УДК 004:512

В.М. Ільман, В.І. Шинкаренко

СТРУКТУРНИЙ ПІДХІД ДО ПРОБЛЕМИ ВІДТВОРЕННЯ ГРАМАТИК

Для опису формальних граматик застосовано структурний підхід. Наведені деякі властивості граматичних структур та їх підструктур. Розглянуто проблему відтворення формально граматичних структур за мовними підструктурами.

Вступ

Як правило, в теорії формальних граматик [1, 2] розглядаються питання властивостей граматик, породжених ними мов, належності граматичних конструкцій до цих мов, тобто застосовується підхід «від граматик до мовних конструкцій». Але в багатьох інтелектуальних предметних областях виникає потреба відтворення, за заданими окремими граматичними конструкціями формальної мови необхідної формальної граматики. Відомо, що алгоритмічно така проблема частково розв'язується, при чому не однозначно. На даний момент існує досить широкий спектр алгоритмів відтворення формальних граматик за заданим набором конструкцій деякої припустимої мови [3]. Виходячи з цього виникає потреба доповнення формальних граматик атрибутами алгебри. Тому запропоновано до застосування математичний об'єкт формальна граматична структура [4, 5], в межах якої, для заданої предметної області, можливо створювати мовні конструкції проводити дослідження їх властивостей, визначати будову-структури та інше.

У матеріалах даної роботи розглянуто формальні породжувальні граматики з позицій формальних структур і на цій основі запропоновано новий підхід до розв'язання задачі, відтворення формальних породжувальних граматик.

Задача відтворення формальних породжувальних граматик досить актуальна при дослідженні структур об'єктів, наприклад, стилістики викладення матеріалу, розпізнавання структур графічних об'єктів [3, 6-8] тощо.

Граматичні структури та підструктури

Дамо спочатку декілька важливих та необхідних для подальшого викладення визначень конструктивних об'єктів, понять і позначень.

Розглянемо довільний термінальний алфавіт $A = \{o, a_1, a_2, \dots a_n\}$ з порожнім елементом o, та нетермінальний алфавіт $B = \{\alpha_1, \alpha_2, \dots \alpha_k\}$ і нехай $V = A \bigcup B$ їх словник. Позначимо *F(V)* – вільна мова побудована, наприклад, за допомогою операції конкатенації над словником *V*. Введемо у розгляд сигнатуру Σ як множину *m*-місних операцій $(*^m)$, наприклад, операції заміщення $(\rightarrow^2)$, операції конкатенації $(\boxtimes^2)$ та інших операцій і введемо [4, 5] наступний формальний об'єкт.

Визначення 1. Породжувальною формальною граматичною структурою формальної граматики з сигнатурою Σ і аксіоматикою Λ назвемо упорядковану трійку

$$C = \langle V, \Sigma, \Lambda \rangle, \quad (1)$$

де аксіоматика Λ може складатися з наступних *форм*: аксіом початку, аксіом виводу та інших аксіом і систем продукцій та їх властивостей. Аксіоматика Λ має скінчену кількість аксиом та продукцій.

Введений формальний об'єкт (1) відтворює будову формальної граматики і з одного боку узагальнює під аксіоматикою Λ синтаксис, аксіоми та правила виводу формальних систем математики [9], а з іншого боку - задає специфікацію алгебраїчної структури, яка застосовується при семантичному дослідженні програм [10].

Зрозуміло, що введена структура дозволяє виводити за допомогою операції заміщення різні конструктивні ланцюжки над словником *V* і сигнатурою $\Sigma \setminus \{\rightarrow^2\}$

згідно її визначеної аксіоматики. Ланцюжок $l \in F(V)$ зветься виведеним у формальній структурі *С*, якщо його вивід починається з аксіоми початку і в ньому застосовані операції за правилами та аксіомами аксіоматики. Так формальна структура зі словником $V = \{a,b\} \cup \{\sigma,\alpha\}$, сигнатурою $\Sigma = \{\rightarrow^2, \boxtimes^2\}$ та аксіоматикою:

$$\Lambda = \begin{cases} \text{продукції виводу :} \\ p_1 : \alpha \rightarrow b \mid \alpha = b \quad \text{– аксіома виводу;} \\ p_2 : \sigma \rightarrow a \boxtimes \alpha \quad \text{– аксіома початку;} \\ p_3 : \alpha \rightarrow b \boxtimes \alpha; \\ \text{властивості операцій :} \\ \beta \rightarrow c = c, \beta \rightarrow c \neq c \rightarrow \beta, \gamma \rightarrow c\beta \rightarrow cc = cc; \\ c \boxtimes \beta = c\beta, c\beta \neq \beta c, (c \boxtimes \beta) \boxtimes \gamma = c \boxtimes (\beta \boxtimes \gamma) = c\beta\gamma; \\ \beta, \gamma \in N; c \in F(V); \end{cases}$$

дозволяє вивести ланцюжки

$$a \boxtimes \alpha = a\alpha \in F(V),$$

$$a \boxtimes b = ab \in F(A);$$

$$a \boxtimes b \boxtimes \alpha = ab\alpha \in F(V),$$

$$a \boxtimes b \boxtimes b = abb \in F(A) \text{ і т. д.}$$

Множина виведених у структурі *С* ланцюжків $\{l_i\}$ таких, що $l_i \in F(A)$ утворює формальну мову $L(A) \subset F(A)$.

За визначенням 1 структура *С* є граматичною, тобто є повною у тому розумінні, що, так як і в граматиках, всі символи словника (окрім можливо порожнього) обов'язково використовуються в аксіомах і продукція її аксіоматики. Тому, зрозуміло, що за заданою аксіоматикою однозначно відтворюється формальна граматична структура і її граматика, а також породжується певна формальна мова. В подальшому розглядатимуться тільки граматичні структури з одноелементною сигнатурою заміщення і зі спрощеною аксіоматикою: з аксіомами початку та виводу. Так для вище наведеної аксіоматики спрощена – має вигляд:

$$\Lambda = \begin{cases} \alpha \rightarrow b \quad \text{– аксіома виводу,} \\ \sigma \rightarrow a\alpha \quad \text{– аксіома початку,} \\ \alpha \rightarrow b\alpha \end{cases}$$

за якою породжується формальна мова

$$L(A) = \{ab^k; k \in N\}.$$

Очевидно, для структури (1) в основному зберігаються результати отримані для формальних граматик, наприклад, у класі формально граматичних структур при введені конструктивних обмежень на продукції аксіоматики можливо виділити класи *ВС* - структур і *VС* - структур, які відповідають контекстно залежним і контекстно вільним граматикам відповідно, при чому, *VС* - структура є частковим випадком *ВС* - структури. Крім того для структур (1) можливо ввести поняття функціональної еквівалентності.

