

О. Ю. Несміян, М. А. Павленко

Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна

МЕТОД АНАЛІЗУ ТА ОБРОБКИ ІНФОРМАЦІЇ В СППР АСУ

Предмет: У статті розглянуті питання семантичного та синтаксичного аналізу текстів, процесів декомпозиції речення на складові, побудови синтаксичної структури речення; семантичної інтерпретації фрагментів речення природної мови у фрагменти М-мови; композиції фрагментів М-мови в структуру реферату, що описує проблемне середовище. **Мета:** розробка методу аналізу та обробки інформації в СППР АСУ. **Завдання:** Провести аналіз досліджень останніх років пов'язаних з вирішенням проблеми аналізу змісту мовних структур з погляду на створення систем діалогу з програмним забезпеченням, зроблено висновки щодо доцільності напрямку досліджень. Виконати детальний опис етапів виконання семантичного та синтаксичного аналізу з описом вхідних, проміжних та вихідних даних. Визначити рекомендації щодо варіативності вхідних даних в залежності від структури речень предметної області (тексту викладеного природною мовою). Запропонувати правила формування предикатів для різних частин української мови. Визначити типовий набір службових слів для наповнення еталонного словника. Обґрунтувати виконання наведених операцій щодо об'єднання в одну обчислювальну структуру, яка виконується за визначеними етапами. Розробити процедури проходження кожного з етапів. Запропонувати правила формування синтаксичних груп з визначених пар лексем. **Висновки:** Застосування розробленого методу на основі визначених правил до предикатів дозволить виконати вибірку інформативно суттєвої інформації із набору текстової інформації та представити її у вигляді анотації (реферату).

Ключові слова: СППР, АСУ, речення, словоформа, лексема, синтаксична група, семантичний зв'язок.

Вступ

Постановка проблеми: Спроби формалізувати інтелектуальну діяльність людини призвели до постановки фундаментальної лінгвістичної задачі [1], що полягає в побудові функціональної моделі природної мови. Формально моделі мови розглядаються як компоненти різних прикладних систем. Мета аналізу речень на природній мові – переведення їх на машинну мову обчислювальної системи. При цьому реалізуються наступні функції [2]: розпізнавання правильно побудованих речень природної мови; декомпозиція («розбивка») речення на складові (фрагменти) і побудова відповідної синтаксичної структури речення; семантична інтерпретація фрагментів речення природної мови у фрагменти М-мови; композиція («збірка») фрагментів М-мови в структуру, що описує ситуацію проблемного середовища. Реалізація цих функцій здійснюється на етапах морфологічного і синтаксичного аналізів, семантичної інтерпретації і проблемного аналізу.

Аналіз останніх досліджень і публікацій: Аналіз досліджень останніх років пов'язаних з вирішенням проблеми аналізу змісту мовних структур з погляду на створення систем діалогу з програмним забезпеченням показав, що найбільш поширеними є підходи, засновані на синтаксичному, семантичному аналізах і шаблонах. Перший підхід, який використовує синтаксичні конструкції, - найважчий. Синтаксичне представлення запиту будується на основі підмета, присудка, суфікса тощо, які визначаються за допомогою морфологічних характеристик. Це представлення нічого не говорить про сутність запиту. Другий підхід, заснований на семантиці, набагато ближчий до сутності запиту та є уніфікованим [4, 5]. Саме апарат семантичного аналізу текстової інформації був застосований при виконанні досліджень наведених в даній статті.

Метою статті є розробка методу аналізу та обробки інформації в СППР АСУ.

