

ВОСХОДЯЩИЕ МП-РАСПОЗНАВАТЕЛИ ТИПА «ПЕРЕНОС-СВЁРТКА»

Принцип работы МП-распознавателей типа «ПЕРЕНОС-СВЁРТКА» такой же, как и «ПЕРЕНОС-ОПОЗНАНИЕ»: выполняя ПЕРЕНОС, в магазине образуется основа, которая сворачивается по основывающему правилу.

Отличие в том, что в магазин помещаются не символы грамматики (терминалы и нетерминалы), а их коды.

Код одного и того же символа может быть разным в зависимости от верхнего символа магазина.

Такое кодирование позволяет только по верхнему символу магазина определить, находится ли в верхней части магазина основа, и, поэтому, процедура ОПОЗНАНИЕ не нужна.

Для того, чтобы закодировать символы грамматики и построить МП-распознаватель типа «ПЕРЕНОС-СВЁРТКА», нужно будет построить специальный граф.

Граф будем строить по пополненной грамматике, т. е. в грамматику будем добавлять правило $S' \rightarrow S$, где S — начальный нетерминал исходной грамматики, а S' — начальный нетерминал пополненной грамматики.

К вершинам графа будем обращаться по номеру (0 — номер начальной вершины).

Вершинам графа соответствуют множества пунктов.

Пункт — это правило грамматики, в правой части которого, кроме терминалов и нетерминалов, присутствует ещё точка (.). Точка может находиться в начале правой части, в конце её или между другими символами.

Дуга в графе, выходящая из одной вершины в другую, отмечается символом грамматики.

Правила построения графа.

Начальной вершине 0 графа соответствует множество пунктов M_0 , в которое входят:

- 1) правило для начального нетерминала с точкой в начале правой части;
- 2) если в этом множестве есть пункт вида $A \rightarrow \alpha.V\beta$ (α и β могут быть пустыми), т. е. точка стоит перед нетерминалом, то в множество добавляются все пункты с нетерминалом V в левой части с точкой в начале правой части.

Все пункты, которые можно добавить в множество, добавляются.

Если в множестве пунктов M_i , соответствующем вершине i , есть пункты вида

$$A \rightarrow \alpha_1.X\beta_1,$$

$$A \rightarrow \alpha_2.X\beta_2,$$

.....

.....

.....

$$A \rightarrow \alpha_n.X\beta_n$$

($\alpha_1, \alpha_2, \dots, \alpha_n, \beta_1, \beta_2, \dots, \beta_n$ могут быть пустыми), а X — терминал или нетерминал, то из вершины i проводим дугу в вершину j , которую отмечаем символом X .

Вершине j соответствует множество пунктов M_j , в которое входят:

1) $A \rightarrow \alpha_1.X.\beta_1,$

$$A \rightarrow \alpha_2.X.\beta_2,$$

.....

.....

.....

$$A \rightarrow \alpha_n.X.\beta_n$$

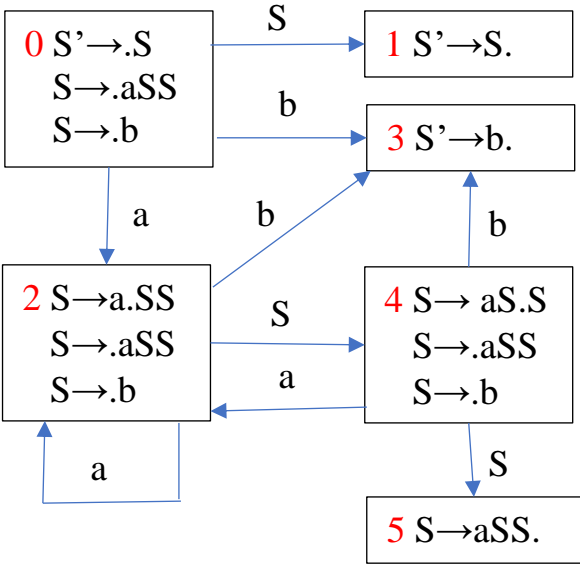
- 2) если в этом множестве есть пункт вида $A \rightarrow \alpha.V\beta$ (α и β могут быть пустыми), т. е. точка стоит перед нетерминалом, то в множество добавляются все пункты с нетерминалом V в левой части с точкой в начале правой части.

