

ЛЕКЦИЯ 4

“КОНЕЧНЫЕ И МАГАЗИННЫЕ АВТОМАТЫ. ПОСТРОЕНИЕ КОНЕЧНЫХ АВТОМАТОВ С МИНИМАЛЬНЫМ ЧИСЛОМ СОСТОЯНИЙ”

ПЛАН

1. Построение конечных автоматов с минимальным числом состояний.
2. Алгоритмы минимизации
3. Автоматы с магазинной памятью и их особенности.
4. Пример преобразования с помощью НП-транслятора

О возможностях построения минимальных детерминированных конечных автоматов (МДКА)

Напоминание.

1. Отношение \mathcal{R} , заданное на множестве A и являющееся *рефлексивным*, *транзитивным* и *симметричным*, называется отношением *эквивалентности*.

2. *Число классов* эквивалентности называется *индексом эквивалентности*.

Если $[x\mathcal{R}y, z \in T^* - \text{любой}] \Rightarrow xz\mathcal{R}yz$, то отношение \mathcal{R} – есть *правый инвариант* на T^* .

Пусть $A = (Q, \Sigma, \delta, q_0, F)$ – конечный автомат.

Отношение эквивалентности \mathcal{R}_M , ассоциированное с конечным автоматом, определяется следующим образом:

$x\mathcal{R}_M y \Leftrightarrow \tilde{\delta}(q_0, x) = \tilde{\delta}(q_0, y)$ и \mathcal{R}_M – правый инвариант на множестве цепочек, допускаемых КА

Для регулярного языка:

$$L \subset T^*, x\mathcal{R}_L y, z \in T^* - \text{любой}, xz \in L, \text{ если } yz \in L.$$

Теорема Майхилла – Нерода. Для любого регулярного языка L существует детерминированный конечный автомат, который имеет переходную функцию и число состояний, равное индексу эквивалентности отношения \mathcal{R}_L , ассоциированного с языком L .

МИНИМИЗАЦИЯ КОНЕЧНЫХ АВТОМАТОВ ПО РЕГУЛЯРНОМУ ВЫРАЖЕНИЮ

• Правила написания регулярных выражений

Что заключено в фигурные скобки от нуля до бесчисленного числа раз, круглые—используются для объединения альтернатив, из которых одна обязательно присутствует.

Пусть имеется алфавит, составленный из символов (букв, литер) $V_T = \{x_1, x_2, \dots, x_n\}$.

1. Множество всевозможных цепочек, составленных из букв $x_i \in V_T$:

$$L = \{x_1 \vee x_2 \vee x_3 \vee \dots \vee x_n\}.$$

2. Множество цепочек, составленных из литер $x_i \in V_T$ и оканчивающихся литерой x_1 .

$$L = \{x_1 \vee x_2 \vee x_3 \vee \dots \vee x_n\} \wedge x_1.$$

3. Множество цепочек, составленных из литер $x_i \in V_T$ начинающихся цепочкой l_1 и оканчивающихся l_2 .

$$L = l_1 \wedge \{x_1 \vee x_2 \vee x_3 \vee \dots \vee x_n\} \wedge l_2.$$

4. Множество однолитерных цепочек (однобуквенных слов) совпадает с алфавитом

$$L = x_1 \vee x_2 \vee x_3 \vee \dots \vee x_n.$$

5. Множество двулитерных цепочек (двухбуквенных слов)

$$L = (x_1 \vee x_2 \vee x_3 \vee \dots \vee x_n) \wedge (x_1 \vee x_2 \vee x_3 \vee \dots \vee x_n).$$

6. Множество m – буквенных слов

$$L = (x_1 \vee x_2 \vee x_3 \vee \dots \vee x_n) \underbrace{\wedge \dots \wedge}_{m-2} (x_1 \vee x_2 \vee x_3 \vee \dots \vee x_n)$$

- Алгоритм, разработанный профессором кафедры кибернетики и вычислительной техники, доктором технических наук Евгением Артёмовичем Бутаковым для минимизации КА Мили по регулярному выражению.

Шаги(Этапы)

1. Разметка мест.
2. Минимизация числа внутренних состояний автомата по разметке.
3. Построение таблицы переходов.
4. Дальнейшая минимизация по таблице переходов.

Место регулярного выражения - это позиция между двумя литерами (символами из алфавита входных сигналов Σ), а также начало и конец выражения.

