Теория автоматов и формальные грамматики

Регулярные грамматики и конечные автоматы

Основные определения

Приведем некоторые базовые определения теории грамматик.

- ***Алфавит*** **(V)** – конечное непустое множество элементов, называемых *символами* (*буквами*).
- ***Цепочкой*** *(или словом)* в алфавите V называется любая конечная последовательность символов этого алфавита.

Пусть задан алфавит V = {a,b,c}. Тогда α = baaa является словом в алфавите V.

- Цепочка, которая не содержит ни одного символа, называется *пустой* цепочкой и обозначается **e**.

Основные определения

- *Длиной цепочки* ***w*** называется число составляющих ее символов (обозначается |*w*|), причём каждый символ считается столько раз, сколько раз он встречается в *w*.
  Например, |baaa| = 4 и |**e**| = 0.
- Обозначим через ***V**** **-** множество, содержащее все цепочки в алфавите V, **включая пустую цепочку е.**
- Обозначим через V**+** - множество, содержащее все цепочки в алфавите *V*, **исключая пустую цепочку е.**

Например, пусть V = {1,0}, тогда:

*V** ={e,0,00,01,10,11, 000,…},

V+ ={0,1,00,01,10,11,000,...}.

Основные определения

- ***Формальный язык*** – это множество слов (строк, цепочек) над конечным алфавитом V.

Например, множество {a, abb} является языком над алфавитом {a,b}, множество $\{a^k b a^l \mid k \le l\}$ является языком над алфавитом {a,b}.

Необходимо различать *пустой язык* L=Ø и язык, содержащий только пустую цепочку: L={**e**}.

Поскольку каждый язык является множеством, можно рассматривать операции объединения, пересечения и разности языков, заданных над одним и тем же алфавитом (обозначения $L1 \cup L2$, $L1 \cap L2$, $L1 - L2$).

Основные определения

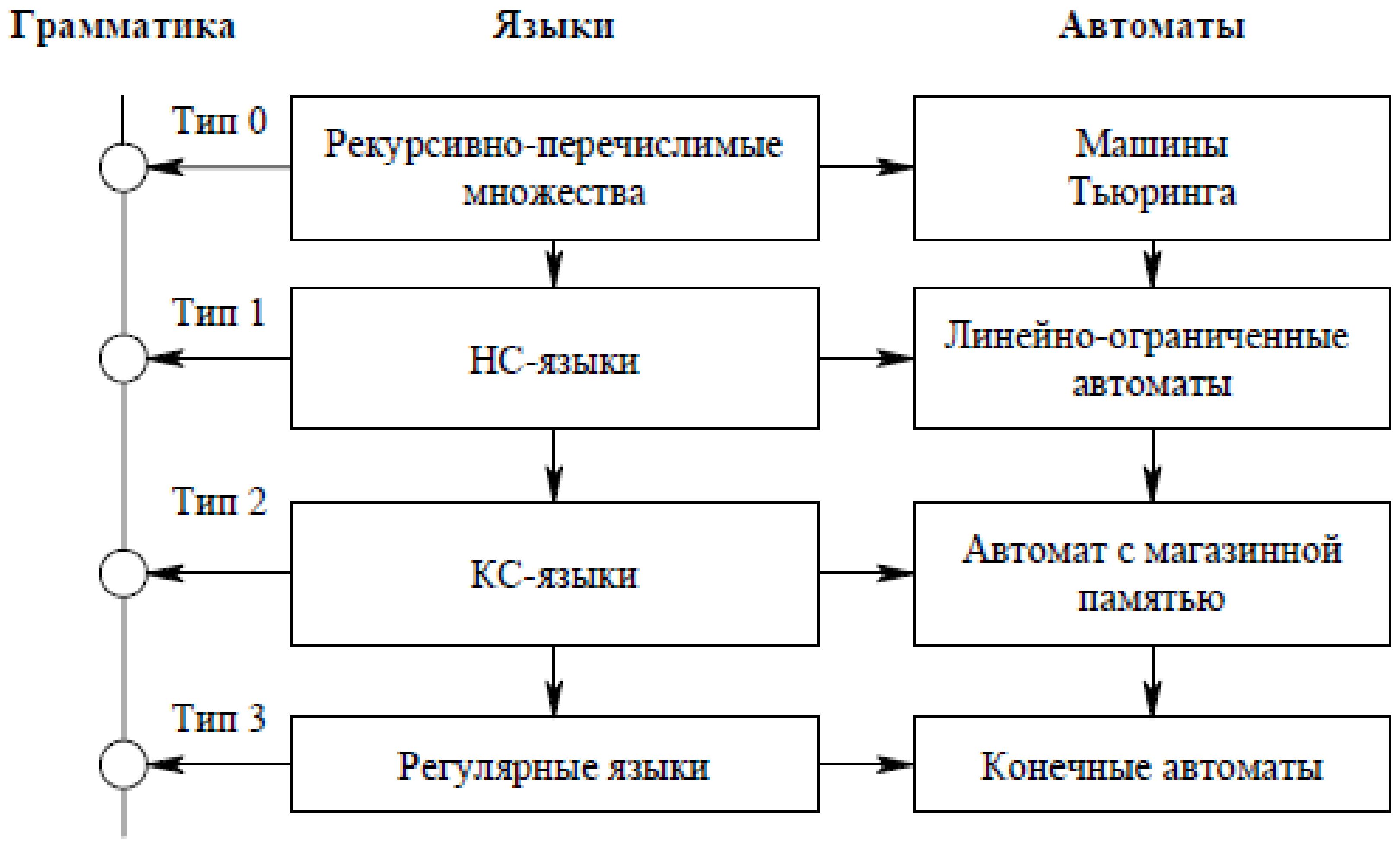
- ***Грамматика*** – система правил, предназначенная для задания множества цепочек и символов данного алфавита.

