "Інтегровані інформаційні технології та системи ПТС'2007" (Київ, 2007):Тези доп. - К.: НАУ,2007.- С. 9-13.