

УДК 510.635:004.891(045)

## ВІДНОВЛЕННЯ ТЕКСТОВОЇ ІНФОРМАЦІЇ З ФОРМАЛЬНИХ ЛОГІКО-ЛІНГВІСТИЧНИХ МОДЕЛЕЙ

А.І. Вавіленкова

*Національний авіаційний університет, Україна*

Логіко-лінгвістична модель [1] є засобом, що дозволяє формалізувати тексти природної мови за єдиним принципом. У свою чергу, аналіз таких моделей дає змогу зробити зворотну операцію – відновити текст. Формальний апарат трансформації тексту у логіко-лінгвістичну модель і навпаки виступає єдиним засобом автоматизації процесу обробки текстової інформації.

Змістовна компонента аналізу електронних документів містить у собі поєднання бази правил, що створена на основі досліджень лінгвістів, та методів обробки масивів текстової інформації [2].

Нехай деякий текстовий фрагмент задано такими логіко-лінгвістичними моделями, з яких потрібно відновити текст:

$$\{L^{S_1} = p_1(x_1, g_1, y_{11}, q_{11}, z_1, r_1, 0) \& p_1(x_1, g_1, y_{12}, q_{11}, z_1, r_1, 0) \&$$

$$p_1(x_1, g_1, y_{11}, q_{12}, z_1, r_1, 0) \& p_1(x_1, g_1, y_{12}, q_{12}, z_1, r_1, 0),$$

$$L^{S_2} = [p_2(x_2, 0, y_{21}, x_1, 0, 0, 0) \& p_2(x_2, 0, y_{22}, 0, z_{21}, 0, 0) \&$$

$$p_2(x_2, 0, y_{23}, q_{22}, 0, 0, 0)] \rightarrow p'_2(y_{11}, 0, y'_2, 0, 0, 0, 0)\},$$

*\{L^{S\_1} = застосовуються (мережі, нейронні, компаніями, різноманітними, технологій, інтелектуальних, 0) \&*

*застосовуються (мережі, нейронні, центрами, різноманітними, технологій, інтелектуальних, 0) \&*

*застосовуються (мережі, нейронні, компаніями, інноваційними, технологій, інтелектуальних, 0) \&*

*застосовуються (мережі, нейронні, центрами, інноваційними, технологій, інтелектуальних, 0),*

$$L^{S_2} = [залежить (те, 0, направленості, мереж, 0, 0, 0) \&$$

$$залежить (те, 0, алгоритмів, 0, роботи, 0, 0) \&$$

$$залежить (те, 0, результатів, наочних, 0, 0, 0)] \rightarrow$$

$$\text{матиме (клієнт, 0, користь, 0, 0, 0, 0)}\}.$$

1) Кількість атомарних предикатів, що описують частини речення  $S$ , які відображають закінчений зміст, дорівнює  $v^{S_1} = 4$ ,  $v^{S_2} = 4$ .

2) Множини, що входять до логіко-лінгвістичної моделі містять такі елементи:

$P^{S_1} = \{застосовуються\}; X_p^{S_1}(h) = \{мережі\}; G_p^{S_1}(x, h) = \{нейронні\};$   
 $Y_p^{S_1}(x, g, h) = \{компаніями, центрами\};$   
 $Q_p^{S_1}(x, g, y, h) = \{різноманітними, інноваційними\};$   
 $Z_p^{S_1}(x, g, y, q, h) = \{технологій\}; R_p^{S_1}(x, g, y, q, z, h) = \{інтелектуальних\};$   
 $P^{S_2} = \{залежить, матиме\}; X_p^{S_2}(h) = \{те, клієнт\};$   
 $Y_p^{S_2}(x, g, h) = \{направленості, алгоритмів, результатів, користь\};$   
 $Q_p^{S_2}(x, g, y, h) = \{мереж, наочних\}; Z_p^{S_2}(x, g, y, q, h) = \{роботи\}.$

3) Кортеж логічних операцій речення  $S_1$  містить такі елементи:  
 $O(S_1) = [\&, \&, \&]$ , для речення  $S_2 - O(S_2) = [\&, \&, \rightarrow]$ .

4) Потужності множин  $|P^{S_1}| = 1; |H_p^{S_1}| = \emptyset; |X_p^{S_1}(h)| = 1;$   
 $|G_p^{S_1}(x, h)| = 1; |Y_p^{S_1}(x, g, h)| = 2; |Q_p^{S_1}(x, g, y, h)| = 2; |Z_p^{S_1}(x, g, y, q, h)| = 1;$   
 $|R_p^{S_1}(x, g, y, q, z, h)| = 1, |P^{S_2}| = 2; |H_p^{S_2}| = \emptyset; |X_p^{S_2}(h)| = 2;$   
 $|G_p^{S_2}(x, h)| = \emptyset; |Y_p^{S_2}(x, g, h)| = 4; |Q_p^{S_2}(x, g, y, h)| = 2;$   
 $|Z_p^{S_2}(x, g, y, q, h)| = 1; |R_p^{S_2}(x, g, y, q, z, h)| = \emptyset.$

5) Потужності множин відношень та суб'єктів першого речення  
 $|P^{S_1}| = 1$  і  $|X_p^{S_1}(h)| = 1$ , тому речення природної мови просте з  
однорідними додатками та означеннями.

6) Потужності множин відношень та суб'єктів другого речення  
 $|P^{S_2}| = 2$  і  $|X_p^{S_2}(h)| = 2$ , тому речення природної мови складне.

7) Так як елементи множини відношень та суб'єктів другого речення природної мови 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.

