Лингвистические основы информатики
Лекция 11
Синтаксический анализ и автомат с магазинной памятью

И. А. Михайлова, Ю. В. Нагребецкая

Уральский федеральный университет
Институт естественных наук и математики
Департамент математики, механики и компьютерных наук

Направления: Математика и компьютерные науки
Компьютерная безопасность
(6 семестр)

Задача синтаксического анализа

Дан КС язык L, заданный КС грамматикой $G = (\Sigma, \Gamma, P, S)$, и цепочка $w \in \Sigma^*$. Проверить, будет ли $w \in L$.

- Любой синтаксический анализатор (в этом курсе их будет несколько) проверяет этот факт, строя, кроме того вывод цепочки w в грамматике G.

В чем идея?

- Попробуйте придумать алгоритм, который по заданной строке (содержащей круглые скобки и какие-то другие символы) проверяет, правильно ли расставлены скобки.

В чем идея?

- Попробуйте придумать алгоритм, который по заданной строке (содержащей круглые скобки и какие-то другие символы) проверяет, правильно ли расставлены скобки.
- Скорее всего алгоритм будет следующим: встречая любой символ, кроме скобок, вы будете его пропускать. Встретив открывающуюся скобку вы должны будете ее запомнить, потому что где-то должна находиться соответствующая ей закрывающая.

В чем идея?

- Попробуйте придумать алгоритм, который по заданной строке (содержащей круглые скобки и какие-то другие символы) проверяет, правильно ли расставлены скобки.
- Скорее всего алгоритм будет следующим: встречая любой символ, кроме скобок, вы будете его пропускать. Встретив открывающуюся скобку вы должны будете ее запомнить, потому что где-то должна находиться соответствующая ей закрывающая.
- Это легко сделать, положив открывающуся скобку в стек. Когда встретите закрывающую скобку, то достаточно будет удалить одну скобку из стека.

В чем идея?

- Попробуйте придумать алгоритм, который по заданной строке (содержащей круглые скобки и какие-то другие символы) проверяет, правильно ли расставлены скобки.
- Скорее всего алгоритм будет следующим: встречая любой символ, кроме скобок, вы будете его пропускать. Встретив открывающуюся скобку вы должны будете ее запомнить, потому что где-то должна находиться соответствующая ей закрывающая.
- Это легко сделать, положив открывающуся скобку в стек. Когда встретите закрывающую скобку, то достаточно будет удалить одну скобку из стека.
- Скобки в строке расставлены верно, если выполняются два условия:
  1. Всякий раз, когда нужно удалить скобку из стека, стек не пуст.
  2. После прочтения всей строки стек пуст.

В чем идея?

- Попробуйте придумать алгоритм, который по заданной строке (содержащей круглые скобки и какие-то другие символы) проверяет, правильно ли расставлены скобки.
- Скорее всего алгоритм будет следующим: встречая любой символ, кроме скобок, вы будете его пропускать. Встретив открывающуюся скобку вы должны будете ее запомнить, потому что где-то должна находиться соответствующая ей закрывающая.
- Это легко сделать, положив открывающуся скобку в стек. Когда встретите закрывающую скобку, то достаточно будет удалить одну скобку из стека.
- Скобки в строке расставлены верно, если выполняются два условия:
  1. Всякий раз, когда нужно удалить скобку из стека, стек не пуст.
  2. После прочтения всей строки стек пуст.
- Этот пример иллюстрирует идею работы автомата с магазинной памятью: он читает строку посимвольно (как и конечный автомат), но способен обращаться к своей “памяти” (стеку).

