

## ГРАФОВЫЙ СПОСОБ АНАЛИЗА СИНТАКСИЧЕСКИХ ДИАГРАММ

Ю.Д. Рязанов, П.В. Крамаренко

*Белгородский государственный технологический университет им. В.Г. Шухова*

Классическая технология проектирования трансляторов предполагает разработку спецификации транслятора в виде КС-грамматики [1] и, используя различные методики [2, 3], преобразование ее в программу транслятора.

Одним из удобных способов описания языков является синтаксическая диаграмма (СД), которую впервые применил Н. Вирт для описания языка Паскаль [4]. Сейчас известны примеры использования СД при ручном проектировании трансляторов [5 — 7]. В работах [8, 9] описаны алгоритмы построения рекурсивных и нерекурсивных программ-распознавателей линейной сложности по детерминированным СД. Там же представлены алгоритмы анализа табличного представления псевдодетерминированных СД [10] с целью определения их принадлежности множеству детерминированных СД, ориентированные на компьютерную реализацию. В этой работе предлагается другой, графовый способ анализа СД, который может быть использован как при автоматическом, так и при ручном проектировании трансляторов.

С целью обеспечения возможности описания алгоритма анализа, определим СД пятеркой  $D = (T, N, S, G, F)$ , где  $T$  — конечное множество терминалов;  $N$  — конечное множество нетерминалов;  $S \in N$  — начальный нетерминал;  $G = (V, E)$  — ориентированный граф, где

$V = V_T \cup V_N \cup V_u \cup V_{вход} \cup V_{выход}$ , где

$V_{вход}$  — множество точек входа,  $|V_{вход}| = |N|$ ;

$V_T$  — множество терминальных вершин;

$V_N$  — множество нетерминальных вершин;

$V_u$  — конечное множество узлов;

$V_{выход}$  — множество точек выхода,  $|V_{выход}| = |N|$ ;

$E = E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5$ , где

$E_1 \subseteq \{(a, b) \mid a \in V_{вход}, b \in V_u\}$  — множество входных дуг;

$E_2 \subseteq \{(a, b) \mid a \in V_u, b \in V_{выход}\}$  — множество выходных дуг;

$E_3 \subseteq \{(a, b) \mid a \in V_u, b \in V_T \cup V_N\}$  — множество дуг, выходящих из узлов;

$E_4 \subseteq \{(a, b) \mid a \in V_T \cup V_N, b \in V_u\}$  — множество дуг, входящих в узлы;

$E_5 \subseteq \{(a, b) \mid a \in V_u, b \in V_u\}$  — множество  $\varepsilon$ -дуг, соединяющих узлы;

$F : V_T \cup V_N \rightarrow T \cup N$  — отображение множества вершин на множество терминалов и нетерминалов.

Каждому нетерминалу соответствует связная компонента графа. Компонента именуется соответствующим нетерминалом, имеет только одну точку входа и одну точку выхода и конечное множество вершин других типов. Точки входа и выхода на диаграмме компоненты не изображаются. Нетерминальная вершина изображается прямоугольником, в который вписан нетерминальный символ в соответствии с отображением  $F$ . Терминальная вершина изображается кружочком, в который вписан терминальный символ в соответствии с отображением  $F$ . Узел изображается на диаграмме жирной точкой. В точку входа не входит ни одна дуга и выходит конечное множество дуг (входные дуги компоненты). Узлы, в которые входят входные дуги, называются начальными. Из точки выхода не выходит ни одна дуга и входит конечное множество дуг (выходные дуги компоненты). Узлы, из которых выходят выходные дуги, называются заключительными. Каждая дуга, за исключением входных и выходных дуг, может выходить из узла и входить в терминальную или нетерминальную вершину или другой узел, либо выходить из терминальной или нетерминальной вершины и входить в узел. В каждую терминальную и нетерминальную вершину входит только одна дуга и выходит только одна дуга. На количество дуг, входящих в узлы и выходящих из них, ограничений нет. На рис. 1 приведен пример синтаксической диаграммы.