Виклад основного матеріалу

Вибір підходу, що базується на семантичному аналізі текстової інформації обумовлений не тільки уніфікацією доступних методів та близькістю до сутності запитів, а й тим, що у ньому використовується синтаксична інформація з семантичних словників, що дозволяє досить варіативно підходити до їх побудови максимально звужуючи предметну область, а відповідно і швидкість обробки. Тобто, кожне слово в словнику має характеристики, що дозволяють визначати смислові відношення між ним та іншими словами, точніше, їх значеннями. Повний опис зв'язків між смислами слів (а одне слово часто має кілька смислів) утворює тезаурус, який представляє собою велику мережу зі словами і їх смислами як вузли. За допомогою таких тезаурусів виконується побудова семантичного подання запиту. Основне завдання при цьому - відсікти непотрібні смисли, постаратися виділити за допомогою синтаксичних зв'язків достовірні семантичні конструкції. Дослідження, що проводяться в даній роботі, пов'язані з розробкою методу та алгоритмів, що реалізують всі перераховані вище кроки. На наш погляд, запити до структурованих джерел даних цілком можуть бути викладені у вигляді одного або декількох простих речень природної мови, в яких відсутні неологізми і для яких нескладно здійснювати синтаксичний аналіз. При синтаксичному аналізі природно-мовного речення будується граф залежностей, який після ряду послідовних перетворень містить інформацію, необхідну для побудови SQL-запиту. За своєю сутністю реалізація кожного з етапів пропонується представити взаємозалежним ланцюгом операцій. Структура запропонованого методу показана на рис. 1.

Відповідно до опису формального апарату БД рішення задачі семантичного та синтаксичного аналізу може бути представлено у вигляді перетворень в просторі ситуацій.



Рис. 1. Структура методу семантико-синтаксичного аналізу та обробки інформації

Відповідно до опису формального апарату БД рішення задачі семантичного та синтаксичного аналізу може бути представлено у вигляді перетворень в просторі ситуацій. Кожна ситуація описується продукційним правилом виду $\langle q, r \rangle$, де q - умова застосовності, r - програма. Тому по кожній моделі даних будується власна система продукцій. Умова застосовності записується як правильно побудована формула логіки предикатів першого порядку.

Опис правил пропонується здійснювати в три етапи. На першому етапі правило описується в змістовній формі у вигляді множини висловлювань. На другому етапі кожен вислів представляється у вигляді відповідного предиката, а на останньому - будується формула для умови застосування q продукції. База правил морфологічного аналізу, по своїй суті, включає в себе продукційні правила по розпізнаванню ситуацій, пов'язаних з ідентифікацією слів, що відносяться до незмінних і змінних частин мови. Логічно передбачити, що вхідна інформація подається у вигляді результатів лексичного аналізу, тобто на вхід системи продукцій подається певна лексема $x_i \in X$, де X - множина лексем природно-мовного запиту. Можливі ситуації, що виникають при проведенні семантико-синтаксичного аналізу, опишемо у вигляді продукційних правил відображених в попередньо задекларованих етапах.

Правило pr_1^M - ситуація, що описує приналежність лексеми до незмінної частини мови: x_i лексема відноситься до незмінних частин мови тоді, коли має місце ситуація, що представляє собою кон'юнкцію таких фактів: лексема x_i належить $s_i \in S$; s_i має значення s_{32} (більшість незмінних слів); лексема x_i перетворюється в флективний клас y_i ; флективний клас y_i належить множині $s_{2j} \in s_2$; s_{2j} має значення s_{25} (множина флективних класів

незмінних слів); флективний клас s_{2j} відображається в набір морфологічної інформації y_2 ; набір морфологічної інформації y_2 належить до s_{74k} (множина незмінних частин мови); s_{74k} має значення s_{740} (множина частин мови незмінних слів).