Все пункты, которые можно добавить в множество, добавляются.

Граф строится так, чтобы в нём не было двух вершин с одинаковыми множествами пунктов.

Пример.

- 0. $S' \rightarrow S$
- 1. $S \rightarrow aSS$
- 2. $S \rightarrow b$



Этот граф можно представить таблицей

вершина	пункты	символ	вершина
0	$S' \rightarrow .S$	S	1
	$S \rightarrow .aSS$	a	2
	$S \rightarrow .b$	b	3
1	$S' \rightarrow S.$		
2	$S \rightarrow a.SS$	S	4
	$S \rightarrow .aSS$	a	2
	$S \rightarrow .b$	b	3
3	$S' \rightarrow b.$		
4	$S \rightarrow aS.S$	S	5
	$S \rightarrow .aSS$	a	2
	$S \rightarrow .b$	b	3
5	$S \rightarrow aSS.$		

МП-распознаватель типа «ПЕРЕНОС-СВЁРТКА» задаётся таблицей T , в которой:

1) строки соответствуют магазинным символам — номерам вершин графа;
2) столбцы соответствуют входным символам — терминалам и концевому маркеру, а также нетерминалам;

3) В клетку $T[i, x]$, где x — терминал, пишем $P(j)$ ($T[i, x] := P(j)$), если в графе из вершины i идёт дуга в вершину j , отмеченная терминалом x .

$P(j)$ обозначает операцию ПЕРЕНОС, которая выполняется следующим образом: $ВТОЛК(j)$, сдвиг.

4) В клетку $T[i, N]$, где N — нетерминал, пишем j ($T[i, x] := j$), если в графе из вершины i идёт дуга в вершину j , отмеченная нетерминалом N .

5) В клетки $T[i, x]$, для всех x — терминал или концевой маркер, пишем $СВ(\#k)$ ($T[i, x] := СВ(\#k)$), если множество M_i , соответствующее вершине i , состоит только из одного пункта: $k. A \rightarrow \alpha$.

$СВ(\#k)$ означает свёртку по k -му правилу грамматики ($k. A \rightarrow \alpha$), которая выполняется следующим образом:

1) из магазина выталкивается количество символов, равное длине цепочки α ;

2) пусть после выполнения всех выталкиваний вверху магазина будет символ i , тогда в магазин вталкивается символ из $T[i, A]$, если там не пусто, если там пусто, то отвергнуть.

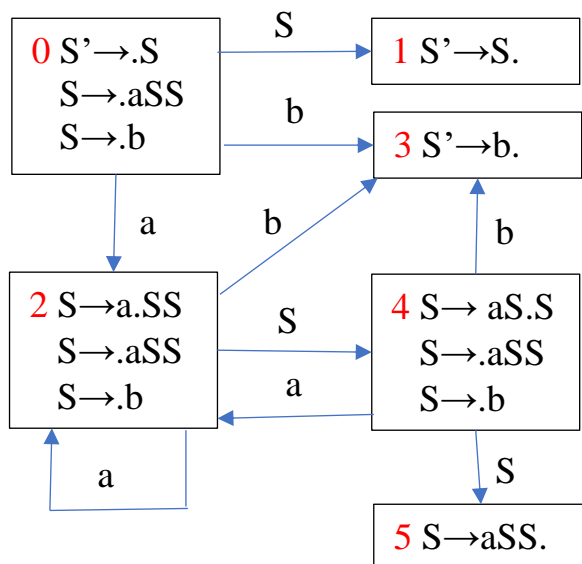
6) В клетку $T[i, \vdash]$ пишем ДОП ($T[i, \vdash] := ДОП$), если множество M_i , соответствующее вершине i , состоит только из одного пункта: $0. S' \rightarrow S$.

7) Пустые клетки таблицы соответствуют ОТВЕРГНУТЬ.

8) Начальное содержимое магазина: $\Delta 0$.

Если вверху магазина i , и множество M_i содержит пункт с точкой в конце правой части, то распознаватель выполняет свёртку по соответствующему правилу.

Если вверху магазина i , и множество M_i содержит пункт с точкой не в конце правой части, то распознаватель выполняет перенос или завершение свёртки.