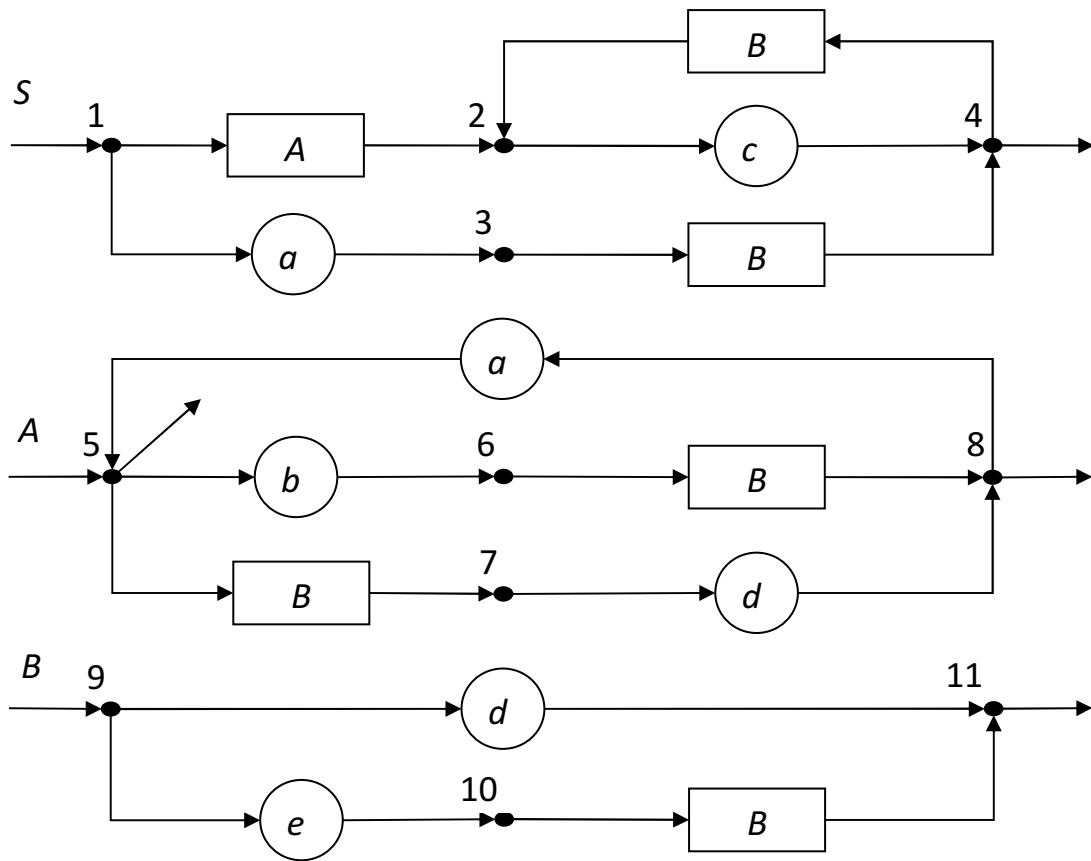


Рис. 1. Синтаксическая диаграмма

СД является псевдодетерминированной, если каждая ее компонента содержит только один начальный узел, не имеет  $\epsilon$ -дуг и любые две дуги, выходящие из любого узла, входят в вершины, содержащие различные символы. В работе [10] показано, что любую СД можно преобразовать в эквивалентную ей псевдодетерминированную СД.

СД является детерминированной, если она псевдодетерминированная и каждый ее узел детерминированный. Узел является детерминированным, если множества выбора (определяется ниже) любых двух дуг, выходящих из него, не пересекаются. Отметим, что дуги, выходящие из заключительного узла псевдодетерминированной компоненты СД, могут идти в терминальную, нетерминальную вершину или в точку выхода.

Процесс вывода терминальной цепочки в СД можно представить как «движение» по дугам от точки входа начальной компоненты к ее точке выхода. При этом если дуга идет в терминальную вершину, то вписанный в нее символ добавляем в терминальную цепочку, если дуга идет в нетерминальную вершину, то переходим в соответствующую компоненту и движемся по ней аналогичным образом до точки выхода, после чего возвращаемся в предыдущую компоненту и продолжаем движение. После прохождения выходной дуги

начальной компоненты в терминальную цепочку добавляем концевой маркер ( $\perp$ ) и вывод заканчивается.

Символ  $x$ , который может быть добавлен в терминальную цепочку непосредственно после прохождения дуги  $e$ , принадлежит множеству выбора дуги  $e$ .

Для определения детерминированности псевдодетерминированной СД нужно найти множества выбора всех дуг. Для этого построим граф анализа СД по следующим правилам.

Вершинами графа являются узлы СД, терминалы и концевой маркер.

Если в СД существует дуга из узла  $u_i$  в терминальную вершину с терминалом  $x$ , то в граф анализа СД добавить дугу из вершины  $u_i$  в вершину  $x$ .