G – грамматика; L(G) – язык этой грамматики.

Классификация по Хомскому

- *Грамматики типа 2.* К этому типу относятся контекстно-свободные грамматики (**КС-грамматики**, бесконтекстные грамматики). Грамматика *G = (VT,VN,P,S)* называется *контекстно-свободной грамматикой* (КС-грамматикой), если ее правила вывода имеют вид: **A→β**, где $A \in V_N$; $\beta \in V+$ для *неукорачивающих КС-грамматик*, $\beta \in V^*$ для *укорачивающих*.
- То есть грамматика допускает появление в левой части правила только нетерминального символа. КС-грамматики широко применяются для описания синтаксиса компьютерных языков (программирования).

Иерархия: Грамматики – Языки - Автоматы



КС-грамматики и МП-автоматы

КС-грамматики

- Контекстно-свободной грамматикой (КС-грамматикой) называется система $G = (V_T, V_N, S, P)$, где V_T и V_N – непересекающиеся конечные множества терминальных и нетерминальных символов (терминалов и нетерминалов) соответственно;
- $S \in V_N$ – начальный символ, называемый аксиомой грамматики, P – конечное множество правил. Каждое правило в КС-грамматике имеет вид $A \rightarrow a$, где $A \in V_N$ и $a \in (V_N \cup V_T)^*$