вершина	пункты	символ	вершина
0	$S' \rightarrow \cdot S$ $S \rightarrow \cdot aSS$ $S \rightarrow \cdot b$	S a b	1 2 3
1	$S' \rightarrow S \cdot$		
2	$S \rightarrow a \cdot SS$ $S \rightarrow \cdot aSS$ $S \rightarrow \cdot b$	S a b	4 2 3
3	$S' \rightarrow b \cdot$		
4	$S \rightarrow aS \cdot S$ $S \rightarrow \cdot aSS$ $S \rightarrow \cdot b$	S a b	5 2 3
5	$S \rightarrow aSS \cdot$		

Таблица МП-распознавателя:

	a	b	\perp	S
0	П(2)	П(3)		1
1			доп	
2	П(2)	П(3)		4
3	СВ(#2)	СВ(#2)	СВ(#2)	
4	П(2)	П(3)		5
5	СВ(#1)	СВ(#1)	СВ(#1)	

Н.С.М.: Δ0

Пример.

0. $S' \rightarrow S$
1. $S \rightarrow aSS$
2. $S \rightarrow b$

Таблица МП-распознавателя:

	a	b	\perp	S
0	П(2)	П(3)		1
1			доп	
2	П(2)	П(3)		4
3	СВ(#2)	СВ(#2)	СВ(#2)	
4	П(2)	П(3)		5
5	СВ(#1)	СВ(#1)	СВ(#1)	

Н.С.М.: $\Delta 0$

Правый вывод:

$S'_0 \Rightarrow S_1 \Rightarrow aSS_2 \Rightarrow aS_1b \Rightarrow aaSS_2b \Rightarrow aaS_2bb \Rightarrow aabbb$

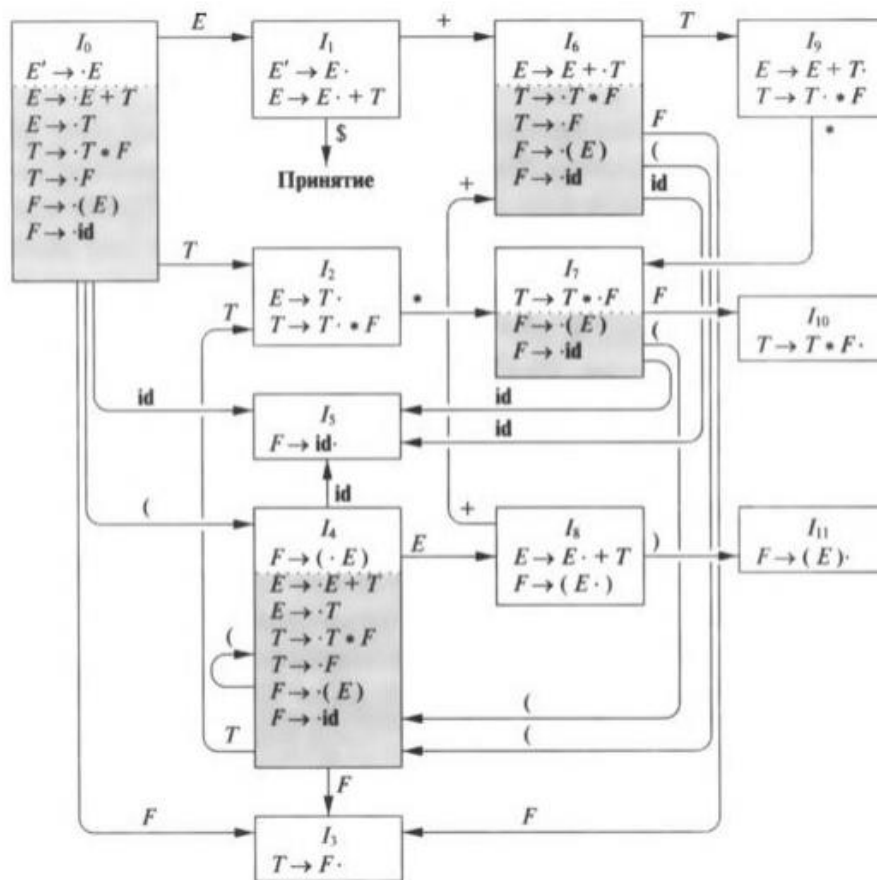
Протокол обработки цепочки aabbb МП-распознавателем:

магазин	цепочка	действие
$\Delta 0$	aabbb	П(2)
$\Delta 02$	abbb	П(2)
$\Delta 022$	bbb	П(3)
$\Delta 0223$	bb	СВ(#2)
$\Delta 0224$	bb	П(3)
$\Delta 02243$	b	СВ(#2)
$\Delta 02245$	b	СВ(#1)
$\Delta 024$	b	П(3)
$\Delta 0243$		СВ(#2)
$\Delta 0245$		СВ(#1)
$\Delta 01$		допустить

Грамматика называется LR(0)-грамматикой, если каждое множество пунктов, содержащее пункт вида $A \rightarrow \alpha \cdot$, состоит из единственного пункта.