Формальное определение МП-автомата

- Автоматом с магазинной памятью (МП-автоматом) назвается $\mathcal{M} = (Q, \Sigma, \Delta, \delta, q_0, F, \gamma_0)$, где

Формальное определение МП-автомата

- Автоматом с магазинной памятью (МП-автоматом) назвается $\mathcal{M} = (Q, \Sigma, \Delta, \delta, q_0, F, \gamma_0)$, где
- Q — множество его состояний (конечное множество),

Формальное определение МП-автомата

- Автоматом с магазинной памятью (МП-автоматом) назвается $\mathcal{M} = (Q, \Sigma, \Delta, \delta, q_0, F, \gamma_0)$, где
- Q — множество его состояний (конечное множество),
- Σ — входной алфавит, Δ — стековый алфавит,

Формальное определение МП-автомата

- Автоматом с магазинной памятью (МП-автоматом) назвается $\mathcal{M} = (Q, \Sigma, \Delta, \delta, q_0, F, \gamma_0)$, где
- Q — множество его состояний (конечное множество),
- Σ — входной алфавит, Δ — стековый алфавит,
- q_0 — начальное состояние, F — множество конечных состояний,

Формальное определение МП-автомата

- Автоматом с магазинной памятью (МП-автоматом) назвается $\mathcal{M} = (Q, \Sigma, \Delta, \delta, q_0, F, \gamma_0)$, где
- Q — множество его состояний (конечное множество),
- Σ — входной алфавит, Δ — стековый алфавит,
- q_0 — начальное состояние, F — множество конечных состояний,
- $\gamma_0 \in \Delta^*$ — начальное содержимое стека,

Формальное определение МП-автомата

- Автоматом с магазинной памятью (МП-автоматом) назвается $\mathcal{M} = (Q, \Sigma, \Delta, \delta, q_0, F, \gamma_0)$, где
- Q — множество его состояний (конечное множество),
- Σ — входной алфавит, Δ — стековый алфавит,
- q_0 — начальное состояние, F — множество конечных состояний,
- $\gamma_0 \in \Delta^*$ — начальное содержимое стека,
- δ — частичная функция из $Q \times \Delta \times \Sigma$ в $Q \times \Delta^* \times \{\rightarrow, \_\}$,— множество переходов (множество команд), опишем его на следующем слайде.

Формальное определение МП-автомата

- Автоматом с магазинной памятью (МП-автоматом) назвается $\mathcal{M} = (Q, \Sigma, \Delta, \delta, q_0, F, \gamma_0)$, где
- Q — множество его состояний (конечное множество),
- Σ — входной алфавит, Δ — стековый алфавит,
- q_0 — начальное состояние, F — множество конечных состояний,
- $\gamma_0 \in \Delta^*$ — начальное содержимое стека,
- δ — частичная функция из $Q \times \Delta \times \Sigma$ в $Q \times \Delta^* \times \{\rightarrow, \_\}$,— множество переходов (множество команд), опишем его на следующем слайде.

Замечание

Символы входного алфавита Σ будем обозначать строчными буквами $a, b \in \Sigma$ как и терминалы грамматики (как вы увидите ниже, это они и будут). Символы стекового алфавита Δ будем обозначать заглавными буквами $X, Y \in \Delta$, но эти символы не стоит путать с нетерминалами.

Переходы МП-автомата

МП-автомат действует по-тактово и на каждом такте

- находится в некотором состоянии $q \in Q$,
- на вершине стека наблюдает символ $X \in \Delta$,
- во входной строке видит символ $a \in \Sigma$.

Переходы МП-автомата

МП-автомат действует по-тактово и на каждом такте

- находится в некотором состоянии $q \in Q$,
- на вершине стека наблюдает символ $X \in \Delta$,
- во входной строке видит символ $a \in \Sigma$.

После выполнения команды автомат

- перейдет в состояние $q' \in Q$,
- вытолкнет из стека символ X и втолкнет цепочку $\gamma \in \Delta^*$,
- во входной строке либо останется на месте, либо сдвинется на один символ вправо.