Места регулярного выражения имеют следующие типы (названия).

Начальное место — начало выражения

Конечное место — конец выражения.

Основное место — место, слева от которого стоит литера, а также начальное место.

Предосновное место - место, справа от которого стоит литера.

Разметка мест.

1. Первоначально осуществляется *сквозная нумерация* всех основных мест, которые предварительно размечаются короткими вертикальными линиями на записи регулярного выражения.

2. Отмечаются длинными вертикальными линиями на записи регулярного выражения *предосновные* места, *которые не совпадают с основными*.

3. Выставляются индексы *всех предосновных* мест, *не совпадающих* с основными местами, при этом одному и тому же предосновному месту может соответствовать множество (несколько) индексов.

Правила построения системы индексов

1. Индекс перед любыми скобками распространяется на все начальные места дизъюнктивных членов, записанных в этих скобках.

2. Индекс конечного места любого дизъюнктивного члена, заключённого в любые скобки, распространяется на место, непосредственно следующее за этими скобками.

3. Индекс места перед итерационными скобками распространяется на место, непосредственно следующее за этими скобками.

4. Индекс места за итерационными скобками распространяется в начальные места всех дизъюнктивных членов в этих скобках.

5. Индекс конечного места любого дизъюнктивного члена, заключённого в итерационные скобки, распространяется на начальные места всех дизъюнктивных членов, заключённых в эти скобки.

6. Индексы места, справа и слева от которого стоят буквы никуда не распространяются.

7. Индекс конечного места распространяется на те же места, на которые и индекс начального места.

Правилом 7 можно не пользоваться, когда предполагается программная реализация конечного автомата.

Правило минимизации

Если *несколько* предосновных мест *отмечено одинаковой совокупностью* индексов и *справа* от этих мест записаны *одинаковые литеры*, то *основные* места, расположенные *справа от этих литер*, можно отметить *одинаковыми* индексами.

Возможна цикличность применения правил разметки с правилами минимизации.

Пример. Минимизируем выражение $\{1\}1 \vee \{1\}1\{0 \vee 1\}1$.

$$\begin{array}{ccccccc}
 \begin{array}{c} | \{ | 1 | \} | 1 | \\ 0 \quad 1^{1)} \quad 2^{1)} \end{array} & \vee & \begin{array}{c} | \{ | 1 | \} | 1 | \{ 0 | \vee | 1 | \} | 1 | \\ 0 \quad 3^{2)} \quad 4^{2)} \quad 5 \quad 6^{3)} \quad 7^{3)} \end{array} \\
 \begin{array}{cc} 0 & 0 \\ 1 & 1 \end{array} & & \begin{array}{ccccc} 0 & 0 & 4 & 4 & 4 \\ 3 & 3 & 5 & 5 & 5 \\ & & 6 & 6 & 6 \end{array}
 \end{array}$$

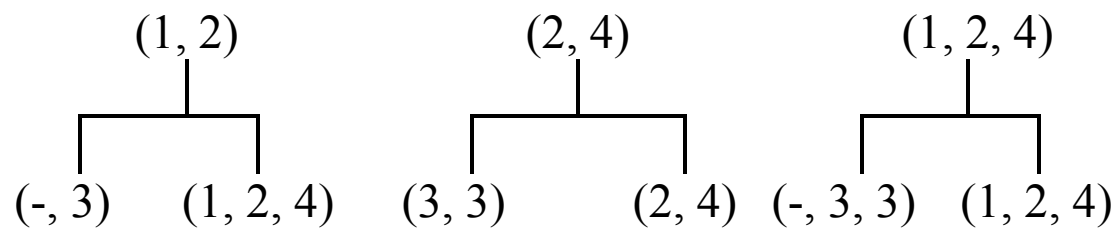
Под упрощение подпадают три пары позиций (1, 2), (3, 4) и (5, 6). После минимизации

$$\begin{array}{ccccccc}
 \begin{array}{c} | \{ | 1 | \} | 1 | \\ 0 \quad 1 \quad 1 \end{array} & \vee & \begin{array}{c} | \{ | 1 | \} | 1 | \{ 0 | \vee | 1 | \} | 1 | \\ 0 \quad 2 \quad 2 \quad 3 \quad 4 \quad 4 \end{array} \\
 \begin{array}{cc} 0 & 0 \\ 1 & 1 \end{array} & & \begin{array}{ccccc} 0 & 0 & 2 & 2 & 2 \\ 2 & 2 & 3 & 3 & 3 \\ & & 4 & 4 & 4 \end{array}
 \end{array}$$

Получаем таблицу переходов НКА

	0	1	2	3	4
0	–	–	3	3	3
1	1,2	1	2, 4	4	4

Приведём НКА к ДКА.