8) Логіко-лінгвістична модель першого речення природної мови  $S_1$  складається з чотирьох атомарних предикатів:

$$\{L^{S_1} = p_1(x_1, g_1, y_{11}, q_{11}, z_1, r_1, 0) \& p_1(x_1, g_1, y_{12}, q_{11}, z_1, r_1, 0) \& p_1(x_1, g_1, y_{11}, q_{12}, z_1, r_1, 0) \& p_1(x_1, g_1, y_{12}, q_{12}, z_1, r_1, 0)\}.$$

Словосполучення, відновлені з атомарного предикату  
 $p_1(x_1, g_1, y_{11}, q_{11}, z_1, r_1, 0)$ : «нейронні мережі, застосовуються»

компаніями, різноманітними компаніями, компаніями для технологій, інтелектуальних технологій».

Словосполучення, відновлені з атомарного предикату  $p_1(x_1, g_1, y_{12}, q_{11}, z_1, r_1, 0)$ : «нейронні мережі, застосовуються компаніями, різноманітними центрами, центрами для технологій, інтелектуальних технологій».

Словосполучення, відновлені з атомарного предикату  $p_1(x_1, g_1, y_{11}, q_{12}, z_1, r_1, 0)$ : «нейронні мережі, застосовуються компаніями, інноваційними компаніями, компаніями для технологій, інтелектуальних технологій».

Словосполучення, відновлені з атомарного предикату  $p_1(x_1, g_1, y_{12}, q_{12}, z_1, r_1, 0)$ : «нейронні мережі, застосовуються компаніями, інноваційними центрами, центрами для технологій, інтелектуальних технологій».

Логіко-лінгвістична модель другого речення природної мови  $S_2$  також складається з чотирьох атомарних предикатів:

$$L^{S_2} = [p_2(x_2, 0, y_{21}, x_1, 0, 0, 0) \& p_2(x_2, 0, y_{22}, 0, z_{21}, 0, 0) \& p_2(x_2, 0, y_{23}, q_{22}, 0, 0, 0)] \rightarrow p'_2(y_{11}, 0, y'_2, 0, 0, 0, 0) \}$$

Словосполучення, відновлені з атомарного предикату  $p_2(x_2, 0, y_{21}, x_1, 0, 0, 0)$ : «залежить від направленості, направленості мереж».

Словосполучення, відновлені з атомарного предикату  $p_2(x_2, 0, y_{22}, 0, z_{21}, 0, 0)$ : «залежить від алгоритмів, алгоритмів роботи».

Словосполучення, відновлені з атомарного предикату  $p_2(x_2, 0, y_{23}, q_{22}, 0, 0, 0)$ : «залежить від результатів, наочних результатів».

Словосполучення, відновлені з атомарного предикату  $p'_2(y_{11}, 0, y'_2, 0, 0, 0, 0)$ : «матиме користь».

9) Відновлений порядок слів у реченні природної мови  $S_1$  за схемою  $g_{\delta_1}^{(\lambda)} \rightarrow x_{\delta_1}^{(\lambda)} \rightarrow p_{\delta_1}^{(\lambda)} \rightarrow q_{\delta_1}^{(\lambda)} \rightarrow y_{\delta_1}^{(\lambda)} \rightarrow r_{\delta_1}^{(\lambda)} \rightarrow z_{\delta_1}^{(\lambda)} \rightarrow h_{\delta_1}^{(\lambda)}$ : «нейронні мережі застосовуються різноманітними інноваційними компаніями та центрами для інтелектуальних технологій».

10) Відновлений порядок слів у реченні природної мови  $S_2$  за схемою  $g_{\delta_2}^{(\lambda)} \rightarrow x_{\delta_2}^{(\lambda)} \rightarrow p_{\delta_2}^{(\lambda)} \rightarrow q_{\delta_2}^{(\lambda)} \rightarrow y_{\delta_2}^{(\lambda)} \rightarrow r_{\delta_2}^{(\lambda)} \rightarrow z_{\delta_2}^{(\lambda)} \rightarrow h_{\delta_2}^{(\lambda)}$ :

«те залежить від направленості мереж, алгоритмів роботи та наочних результатів, внаслідок чого клієнт матиме користь».

Оригінальний текстовий фрагмент, для якого були побудовані задані логіко-лінгвістичні моделі має вигляд: "Нейронні мережі застосовуються різноманітними інноваційними компаніями та центрами для інтелектуальних технологій. Від їх направленості, алгоритмів роботи та наочних результатів залежить те, чи матиме користь клієнт".

Таким чином, аналіз логіко-лінгвістичних моделей текстових документів представляє собою складний процес отримання інформації про структуру та зміст тексту, що розглядається, на основі виявлення закономірностей і тенденцій синтаксичної, семантичної та лексичної побудови тексту. Результатом проведення аналізу логіко-лінгвістичної моделі текстового документу є відновлений текст [3].

### Література

1. Вавіленкова А. І. Структуризація текстової інформації за допомогою логіко-лінгвістичного моделювання. *Системи підтримки прийняття рішень. Теорія і практика: зб. доп. наук.-практ. конф. з міжнар. участю* (м. Київ, 5 червня 2017 р.). Київ: ПІММС НАНУ, 2017. С. 101 – 104.
2. Вавіленкова А. І. Логіка роботи системи порівняльного аналізу електронних текстових документів. *International Scientific Journal Acta Universitatis Pontica Euxinus. Special number for XIII international conference «Strategy of quality in industry and education»* (Varna, Bulgaria, 5 – 8 June 2017). Varna, Bulgaria, 2017. Vol. II. P. 393 – 396.
3. Вавіленкова А. І. Аналіз і синтез логіко-лінгвістичних моделей речень природної мови: монографія. Київ: ТОВ «СІК ГРУП УКРАЇНА», 2017. 152 с.