

# Формальные языки

алфавит  $\Sigma$  - конечное непустое мн-во

слово, цепочка, строка  $\Sigma^* = \bigcup_{k=0}^{\infty} \Sigma^k$

формальный язык  $L \subseteq \Sigma^*$

⑬ мн-во формальных языков (даже: языков)  $2^{\Sigma^*}$  бесконечно

„Описание” языка - слово конечной длины

Пр.:

$$\Sigma = \{0, 1\}$$

$L$  = „слова, содержащие четное число единиц”

$$0101 \in L$$

$$0111 \notin L$$

⑬ всего „описаний” счетное мн-во

Перечисление слов  $\{01, 011, 10, 1010\}$

- метод распознавания ( $w \in L$  ?)
- метод порождения (описание слова  $L$ )

Описание языка через метод порождения:

Е-н.с.п

$A, B$  - н.с.п  $\Rightarrow AB$  - н.с.п

$A$  - н.с.п  $\Rightarrow (A)^*$  - н.с.п

$\exists C^{++}$ , программа  $P$

$$L = \{w \mid p(w) = 1\}$$

## Регулярные/автоматные языки

Конкатенация:  $L, B \in \Sigma^*$   $LB$

$$1) L \in \Sigma^k, B \in \Sigma^e \quad LB \in \Sigma^{k+e}$$

$$2) \gamma = LB \quad \gamma_i = \begin{cases} L_i & i \leq k \Rightarrow L_i \\ B_{i-k} & i > k \Rightarrow B_{i-k} \end{cases}$$

$$3) (LB)^\gamma = L(B^\gamma) \quad L\varepsilon = \varepsilon L = L$$

$$AB = \{x \mid x = yz, y \in A, z \in B\}$$

Пр:

$$A = \{0, 01\}, B = \{0, 10\}$$

$$AB = \{00, 010, 0110\}$$

Операции:

$$1) A \cup B$$

$$2) AB$$

$$\sum^k A^k = \underbrace{AA \dots A}_k$$

$$A^0 = \{\varepsilon\}$$

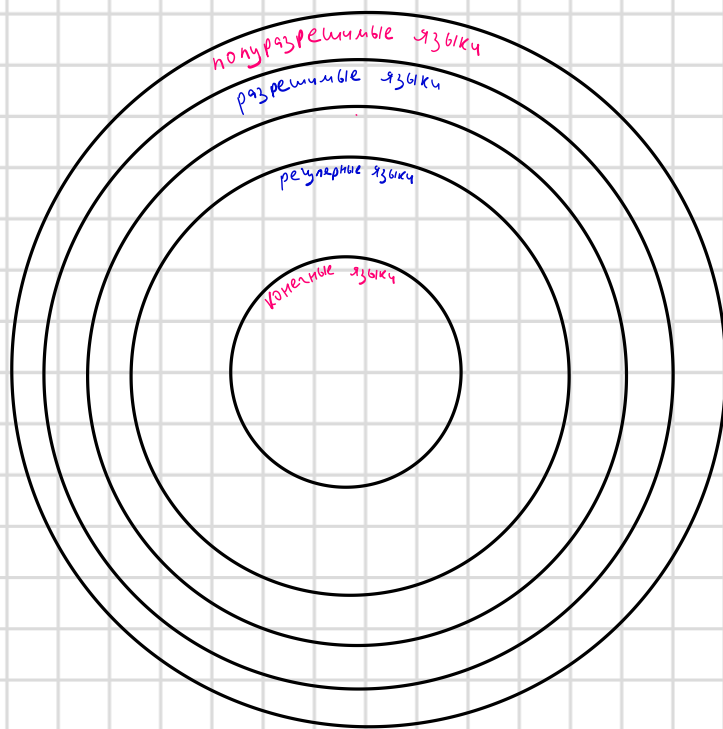
$$3) \text{Замыкание Кнута} \quad A^* = \bigcup_{k=0}^{\infty} A^k$$

$$\text{Reg}_0 = \{\emptyset, \{\varepsilon\}, \{c\} \text{ for } c \in \Sigma\}$$

Базовые регулярные языки

$$\text{Reg}_{i+1} = \text{Reg}_i \cup \{A \cup B, AB, A^* \mid A, B \in \text{Reg}_i\}$$

$$\emptyset^* = \{\varepsilon\}$$



$$Reg_1 = \{ \emptyset, \{\epsilon\}, \{a\}, \{b\}, \dots, \{a, b\}, \{a, \epsilon\}, \dots, \{ab\}, \{aa\}, \{\epsilon, a, a, a, \dots\}, \{\epsilon, b, b, \dots, b\} \dots \}$$