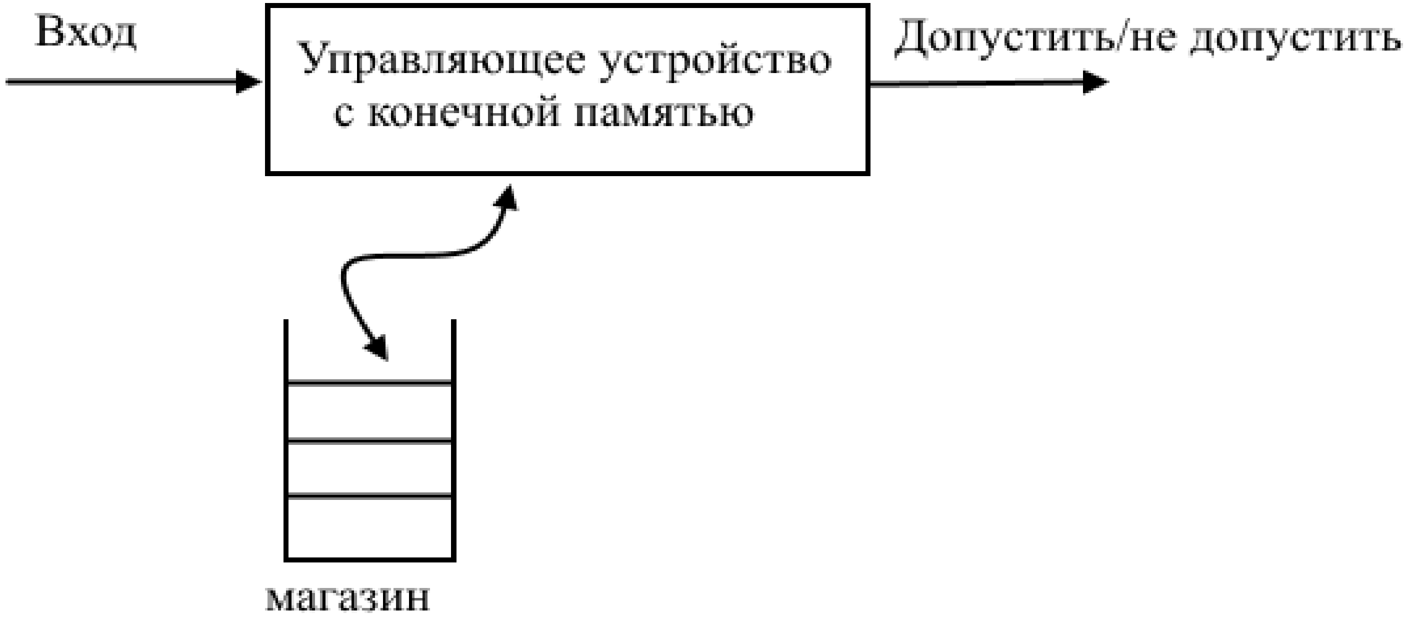
КС-грамматики

- **Пример:** $G_1 = (\{r,t\}, \{S\}, S, P)$ (КС-грамматика), P состоит из правил вида $S \to rStS \mid tSrS \mid \varepsilon$.
- Грамматика порождает множество всех слов в алфавите $\{r,t\}$, содержащих одинаковое число букв r и t.
- Слово $a = trttrr$. Процесс его вывода из аксиомы S :

$$S \Rightarrow tSrS \Rightarrow trStSrS \Rightarrow trtSrS \Rightarrow trtSr \Rightarrow trttSrSr \Rightarrow trttrr \in L.$$

МП-автомат

- Автомат с магазинной памятью (МП-автомат)– это НКА, имеющий дополнительную потенциально неограниченную ленту памяти (магазин). В начальный момент времени магазин содержит начальный символ Z_0.



МП-автомат

МП-автоматом называется система $S = (Q, X, M, \delta, q0, Z0, F)$,

Q – конечное множество состояний,

X – конечный входной алфавит,

M – конечный алфавит магазинных символов,

$\delta: Q \times (M \cup \{\varepsilon\}) \times X \rightarrow 2^{Q \times X^*}$ - функция переходов,

$q0 \in Q$ – начальное состояние,

Z_0 – начальный магазинный символ, так называемый маркер дна,

$F \subseteq Q$ – множество заключительных состояний.

МП-автомат

Конфигурация МП-автомата S` – это тройка (q,w,a) $\in$ {Q x X* x M*},

q – текущее состояние,

w – оставшаяся часть входного слова, при w= ε – слово прочитано,

a – содержимое магазина, при a= ε магазин пуст.

Начальная конфигурация – конфигурация вида (q0, w, Z0), w$\in$ X*.

Заключительная конфигурация вида (q, ε, a), где q$\in$ F и a$\in$ M*.

МП-автомат

- Отношение $\vdash$ называется **тактом работы** МП-автомата.
- Определяется такт следующим образом:

$(q, rw, Za) \vdash (q', w, ya)$, если множество $\delta\ (q, r, Z)$ содержит (q', y), где $q, q' \in Q$, $r \in X \cup \{\varepsilon\}$, $w \in X^*$, $Z \in M$ и $a, y \in M^*$.

Если магазин пуст, то следующий такт невозможен.

МП-автоматы и КС-грамматики

- $C \vdash^{0} C'$ означает, что конфигурация $C=C'$,
- $C_0 \vdash^{k} C_k$, $k>=1$ означает, что существуют такие конфигурации $C_1, \ldots, C_{k-1}$, что $C_i \vdash C_{i+1}$, для всех $0<=i<=k$.
- Отношение $\vdash^{*}$ - рефлексивное транзитивное замыкание отношения $\vdash$ ($C \vdash^{*} C'$), означающее $C \vdash^{k} C'$, для $k>=0$.
- Отношение $\vdash^{+}$ - транзитивное замыкание отношения $\vdash$ ($C \vdash^{+} C'$), означающее $C \vdash^{k} C'$, для $k>=1$.

МП-автоматы и КС-грамматики

- МП-автомат допускает слово w, если

$(q0, w, Z0) \vdash^* (q, \varepsilon, a)$ для некоторых **$q \in F$,** $a \in X^*$.