Переходы МП-автомата

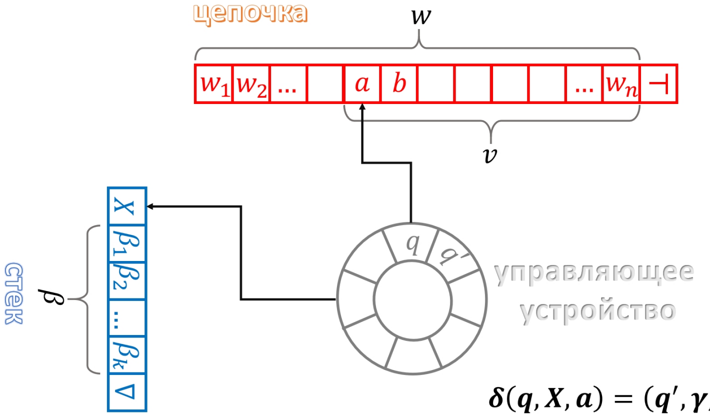
Такой переход соответствует записи $\delta(q, X, a) = (q', \gamma, \_)$ (см. рис. 1–2) (если автомат не сдвигается во входной строке) или $\delta(q, X, a) = (q', \gamma, \rightarrow)$ (если сдвигается) (см. рис. 3–4).

Переходы МП-автомата

Такой переход соответствует записи $\delta(q, X, a) = (q', \gamma, \_)$ (см. рис. 1–2) (если автомат не сдвигается во входной строке) или $\delta(q, X, a) = (q', \gamma, \rightarrow)$ (если сдвигается) (см. рис. 3–4).

Чтобы избежать ситуаций, когда стек пуст или входная строка закончилась, добавляют дополнительный стековый символ дно стека ∇ (его нельзя вытолкнуть из стека) и входной символ конца строки $\dashv$. Далее всегда считается, что Δ и Σ содержат эти символы (см. рис. 1,2).