Визначення 2. Дві граматичні структури C_1 і C_2 функціонально еквівалентні (слабко), якщо вони породжують одну і ту ж мову, тобто $L(A_1) = L(A_2)$.

Визначення 3. Структура *С* зветься нескороченою, якщо продукції і аксіоми її аксіоматики задовольняють властивості $(x \rightarrow y; |x| \leq |y|, x, y \in F(V))$, де $|l|$ - довжина ланцюжка l.

Визначення 4. Структура *С* зветься o- вільною граматичною структурою, якщо її аксіоматика не містить у собі продукцій і аксіом типу $(x \rightarrow o; x \in F(V))$.

Не складно довести, що в будь якому класі еквівалентності граматичних структур завжди існує нормальна структура C_h, тобто структура продукції і аксіоми аксіоматики, якої мають властивість $(x \rightarrow y; x \in F(N))$. А також, що у всякому класі еквівалентності з нескороченою однозначною структурою існує однозначна o- вільна *ВС*- граматична структура.

Але для формальних граматичних структур можливо встановити нові алгебраїчні результати, наприклад, структура (1) є універсальною відносно вільної мови *F(V)* у тому розумінні, що структура визначена на словникові $V \subset F(V)$ і породжує множину ланцюжків *L(V)*, по операції заміщення за аксіоматикою Λ, таку, що має місце ланцюг за включенням $L(A) \subset L(V) \subset F(V)$. Серед множини формальних структур можливо виділити гомоморфні та *ізоморфні* структури. Так формальні структури C і $C_1 = \langle V_1, \Sigma_1, \Lambda_1 \rangle$ ізоморфні, якщо для будь яких елементів $b \in V$ і будь яких операцій $(*) \in \Sigma$, $(*_1) \in \Sigma_1$ та форм аксіоматик $\varphi \in \Lambda$, $\varphi_1 \in \Lambda_1$ існує взаємно однозначне відображення ρ множини V на множину V_1 і форм аксіоматики Λ на форми аксіоматики Λ_1 таке, що

$$\begin{aligned}&\rho(*(b)) = *_1(\rho(b), \quad \rho(\varphi(b,*(b))) = \\ &= \varphi_1(\rho(b),*_1(\rho(b)))\end{aligned}.$$

Так як нас цікавить питання відтворення формальних структур, то розглянемо їх будову за допомогою формальних граматичних підструктур.

Визначення 5. Структуру $C_1 = \langle V_1, \{\to^2\}, \Lambda_1 \rangle$ назвемо підструктурою структури $C = \langle V, \{\to^2\}, \Lambda \rangle$, якщо виконуються включення $V_1 \subseteq V$ і $\Lambda_1 \subseteq \Lambda$ - по спрощеній аксіоматиці, і скорочено це позначимо $C_1 \preceq C$. Підструктура C_1 є порожньою підструктурою, якщо вона породжує тільки порожню мову $L = \varnothing$, тобто для підструктури задовольняється хоча б одна з умов

1) $V_1 = \{o\} = \varnothing$;
2) $\Lambda_1 = \{x_i \to o; i \in J\}$;
3) ($V_1 = \{o\}$ і $\Lambda_1 = \{x_i \to o; i \in J\}$);
4) аксіоматика Λ_1 не може породжувати ні одного ланцюжка мови $L(V_1)$.

Зауваження 1. Очевидно, наведене визначення порожньої підструктури за умовами 1) – 4) еквівалентні згідно визначенню 2.

Зрозуміло, що довільна не порожня підструктура граматичної структури C частково зберігає за собою той же тип, який має структура C, тобто як сама структура так і її підструктури можуть належати до одного з класів, наприклад, BC, VC - структур, крім того ця підструктура може частково породжувати або зовсім не породжувати ні одного ланцюжка мови $L(A)$.

Породжувальні, повні та утворюючі граматичні підструктури

Визначення 6. Підструктуру C^* формальної граматичної структури (1) назвемо породжувальною підструктурою, якщо існує вивід *W(l)* ланцюжка *l* у структурі $C^* \preceq C$, такий, що $l \in L(A)$. Породжувальну підструктуру $C^* \preceq C$, в якій виводиться тільки один ланцюжок $l \in L(A)$ формальної мови граматичної структури C назвемо структурою ланцюжка *l* формальної мови *L(A)*.

Отже довільна підструктура граматичної структури *C* тоді і тільки тоді породжувальна, коли її аксіоматика Λ^* містить у собі хоча б по одній аксіомі виводу та початку і сукупність продукцій, пов'язаних хоча б з одним виводом $W(l)$ ланцюжка $l \in L(A)$.

Так як під виводом *W(l)* ланцюжка $l \in L(A)$ розуміється упорядкована послідовність безпосередньо виведених у структурі C_l^* проміжних ланцюжків [1, 2], то взагалі між структурою ланцюжка і його виводом існує тільки гомоморфне відношення, і тому вивід *W(l)* задає будову ланцюжка *l*.

Не складно бачити, що будь яка підструктура C_i структури є частково універсальною відносно вільної мови $F(V_i) \subseteq F(V)$, тобто і відносно мови $F(V)$.

