УДК 004.519.7 (045)

ЗМІСТОВНИЙ АНАЛІЗ ЛОГІКО-ЛІНГВІСТИЧНИХ МОДЕЛЕЙ ПРЕДСТАВЛЕННЯ ЗНАНЬ

А.І. Вавіленкова
Національний авіаційний університет
e-mail: a_vavilenkova@mail.ru

Незалежно від типу та стилю написання текстової інформації, яка досліджується, до кожного речення природної мови може входити тільки п'ять можливих типів членів речення: підмет, присудок, додаток, обставина та означення [1]. Це означає, що узагальнивши інформацію про структуру речень природної мови, дослідивши залежності між простими реченнями в контексті складних та формалізувавши правила формування змістовних зв'язків між словами речень, можна створити єдиний шаблон, за допомогою якого буде описане довільне речення природної мови. Математичним апаратом, що дозволяє формалізувати процес виявлення зв'язків між словами та простими реченнями природної мови, а також створити такий шаблон, є логіка предикатів.

Загальна форма логіко-лінгвістичної моделі речення природної мови довільної складності має вигляд:

$$L^S = \bigwedge_{p \in P^S} \bigwedge_{h \in H_p^S} L_p^S(h), \tag{1}$$

$$L_p^S(h) = \bigwedge_{x \in X_p^S(h)} \bigwedge_{g \in G_p^S(x,h)} L_p^S(x,g,h), \tag{2}$$

$$L_p^S(x,g,h) = \bigwedge_{y \in Y_p^S(x,g,h)} \bigwedge_{q \in Q_p^S(x,g,y,h)} L_p^S(x,g,y,q,h), \tag{3}$$

$$L_p^S(x,g,y,q,h) = \bigwedge_{z \in Z_p^S(x,g,y,q,h)} \bigwedge_{r \in R_p^S(x,g,y,q,z,h)} L_p^S(x,g,y,q,z,r,h), \tag{4}$$

де S – речення природної мови;

p – відношення, що пов'язує суб'єкти, об'єкти та предмети відношень у реченні S, $p \in P^S$ – множина відношень, що входять до речення S;

h – характеристика p-го відношення речення S, $h \in H_p^S$ – множина характеристик p-го відношення у реченні S;

$L_p^S(h)$ – предикат (предикативний вираз) [2], який описує p-е відношення з h-ю характеристикою і пов'язує суб'єкти, об'єкти та предмети відношення p в реченні S;

x – суб'єкт речення S, $x \in X_p^S(h)$ – множина суб'єктів, що пов'язані з об'єктами речення S p- им відношенням, яке володіє h-ю характеристикою;

g – характеристика суб'єкта x речення S, $g \in G_p^S(x,h)$ – множина характеристик суб'єкта $x \in X_p^S(h)$;

$L_p^S(x,g,h)$ – предикат (предикативний вираз), який описує p-е відношення з h-ю характеристикою між суб'єктом $x \in X_p^S(h)$ з характеристикою $g \in G_p^S(x,h)$, об'єктами та предметами p-го відношення в реченні S;

y – об'єкт речення S, $y \in Y_p^S(x,g,h)$ – множина об'єктів, що пов'язані з суб'єктами речення S p-им відношенням, яке володіє h-ю характеристикою;

q – характеристика об'єкта y речення S, $q \in Q_p^S(x,g,y,h)$ – множина характеристик об'єкта $y \in Y_p^S(x,g,h)$;

$L_p^S(x,g,y,q,h)$ – предикат (предикативний вираз), який описує p-е відношення з h-ю характеристикою між суб'єктом $x \in X_p^S(h)$ з характеристикою $g \in G_p^S(x,h)$ і об'єктом $y \in Y_p^S(x,g,h)$ з характеристикою $q \in Q_p^S(x,g,y,h)$ та предмети p-го відношення в реченні S;

z – предмет p-го відношення речення S, $z \in Z_p^S(x,g,y,q,h)$ – множина предметів p-го відношення, яке володіє h-ю характеристикою, між суб'єктом $x \in X_p^S(h)$ з характеристикою $g \in G_p^S(x,h)$ та об'єктом $y \in Y_p^S(x,g,h)$ з характеристикою $q \in Q_p^S(x,g,y,h)$;

r – характеристика предмета p-го відношення речення S, $r \in R_p^S(x,g,y,q,z,h)$ – множина характеристик предмета $z \in Z_p^S(x,g,y,q,h)$;

$L_p^S(x,g,y,q,z,r,h)$ – простий, неділимий предикат, який описує частину речення, яка має закінчений зміст, та відображає в реченні S p-е відношення з h-ю характеристикою між суб'єктом $x \in X_p^S(h)$ з характеристикою $g \in G_p^S(x,h)$ і об'єктом $y \in Y_p^S(x,g,h)$ з характеристикою $q \in Q_p^S(x,g,y,h)$, предмет якого $z \in Z_p^S(x,g,y,q,h)$ володіє характеристикою $r \in R_p^S(x,g,y,q,z,h)$.