Переходы МП-автомата. Иллюстрация



$$\delta(q, X, a) = (q', \gamma, \_)$$

Рис. 1

Переходы МП-автомата. Иллюстрация

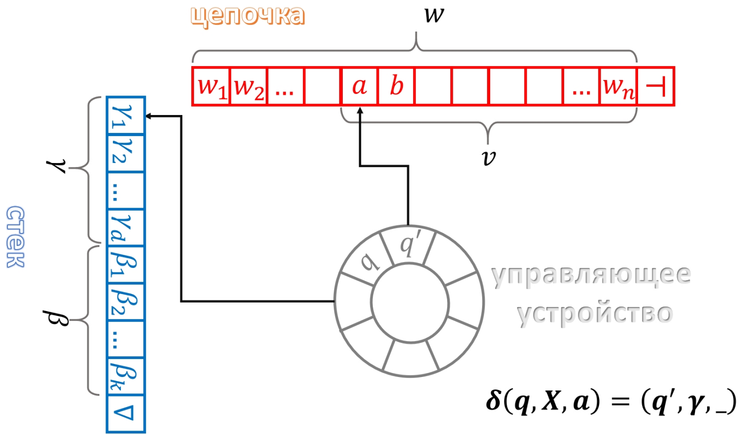


Рис. 2

Переходы МП-автомата. Иллюстрация

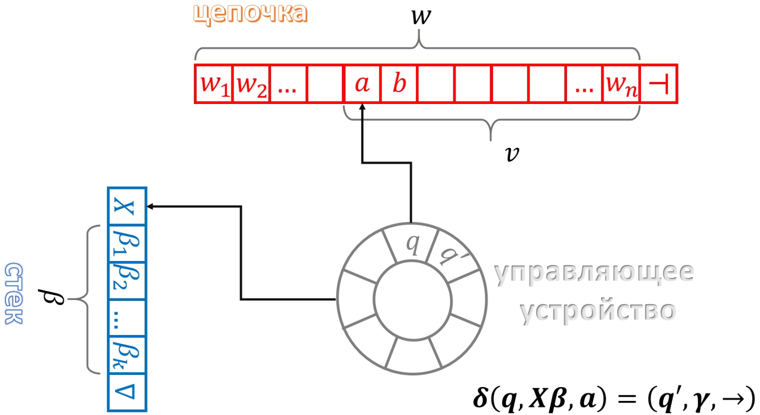


Рис. 3

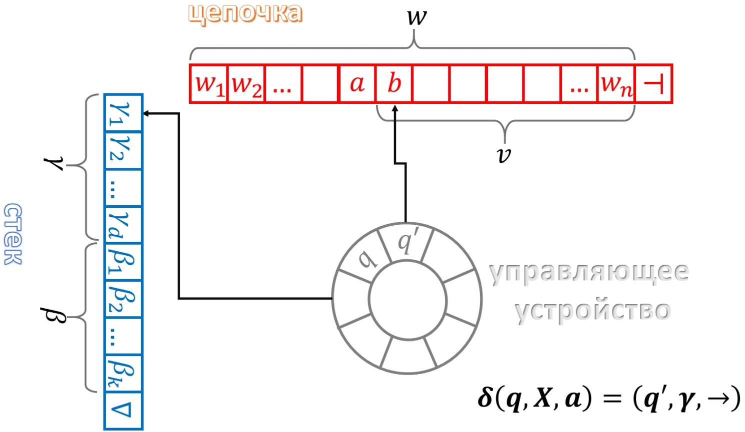

цепочка
w
w_1 w_2 ... a b ... w_n ⊣
v
стек
γ
γ_1 γ_2 ... γ_d β_1 β_2 ... β_k ∇
β
q q'
управляющее устройство
$\delta(q, X, a) = (q', \gamma, \rightarrow)$

Рис. 4

Пример. Управляющая таблица МП-автомата

Множество переходов записывается в виде управляющей таблицы (слева состояния автомата и стековый алфавит, сверху входной афавит)

		a	b	$\dashv$
q_0	a	$(q_0, aa, \rightarrow)$	$(q_1, \varepsilon, \rightarrow)$	
	∇	$(q_0, a, \rightarrow)$		
q_1	a		$(q_1, \varepsilon, \rightarrow)$	
	∇			$(q_2, \varepsilon, \_)$
q_2	a			
	∇			

- Из таблицы видно, что $\delta(q_0, a, a) = (q_0, aa, \rightarrow)$ и $\delta(q_1, \nabla, \dashv) = (q_2, \varepsilon, \_)$,
- Кроме того, $\delta(q_0, \nabla, b) = \emptyset$, так как ячейка пуста.
- Для каждой тройки (состояние, символ стека, входной символ) в примере имеется не более одной команды, поэтому данный автомат является детерминированным МПА.

Конфигурации МП-автомата

- Пусть $w \in \Sigma^*$ — цепочка, которая подается на вход МП-автомата $\mathcal{M} = (Q, \Sigma, \Delta, \delta, q_0, F, \gamma_0)$. При обработке этой цепочки на i-ом шаге работы получим конфигурацию МП-автомата $[q_i, \beta_i, v_i]$, где q_i — состояние, в котором находится автомат, β_i — содержимое стека, v_i — необработанная часть цепочки w (ее суффикс) (см. рис.5).

Конфигурации МП-автомата

- Пусть $w \in \Sigma^*$ — цепочка, которая подается на вход МП-автомата $\mathcal{M} = (Q, \Sigma, \Delta, \delta, q_0, F, \gamma_0)$. При обработке этой цепочки на i-ом шаге работы получим конфигурацию МП-автомата $[q_i, \beta_i, v_i]$, где q_i — состояние, в котором находится автомат, β_i — содержимое стека, v_i — необработанная часть цепочки w (ее суффикс) (см. рис.5).
- Если $[q_i, \beta_i, v_i]$ — конфигурация на i-ом шаге, $[q_{i+1}, \beta_{i+1}, v_{i+1}]$ — конфигурация на $(i+1)$-ом шаге, то записывают $[q_i, \beta_i, v_i] \models [q_{i+1}, \beta_{i+1}, v_{i+1}]$ (см. рис.6).

Конфигурации МП-автомата

- Пусть $w \in \Sigma^*$ — цепочка, которая подается на вход МП-автомата $\mathcal{M} = (Q, \Sigma, \Delta, \delta, q_0, F, \gamma_0)$. При обработке этой цепочки на i-ом шаге работы получим конфигурацию МП-автомата $[q_i, \beta_i, v_i]$, где q_i — состояние, в котором находится автомат, β_i — содержимое стека, v_i — необработанная часть цепочки w (ее суффикс) (см. рис.5).
- Если $[q_i, \beta_i, v_i]$ — конфигурация на i-ом шаге, $[q_{i+1}, \beta_{i+1}, v_{i+1}]$ — конфигурация на $(i+1)$-ом шаге, то записывают $[q_i, \beta_i, v_i] \models [q_{i+1}, \beta_{i+1}, v_{i+1}]$ (см. рис.6).
- Если $[q_i, \beta_i, v_i] \models^* [q_j, \beta_j, v_j]$, то последняя конфигурация получена при помощи нескольких тактов работы МП-автомата.