Нехай C_1 і C_2 довільні підструктури структури *C*, під їх об'єднанням і перетином будемо розуміти $C_1 \cup C_2 = \langle V_1 \bigcup V_2, \Sigma, \Lambda_1 \bigcup \Lambda_2 \rangle$ і $C_1 \cap C_2 = \langle V_1 \bigcap V_2, \Sigma, \Lambda_1 \bigcap \Lambda_2 \rangle$, де $\bigcup, \bigcap$ - звичайні теоретико-множинні операції. Причому перетин вважається порожнім, якщо $V_1 \bigcap V_2 = \varnothing$ і $\Lambda_1 \bigcap \Lambda_2 = \varnothing$, або

($V_1 \bigcap V_2 = \varnothing$ і $\Lambda_1 \bigcap \Lambda_2 = \varnothing$), тоді, справедлива така теорема.

Теорема 1. Не порожній перетин (об'єднання) сукупності підструктур $\{C_i, i \in I\}$ o- вільної граматичної структури C також утворюють підструктуру даної структури.

Крім того сукупність усіх підструктур структури *C* є структурою – решіткою на операціях об'єднання та перетину.

Між двома підструктурами C_1 і C_2 можливо ввести відношення включення: за словником $C_1(V_1) \prec C_2(V_2); V_1 \subseteq V_2$, за аксіоматикою $C_1(\Lambda_1) \prec C_2(\Lambda_2); \Lambda_1 \subseteq \Lambda_2$. Під включенням підструктур $C_1 \prec C_2$ в подальшому розуміється включення за словником і за аксіоматикою цих підструктур. Отже між двома підструктурами C_1 і C_2 існує відношення включення ($\prec$) тоді і тільки тоді, коли мають місце включення $V_1 \subset V_2, \Lambda_1 \subseteq \Lambda_2$ або $V_1 \subseteq V_2, \Lambda_1 \subset \Lambda_2$, або наступне - $V_1 \subset V_2, \Lambda_1 \subset \Lambda_2$.

Визначення 7. Сукупність (не всіх порожніх) підструктур C_i, для яких виконується умова $\{C_i;\ i \in I, \bigcup_i C_i = C\}$ називається системою утворюючих підструктур структури *C*. Якщо система утворюючих підструктур структури C складається з двох підструктур C_1 і C_2, таких, що $C_1 \cup C_2 = C$ і $C_1 \cap C_2 = \varnothing$, то підструктура C_2 є доповненням підструктури C_1 до формальної структури *C*.

Система утворюючих підструктур називається повною системою у тому розумінні, що вона повністю відтворює формальну структуру *C* і в ній нема зайвих підструктур, які не впливають на відтворення структури *C*. Виходячи з того, що словник *V* і аксіоматика Λ формальної граматичної структури C є скінченими приходимо до висновку, що система утворюючих підструктур $\{C_i\}$ також є скінченою множиною.

Лема 1. Нехай підструктура C_2 є доповненням підструктури C_1 до формальної структури *C*, тоді множину всіх підструктур $\{C_i\}$ структури *C* можна розбити на три класи підструктур:

$$K_1 = \{C_j; C_j \prec C_1\},$$

$$K_2 = \{C_j; C_j \prec C_2\},$$

$$K_3 = \{C_j; C_j \cap C_1 \in K_1, C_j \cap C_2 \in K_2\}.$$

Наслідок 1. Очевидно, лема 1 має місце і в тому випадку, коли доповнення $C_2 = \bigcup_k C_k$, де $C_k \nprec C_1$ - підструктури формальної структури *C*.

Результат леми розбиття на класи є корисним при визначені будови множини підструктур формальної структури, зокрема будови її системи утворюючих підструктур.

Теорема 2. У будь якій системі утворюючих підструктур $S = \{C_i;\ i \in I, \bigcup_i C_i = C\}$ формальної граматичної структури C можливо виділити систему $S^* \subseteq S$ утворюючих породжувальних підструктур або побудувати на системі S систему підструктур S^* таку, що $S^* = \{C_j^*;\ j \in J \subset I, \bigcup_j C_j^* = C\}$.

Зазначимо, що теорема стосується як існування системи підструктур S^*, так і існування процедури побудови такої підсистеми,

За умовою теореми система S є утворюючою структури *C*, аксіоматики підструктур яких містять форми $(p_i; i \in I)$ необхідні для виведення у структурі *C* всіх ланцюжків мови *L(A)*. Тому для доведення теореми розіб'ємо скінчену систему S на дві підмножини S_1 - яка складається з породжувальних підструктур і S_2 - з не породжувальних підструктури так, що $S_1 \bigcup S_2 = S$ і $S_1 \bigcap S_2 = \varnothing$. Якщо система S_1 є утворюючою, тобто $S^* = S_1$, тоді теорема доведена. У протилежному випадку за теоремою 1 на підструктурах систем S_1 і S_2 за допомогою суперпозицій f_i над операціями ($\cup$) і ($\cap$) утворимо підструктури C_k на окремих формах $p_i = f_i(C_j; j \in J \subset I)$ або їх комбінаціях. Очевидно, в силу повноти множини форм

$(p_i;\ i \in I)$, об'єднання $\underset{k}{\cup} C_k$ зі словником V є породжувальною підструктурою з одного боку і утворюючою системою системи C з другого боку. Доповнюючи породжувальні підструктури системи S_1 підструктурою $\underset{k}{\cup} C_k$ знову отримаємо систему утворюючих підструктур граматичної структури C. Зрозуміло, що це доведення справедливе і у випадку, коли окремо одна з систем S_1 і S_2 є порожньою.

Серед сукупності підструктур $\{C_i\}$ граматичної структури C існують такі підструктури C_j, аксіоматика яких Λ_j повністю відтворює їх структуру, тобто $C_j = \left\langle V_j, \Sigma, \Lambda_j \right\rangle$. Назвемо такі підструктури повними підструктурами $\overline{C}_j$, а відповідні їм аксіоматики відтворюючими аксіоматиками $\overline{\Lambda}_j$.

Теорема 3. На будь-якій системі утворюючих підструктур S структури C можна побудувати систему утворюючих повних підструктур
$\overline{S} = \{\overline{C}_j;\ j \in J,\ \underset{j}{\cup} \overline{C}_j = C\}$.