- Язык L(S) допускаемый МП-автоматом S – множество всех входных цепочек, допускаемых автоматом.

$L(S) = \{w \mid w \in X^*, (q0, w, Z0) \vdash^* (q, \varepsilon, a), q \in F, a \in M^*\}$.

МП-автоматы и КС-грамматики

- По КС-грамматике G возможно построить МП-автомат S_G такой, что $L_\varepsilon(S_G)= L(G)$.
- Для КС-грамматики $G=(V_T, V_N, P, S)$ строится МП-автомат $S_G=(\{q\}, V_T, V_T \cup V_N, \delta, q, S`, \varnothing)$ который допускает язык с опустошением всего магазина.
- Функция переходов δ будет определяться следующим образом:

$\delta(q, \varepsilon, M) = \{(q, a) \mid M \rightarrow a \in P\}$ для всех $M \in V_N$;

$\delta(q, r, r) = \{(q, \varepsilon)\}$ для всех $a \in V_T$.

МП-автоматы и КС-грамматики

- **Пример:** Необходимо построить МП-автомат, допускающий язык $L=\{r^n t^n | n>=1\}$.
- Автомат задается формальной системой $S=(\{q_0,q_1,q_2,q_3\}, \{r,t\}, \{Z_0, 0\}, \delta, q_0, \{q_3\})$.

Функция переходов МП-автомата

Q	$X \cup \{\varepsilon\}$	M	δ
q_0	r	Z_0	$(q_1, 0Z_0)$
q_1	r	0	$(q_1, 00)$
q1	t	0	(q_2, ε)
q_2	t	0	(q_2, ε)
q_2	ε	Z_0	(q_3, Z_0)

МП-автоматы и КС-грамматики

- Входное слово rrtt. Последовательность конфигураций автомата S следующая:
- $(q_0, rrtt, Z_0) \vdash$
- $(q_1, rtt, 0Z_0) \vdash$
- $(q_1, tt, 00Z_0) \vdash$
- $(q_2, t, 0Z_0) \vdash$
- $(q_2, \varepsilon, Z_0) \vdash$
- (q_3, ε, Z_0)

МП-автоматы и КС-грамматики

- **Пример 2**
- $G=(\{r, +, *, (,)\}, \{E, T, F\}, P, E)$,
- правила вывода (P) имеют вид: $E\rightarrow E+T$, $E\rightarrow T$, $T\rightarrow T*F$, $T\rightarrow F$, $F\rightarrow (E)$, $F\rightarrow r$.
- Необходимо построить автомат S_G.

  $S_G=(\{q\}, X, M, \delta, q, E, \varnothing)$, где $X=\{r, +, *, (,)\}$.

МП-автоматы и КС-грамматики

- При входе r*(r+r) для S_G возможна последовательность конфигураций
- (q, r*(r+r), E) ├─ (q, r*(r+r), T) ├─
- (q, r*(r+r), T*F) ├─ (q, r*(r+r), F*F) ├─
- (q, r*(r+r), r*F) ├─ (q, *(r+r), *F) ├─
- (q, (r+r), F) ├─ (q, (r+r), (E)) ├─
- (q, r+r), E+T)) ├─ (q, r+r), T+T)) ├─
- (q, r+r), F+T)) ├─ (q, r+r), r+T)) ├─
- (q, +r), +T)) ├─ (q, r), T)) ├─
- (q, r), F)) ├─ (q, r), r)) ├─
- (q,),)) ├─ (q, ε, ε).

Функция переходов МП-автомата S_G

Q	$X \cup \{\varepsilon\}$	M	δ
q	ε	E	{(q, E+T), (q, T)}
q	ε	T	{(q, T*F), (q, F)}
q	ε	F	{(q, E), (q, r)}
q	r	r	(q, ε)
q	+	+	(q, ε)
q	*	*	(q, ε)
q	(	(	(q, ε)
q	)	)	(q, ε)

МП-автоматы и КС-грамматики

- Приведенная последовательность соответствует **левому выводу** слова r*(r+r) в КС-грамматике G. $E \Rightarrow T \Rightarrow T*F \Rightarrow F*F \Rightarrow r*F \Rightarrow r*(E) \Rightarrow r*(E+T) \Rightarrow r*(T+T) \Rightarrow \Rightarrow r*(F+T) \Rightarrow r*(r+T) \Rightarrow r*(r+F) \Rightarrow r*(r+r) \in L(G)$.