Получим

	0	1	2	3	4	$(1, 2)$	$(2, 4)$	$(1, 2, 4)$
0	-		3	3	3	3	3	3
1	$(1,2)$	1	$(2, 4)$	4	4	$(1, 2, 4)$	$(2, 4)$	$(1, 2, 4)$

В ДКА ряд состояний не достижим, это 1, 2, (2, 4).Исключим их
Перенумеруем состояния КА

№ Старый	№ Новый
0	0
3	1
4	2
1, 2	3
1, 2, 4	4

Окончательно

	0	1	2	3	4
0	—	1	1	1	1
1	3	2	2	4	4

МИНИМИЗАЦИЯ КОНЕЧНОГО АВТОМАТА ПО ТАБЛИЦЕ ПЕРЕХОДОВ

Алгоритм

1. Ввести дополнительное множество переменных (сигналов), которое называется множеством значений выхода. Каждому из состояний КА ставится в соответствие сигнал, который появляется на выходе конечного автомата в это состоянии.
2. Поместить данные сигналы над соответствующими столбцами таблицы переходов
3. Минимизировать число состояний ДКА путём объединения столбцов с одинаковым содержимым при условии равенства их выходных сигналов.

Пример. Рассмотрим функцию переходов предыдущего раздела.

Введём сигналы: p – рабочее состояние, z – заключительное состояние цепочка допущена.

Сигналы	p	p	z	z	z
	0	1	2	3	4
0	–	1	1	1	1
1	3	2	2	4	4

Одинаковы столбцы, соответствуют состояниям (1 и 2), (3 и 4) и имеют одинаковые сигналы. Обозначим: $0 \rightarrow 0$, $1 \rightarrow 1$, $2 \rightarrow 2$, $(3, 4) \rightarrow 3$.

Получим:

Сигналы	<i>p</i>	<i>p</i>	<i>z</i>	<i>z</i>
	0	1	2	3
0	—	1	1	1
1	3	2	2	3

МИНИМИЗАЦИЯ КОНЕЧНЫХ АВТОМАТОВ ПО ТАБЛИЦАМ ПЕРЕХОДОВ И ВЫХОДОВ

Данный способ минимизации предназначен для минимизации автоматов Мили.

1. Необходимо ввести множество выходных сигналов, соответствующих состояниям конечного автомата аналогично тому, как это выполняется при минимизации по таблице переходов.

2. Используя функцию переходов КА и соответствие состояний переменных выхода, выдаваемых КА при переходе в очередное состояние, построить функцию выхода КА.

3. Если **столбцы** таблицы функции **выходов** КА **совпадают**, то это указывает на **неразличимость** состояний, соответствующих столбцам, **по выходу** (по выходному сигналу).

Необходимо выписать все возможные сочетания пар из состояний, не различимых по выходному сигналу.

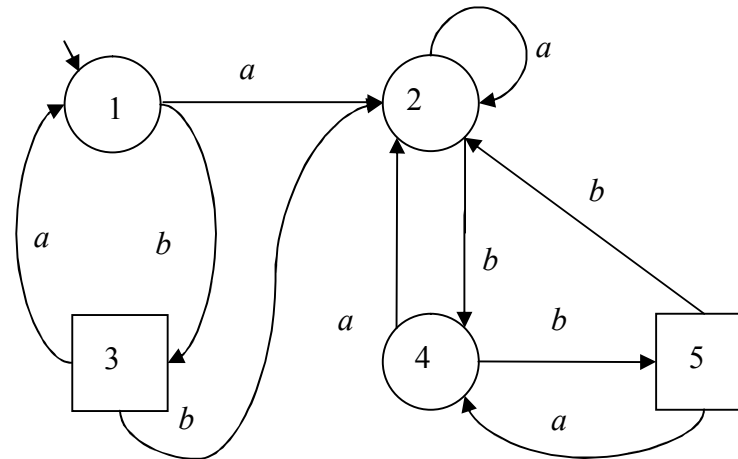
4. Каждая выписанная пара подвергается анализу тем же способом, как и при приведении НКА к ДКА.