Даній ситуації відповідають такі предикати:

- основному твердженню «існує лексема $x_i \in X$, яка знаходиться в відношенні кваліфікації з множиною s_{32} » відповідає предикат -

$$(\exists x_i : X) P_s(x_i, s_{33}, t_2^r);$$

- фактам відповідає такий набір предикатів:

$$P_f(t_{in}(x_i), y_1, f_1), P_E(y_1, s_{33}),$$

$$P_f(t_{in}(x_i, H_1(u, v)), y_2, f_4), P_f(t_{in}(y_2), y_3, f_1),$$

$$P_E(y_3, s_{25}), P_f(t_{in}(y_4, H_2(u, v)), y_4, f_4),$$

$$P_f(t_{in}(y_4), y_5, f_1), P_E(y_5, s_{742}).$$

Таким чином, елементи продукції pr_1^M матимуть такий вигляд:

$$q_1^M = P_f(t_{in}(x_i), y_1, f_1) \wedge P_E(y_1, s_{33}) \wedge$$

$$P_f(t_{in}(x_i, H_1(u, v)), y_2, f_4) \wedge P_f(t_{in}(y_2), y_3, f_1) \wedge$$

$$P_E(y_3, s_{25}) \wedge P_f(t_{in}(y_4, H_2(u, v)), y_4, f_4) \wedge$$

$$P_f(t_{in}(y_4), y_5, f_1) \wedge P_E(y_5, s_{740}) \leftrightarrow$$

$$(\exists x_i : X) P_s(x_i, s_{33}, t_2^r)$$

$$r_1^M = add[v_{il}, y_5]$$

Структура r_1^M присвоює елементу v_{il} вектора $V_i \in V$ назву отриманої частини мови.

При описі наступних правил сумістимо змістовний і формальний опис отриманих рішень.

Продукція pr_2^m відноситься до ситуації, що описує приналежність лексеми до частини мови іменник: x_s лексема відноситься до частини мови іменник тоді, коли має місце ситуація, що представляє собою таку кон'юнкцію фактів:

- лексема x_i має закінчення y_1 потужністю

$$|e| = 3 - P_f(t_{in}(x_i, 3), y_1, f_2);$$

- закінчення y_1 належить множині $s_{1j} \in S$ -

$$P_f(t_{in}(y_1), y_2, f_1)$$

- множина s_{1j} має значення s_{11} (множина закінчень іменників) -

$$P_E(y_2, s_{11});$$

- лексема x_i має основу y_2 -

$$P_f(t_{in}(x_i, 3), y_3, f_3);$$

- основа y_2 лексеми x_i належить множині

$$s_{3k} \in S - P_f(t_{in}(y_3), y_4, f_1);$$

- множина s_{3k} має значення s_{31} (множина основ) -

$$P_E(y_4, s_{31});$$

або

- лексема x_i має закінчення y_1 потужністю

$$|e| = 2 - P_f(t_{in}(x_i, 2), y_1, f_2);$$

– закінчення y_1 належить множині $s_{1j} \in S$ -
 $P_f(t_{in}(y_1), y_2, f_1)$;
 – множина s_{1j} має значення s_{11} (множина закінчень іменників) - $P_E(y_2, s_{11})$;
 – лексема x_i має основу
 $y_2 - P_f(t_{in}(x_1, 2), y_3, f_3)$;
 – основа y_2 лексеми x_i належить множині $s_{3k} \in S$ - $P_f(t_{in}(y_3), y_4, f_1)$;
 – множина s_{3k} має значення s_{31} (множина основ) - $P_E(y_4, s_{31})$;
 або
 – лексема x_i має закінчення y_1 потужністю $|e|=1$ - $P_f(t_{in}(x_i, 2), y_1, f_2)$;
 – закінчення y_1 належить множині $s_{1j} \in S$ -
 $P_f(t_{in}(y_1), y_2, f_1)$;
 – множина s_{1j} має значення s_{11} (множина закінчень іменників) - $P_E(y_2, s_{11})$;
 – лексема x_i має основу
 $y_2 - P_f(t_{in}(x_1, 1), y_3, f_3)$;
 – основа y_2 лексеми x_i належить множині $s_{3k} \in S$ - $P_f(t_{in}(y_3), y_4, f_1)$;
 – множина s_{3k} має значення s_{31} (множина основ) - $P_E(y_4, s_{31})$;
 або
 – лексема x_i має закінчення y_1 потужністю $|e|=0$ - $P_f(t_{in}(x_i, 0), y_1, f_2)$;
 – закінчення y_1 належить множині $s_{1j} \in S$ -
 $P_f(t_{in}(y_1), y_2, f_1)$;