Для доведення теореми розглянемо довільну підструктуру $C_i \in S$. Якщо ця структура не є o- вільною і на її словникові V_i за аксіоматикою Λ_i можливо вивести тільки порожній ланцюжок тоді за визначенням 5 за умов 1), 3) і 4) підструктура - порожня. За зауваженням 1 замінимо структуру C_i еквівалентною структурою з умовою 2) так, що аксіоматика $\overline{\Lambda}_j$ буде складатися тільки з продукцій вигляду $x_i \to o$, після заміни правих частин продукцій порожнім символом o, а словник $\overline{V}_j$ створимо з різних символів аксіоматики $\overline{\Lambda}_j$. Таким чином у цьому випадку маємо $\overline{C}_j \in S$.

Нехай тепер структура C_i не порожня, тоді приймемо аксіоматику Λ_i за аксіоматику $\overline{\Lambda}_j$, при цьому можливі такі випадки:

1) аксіоматика Λ_i повністю відтворює структуру C_i, тобто $V_i = \overline{V}_j$;

2) аксіоматика Λ_i не повністю відтворює структуру C_i, але $\overline{V}_j \subset V_i$.

З чого за сукупністю випадків 1) і 2) слідує включення, $\overline{C}_j \prec C_i$.

Якщо для аксіоматики $\Lambda_i = \overline{\Lambda}_i$ маємо $V_i \subset \overline{V}_i$, то розбиваючи аксіоматику $\Lambda_i = \overline{\Lambda}_j \bigcup \Lambda_{i-j}$, так, щоб $\overline{V}_j \subseteq V_i$ отримаємо і в цьому випадку включення $\overline{C}_j \prec C_i$. При цьому завершується доведення теореми.

Зауваження 2. За результатом теореми 2 маємо для систем утворюючих підструктур наступне включення $\overline{S}^* \subseteq \overline{S}$ і за теоремою 3 – ланцюг за включенням $\overline{S}^* \subseteq \overline{S} \subseteq S$.

Максимальні граматичні підструктури. Умови повноти системи утворюючих підструктур

З'ясуємо тепер питання критеріїв, за якими можливо встановити існування систем утворюючих підструктур формальної граматичної структури і визначимо ефективні критерії, за якими можливо відтворити граматичну структуру за структурами утворюючих ланцюжків заданої формальної мови.

Для розв'язку проблема існування критеріїв про знаходження систем утворюючих підструктур скористуємося алгебраїчним підходом [11].

Визначення 8. Нехай C довільна граматична структура (1), тоді підструктура C_m структури C називається максимальною підструктурою $C_m \prec C$, якщо не існує такої підструктури $C_1 \prec C$, для якої мало б місце власне включення $C_m \prec C_1$.

Позначимо p_i довільну продукцію аксіоматики Λ структури C. Тепер, підструктура C_m буде максимальною відносно структури C тоді і тільки тоді, коли для будь якого елемента $v \vec{\in} C \triangleleft C_m$ такого, що $v \in V \bigcup \{p_i, i \in I\}$ має місце $C_m \boxplus \{v\} = C$.

Тут під різницею $C \lhd C_m$ структур розуміється підструктура: $\langle V \setminus V_m, \Sigma, \Lambda \rangle$ або $\langle V, \Sigma, \Lambda \setminus \Lambda_m \rangle$ формальної граматичної структури *C*, а операція ($\boxplus$) визначає приєднання елемента *v* до словника або аксіоматики підструктури C_m до структури *C*.

Позначимо *M* множину всіх підструктур максимальних щодо структури *C*. Зрозуміло, що граматична структура *C* має скінчену кількість максимальних підструктур. Для подальшого необхідна наступна лема розширення будь-якої підструктури граматичної структури $C = \langle V, \Sigma, \Lambda \rangle$.

Лема 2. Будь яку підструктуру $C_1 \prec C$ можливо розширити до максимальної підструктури $C_m \in M$ структури *C*.

За ствердженням леми маємо, що для довільної підструктури C_1 структури *C* у множині *M* існує підструктура C_m, що можливе тільки таке включення $C_1 \prec C_m$, бо в протилежному випадку виконується включення $C_1 \succ C_m$ і підструктура C_1 не є власною підструктурою структури *C*. Припустимо, що для підструктури C_1 у множині *M* не існує підструктури $C_m \succ C_1$. Тоді приєднуючи до підструктури C_1 усі елементи $v \in V \bigcup \{p_i, i \in I\}$, яких нема у цій підструктурі крім одного $v^* \vec{\notin} C_1$ за скінчену кількість кроків отримаємо максимальну підструктуру $C_{1,m}$ щодо структури *C*, що призводить до протиріччя з припущенням. Таким чином будь-яку власну підструктуру граматичної структури завжди конструктивно можливо розширити до максимальної підструктури.

Тепер розглянемо критерії, за яким визначається, що система підструктур граматичної структури є утворюючою системою.

Теорема 4. Для того, щоб система підструктур $S = \{C_i; i \in I\}$ граматичної структури *C* була утворюючою необхідно і достатньо, щоб для будь якої підструктури $C_m \in M$ у системі *S* знайшовся хоча б один елемент $v \in \{V_i \bigcup \{p_{i,j}; j \in J\}; i \in I\}$ такий, що $v \vec{\notin} C_m$.

За необхідністю система *S*- утворююча щодо структури *C*, тобто $\bigcup_{i \in I} C_i = C$ і так як для максимальної підструктури $C_m \in M$, за її визначенням існує такий елемент $v \vec{\notin} C_m$, що $v \vec{\in} C \lhd C_m$, то в системі *S* існує хоча б одна підструктура C_j для якої $v \vec{\in} C_j$.

При доведенні достатності розглянемо таку систему *S*, що для будь-якої максимальної підструктури $C_m \in M$ структури *C* в ній існує хоча б один елемент $v \vec{\in} C_i \in S$, такий, що $v \vec{\notin} C_m$. Доведемо, що система *S* є утворюючою, тобто $\bigcup_{i \in I} C_i = C$.