Конфигурации МП-автомата

- Пусть $w \in \Sigma^*$ — цепочка, которая подается на вход МП-автомата $\mathcal{M} = (Q, \Sigma, \Delta, \delta, q_0, F, \gamma_0)$. При обработке этой цепочки на i-ом шаге работы получим конфигурацию МП-автомата $[q_i, \beta_i, v_i]$, где q_i — состояние, в котором находится автомат, β_i — содержимое стека, v_i — необработанная часть цепочки w (ее суффикс) (см. рис.5).
- Если $[q_i, \beta_i, v_i]$ — конфигурация на i-ом шаге, $[q_{i+1}, \beta_{i+1}, v_{i+1}]$ — конфигурация на $(i+1)$-ом шаге, то записывают $[q_i, \beta_i, v_i] \models [q_{i+1}, \beta_{i+1}, v_{i+1}]$ (см. рис.6).
- Если $[q_i, \beta_i, v_i] \models^* [q_j, \beta_j, v_j]$, то последняя конфигурация получена при помощи нескольких тактов работы МП-автомата.
- Для текущей конфигурации $[q, X\beta, av]$ выполним команду $\delta(q, X, a) = (q', \gamma, \rightarrow)$ и получим (см. рис.7–8).

$$[q, X\beta, av] \models [q', \gamma\beta, v]$$

Конфигурации МП-автомата

- Пусть $w \in \Sigma^*$ — цепочка, которая подается на вход МП-автомата $\mathcal{M} = (Q, \Sigma, \Delta, \delta, q_0, F, \gamma_0)$. При обработке этой цепочки на i-ом шаге работы получим конфигурацию МП-автомата $[q_i, \beta_i, v_i]$, где q_i — состояние, в котором находится автомат, β_i — содержимое стека, v_i — необработанная часть цепочки w (ее суффикс) (см. рис.5).
- Если $[q_i, \beta_i, v_i]$ — конфигурация на i-ом шаге, $[q_{i+1}, \beta_{i+1}, v_{i+1}]$ — конфигурация на $(i+1)$-ом шаге, то записывают $[q_i, \beta_i, v_i] \models [q_{i+1}, \beta_{i+1}, v_{i+1}]$ (см. рис.6).
- Если $[q_i, \beta_i, v_i] \models^* [q_j, \beta_j, v_j]$, то последняя конфигурация получена при помощи нескольких тактов работы МП-автомата.
- Для текущей конфигурации $[q, X\beta, av]$ выполним команду $\delta(q, X, a) = (q', \gamma, \rightarrow)$ и получим (см. рис.7–8).

$$[q, X\beta, av] \models [q', \gamma\beta, v]$$

- Для этой же конфигурации $[q, X\beta, av]$ выполним команду $\delta(q, X, a) = (q', \gamma, \_)$ и получим (см. рис.9–10).