Кількість ν^S частин речення S, що мають закінчений зміст і описуються простим предикатом $L_p^S(x,g,y,q,z,r,h)$, розраховується за формулами:

$$\nu^S = \sum_{p \in P^S} \sum_{h \in H_p^S} \nu_p^S(h),$$

$$\nu_p^S(h) = \sum_{x \in X_p^S(h)} \sum_{g \in G_p^S(x,h)} \nu_p^S(x,g,h),$$

$$\nu_p^S(x,g,h) = \sum_{y \in Y_p^S(x,g,h)} \sum_{q \in Q_p^S(x,g,y,h)} \nu_p^S(x,g,y,q,h),$$

$$\nu_p^S(x,g,y,q,h) = \sum_{z \in Z_p^S(x,g,y,q,h)} \left| R_p^S(x,g,y,q,z,h) \right|.$$

Речення – це мінімальна і основна комунікативна одиниця мови, що має бути цілісною і передавати інформацію в усій складності залежностей і зв'язків. Тому в основу змістовного аналізу логіко-лінгвістичних моделей покладено функціональні залежності між головними та другорядними членами речень природної мови.

Нехай логіко-лінгвістична модель речення природної мови задана формулами (1) – (4), тоді алгоритм змістовного аналізу буде складатися з таких кроків.

1) Визначити кількість атомарних предикатів ν^S, що описують частини речення S, які відображають закінчений зміст.

2) Зафіксувати множини P^S, H_p^S, $X_p^S(h)$, $G_p^S(x,h)$, $Y_p^S(x,g,h)$, $Q_p^S(x,g,y,h)$, $Z_p^S(x,g,y,q,h)$, $R_p^S(x,g,y,q,z,h)$.

3) Зафіксувати значення елементів кортежу логічних операцій $O(S)$ речення S:

$$O(S) = [o_k(S), k = \overline{1,m}],$$

де m – загальна кількість логічних операцій, наявних у реченні S.

4) Визначити потужності множин P^S, H_p^S, $X_p^S(h)$, $G_p^S(x,h)$, $Y_p^S(x,g,h)$, $Q_p^S(x,g,y,h)$, $Z_p^S(x,g,y,q,h)$, $R_p^S(x,g,y,q,z,h)$.

5) Якщо хоча б одна із потужностей множин $\left|P^S\right|$ або $\left|X_p^S(h)\right|$ дорівнює одиниці, то речення природної мови просте.

6) Інакше, якщо хоча б одна із потужностей множин $\left|P^S\right| \geq 2$ або $\left|X_p^S(h)\right| \geq 2$, то речення природної мови складне.

7) Визначити вид типової форми логіко-лінгвістичної моделі (1) – (4).

8) Для кожного атомарного предикату $L_p^S(x,g,y,q,z,r,h) = p(x,g,y,q,z,r,h)$ сформувати словосполучення між такими елементами логіко-лінгвістичних моделей між:

- відношенням $p \in P^S$ та характеристикою $h \in H_p^S$ цього відношення;
- відношенням $p \in P^S$ та об'єктом $y \in Y_p^S(x,g,h)$;
- об'єктом $y \in Y_p^S(x,g,h)$ та предметом відношення $z \in Z_p^S(x,g,y,q,h)$;
- суб'єктом $x \in X_p^S(h)$ та його характеристикою $g \in G_p^S(x,h)$;
- об'єктом $y \in Y_p^S(x,g,h)$ та його характеристикою $q \in Q_p^S(x,g,y,h)$;
- предметом відношення $z \in Z_p^S(x,g,y,q,h)$ та його характеристикою $r \in R_p^S(x,g,y,q,z,h)$.

9) Відновити прямий порядок слів у реченні природної мови шляхом об'єднання отриманих в п.8 словосполучень, а також враховуючи логічні операції між атомарними предикатами.

Нехай речення природної мови задане такою логіко-лінгвістичною моделлю:
[Постановляє (Рада, України, 0,0,0,0,0)
Постановляє (Рада, Верховна, 0,0,0,0,0)] →
[Провести (~~О~~, 0, слухання, парламентські, питань, реалізації, 11 грудня 2001 року)&
Реалізація (~~О~~,0, політики, інтеграції, Союзу, Європейського, 11 грудня 2001 року)].