Если в СД существует дуга из узла  $u_i$  в нетерминальную вершину с нетерминалом  $X$ , то в граф анализа СД добавить дугу из вершины  $u_i$  в вершину  $u_j$ , где  $u_j$  — начальный узел компоненты, соответствующей нетерминалу  $X$ .

Если в компоненте, соответствующей нетерминалу  $X$ , существует дуга из узла  $u_i$  в точку выхода (выходная дуга), то в граф анализа СД добавить дугу из вершины  $u_i$  в вершину  $u_j$ , если в СД есть дуга из нетерминальной вершины с нетерминалом  $X$  в узел  $u_j$ .

Процесс построения графа продолжается, пока по указанным правилам можно добавить хотя бы одну дугу.

Граф анализа СД (рис. 1) представлен на рис. 2.

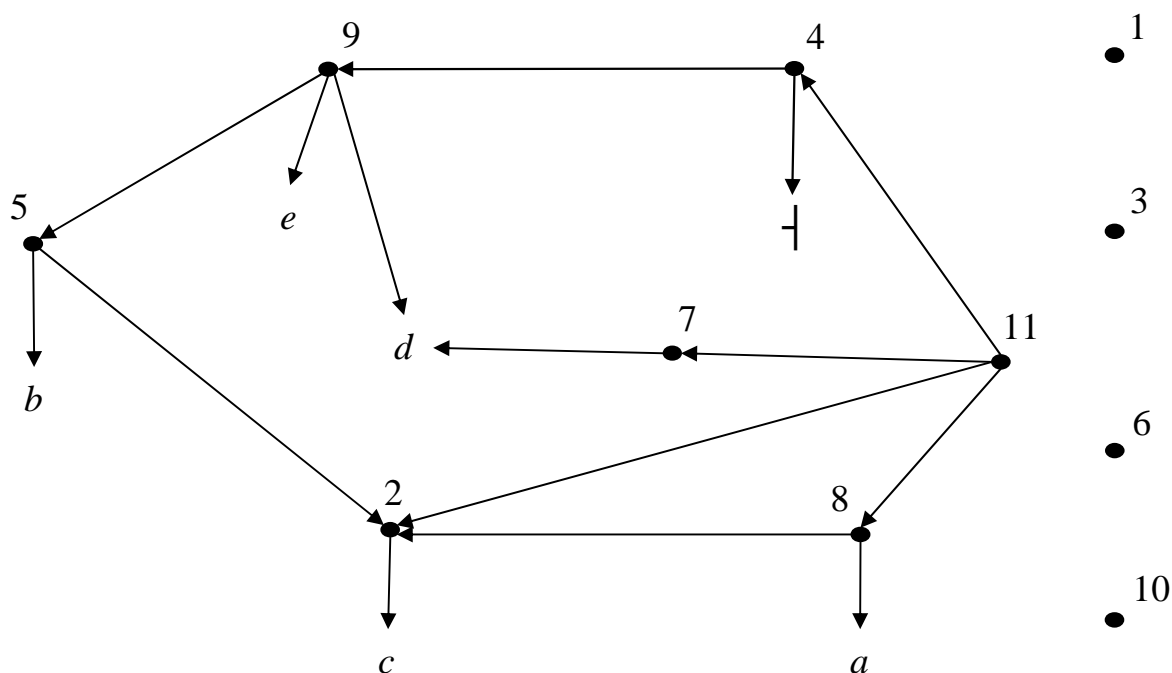


Рис. 2. Граф анализа синтаксической диаграммы

Граф анализа СД обладает важным свойством: если из вершины  $u$  существует путь в вершину  $x$ , то терминал  $x$  может быть добавлен в цепочку в процессе вывода после прохождения узла  $u$ . Следовательно, множество выбора дуги, входящей в нетерминальную вершину с нетерминалом  $X$ , содержит все терминалы, к которым существует путь в графе анализа СД из начального узла компоненты, соответствующей нетерминалу  $X$ , а множество выбора любой выходной дуги в компоненте, соответствующей нетерминалу  $X$ , содержит все терминалы, к которым существует путь в графе анализа СД из вершин, соответствующих узлам СД, в которые входят дуги из нетерминальных вершин с нетерминалом  $X$ . Множество выбора дуги, входящей в терминальную вершину, определяется тривиально и содержит только один терминал, записанный в этой вершине.