$$[q, X\beta, av] \models [q', \gamma\beta, av]$$

Конфигурации МП-автомата. Иллюстрация

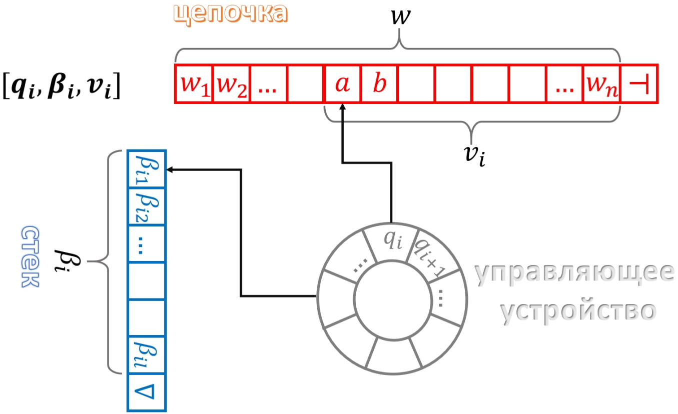


Рис. 5

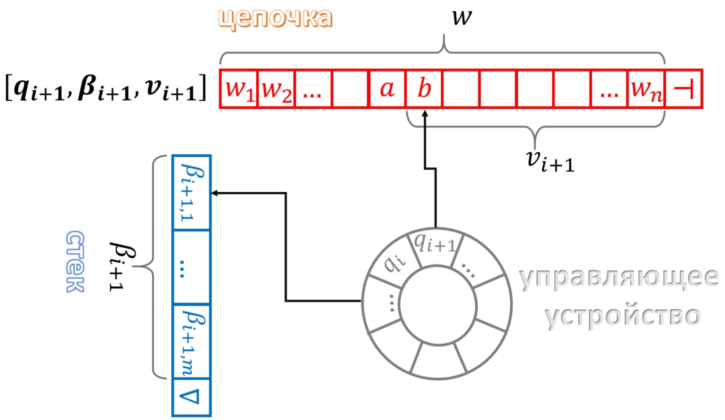
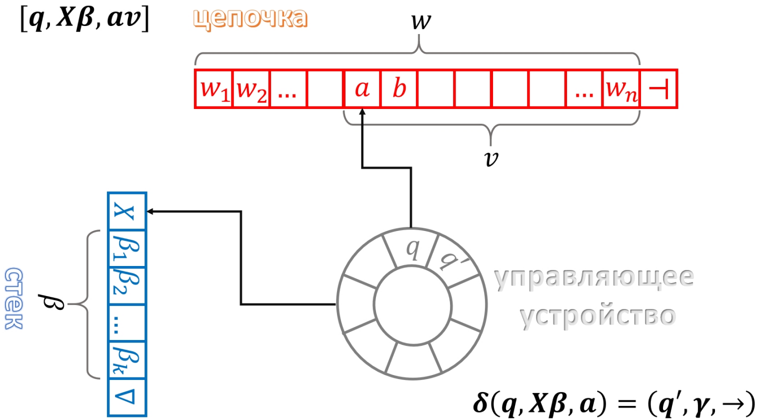
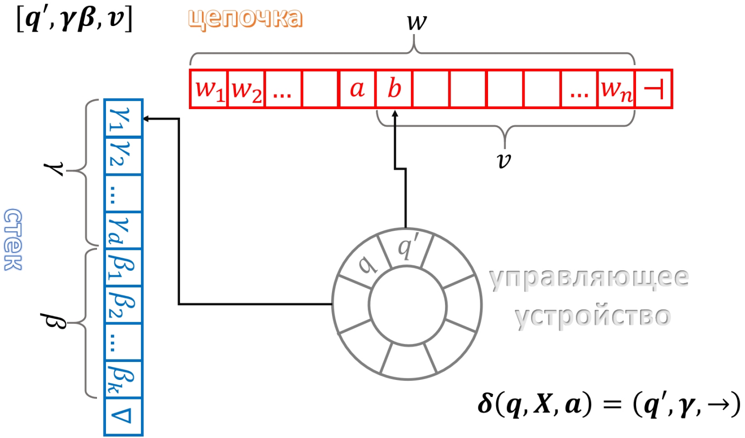

цепочка
w
$[q_{i+1}, \beta_{i+1}, v_{i+1}]$
w_1 w_2 … a b … w_n ⊣
v_{i+1}
стек
β_{i+1}
$\beta_{i+1,1}$
…
$\beta_{i+1,m}$
∇
q_i q_{i+1}
управляющее устройство

Рис. 6



[q, Xβ, av]
цепочка
w
w1 w2 ... a b ... wn ⊣
v
X β1 β2 ... βk ∇
β
стек
q q′
управляющее устройство
δ(q, Xβ, a) = (q′, γ, →)