$q_2^M = ((P_f(t_{in}(x_i, 3), y_1, f_2) \wedge P_f(t_{in}(y_1), y_2, f_1) \wedge P_E(y_2, s_{11}) \wedge P_f(t_{in}(x_1, 3), y_3, f_3) \wedge P_f(t_{in}(y_3), y_4, f_1) \wedge P_E(y_4, s_{31})) \vee$
 $((P_f(t_{in}(x_i, 2), y_1, f_2) \wedge P_f(t_{in}(y_1), y_2, f_1) \wedge P_E(y_2, s_{11}) \wedge P_f(t_{in}(x_1, 2), y_3, f_3) \wedge P_f(t_{in}(y_3), y_4, f_1) \wedge P_E(y_4, s_{31})) \vee$
 $((P_f(t_{in}(x_i, 1), y_1, f_2) \wedge P_f(t_{in}(y_1), y_2, f_1) \wedge P_E(y_2, s_{11}) \wedge P_f(t_{in}(x_1, 1), y_3, f_3) \wedge P_f(t_{in}(y_3), y_4, f_1) \wedge P_E(y_4, s_{31})) \vee$
 $((P_f(t_{in}(x_i, 0), y_1, f_2) \wedge P_f(t_{in}(y_1), y_2, f_1) \wedge P_E(y_2, s_{11}) \wedge P_f(t_{in}(x_1, 1), y_3, f_3) \wedge P_f(t_{in}(y_3), y_4, f_1) \wedge P_E(y_4, s_{31}))) \wedge$
 $P_f(t_{in}(y_3, H_1(u, v)), y_5, f_4) \wedge P_f(t_{in}(y_5), y_6, f_1) \wedge P_E(y_6, s_{21}) \wedge P_f(t_{in}(y_5, H_1(u, v)), y_6, f_4) \wedge P_E(y_6, x_{7411}) \wedge$
 $P_f(t_{in}(y_5, y_1, H(u, v, z)), y_7, f_5) \wedge P_f(t_{in}(y_7), y_8, f_1) \wedge P_E(y_8, s_{751}) \leftrightarrow (\exists x_i : X) P_s(x_i, x_{7411}, t_2^M)$;

$$r_2^M = add[V_i, v_{i1}, y_6] \wedge V(V_i, (v_{i2}, l_1(y_8)), (v_{i3}, l_2(y_8)), (v_{i4}, l_3(y_8)), (v_{i5}, l_4(y_8)))$$

Правило r_2^M присвоює елементу v_{i1} вектора $V_i \in V$ назву отриманої частини мови, v_{i2} - ознака елемента живої природи за допомогою функції l_1 , яка виділяє перший елемент кортежу змінної МІ, елементу v_{i3} та привласнює значення роду за допомогою функції l_2 , яка виділяє другий елемент кортежу змінної МІ, елементу v_{i4} присвоює зна-

