

АЛГОРИТМ ПОБУДОВИ МАТЕМАТИЧНОЇ МОДЕЛІ ТЕКСТУ

Вавіленкова А.І.

Національний авіаційний університет, м. Київ

Текст – це комплекс взаємопов’язаних одне з одним речень, що володіє певною автономністю по відношенню до аналогічних комплексів. Цілісність тексту перетворює його в систему, в якій елементи залежать і передбачають один одного. Кожен текст $t \in T$ із множини електронних документів можна представити у вигляді впорядкованої четвірки:

$$t = \langle C, F, B, A \rangle,$$

де $C = \{c_1, \dots, c_i, \dots, c_n\}$ – множина існуючих типів текстів, $i = \overline{1, n}$, n – кількість типів;

$F = \{f_1, \dots, f_j, \dots, f_m\}$ – множина складних синтаксичних частин тексту, $j = \overline{1, m}$, m – кількість складних синтаксичних частин;

B – текстова база, що складається з набору ключових слів тексту та взаємопов’язаних пропозицій;

$A = \{a_1, \dots, a_k, \dots, a_q\}$ – множина абзаців тексту, $k = \overline{1, q}$, q – кількість абзаців.

Тоді для побудови змістовної моделі тексту необхідно здійснити ряд кроків.

1. Визначити тип зв’язного тексту, в межах якого формується глобальний зв’язок. Кожен тип тексту має свої мовні та когнітивні особливості, на основі чого для кожного з них можна побудувати текстову базу та ситуаційну модель.

2. Виявити можливі типи зв’язків між складними синтаксичними частинами, тобто побудувати схематичну структуру тексту.

3. Визначити тип кожного з абзаців.

4. Визначити тип зв’язку між реченнями в кожному абзаці. Зв’язок між реченнями у тексті буває ланцюговим або паралельним. Ланцюговий зв’язок – думка проходить від одного речення до іншого та здійснюється зазвичай при повторі слова з попереднього речення та розгортанні її у наступному реченні. Паралельний зв’язок – полягає у вживанні речень, в яких однаковий порядок слів, однотипні граматичні форми вираження членів речення.

5. Визначити засоби побудови та відновлення логічних зв’язків у кожному абзаці $a_k \in A$. Задача дослідження – встановити види такого зв’язку та визначити правила його передачі.

Якщо об’єднати основні положення структурного аналізу, лінгвістики тексту та математичного моделювання, то можна формалізувати процес створення змістовної моделі тексту.