Припустимо, що система *S* не є системою утворюючих підструктур - $\bigcup_{i \in I} C_i \neq C$, тоді користуючись результатами леми 2, будь-яку підструктуру $C_i \in S$ розширимо до максимальної підструктури $C_{i,m} \in M$ структури *C* з чого маємо включення $C_i \prec C_{i,m}$. Але за умовою у підструктурі C_i існує такий елемент v_i, для якого $v_i \vec{\notin} C_{i,m}$, що призводить до порушення включення $C_i \prec C_{i,m}$. Таким чином наше припущення про те, що сукупність підструктур *S* не є системою утворюючих підструктур породжувальної граматичної структури *C* не вірне і теорема доведена.

Зауваження 3. Нескладно бачити, що теорема 4 виконується також для систем породжувальних підструктур S^* і повних під структур $\overline{S}$ граматичної структури *C*.

Алгоритм побудови утворюючих підструктур. Мінімальні граматичні підструктури

Визначення 9. Виведені ланцюжки у структурах $\overline{C}_i^*$ системи утворюючих підструктур $\overline{S}^*$ структури *C* називається зразком s_l формальної мови *L(A)*.

Отже за результатами леми 1 та теореми 4 можливо запропонувати наступний алгоритм побудови системи утворюючих підструктур формальної структури C і дослідити будову цієї утворюючої системи підструктур:

крок 1) побудувати множину максимальних підструктур М структури С;

крок 2) за елементами $v_i \vec{\in} C$, які не входять до максимальних підструктур побудувати систему утворюючих підструктур, що містять у собі елементи v_i;

крок 3) на системі утворюючих підструктур формальної системи побудувати структурний граф залежності підструктур;

крок 4) виділити у системі утворюючих підструктур підструктуру C_1 і доповнену до неї підструктуру C_2, відносно яких за лемою 1 побудувати три класи K_1, K_2 та K_3;

крок 5) з класів K_1 і K_2 видиліти максимальні за включенням підструктури, тобто такі $C_k \in K_i$, що для всіх $C_j \in K_i;\ j \neq k \quad C_j \prec C_k$;

крок 6) на об'єднані незалежних структур класів K_1 і K_2 побудувати повну систему утворюючих підструктур формальної структури.

Приклад 1. Застосуємо наведений алгоритм до формальної структури C з аксіоматикою:

$$\Lambda = \begin{cases} p_1 : \sigma \to a\alpha - \text{аксіома початку} \\ p_2 : \alpha \to a\alpha \\ p_3 : \alpha \to b\beta \\ p_4 : \beta \to b\beta \\ p_5 : \beta \to c\gamma \\ p_6 : \gamma \to c\gamma \\ \text{аксіоми виводу} \\ p_7 : \gamma \to c \\ p_8 : \beta \to c \end{cases} \qquad \text{та мовою } L = \{a^k b^n c^m;\ k, n, m \in N\}. \qquad (2)$$

І побудуємо породжувальну систему $\overline{S}^*$ граматичної структури на аксіоматиці (2).

Крок 1. Так як система $\overline{S}^*$ повна, тому для неї і максимальної множини M достатньо скористуватися відповідними аксіоматиками. Так множина аксіоматик для сукупності M є $\{\Lambda \setminus p_8, \Lambda \setminus p_7, \Lambda \setminus p_6, \Lambda \setminus p_5, \Lambda \setminus p_4, \Lambda \setminus p_2\}$. Підструктури з аксіоматиками $\Lambda \setminus p_1$ та $\Lambda \setminus p_3$ не залучені до множини M бо вони не породжувальні.

Крок 2. Система утворюючих підструктур повинна включати такі елементи: $p_2, p_4, p_5, p_6, p_7, p_8$. Такою системою, наприклад, буде утворююча система з аксіоматиками:

$$\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4, \Lambda_5\}, \qquad (3)$$

де $\Lambda_1 = \{p_1, p_3, p_8\}$,

$\Lambda_2 = \{p_1, p_2, p_3, p_8\}$,

$\Lambda_3 = \{p_1, p_3, p_4, p_8\}$,

$\Lambda_4 = \{p_1, p_3, p_5, p_7\}$ і

$\Lambda_5 = \{p_1, p_3, p_5, p_6, p_7\}$.

Крок 3. Структурний граф системи утворюючих підструктур формальної структури C з відповідними аксіоматиками за включеннями: $\overline{C}_4^* \prec \overline{C}_5^* \prec C$, $\overline{C}_1^* \prec \overline{C}_3^* \prec C$ і $\overline{C}_1^* \prec \overline{C}_2^* \prec C$; наведений на рис. 1.

Крок 4. Серед системи підструктур з аксіоматиками (3) виберемо структуру $\overline{C}_5^*$, для якої доповненою до формальної структури C буде структура $\overline{C}_2^* \cup \overline{C}_3^*$. Так

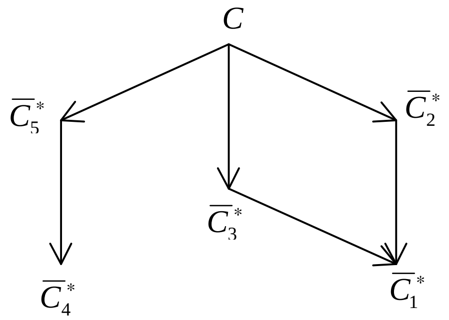

Рис. 1. Структурний граф системи утворюючих підструктур

як $\overline{C}_5^* \cup \overline{C}_2^* \cup \overline{C}_3^* = C$ і аксіоматика структури $\overline{C}_5^* \cap (\overline{C}_2^* \cup \overline{C}_3^*)$ складається з продукцій $\{p_1, p_3\}$, які дозволяють породжувати лише порожню мову тому система утворюючих підструктур розбивається на класи $K_1 = \{\overline{C}_4^*, \overline{C}_5^*\}$, $K_2 = \{\overline{C}_1^*, \overline{C}_2^*, \overline{C}_3^*\}$ і $K_3 = \varnothing$.

Крок 5. За структурним графом максимальними за включенням підструктурами класу K_1 є $\overline{C}_5^*$, а класу K_2 - $\overline{C}_2^*$, $\overline{C}_3^*$.

Крок 6. На незалежних підструктурах класів K_1 та K_2 можливо отримати повну систему утворюючих породжувальних підструктур $\{\overline{C}_2^*, \overline{C}_3^*, \overline{C}_5^*\}$.