Целью анализа является определение групп состояний, неразличимых **по входному символу** и **выходному сигналу**, в которые можно попасть из анализируемой пары состояний.

5. Если из неразличимых состояний, полученных в п. 4, можно составить замкнутую цепочку, то их можно объединить в одно. Например $(1, 2) \rightarrow (2, 6) \rightarrow (6, 1)$ можно укрупнить до $(1, 2, 6)$.

6. Построить функцию переходов минимального ДКА. В качестве состояний МДКА будут использованы: устойчивые пары состояний, цепочки состояний, обособленные состояния исходного КА.

Пример.



1. Обозначим сигналы (состояния) выхода: x_1 – рабочее состояние, x_2 – состояние, соответствующее окончанию работы. Восстановим по графу КА функцию переходов.

Состояния	1	2	3	4	5
Значения выхода	x_1	x_1	x_2	x_1	x_2
a	2	2	1	2	4
b	3	4	2	5	2

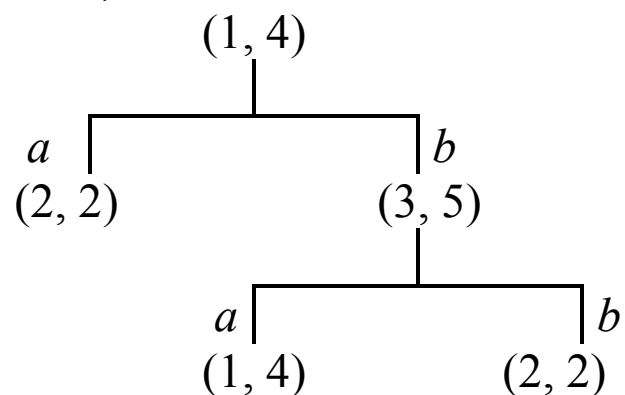
2. Используя таблицу переходов, составим функцию выхода

Состояния	1	2	3	4	5
Значения выхода	x_1	x_1	x_2	x_1	x_2
a	x_1	x_1	x_1	x_1	x_1
b	x_2	x_1	x_1	x_2	x_1

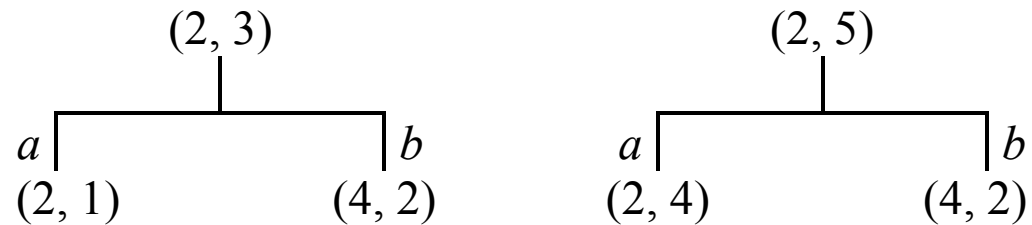
3. Множества состояний, которые не различимы по выходу, это $\{2, 3, 5\}$ и $\{1, 4\}$. Представим их парами $\{(2, 3), (3, 5), (2, 5)\}$ и $(1, 4)$.

Пары $(2, 3) \rightarrow (3, 5) \rightarrow (2, 5)$ образуют цепочку, и, если последующий анализ покажет их устойчивость, могут быть объединены в $(2, 3, 5)$.

4. Осуществим поиск пар, которые неразличимы по выходному сигналу и по входной строке. Анализ оформим древовидно, для наглядности.



Пары $(1, 4)$ и $(3, 5)$ – устойчивы, не различимы ни по входной строке, ни по выходному сигналу.