Рис. 7



$[q', \gamma\beta, v]$
цепочка
w
w_1 w_2 … a b … w_n ⊣
v
γ
γ_1 γ_2 … γ_d
β
β_1 β_2 … β_k ∇
стек
q q'
управляющее устройство
$\delta(q, X, a) = (q', \gamma, \rightarrow)$

Рис. 8

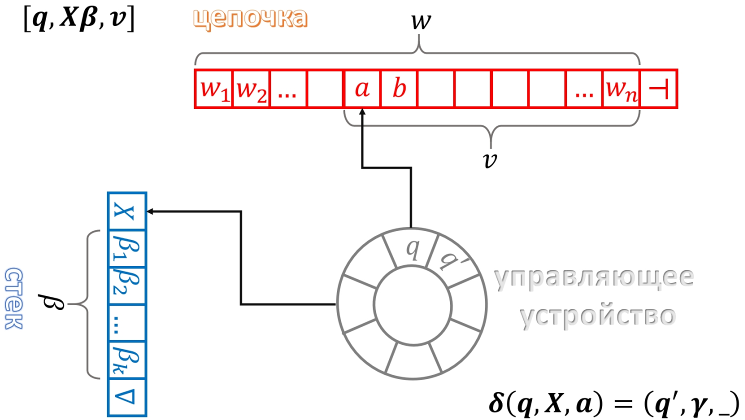
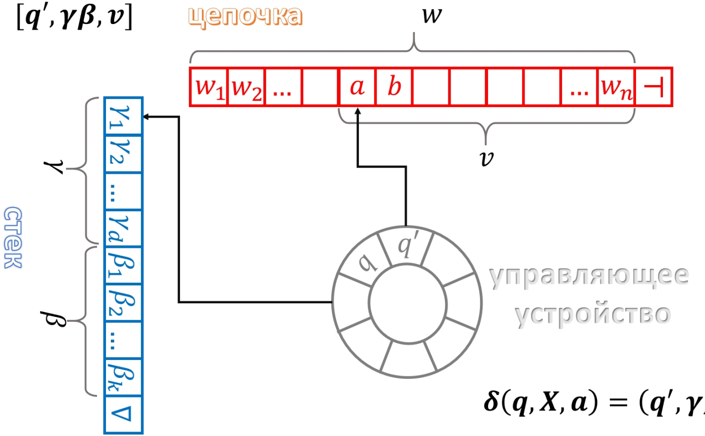

$[q, X\beta, v]$
цепочка
w
w_1 w_2 … a b … w_n $\dashv$
v
X β_1 β_2 … β_k ∇
стек
β
q q'
управляющее устройство
$\delta(q, X, a) = (q', \gamma, \_)$

Рис. 9

$[q', \gamma\beta, v]$



$$\delta(q, X, a) = (q', \gamma, \_)$$

Рис. 10

Распознаваемость МПА при помощи заключительных состояний

- Для автомата $\mathcal{M} = (Q, \Sigma, \Delta, \delta, q_0, F, \gamma_0)$ и слова $w \in \Sigma^*$ начальной конфигурацией будет $[q_0, \gamma_0, w]$.
- Говорят, что слово $w \in \Sigma^*$ распознается (или допускается) МПА $\mathcal{M}$ при помощи заключительных состояний, если существует последовательность $[q_0, \gamma_0, w] \models^* [f, \beta, \dashv]$, где $f \in F$. То есть когда цепочка полностью прочитана МП автомат находится в конечном состоянии (см. рис.11–12).
- Язык, распознаваемый МПА $\mathcal{M}$ — это множество слов над алфавитом Σ, которые распознает автомат, то есть $L(\mathcal{M}) = \{w \in \Sigma^* \mid \exists [q_0, \gamma_0, w] \models^* [f, \beta, \dashv], f \in F\}$