Покажемо, що побудована таким чином повна система утворюючих підструктур не єдина.

Для цього спочатку побудуємо повний граф системи утворуючих підструктур . Граф вважається повним у тому розумінні, що включає всі можливі повні (за визначенням 8) породжуючі підструктури формальної структури *C*. Граф будується за рекурентною процедурою: спочатку знаходяться всі максимальні підструктури формальної структури *C*, потім - максимальні підструктури для них і т. д.

Повний граф для прикладу 1 показано на рис. 2 (для наочності та спрощення представлення деякі вершини графа зазначені кілька разів).

Аксіоматики відповідних підструктур, для якого:

$\Lambda_{1,1} = \{p_1, p_2, p_3, p_4, p_5, p_6, p_7\}$,

$\Lambda_{1,2} = \{p_1, p_2, p_3, p_4, p_8\}$,

$\Lambda_{1,3} = \{p_1, p_2, p_3, p_4, p_5, p_7, p_8\}$,

$\Lambda_{1,4} = \{p_1, p_2, p_3, p_5, p_6, p_7, p_8\}$,

$\Lambda_{1,5} = \{p_1, p_3, p_4, p_5, p_6, p_7, p_8\}$,

$\Lambda_{2,1} = \{p_1, p_3, p_4, p_5, p_6, p_7\}$,

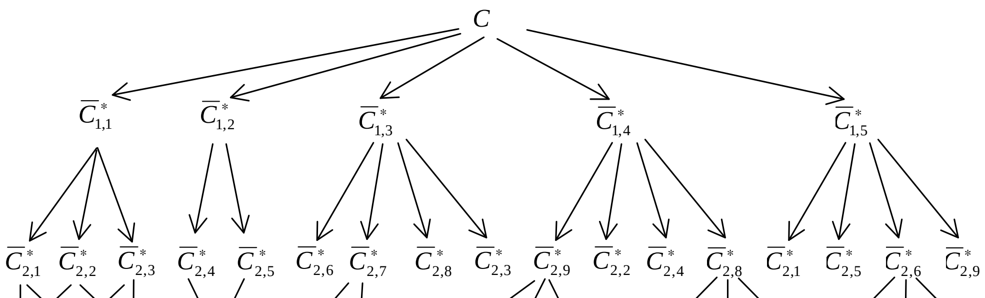

Рис. 2. Повний структурний граф системи утворюючих підструктур

$\Lambda_{2,2} = \{p_1, p_2, p_3, p_5, p_6, p_7\}$,

$\Lambda_{2,3} = \{p_1, p_2, p_3, p_4, p_5, p_7\}$,

$\Lambda_{2,4} = \{p_1, p_2, p_3, p_8\}$,

$\Lambda_{2,5} = \{p_1, p_3, p_4, p_8\}$,

$\Lambda_{2,6} = \{p_1, p_3, p_4, p_5, p_7, p_8\}$,

$\Lambda_{2,7} = \{p_1, p_2, p_3, p_4, p_8\}$,

$\Lambda_{2,8} = \{p_1, p_2, p_3, p_5, p_7, p_8\}$,

$\Lambda_{2,9} = \{p_1, p_3, p_5, p_6, p_7, p_8\}$,

$\Lambda_{3,1} = \{p_1, p_3, p_5, p_6, p_7\}$,

$\Lambda_{3,2} = \{p_1, p_3, p_4, p_5, p_7\}$,

$\Lambda_{3,3} = \{p_1, p_2, p_3, p_5, p_7\}$,

$\Lambda_{3,4} = \{p_1, p_3, p_8\}$,

$\Lambda_{3,5} = \{p_1, p_3, p_5, p_7, p_8\}$,

$\Lambda_{4,1} = \{p_1, p_3, p_5, p_7\}$.

За основні утворюючі підструктури (утворюючий базис) можливо взяти кінцеві породжувальні структури графа $C_{4,1}$ і $C_{3,4}$ (рис. 2) з відповідними аксіоматиками (тобто такі, що не мають підструктур за відношенням включення) та додати до них підструктури з аксіоматиками

$\Lambda \setminus \Lambda_{4,1} \setminus \Lambda_{3,4}$. Отже отримаємо нову систему утворюючих підструктур з аксіоматиками $\{\Lambda_{4,1}, \Lambda_{3,4}, \Lambda \setminus \Lambda_{4,1} \setminus \Lambda_{3,4}\}$. Таким чином на структурах двох простих ланцюжків $l_{4,1} = abc$ і $l_{3,4} = abcc$ та рекурсивних продукціях p_2, p_4 і p_6 аксіоматики (2) відтворюється формальна граматична структура *C*.

Визначення 10. Систему утворюючих підструктур структури *C* побудовану на кінцевих породжувальних підструктурах структурного графа назвемо мінімальною системою утворюючих підструктур S_o структури *C*.

Зауваження 4. Запропоновану методику побудови повної системи утворюючих підструктур також зручно застосовувати тому випадку коли відомі дерева виводів утворюючих ланцюжків, при цьому слід звернути увагу на те, що однозначне відтворення формальних систем можливе тільки для *VC* - структур.

Відтворення граматичних структур за зразками формальної мови

Перш ніж перейти до розгляду задачі відтворення формальних граматик за заданими структурами ланцюжків зразка, введемо деякі припущення:

1) рішення щодо ланцюжків зразка формальної мови, кількості його елементів та повноти приймаються поза межами даної методи;

2) виходячи з того, що вибраному зразку може відповідати декілька мов, рішення щодо прийнятної формальної мови приймається поза межами даної методи;

3) формальні структури ланцюжків зразку приймаються за систему утворюючих підструктур граматичної структури прийнятної мови, тобто мають загальну граматичну структуру;

4) структури ланцюжків зразка утворюють систему $\overline{S}^*$.

Нехай задано зразок $s_l = \{l_i; i \in I, l_i \in L(A)\}$ і система структур ланцюжків цього зразка $S = \{\overline{C}_i^*; i \in I\}$. Тепер можливо поставити наступну задачу: на заданій парі $\langle s_l, S \rangle$ побудувати мінімальну систему утворюючих підструктур S_o (мінімальну формальну граматичну структуру C_o).