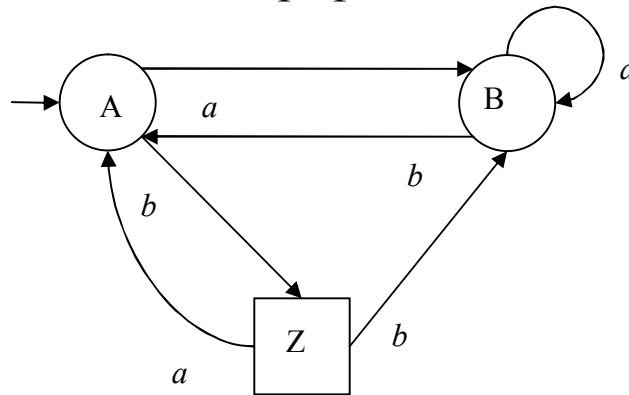
Пара $(2, 3)$ неустойчива, она различается как по входной строке, так и по выходному сигналу. Состояния $(2, 5)$ хотя и не различимы по входной строке, отличаются по выходным сигналам, выдаваемым конечным автоматом в состояниях $(2, 4)$.

5. Из устойчивых пар состояний $(1, 4)$ и $(3, 5)$ не может быть построена замкнутая цепочка, состояние 2 – обособленное, они и составят множество состояний МДКА.

6. Состояния $(1, 4)$ и $(3, 5)$ являются состояниями неразличимости, 2 – отдельное.
Таблица:

Состояние	$(1, 4)$	2	$(3, 5)$
Обозначение	A	B	Z
a	B	B	A
b	Z	A	B

Конечному автомату будет соответствовать граф вида



и формальная праволинейная грамматика

Z	::=	Ab b
A	::=	Za Bb
B	::=	Ba Zb Aa

МАГАЗИННЫЕ АВТОМАТЫ

Магазинные автоматы \equiv автоматы с магазинной памятью \equiv МП-автоматы.

Недетерминированным МП-автоматом называется семёрка следующих объектов (компонентов):

$$A = (\Sigma, Q, \Theta, \delta, q_0, z_0, F),$$

где Σ – непустое конечное множество входных символов (входной алфавит);

Q – конечное множество внутренних состояний МП-автомата (алфавит состояний);

Θ – конечное множество (алфавит) магазинных символов;

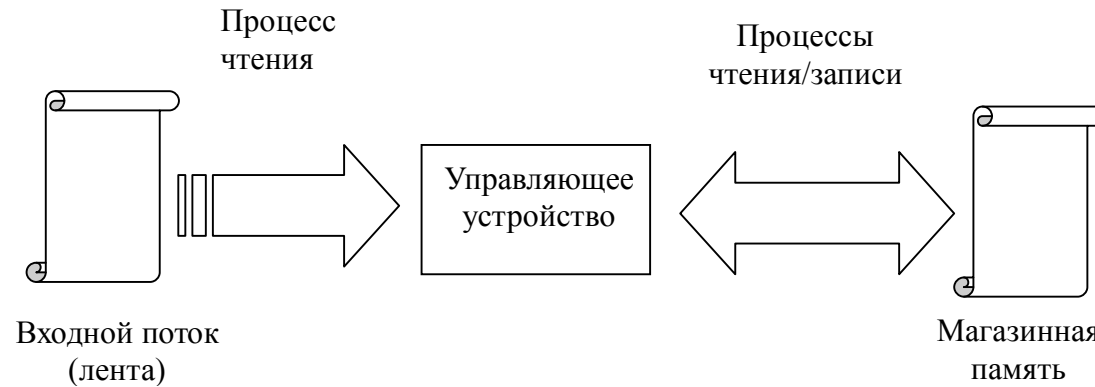
δ – отображение подмножества, определяемого декартовым произведением $Q \otimes \Sigma \otimes \Theta$ во множество подмножеств $Q \otimes \Theta$: $\delta: Q \otimes \Sigma \otimes \Theta \rightarrow Q \otimes \Theta$.

$q_0 \in Q$ – начальное состояние МП-автомата;

$z_0 \in \Theta$ – начальный (первый, инициированный) символ в магазинной памяти;

$F \subset Q$ – множество заключительных состояний автомата.

МП-автомат



Множество правил перехода, которое реализует один такт работы автомата, ***называется управляющим устройством.***

Конфигурация:

1. Текущим состоянием (его номером) МП-автомата.
2. Символом на вершине магазина МП-автомата.
3. Символом на входе (текущим входным символом) МП-автомата.

- Такт работы МП-автомата может включать следующие операции.

1. Операции над магазином МП-автомата. Это:

ЗАПИСАТЬ \equiv ВТОЛКНУТЬ \equiv ПОМЕСТИТЬ;

ВЫТОЛКНУТЬ;

НЕ ИЗМЕНЯТЬ СОДЕРЖИМОГО МАГАЗИНА.