Опр.: Семейство регулярных языков:  $Reg = \bigcup_{k=0}^{\infty} Reg_k$

Лемма (замкнутость регулярных языков)

$$\exists A, B \in Reg \Rightarrow \begin{cases} \bullet A \cup B \in Reg \\ \bullet AB \in Reg \\ \bullet A^* \in Reg \end{cases}$$

$$\triangleright A \in Reg_i, B \in Reg_j$$

$$A \cup B, AB, A^* \in Reg_{\max(i,j)+1} \subset Reg \quad \blacktriangleleft$$

$$X \in Good \quad X \subset 2^{\Sigma^*} \quad \text{Хорошее семейство языков}$$

$$X = \text{set} \langle \text{lang} \rangle$$

$$Good: \text{set} \langle \text{set} \langle \text{lang} \rangle \rangle$$

$$1) Reg_0 \subset X$$

$$2) X \text{ замкнуто относительно } AB, A \cup B, A^* \quad (A, B \in X \Rightarrow A \cup B, AB, A^* \in X)$$

Теорема ("Хорошесть" регулярных)

$$Reg = \bigcap_{X \in Good} X$$

$\triangleright$  потом  $\blacktriangleleft$

Обозначение:

$$1) \emptyset - \text{пустое мн-во}$$

$$2) \epsilon - \text{пуст. строка}$$

$$3) c - \text{символ}$$

- 4)  $\exists$  язык  $A \cup B$  on negative  $L$   $B$   $\left. \begin{array}{l} A \\ B \end{array} \right\}$
- $AB \rightarrow L B$  (mid)
  - $A \cup B \rightarrow L B$  (min)
  - $A^* \rightarrow L^*$  (make priority)

Пр:

$$\Sigma = \{0, 1\}$$

$$(0/11)^*$$

1)  $0 = \{0\}$

3)  $0/11 = \{0, 11\}$

$$1 = \{1\}$$

2)  $11 = \{11\}$

$(^* / AB)^*$  ← академические регулярные выражения

$$L^+ = L \cdot L^*$$

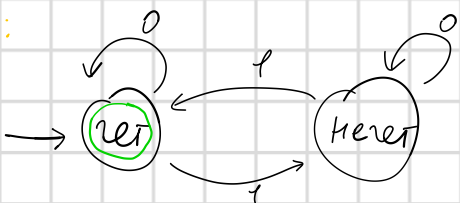
$$L^k = \underbrace{L L \dots L}_k$$

Опишем „слова, содержащие четное число единиц“

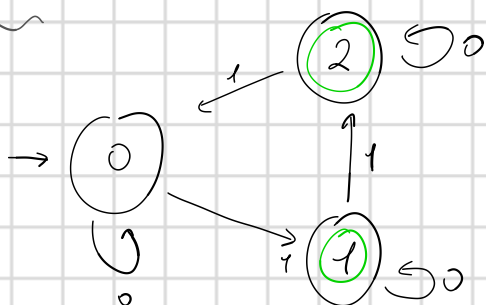
$$0^* / (0^* 1 0^* 1 0^*)^*$$

## Детерминированный конечный автомат

4:



четн. кол-во единиц



число не делится на 3

$$A = \langle \underbrace{\Sigma}_{\text{алфавит}}, \underbrace{Q}_{\text{конечное мн-во состояний}}, \underbrace{q \in Q}_{\text{начальное состояние}}, \underbrace{T \subset Q}_{\text{конечное состояние}}, \underbrace{\delta: Q \times \Sigma \rightarrow Q}_{\text{функция переходов}} \rangle$$

$$(i) \xrightarrow{c} \delta(i, c)$$

$$\exists X \quad |X| = n$$

$$\text{Snap} = Q \times \Sigma^* \quad (\text{мн-во всех возможных описаний автомата})$$

$$\vdash: \langle q, \alpha \rangle \vdash \langle r, \beta \rangle \quad (\text{отношение "переход за один шаг"})$$

$$1) \alpha = c\beta, c \in \Sigma$$

$$2) r = \delta(q, c)$$

Пр:

$$\langle 4, 0101 \rangle \vdash \langle 4, 101 \rangle \vdash \langle 4, 01 \rangle \vdash \langle 4, 1 \rangle \vdash \langle 4, \epsilon \rangle$$

$$L(A) = \{w \mid \langle s, w \rangle \vdash^* \langle t, \epsilon \rangle, t \in T\} \quad \leftarrow \text{язык автомата } A$$