– множина s_{1j} має значення s_{11} (множина закінчень іменників) - $P_E(y_2, s_{11})$;
 – лексема x_i має основу
 $y_2 - P_f(t_{in}(x_1, 0), y_3, f_3)$;
 – основа y_2 лексеми x_i належить множині $s_{3k} \in S$ - $P_f(t_{in}(y_3), y_4, f_1)$;
 – множина s_{3k} має значення s_{31} (множина основ) - $P_E(y_4, s_{31})$;
 і для всіх попередніх фактів:
 – основа y_2 відображається у флективний клас (ФК) $y_3 - P_f(t_{in}(y_3, H_1(u, v)), y_5, f_4)$;
 – флективний клас y_3 лексеми x_i належить множині $s_{2p} \in S$ - $P_f(t_{in}(y_5), y_6, f_1)$;
 – множина s_{2p} має значення s_{21} (багато ФК іменників) - $P_E(y_6, x_{21})$;
 – флективний клас y_3 відображається в постійну морфологічну інформацію (МІ) $y_4 - P_f(t_{in}(y_5, H_1(u, v)), y_6, f_4)$;
 – МІ y_4 має значення x_{7411} (частина мови іменник) - $P_E(y_4, s_{31})$;
 – Флективний клас y_3 та закінчення y_1 зображає в наборі змінної морфологічної інформації $y_5 - P_f(t_{in}(y_5, y_1, H(u, v, z)), y_7, f_5)$;
 – Набір МІ y_5 має множина s_{75t} (множина наборів змінної МІ) - $P_f(t_{in}(y_7), y_8, f_1)$;
 – Множина s_{75t} має значення s_{751} (множина наборів змінної МІ іменників) - $P_E(y_8, s_{751})$.

Тоді продукція pr_2^M може бути представлена у вигляді пари $\langle q_2^M, r_2^M \rangle$, де:

чення числа через функцію l_3 , яка виділяє третій елемент кортежу змінної МІ, і елементу v_{i5} присвоює значення відмінка через функцію l_4 , яка виділяє четвертий елемент кортежу змінної МІ.

Предикат pr_3^M - приналежність лексеми до частини мови прикметник може бути представлений у вигляді пари $\langle q_3^M, r_3^M \rangle$, та мати такий вигляд:

$$r_3^M = \text{add} \left[(V_i, v_{i1}, y_6) \wedge V \left(V_i, (v_{i3}, l_1(y_8)), (v_{i4}, l_2(y_8)), (v_{i5}, l_3(y_8)) \right) \right], \text{ де}$$

$$q_3^M = \left((P_f(t_{in}(x_i, 3), y_1, f_2) \wedge P_f(t_{in}(y_1), y_2, f_1) \wedge P_E(y_2, s_{12}) \wedge P_f(t_{in}(x_i, 3), y_3, f_3) \wedge P_f(t_{in}(y_3), y_4, f_1) \wedge P_E(y_4, s_{31})) \vee \right. \\ \left. (P_f(t_{in}(x_i, 2), y_1, f_2) \wedge P_f(t_{in}(y_1), y_2, f_1) \wedge P_E(y_2, s_{12}) \wedge P_f(t_{in}(x_i, 2), y_3, f_3) \wedge P_f(t_{in}(y_3), y_4, f_1) \wedge P_E(y_4, s_{31})) \right) \wedge \\ P_f(t_{in}(y_3, H_1(u, v)), y_5, f_4) \wedge P_f(t_{in}(y_5), y_6, f_1) \wedge P_E(y_6, s_{22}) \wedge P_f(t_{in}(y_5, H_1(u, v)), y_6, f_4) \wedge P_E(y_6, x_{7412}) \wedge \\ P_f(t_{in}(y_5, y_1, H(u, v, z)), y_7, f_5) \wedge P_f(t_{in}(y_7), y_8, f_1) \wedge P_E(y_8, s_{752}) \leftrightarrow P_4(y_6, x_{7412}) \wedge P_f(t_{in}(y_5, y_1, H(u, v, z)), y_7, f_5) \wedge \\ P_f(t_{in}(y_7), y_8, f_1) \wedge P_E(y_8, s_{752}) \leftrightarrow (\exists x_i : X) P_s(x_i, x_{7412}, t_2^r);$$