2. Операции над состоянием МП-автомата. Их две:

ПЕРЕЙТИ;

ОСТАТЬСЯ.

Операции над входом МП-автомата:

СДВИГ (на одну позицию);

ДЕРЖАТЬ (то есть зафиксировать ленту до следующего шага).

МП-автомат называется МП-распознавателем если у него два выхода (выходных сигнала):

ДОПУСТИТЬ;

ОТВЕРГНУТЬ.

Пример. Задан язык вида $L = \{0^n 1^n, n \geq 1\}$.

Порождающая КС-грамматика: $G[Z] = [V_N = \{Z\}; V_T = \{0, 1\}; R = \{Z \rightarrow 01 | 0Z1\}; Z]$.

МП-автомат для анализа предложений такого языка имеет описание:

$\Sigma = \{0, 1, \blacktriangleleft\}$, \blacktriangleleft - признак (маркер) конца строки;

$\Theta = \{z, \#\}$, $\#$ - выталкиватель магазина;

$Q = \{q_0, q_1\}$.

Устройство управления:

q_0				
На вершине стека		Вход		
		0	1	\blacktriangleleft
	Z	ПЕРЕЙТИ q_0	ПЕРЕЙТИ q_1	
		ЗАПИСАТЬ Z СДВИГ	ВЫТОЛКНУТЬ СДВИГ	ОТВЕРГНУТЬ
	#	ПЕРЕЙТИ q_0	ОТВЕРГНУТЬ	ОТВЕРГНУТЬ
		ЗАПИСАТЬ Z СДВИГ		

q_1				
На вершине стека		Вход		
		0	1	\blacktriangleleft
	Z	ОТВЕРГНУТЬ	ПЕРЕЙТИ q_1	
			ВЫТОЛКНУТЬ СДВИГ	ОТВЕРГНУТЬ
	#	ОТВЕРГНУТЬ	ОТВЕРГНУТЬ	ДОПУСТИТЬ

Трасса работы МП-автомата при разборе цепочки **000111** представлена ниже

№	Магазин	Состояние	Входная лента
1	#	q_0	000111 \blacktriangleleft
2	#Z	q_0	00111 \blacktriangleleft
3	#ZZ	q_0	0111 \blacktriangleleft
4	#ZZZ	q_0	111 \blacktriangleleft
5	#ZZ	q_1	11 \blacktriangleleft
6	#Z	q_1	1 \blacktriangleleft
7	#	q_1	\blacktriangleleft (допустить)

То есть, цепочка **000111** допускается МП-автоматом.

Описание устройства управления МП-автоматом – в виде набора команд:

$$(q_i^r, a_j^r, z_k^r) \rightarrow \{(q_1^{r+1}, z_1^{r+1}), (q_2^{r+1}, z_2^{r+1}), \dots, (q_m^{r+1}, z_m^{r+1})\},$$

где $q_i^r \in Q$, r – номер такта работы МПА с продвижением, $a_j^r \in \Sigma$, $z_k^r, z_k^{r+1} \in \Theta$ или без продвижения

$$(q_i^r, \lambda, z_k^r) \rightarrow \{(q_1^{r+1}, z_1^{r+1}), (q_2^{r+1}, z_2^{r+1}), \dots, (q_m^{r+1}, z_m^{r+1})\}$$

где λ - пустая цепочка.

Теорема 1. Пусть $L\{G\}$ – КС язык, порождаемый грамматикой $G[S] = \langle V_N, V_T, R, S \rangle$ в нормальной форме Грейбах. Тогда существует недетерминированный нисходящий МП-автомат M , такой, что допускает все слова языка $L\{G\}$ и только их, причём автомат $M = \{\Sigma, Q, \Theta, \delta, q_0, Z_0, F\}$ строится следующим образом:

1. $\Sigma = V_T$;
2. $Q = \{q_1\}$;
3. $\Theta = V_N$;
4. $q_0 = q_1$;
5. $Z_0 = S$;
6. $F = \{\}$;
7. $(q_1, \gamma) \in \delta(q_1, a, \mathbf{B})$ всегда, когда подстановка $\mathbf{B} \rightarrow a\gamma$ принадлежит множеству правил R грамматики G , где $\mathbf{B} \in V_N$, $a \in V_T$, $\gamma \in (V_N \cup V_T)^*$.