Алгоритм розв'язку задачі спирається на вище наведену схему побудови системи утворюючих підструктур і полягає в наступному:

1. провести аналіз системи структур *S* на виконання пункту 3) щодо впорядкування підструктур системи *S* загальній граматичній структурі, для цього слід скористатися операцією посимвольного перетину ланцюжків і внести корективи у відповідні структури ланцюжків зразка. Під посимвольним перетином тут мається на увазі однакові послідовності алфавітних символів присутні у кожному з ланцюжків (таких послідовностей може бути декілька);
2. з урахуванням результату кроку 1), побудувати структурні графи виводів ланцюжків заданого зразка;

3. за результатами кроку 2), на системі S побудувати впорядковану систему S_y і прийняти її за систему утворюючих підструктур граматичної структури;

4. на системі утворюючих підструктур формальної системи побудувати структурний граф залежності підструктур;

5. включити кінцеві підструктури $\overline{C}^*_{jk} \in S_y$ структурного графа до мінімальної системи утворюючих підструктур S_o;

6. за допомогою багатократного застосування операцій $(\cap, \triangleleft)$ над кінцевими структурами $\overline{C}^*_{jk}$ структурного графа й іншими структурами системи S_y та структурами результатів цих операцій (або над структурними графами виводів ланцюжків цих структур) виділити прості структури, які відповідають різним продукціям p_k структур системи S_y;

7. доповнити систему S_o простими структурами.

Відтворену таким чином формальну структуру C назвемо мінімальною структурою C_o.

Зауваження 5. Виходячи з припущення 4) всі дії та перетворення за алгоритмом можна виконувати над аксіоматиками відповідних структур і підструктур.

Для порівняння результату побудови мінімальної граматичної структури скористуємося прикладом 7.5 роботи [3].

Приклад 2. Нехай задано зразок $s_l = \{l_i; i = 1\ldots 7\}$ з ланцюжками: $l_1 = ca^3b$ ($l_1 = caaab$), $l_2 = b^2a^2b$, $l_3 = ca^2b$, $l_4 = b^2ab$, $l_5 = cab$, $l_6 = b^3$ і $l_7 = cb$; та їх відповідними структурами з аксіоматиками:

$$\Lambda_1 = \begin{cases} \sigma \to c\alpha, \\ \alpha \to a\beta, \\ \beta \to a\gamma, \\ \gamma \to ab \end{cases};$$

$$\Lambda_2 = \begin{cases} \sigma \to bb\delta, \\ \delta \to a\eta, \\ \eta \to ab \end{cases};$$

$$\Lambda_3 = \begin{cases} \sigma \to c\varphi, \\ \varphi \to a\psi, \\ \psi \to ab \end{cases};$$

$$\Lambda_4 = \begin{cases} \sigma \to bb\mu, \\ \mu \to ab \end{cases};$$

$$\Lambda_5 = \begin{cases} \sigma \to c\lambda, \\ \lambda \to ab \end{cases};$$

$$\Lambda_6 = \{\sigma \to bbb;$$

$$\Lambda_7 = \{\sigma \to cb;$$

в яких перша продукція – аксіома початку, а остання – аксіома виводу.

Крок 1. Аналізуючи наведені структури за посимвольною операцією (на множинах символів) перетину ланцюжків, наприклад, $l_6 \cap l_7 = l_5 \cap l_6 = b$ з'ясовуємо, що в усіх аксіомах виводу повинна бути продукція виводу символу b. Отже аксіоми виводу в аксіоматиках Λ_1, …, Λ_5 слід замінити продукціями $\xi \to a\tau, \tau \to b$; і нетермінали $\gamma, \eta, \psi, \mu, \lambda$ у відповідних аксіоматиках замінити символом ξ, а у шостій і сьомій аксіоматиках замінити аксіоми початку відповідними парами продукцій

$$\Lambda_6 = \begin{cases} \sigma \to bb\tau - \textit{аксіома початку} \\ \tau \to b - \textit{аксіома виводу} \end{cases};$$

$$\Lambda_7 = \begin{cases} \sigma \to c\tau - \textit{аксіома початку} \\ \tau \to b - \textit{аксіома виводу} \end{cases}.$$

Крок 2. Тепер маємо можливість побудувати структурні графи виводів ланцюжків зразка (рис. 3).

Крок 3. Аналіз підструктур за структурними графами виводів, з урахуванням перейменування нетермінального алфавіту

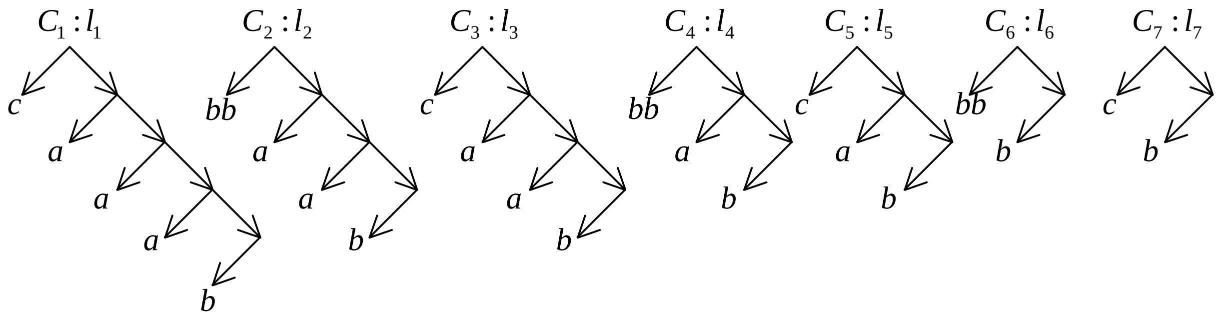

Рис. 3. Структурні графи виводів ланцюжків зразка

надає можливість виявити наступні відношення за включенням між підструктурами (збережені попередні позначки): $\overline{C}_6^* \prec \overline{C}_4^* \prec \overline{C}_2^*$ та $\overline{C}_7^* \prec \overline{C}_5^* \prec \overline{C}_3^* \prec \overline{C}_1^*$.