Предикат pr_4^M – приналежність лексеми до частини мови дієслово, може бути представлений у вигляді пари $\langle q_4^M, r_4^M \rangle$, та мати такий вигляд:

$$r_4^M = \text{add} \left[(V_i, v_{i1}, y_6) \wedge V \left(V_i, (v_{i4}, l_1(y_8)), (v_{i6}, l_2(y_8)), (v_{i7}, l_3(y_8)) \right) \right], \text{ де}$$

$$q_4^M = \left((P_f(t_{in}(x_i, 3), y_1, f_2) \wedge P_f(t_{in}(y_1), y_2, f_1) \wedge P_E(y_2, s_{13}) \wedge P_f(t_{in}(x_i, 3), y_3, f_3) \wedge P_f(t_{in}(y_3), y_4, f_1) \wedge P_E(y_4, s_{31})) \vee \right. \\ \left. (P_f(t_{in}(x_i, 2), y_1, f_2) \wedge P_f(t_{in}(y_1), y_2, f_1) \wedge P_E(y_2, s_{13}) \wedge P_f(t_{in}(x_i, 2), y_3, f_3) \wedge P_f(t_{in}(y_3), y_4, f_1) \wedge P_E(y_4, s_{31})) \vee \right. \\ \left. (P_f(x_i, 1), y_1, f_2) \wedge P_f(t_{in}(y_1), y_2, f_1) \wedge P_E(y_2, s_{13}) \wedge P_f(t_{in}(x_i, 1), y_3, f_3) \wedge P_f(t_{in}(y_3), y_4, f_1) \wedge P_E(y_4, s_{31})) \right) \wedge \\ \left. (P_f(t_{in}(y_3, H_1(u, v)), y_5, f_4) \wedge P_f(t_{in}(y_5), y_6, f_1) \wedge P_E(y_6, s_{23}) \wedge P_f(t_{in}(y_5, H_1(u, v)), y_6, f_4) \wedge P_E(y_6, x_{7413}) \wedge \right. \\ \left. P_f(t_{in}(y_5, y_1, H(u, v, z)), y_7, f_5) \wedge P_f(t_{in}(y_7), y_8, f_1) \wedge P_E(y_8, s_{753}) \leftrightarrow (\exists x_i : X) P_s(x_i, x_{7413}, t_2^r);$$

Предикат pr_5^M – приналежність лексеми до частини мови прислівник може бути представлений у вигляді пари $\langle q_5^M, r_5^M \rangle$, та мати наступний вигляд:

$$r_5^M = \text{add} \left[(V_i, v_{i1}, y_6) \wedge V \left(V_i, (v_{i3}, l_1(y_8)), (v_{i7}, l_2(y_8)) \right) \right], \text{ де}$$

$$q_5^M = \left(\left(P_f(t_{in}(x_i, 1), y_1, f_2) \wedge P_f(t_{in}(y_1), y_2, f_1) \wedge P_E(y_2, s_{15}) \wedge P_f(t_{in}(x_i, 1), y_3, f_3) \wedge P_f(t_{in}(y_3), y_4, f_1) \wedge P_E(y_4, s_{31})) \vee \right. \right. \\ \left. \left. P_f(t_{in}(x_i, 0), y_1, f_2) \wedge P_f(t_{in}(y_1), y_2, f_1) \wedge P_E(y_2, s_{15}) \wedge P_f(t_{in}(x_i, 0), y_3, f_3) \wedge P_f(t_{in}(y_3), y_4, f_1) \wedge P_E(y_4, s_{31})) \right) \wedge \right. \\ \left. P_f(t_{in}(y_3, H_1(u, v)), y_5, f_4) \wedge P_f(t_{in}(y_5), y_6, f_1) \wedge P_E(y_6, s_{23}) \wedge P_f(t_{in}(y_5, H_1(u, v)), y_6, f_4) \wedge P_E(y_6, x_{7413}) \wedge \right. \\ \left. P_f(t_{in}(y_5, y_1, H(u, v, z)), y_7, f_5) \wedge P_f(t_{in}(y_7), y_8, f_1) \wedge P_E(y_8, s_{755}) \leftrightarrow (\exists x_i : X) P_s(x_i, x_{7415}, t_2^r). \right.$$

Активация продукційних правил здійснюється на основі використання методу резолюцій, описаного в [3]. У базі правил морфологічного аналізу умови застосовності q_i^M предикатів pr_i^M повинні зберігатися в скулемівській стандартній формі і складати множину диз'юнктив. Якщо в результаті побудови дерева виведення буде знайдений порожній диз'юнктив, то формується приналежність $r_i^M \in pr_i^M$, в іншому випадку вибирається наступний предикат. Таким чином, здійснюється перетворення $\Psi_1 : T \rightarrow X$, де T - текст природної мови у вигляді множини лексем, X - множина лексем з векторами морфологічної інформації v .