Допуск слова МП-автоматом. Иллюстрация

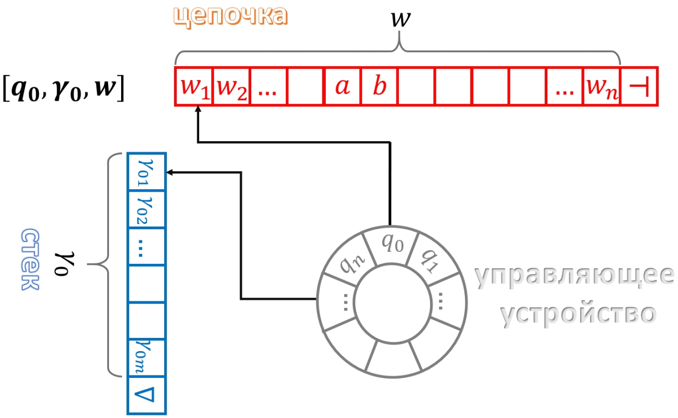


Рис. 11

Допуск слова МП-автоматом. Иллюстрация

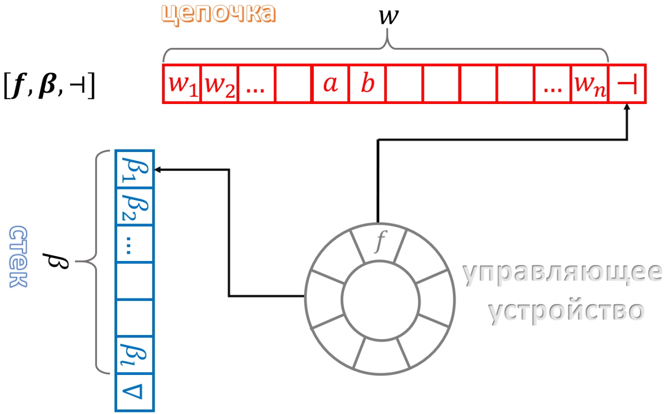


Рис. 12

Пример

		a	b	$\dashv$
q_0	a	$(q_0, aa, \to)$	$(q_1, \varepsilon, \to)$	
	∇	$(q_0, a, \to)$		
q_1	a		$(q_1, \varepsilon, \to)$	
	∇			$(q_2, \varepsilon, \_)$
q_2	a			
	∇			

Рассмотрим управляющую таблицу автомата, у которого $\gamma_0 = \varepsilon, F = \{q_2\}$.

Проверим, распознает ли он слово $w = a^2 b^2$

Пример (продолжение)

Проиллюстрируем протокол обработки слова a^2b^2:

№ такта	состояние	содержимое стека	позиция указателя
1	q_0	∇	$\diamond$aabb$\dashv$
2	q_0	a∇	a$\diamond$abb$\dashv$
3	q_0	aa∇	aa$\diamond$bb$\dashv$
4	q_1	a∇	aab$\diamond$b$\dashv$
5	q_1	∇	aabb$\diamond\dashv$
6	q_2	∇	aabb$\diamond\dashv$

Получим последовательность конфигураций

$$[q_0, \varepsilon, aabb \dashv] \models [q_0, a, abb \dashv] \models [q_0, aa, bb \dashv] \models$$

$$\models [q_1, a, b \dashv] \models [q_1, \varepsilon, \dashv] \models [q_2, \varepsilon, \dashv]$$

Следовательно, $a^2b^2 \in L(\mathcal{M})$.

Пример (продолжение)

Проиллюстрируем протокол обработки слова a^2b^2:

№ такта	состояние	содержимое стека	позиция указателя
1	q_0	∇	$\diamond$aabb$\dashv$
2	q_0	a∇	a$\diamond$abb$\dashv$
3	q_0	aa∇	aa$\diamond$bb$\dashv$
4	q_1	a∇	aab$\diamond$b$\dashv$
5	q_1	∇	aabb$\diamond$$\dashv$
6	q_2	∇	aabb$\diamond$$\dashv$

Получим последовательность конфигураций

$$[q_0, \varepsilon, aabb \dashv] \models [q_0, a, abb \dashv] \models [q_0, aa, bb \dashv] \models$$

$$\models [q_1, a, b \dashv] \models [q_1, \varepsilon, \dashv] \models [q_2, \varepsilon, \dashv]$$

Следовательно, $a^2b^2 \in L(\mathcal{M})$.
Распознает ли автомат слова $a^3b^3, abab, \varepsilon$? Какой язык распознает автомат?