Крок 4. За результатами кроку 3 нескладно побудувати структурний граф (рис. 4) залежності підструктур по включеннях системи $S_y = \{\overline{C}_i^*;\ i = 1 \ldots 7\}$ формальної граматичної структури C.

Крок 5. Кінцевими структурами графа є підструктури $\overline{C}_6^*$ і $\overline{C}_7^*$.

Крок 6. За операцією $(\cap)$ на кінцевих структури $\overline{C}_6^*$ і $\overline{C}_7^*$ та на інших структурах системи S_y (див. рис. 3) виділяється тільки одна рекурсивна продукція $\tau \to \alpha\tau$.

Крок 7. Поєднавши аксіоматику кінцевих підструктур з продукцією винайдену на кроці 6 маємо аксіоматику мінімальної граматичної структури

$$\Lambda_o = \begin{cases} \textit{аксіоми початку}: \\ \sigma \to bb\tau \\ \sigma \to c\tau \\ \tau \to b - \textit{аксіома виводу} \\ \tau \to a\tau \end{cases}$$

за якою породжується формальна мова $L(A) = \{\{bba^k b\} \cup \{ca^n b\};\ k, n = 0, 1, 2 \ldots\}$.

Висновки

У матеріалах даної роботи розглянуто формальні граматичні породжувальні структури та їх підструктури. Визначені їх властивості та з'ясована можливість представлення граматичних структур через системи утворюючих підструктур. Отримано критерій повноти систем утворюючих підструктур формальної граматичної структури і запропоновано алгоритмічне розв'язання проблеми побудови систем утворюючих підструктур. Таким чином розроблений формальний апарат, який дозволяє досліджувати такі об'єкти, як граматики, включаючи графічні граматики зі специфічними операціями поєднання терміналів та нетерміналів у ланцюжки.

Запропонований апарат зі зміною операцій та оксіоматики дозволяє досліджувати й інші об'єкти, такі як графи, алгоритми, тощо.

Запропонована методика алгоритмічного розв'язання задачі відтворення граматичних структур та породжених ними мов за структурами їх мовних конструкцій дозволить вирішувати задачі розпізнавання структур графічних об'єктів на відміну задачі розпізнавання самих графічних об'єктів за наданими структурами, скажімо у вигляді графічних граматик.

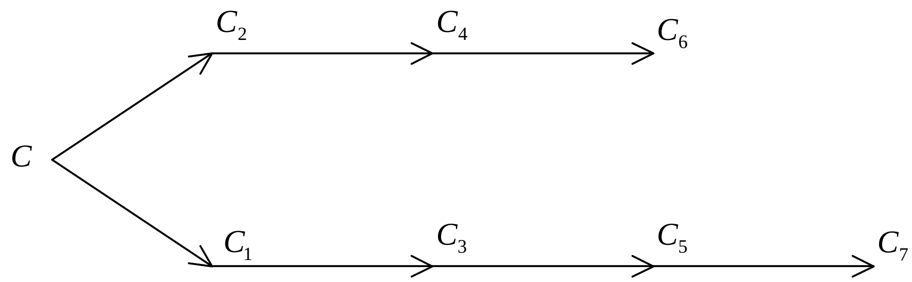

Рис. 4. Структурний граф залежності підструктур

1. *Гинзбург С.* Математическая теория контекстно-свободных языков. - М.: Мир, 1970. – 328 с.
2. *Гладкий А.В.* Формальные грамматики и языки. - М.: Наука, 1973. – 368 с.
3. *Фу К.С.* Структурные методы в распознавании образов. - М.: Мир, 1977. – 318 с.
4. *Ильман В.М., Разумов С.Ю.* Структурный подход в формальных системах // Проблеми математичного моделювання. Міжнар. науково-метод. конф. Тези доповідей. Дніпродзержинськ, 2005. – С. 147 – 148.
5. *Ільман А.В., Ільман В.М.* Формально – структурне моделювання економічних систем. // Вісн. Дніпропетровського національного у-ту залізничного транспорту. – Вип.. 10. – Д.: Вид-во Дніпропетр. нац. ун-ту залізн. трансп., 2006. с. 173 – 180.
6. *Павлюк О.В., Савчинський Б.Д.* Ефективний синтаксический аналіз та розпізнавання структурованих зображень // Управляющие системы и машины. – 2005. – № 5 . – С. 13-24.
7. *Ту Дж., Гонсалес Р.* Принципы распознавания образов. - М.: Мир, 1978. – 411с.
8. Котенко И.В. Восстановление формальных грамматик, задающих сценарии компьютерных атак, по прецендентам // Искусственный интеллект. – 2002. – № 3 . – С. 581-589.
9. *Смальян Р.* Теория формальных систем. - М.: Наука, 1981. – 209 с.
10. *Калинин А.Г., Мацкевич И.В.* Универсальные языки программирования. Семантический подход. М.: Радио и связь, 1991. – 400 с.
11. *Глушков В.М., Цейтлин Г.Е., Ющенко Е.Л.* Алгебра, языки, программирование. - Киев: Наук. думка, 1978. – 320 с.

Отримано 11.12.2006

Про авторів:

Ільман Валерій Михайлович,
кандидат фіз.-мат. наук, доцент кафедри «Комп'ютерні інформаційні технології» Дніпропетровського національного університету залізничного транспорту імені академіка В. Лазаря на,

Шинкаренко Віктор Іванович,
канд. техн. наук, доцент кафедри «Комп'ютерні інформаційні технології» Дніпропетровського національного університету залізничного транспорту імені академіка В. Лазаряна.

Місце роботи авторів:

Дніпропетровський національний університет залізничного транспорту
імені академіка В. Лазаряна.
м. Дніпропетровськ, вул. Лазаряна 2,
каф. КІТ, ДНУЗТ
тел. 8-056-373-15-35.

Дніпропетровський національний університет залізничного транспорту
імені академіка В. Лазаряна.
м. Дніпропетровськ, вул. Лазаряна 2,
каф. КІТ, ДНУЗТ
тел. 8-056-373-15-35.
e-mail: ccp@diit-70.dp.ua