Синтаксичний аналіз пропонується виконувати в три етапи. На першому етапі здійснюється нормалізація лексем природно-мовного речення для виділення синтаксичних груп, до яких відносяться іменник + прийменник. На цьому ж етапі здійснюється видалення несуттєвих лексем з початкової множини, таких як службові частини мови (прийменники, сполучники, частки і т.п.). В результаті будуть сформовані дві множини: нова вихідна множина лексем X і L - множина синтаксичних груп у вигляді векторів пов'язаних лексем. Синтаксичний

зв'язок, що відноситься до типу підпорядкування, передає поєднання двох слів, в якому одне виступає як головне, а інше - як залежне. Тому завданням другого етапу є виявлення синтаксичного зв'язку між двома лексемами множини X і множини векторів L , розбиття лексем на множину головних слів L_1 і множину залежних слів L_2 , причому $L_1 \cap L_2 \neq \emptyset$, і формування множини поєднаних пар лексем $D = \{(x_i, x_j) \mid x_i \in L_1, x_j \in L_2\}$. Для знаходження кореневої вершини необхідно: об'єднати множини L_1 і L_2 : $L_3 = L_1 \cup L_2$; знайти різниці множин L_3 і L_2 : $L_4 = L_3 \setminus L_2$, де L_4 - одноелементна множина кореневих вершин. Таким чином, формується один або кілька графів залежностей $G = \langle X, D \rangle$, де X - множина вершин графа G , яка становить множину лексем $X = \{x_i \mid i = 1, n\}$, а D - множина дуг.

Підсумковий граф залежностей G буде відповідати таким вимогам:

- граф G є неповним графом, тобто не містить петель і циклів;
- граф G є зв'язним.

Граф синтаксичного розбору не може мати петель, так як розроблені правила враховують як мор-

фологічну інформацію лексем, так і порядок лексем в природно-мовному реченні. Друга вимога задовольняється введенням другого обмеження на природну мову запиту, пов'язана з тим, що в кожному наступному реченні запиту має бути присутній іменник, введений в одному з попередніх речень. Грунтуючись на актуальних для системи відбору видах мовної інформації виділимо три групи правил, пов'язаних з аналізом ситуацій по: виявленню синтаксичних груп, синтаксичного зв'язку між парою лексем і синтаксичного зв'язку між парою лексем і синтаксичною групою.

Висновки

Таким чином запропонований метод семантико синтаксичного аналізу, що побудований на основі морфологічного, синтаксичного і проблемного аналізів надає по суті формалізований апарат реалізації семантичного та синтаксичного аналізу. Формалізований апарат дозволив знайти єдиний підхід опису всіх моделей лінгвістичного транслятора. Кожна ситуація описується продукційним правилом з формуванням відповідного предиката.

СПИСОК ЛІТЕРАТУРИ

1. Андреев А.М. Лингвистический процессор для ИПС / А.М. Андреев, Д.В. Березкин, А.В. Брик // Компьютерная хроника. – 1998. – №11. – С. 79 – 100.
2. Дударь З.В. Семантическая нейронная сеть как формальный язык описания и обработки смысла текстов на естественном языке / З.В. Дударь, Д.Е. Шуклин // Радиоэлектроника и информатика. – 2000. – №3. – С. 72 – 76.
3. Математическая логика и автоматическое доказательство теорем: Пер. с англ. / Под ред. С.Ю. Маслова / Ч. Чечень, Р. Ли. – М.: Наука. Главная редакция физико-математической литературы, 1983. – 360 с.
4. Aho, A. V., Sethi, R. and Ullman, J. D. (1986), "Compilers: principles, techniques, and tools". Addison Wesley Longman Publishing Co., Inc. Boston, MA, USA.
5. DeRemer, Frank and Penello, Thomas (October 1982). "Efficient Computation of LALR(1) Look-Ahead Sets". Transactions on Programming Languages and Systems (ACM) 4 (4): 615–649.
6. Kluegl, P., Atzmueller, M. and Puppe, F. (2009), "Meta-level Information Extraction". KI 2009: Advances in Artificial Intelligence. Vol. 5803, pp. 233–240.