Рассмотрим другой пример.

0. $E' \rightarrow E$
1. $E \rightarrow E + T$
2. $E \rightarrow T$
3. $T \rightarrow T * F$
4. $T \rightarrow F$
5. $F \rightarrow (E)$
6. $F \rightarrow id$



Рассмотрим вершину 1.

Множество M_1 содержит 2 пункта:

0. $E' \rightarrow E$.
1. $E \rightarrow E + T$

один из которых содержит точку в конце правой части. Поэтому рассматриваемая грамматика не относится к классу LR(0)-грамматик.

По пункту 0. $E' \rightarrow E$. должны выполнить свёртку по 0-му правилу, а по пункту 1. $E \rightarrow E + T$ — перенос, если на входе $+$.

Ситуацию, когда в одной вершине должны выполнить и перенос и свёртку, назовём конфликтной. Конфликт в данном случае будет разрешён, если при входном символе $+$ (когда нужно выполнить перенос) не имеет смысла выполнять свёртку по 0-му правилу.

Вспомним, что свёртку по правилу $A \rightarrow \alpha$ имеет смысл выполнять, если входной символ принадлежит множеству $\text{СЛЕД}(A)$.

С учётом этого изменим правила построения таблицы МП-распознавателя типа «ПЕРЕНОС-СВЁРТКА»:

1) строки соответствуют магазинным символам — номерам вершин графа;
2) столбцы соответствуют входным символам — терминалам и концевому маркеру, а также нетерминалам;

3) В клетку $T[i, x]$, где x — терминал, пишем $\Pi(j)$ ($T[i, x] := \Pi(j)$), если в графе из вершины i идёт дуга в вершину j , отмеченная терминалом x .

$\Pi(j)$ обозначает операцию ПЕРЕНОС, которая выполняется следующим образом: $\text{ВТОЛК}(j)$, сдвиг.

4) В клетку $T[i, N]$, где N — нетерминал, пишем j ($T[i, x] := j$), если в графе из вершины i идёт дуга в вершину j , отмеченная нетерминалом N .

5) В клетку $T[i, x]$, где x — терминал или концевой маркер, пишем $\text{СВ}(\#k)$ ($T[i, x] := \text{СВ}(\#k)$), если в множестве M_i , соответствующем вершине i , есть пункт k . $A \rightarrow \alpha$. и $x \in \text{СЛЕД}(A)$.

$\text{СВ}(\#k)$ означает свёртку по k -му правилу грамматики (k . $A \rightarrow \alpha$), которая выполняется следующим образом:

1) из магазина выталкивается количество символов, равное длине цепочки α ;

2) пусть после выполнения всех выталкиваний вверху магазина будет символ i , тогда в магазин вталкивается символ из $T[i, A]$, если там не пусто, если там пусто, то отвергнуть.

6) В клетку $T[i, \perp]$ пишем ДОП ($T[i, \perp] := \text{ДОП}$), если множество M_i , соответствующее вершине i , содержит пункт 0. $S' \rightarrow S$.

7) Пустые клетки таблицы соответствуют ОТВЕРГНУТЬ.

8) Начальное содержимое магазина: $\Delta 0$.

Если грамматика такая, что, применяя описанные выше правила построения таблицы распознавателя, получается таблица, в которой в каждой клетке записано не более одного действия, то такая грамматика принадлежит классу SLR-грамматик.

Если в одну клетку нужно записать и перенос и свёртку, то говорят о наличии конфликта «перенос-свёртка».