Результат определения множества выбора дуг с использованием графа анализа СД представлен в таблице 1. Строки соответствуют дугам СД. В первом столбце таблицы записаны обозначения дуг СД. Дуга, выходящая из узла  $u$  в вершину с символом  $x$  обозначается как  $u, x$ . Если  $x$  — терминал, то во втором столбце записывается узел  $u$ , если нетерминал, то во втором столбце записывается начальный узел компоненты, соответствующей этому нетерминалу. Если в первом столбце записана выходная дуга  $u, \text{выход}$ , то во втором столбце записываются узлы, в которые входят дуги из нетерминальных вершин, в которых записан нетерминал, соответствующий компоненте, которой принадлежит узел  $u$ . В третьем столбце записывается множество выбора дуги, которое определяется как множество терминалов, достижимых в графе анализа СД из узлов, записанных во втором столбце таблицы. Множество выбора выходной дуги начальной компоненты обязательно содержит концевой маркер  $\perp$ , может быть, другие символы.

Таблица 1

**Множества выбора дуг, выходящих из узлов СД**

Дуга	Узлы	Множества выбора
1 A	5	$b, d, e, c$
1 a	1	$a$
2 c	2	$c$
3 B	9	$d, e$
4 B	9	$d, e$
4 выход		$\perp$
5 b	5	$b$

5 B	9	d, e
5 выход	2	c
6 B	9	d, e
7 d	7	d
8 a	8	a
8 выход	2	c
9 d	9	d, e
9 e	9	d, e
10 B	9	d, e
11 выход	2, 4, 7, 8	c, d, e, a, $\frac{1}{2}$

Анализируя таблицу видим, что множества выбора любой пары дуг, выходящих из общего узла, не пересекаются, следовательно СД (рис.1) детерминированная.

Итак, в работе введено понятие графа анализа СД, способ его построения и использования для определения детерминированности СД. Рассмотрен пример анализа СД предложенным способом. Этот способ может быть использован как при ручном, так и при автоматизированном построении трансляторов на основе СД.

#### Список литературы:

1. *Chomsky N.* Three models for the description of language, IRE Trans. On Information Theory IT-2:3, 1956, pp. 113-124. *Хомский Н.* Три модели для описания языка // Кибернетический сборник. — М.: ИЛ, 1961. Вып.2. — С. 237 — 266.
2. *Ахо А.* Теория синтаксического анализа, перевода и компиляции / Ахо А., Ульман Дж. — М.: Мир, 1978. Т. 1, 612 с. — т. 2, 487 с.
3. *Льюис Ф.* Теоретические основы проектирования компиляторов / Льюис Ф., Розенкранц Д., Стирнз Р. — М.: Мир, 1979. — 656 с.
4. *Йенсен К.* Паскаль. Руководство для пользователя и описание языка / Йенсен К., Вирт Н. — М: Финансы и статистика, 1982. — 151 с.
5. *Легалов А. И.* Формальные языки и трансляторы / Легалов А. И., Швец Д. А., Легалов И. А. — Красноярск: Сибирский федеральный университет, 2007. — 213 с.
6. *Свердлов С. З.* Языки программирования и методы трансляции / Свердлов С. З. — СПб.: Питер, 2007. — 638 с.

7. *Карпов Ю. Г.* Теория и технология программирования. Основы построения трансляторов / Карпов Ю. Г. — СПб.: БХВ-Петербург, 2005. — 272 с.
8. *Рязанов, Ю.Д.* Псевдодетерминированные синтаксические диаграммы / Ю.Д. Рязанов, М.Н. Севальнева // Прикладная математика, управление и информатика: сборник трудов Междунар. молодеж. конф., Белгород, 3–5 октября 2012 г. : в 2 т. — Белгород : ИД «Белгород», 2012. — Т2. — С. 546–553.
9. *Поляков В.М.* Алгоритм построения нерекурсивных программ-распознавателей линейной сложности по детерминированным синтаксическим диаграммам / Поляков В.М., Рязанов Ю.Д. // Вестник БГТУ им. В.Г. Шухова, № 6, 2013, с. 194 — 199.
10. *Рязанов, Ю.Д.* Псевдодетерминированные синтаксические диаграммы / Ю.Д. Рязанов, М.Н. Севальнева // Прикладная математика, управление и информатика: сборник трудов Междунар. молодеж. конф., Белгород, 3–5 октября 2012 г. : в 2 т. — Белгород : ИД «Белгород», 2012. — Т2. — С. 546–553.