Теорема 2. Пусть $L\{M\}$ – язык, допускаемый магазинным автоматом $M=\{\Sigma, Q, \Theta, \delta, q_0, Z_0, F\}$. Тогда существует такая КС-грамматика $G[S]=\langle V_N, V_T, R, S \rangle$, что $L\{G\} = L\{M\}$, где

1. $V_T = \Sigma$;
2. V_N – множество вида (q, \mathbf{B}, p) , $q, p \in Q$, и $\mathbf{B} \in \Theta$;
3. Подстановка $\mathbf{S} \rightarrow (q_0, Z_0, q)$ принадлежит множеству правил R при любом $q \in Q$;
4. Подстановка $(q, \mathbf{B}, p) \rightarrow (q_1, \mathbf{B}_1, q_2)(q_2, \mathbf{B}_2, q_3) \dots (q_m, \mathbf{B}_m, q_{m+1})$ принадлежит множеству R для всех $q, q_1, q_2, \dots, q_{m+1} \in Q$, $p = q_{m+1}$, для любого $a \in \{\Sigma \cup \varepsilon\}$ и любых $\mathbf{B}, \mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_m \in \Theta$, таких, что $(q_1, \mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_m) \in \delta(q, a, \mathbf{B})$. Если же $m=0$, то $q_1=p$, $p \in \delta(q, a, \mathbf{B})$ $(q, \mathbf{B}, p) \rightarrow a$ принадлежит множеству R .

Пример Задана грамматика $G[S]$ вида

$$\begin{array}{ll} V_N = \{\mathbf{S}, & \mathbf{B}, \mathbf{S} \rightarrow a\mathbf{B} | b\mathbf{D}; & \text{Аксиома грамматики} \\ & \mathbf{D}\}; & \mathbf{S}. \\ V_T = \{a, b\}; & \mathbf{D} \rightarrow a\mathbf{S} | b\mathbf{D}\mathbf{D} | a; \\ & \mathbf{B} \rightarrow a\mathbf{B}\mathbf{B} | b\mathbf{S} | b; \end{array}$$

В соответствии с теоремами имеем:

$$\Sigma = \{a, b\}, Q = \{q_1\}, \Theta = \{\mathbf{S}, \mathbf{B}, \mathbf{D}\}, q_0 = q_1, Z_0 = \mathbf{S}.$$

Синтезированное устройство управления представляет собой

№	Команда МП-автомата	Продукция грамматики
1.	$(q_1, a, S) \rightarrow (q_1, B);$	$S \rightarrow aB;$
2.	$(q_1, b, S) \rightarrow (q_1, D);$	$S \rightarrow bD;$
3.	$(q_1, a, D) \rightarrow \{(q_1, S),$ $(q_1, \lambda)\};$	$D \rightarrow aS;$
		$D \rightarrow a;$
4.	$(q_1, b, D) \rightarrow (q_1, DD);$	$D \rightarrow bDD;$
5.	$(q_1, a, B) \rightarrow (q_1, BB);$	$B \rightarrow aBB;$
6.	$(q_1, b, B) \rightarrow \{(q_1, S);$ $(q_1, \lambda)\}$	$B \rightarrow bS;$
		$B \rightarrow b.$

ТРАНСЛЯЦИЯ С ПОМОЩЬЮ МП-автомата

МП-автомат называется МП-транслятором, если он порождает выходную цепочку. Для этого добавляют операцию над выходом вида

ВЫДАТЬ(s),

где s – цепочка литер (символов, кодов), которую необходимо выдать.

Пример. Пусть необходимо перевести произвольную последовательность цифр “0” и “1” в цепочку “ $1^n 0^m$ ”, где n – число единиц, а m – число нулей.

Пример преобразования: $011011 \Rightarrow 111100$.

Пример устройства управления МП-транслятора.

	0	1	◀
0	ЗАПИСАТЬ (0) СДВИГ	ВЫДАТЬ (1) СДВИГ	ВЫТОЛКНУТЬ ВЫДАТЬ (0) ДЕРЖАТЬ
#	ЗАПИСАТЬ (0) СДВИГ	ВЫДАТЬ (1) СДВИГ	ДОПУСТИТЬ

В таблице управляющего устройства обозначено # – выталкиватель магазина, ◀ – концевой маркер входной строки.