Рецензент: д-р техн. наук, проф. С. Г. Семенов,

Національний технічний університет "Харківський політехнічний інститут", Харків

Received (Надійшла) 22.07.2018

Accepted for publication (Прийнята до друку) 29.08.2018

Метод анализа и обработки информации в СППР АСУ

А. Ю. Несмиян, М. А. Павленко

Предмет: В статье рассмотрены вопросы семантического и синтаксического анализа текстов, процессов декомпозиции предложения на составляющие, построения синтаксической структуры предложения; семантической интерпретации фрагментов предложения естественного языка во фрагменты М-языка; композиции фрагментов М-языка в структуру реферата, описывающую проблемную среду. **Цель:** разработка метода анализа и обработки информации в СППР АСУ. **Задача:** Провести анализ исследований последних лет, связанных с решением проблемы анализа содержания языковых структур с точки зрения создания систем диалога с программным обеспечением, сделать выводы о целесообразности направления исследований. Выполнить детальное описание этапов выполнения семантического и синтаксического анализа с описанием входных, промежуточных и выходных данных. Определить рекомендации по вариативности входных данных в зависимости от структуры предложений предметной области (текста, изложенного естественным языком). Предложить правила формирования предикатов для различных частей украинского языка. Определить типовой набор служебных слов для наполнения эталонного словаря. Обосновать выполнение каждой из приведенных операций по объединению в одну вычислительную структуру, которая выполняется по определенным этапам. Разработать процедуры прохождения каждого из этапов. Предложить правила формирования синтаксических групп по определенным парам лексем. **Выводы:** Применение разработанного метода на основе определенных правил в предикатах позволит выполнить выборку информативно существенной информации из набора текстовой информации и представить ее в виде аннотации (реферата).

Ключевые слова: СППР, АСУ, предложения, словоформа, лексема, синтаксическая группа, семантическая связь.

Method of analysis and processing information in the DSS ACS

O. Nesmiian, M. Pavlenko

Subject: The article contains investigation of the questions related to semantic and syntactic analysis of the texts, decomposition of sentences into component processes, the construction of the syntactic structure of sentences; semantic interpretation of the fragments of a natural language sentence into fragments of an M-language; the composition of the fragments of the M-language in the structure of the abstract, describing the problem environment. **Purpose:** development of a method for analyzing and processing information on the DSS ACS. **Objective:** To analyze the research of recent years related to the solution of analyzing the content of the language structure problem from the point of view of creating dialogue systems software, to make a conclusion about the appropriateness of research directions. Perform a detailed description of the stages of performing semantic and parsing with a description of the input, intermediate and output data. Determine the recommendations for the input data variability depending on the structure of the sentences of the subject area (the text described in natural language). Suggest rules for the formation of predicates for different parts of the Ukrainian language. Determine the typical set of service awards for filling the reference dictionary. Justify the implementation of each of the above operations for combining into a single computing structure, which is performed at certain stages. Develop procedures for passing each of the stages. Suggest rules for the formation of syntactic groups for certain pairs of tokens. **Conclusions:** The application of the developed method based on certain rules in predicates will allow to sample informatively significant information from a set of textual information and present it as an annotation (abstract).

Keywords: DSS, ACS, sentence, wordform, lexeme, syntactic group, semantic relation.