1) Кількість атомарних предикатів, що описують частини речення S, які відображають закінчений зміст, дорівнює $v^S = 4$.

2) Множини, що входять до логіко-лінгвістичної моделі містять такі елементи:
P^S = {Постановляє, Постановляє, Провести, Реалізація};
H_p^S = {*11 грудня 2001 року, 11 грудня 2001 року* };
$X_p^S(h)$ = {Рада, Рада, ~~О~~, ~~О~~};
$G_p^S(x,h)$ = {України, Верховна};
$Y_p^S(x,g,h)$ = {слухання, політики};
$Q_p^S(x,g,y,h)$ = {парламентські, інтеграції};
$Z_p^S(x,g,y,q,h)$ = {питань, Союзу};
$R_p^S(x,g,y,q,h)$ = {реалізації, Європейського}.

3) Кортеж логічних операцій речення S містить такі елементи: $O(S) = [\&, \rightarrow, \&]$.

4) Потужності множин $\left|P^S\right| = 4$; $\left|H_p^S\right| = 2$; $\left|X_p^S(h)\right| = 4$; $\left|G_p^S(x,h)\right| = 2$; $\left|Y_p^S(x,g,h)\right| = 2$; $\left|Q_p^S(x,g,y,h)\right| = 2$; $\left|Z_p^S(x,g,y,q,h)\right| = 2$; $\left|R_p^S(x,g,y,q,z,h)\right| = 2$.

5) Потужності множин $\left|P^S\right| = 4$ і $\left|X_p^S(h)\right| = 4$, тому речення природної мови складне.

6) Так як елементи множини відношень та суб'єктів містять як різні елементи, так і такі, що дублюються, то це свідчить про наявність однорідних членів в простих реченнях, які входять до складного. Операція імплікації між простими предикатами у логіко-лінгвістичній моделі вказує на те, що речення складнопідрядне [3].

7) Логіко-лінгвістична модель складається з чотирьох атомарних предикатів:

$$[p_1(x_1, g_1, 0,0,0,0,0) \,\&\, p_1(x_1, g_2, 0,0,0,0,0)] \rightarrow$$
$$[p_2(\otimes, 0, y_2, q_2, z_2, r_2, h_2) \,\&\, r_2(\otimes, 0, y_3, q_3, z_3, r_3, h)].$$

Словосполучення, відновлені з атомарного предикату $p_1(x_1, g_1, 0,0,0,0,0)$: «*Рада України*».

Словосполучення, відновлені з атомарного предикату $p_1(x_1, g_2, 0,0,0,0,0)$: «*Верховна Рада*».

Словосполучення, відновлені з атомарного предикату $p_2(\otimes, 0, y_2, q_2, z_2, r_2, h_2)$: *«провести слухання», «парламентські слухання», «слухання питань», «питань реалізації», «провести 11 грудня 2001 року»*.

Словосполучення, відновлені з атомарного предикату $r_2(\oplus, 0, y_3, q_3, z_3, r_3, h_2)$: *«реалізація політики», «політики інтеграції», «політики Союзу», «Європейського Союзу», «реалізація 11 грудня 2001 року»*.

8) Відновлений порядок слів у реченні: *«Верховна Рада України постановляє: провести парламентські слухання питань реалізації політики інтеграції Європейського Союзу 11 грудня 2001 року»*.

Для сучасних засобів аналітичної обробки текстів характерною є функція оперативного аналізу інформації, отриманої на запит для вибору подальшого напрямку дослідження документу. До таких засобів відносять технології виділення фактографічної інформації про об'єкти з урахуванням посилань на них: нечіткий пошук, кластерний аналіз сховищ та підбір документів, виділення ключових тем, побудова анотацій [4]. На відміну від існуючих методів, змістовний аналіз логіко-лінгвістичних моделей дав можливість розробити принципи відновлення логічних зв'язків між синтаксичними текстовими складовими.

Список літератури

1. Vavilenkova A.I. A self-system to identify conceptual relationships in text / A.I. Vavilenkova // Proceedings of the National Aviation University. – 2015. – № 1(62). – P. 63–69.

2. Джарратано Д. Экспертные системы: принципы разработки и программирование: пер. с англ. – 4-е изд. – М.: ООО «Вильямс», 2007. – 1152 с.

3. Вавіленкова А.І. Теоретичні основи аналізу електронних текстів: [монографія] / А.І. Вавіленкова, Д.В.Ланде, О.Є. Литвиненко. – К.: НАУ, 2014. – 258с.

4. Барсегян А.А. Методы и модели анализа данных: OLAP и Data Mining / Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. – С.-Пб.:БХВ-Петербрг, 2007. – 384 с.