Конфликт «перенос-свёртка» возникает, если в множестве M_i есть такие два пункта

$$A \rightarrow \alpha. \text{ и } B \rightarrow \beta.x\gamma, \text{ что } x \in \text{СЛЕД}(A).$$

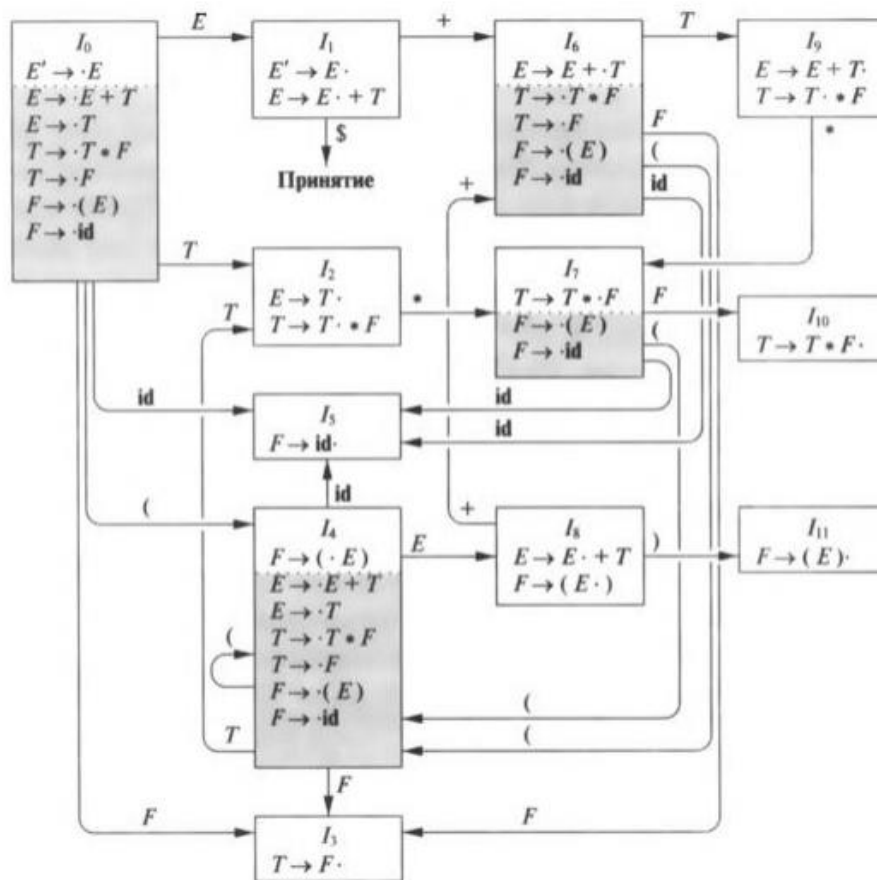
Если в одну клетку нужно записать как минимум две свёртки, то говорят о наличии конфликта «свёртка-свёртка».

Конфликт «свёртка-свёртка» возникает, если в множестве M_i есть такие два пункта

$$A \rightarrow \alpha. \text{ и } B \rightarrow \beta., \text{ что } \text{СЛЕД}(A) \cap \text{СЛЕД}(B) \neq \emptyset.$$

Пример (продолжение).

0. $E' \rightarrow E$
1. $E \rightarrow E + T$
2. $E \rightarrow T$
3. $T \rightarrow T * F$
4. $T \rightarrow F$
5. $F \rightarrow (E)$
6. $F \rightarrow id$



Для построения таблицы потребуются множества СЛЕД для каждого нетерминала:

Нетерминал	СЛЕД
E'	$\{ \}$
E	$\{ +,) \}$
T	$\{ +,), * \}$
F	$\{ +,), * \}$

Таблица МП-распознавателя

	id	+	*	()	\neg	E	T	F
0	П(5)			П(4)			1	2	3
1		П(6)				ДОП			
2		СВ(#2)	П(7)		СВ(#2)	СВ(#2)			
3		СВ(#4)	СВ(#4)		СВ(#4)	СВ(#4)			
4	П(5)			П(4)			8	2	3
5		СВ(#6)	СВ(#6)		СВ(#6)	СВ(#6)			
6	П(5)			П(4)				9	3
7	П(5)			П(4)					10
8		П(6)			П(11)				
9		СВ(#1)	П(7)		СВ(#1)	СВ(#1)			
10		СВ(#3)	СВ(#3)		СВ(#3)	СВ(#3)			
11		СВ(#5)	СВ(#5)		СВ(#5)	СВ(#5)			

Н.С.М.: Δ0