Теорема (Книги)

$$\text{Reg} = \text{Aut}$$

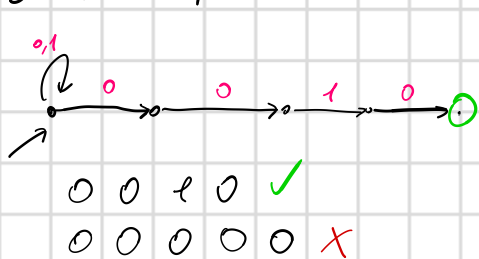
$$\text{Aut} = \{X \mid \exists \text{ ДКА } A: X = L(A)\}$$

План зок-ва:

$$\text{для } L \exists \text{ рег. выражение} \Leftrightarrow \text{для } L \exists \text{ ДКА}$$

$$\text{рег. выражение} \xrightarrow{\text{рекурсивно}} \text{E-НКА} \xrightarrow{\text{E-замык.}} \text{НКА} \xrightarrow{\text{ал. Томсона}} \text{ДКА}$$

$$\text{ДКА} \xrightarrow{\text{д.п.}} \text{рег. вып.}$$



Неретерминированный К.А.

$\triangleright \Rightarrow \text{Reg} \subset \text{Aut}$

$$HKA = \langle \Sigma, Q, S \subseteq Q, T \subseteq Q, \delta: Q \times \Sigma \rightarrow 2^Q \text{ (set <int>)} \rangle$$
$$DKA = \langle \dots, \delta: Q \times \Sigma \rightarrow Q \rangle$$
$$L_A = \{ \omega \mid \omega - \text{можно принять } A \}$$

$w \in L_A$  : Нужно показать путь

$w \notin L_A$ : Ну пусть путь не пройдет в Терм. со ст.

Эмүлүгүм ДКА

$\mathcal{I} : \text{int} \longleftrightarrow \bullet$

T: array<bool>

$$T[q] = \text{true} \iff \textcircled{q}$$
$$\delta[q][c] = r \quad q \xrightarrow{c} r$$

bool check(a):

```
int cur = 5
```

```
for c in a:
```

$$cur = s[cur][c]$$

```
return T[cur]
```

 $|a| + 1$ 

Hand-drawn diagram of a 2D array with row and column indices. The array has 4 rows and 6 columns. Row 0 contains 0s, 1s, and 0s. Row 1 is empty. Row 2 contains 0s, 1s, and 0s. Row 3 is empty. The first column is labeled '1' and the last column is labeled 'Q'. The first row is labeled 'Row' and the second row is labeled 'K-1'.

$\text{can}[K][q]$  - можно прийти в  $q$ , считав  $K$  символов

A diagram showing a vector  $c$  pointing from point  $r$  to point  $q$ .

$$\text{can}[K][q] = 1 \quad \text{if} \quad \text{can}[K-1][r] = 1 \quad \&\& \quad \mathcal{J}[r][c] \ni q$$

Row next(<sup>вектор</sup>Row, X, c):

Row ans;

for r in Q:

if X[r]=1

for q in d[r, c]:

ans[q]=r

check(a):

Row cur = {s}

for c in a:

cur = next(cur, c)

return (cur ∩ T) ≠ ∅

DKA

$Q_{DKA} = 2^{Q_{HKA}}$

$2^A = \text{мн-во}$   $2^B = \text{мн-во}$   $2^{A \cup B} = \text{мн-во}$

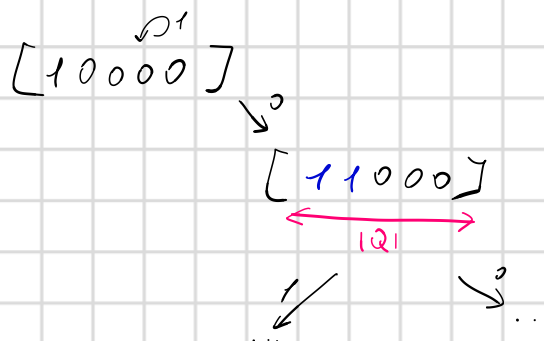
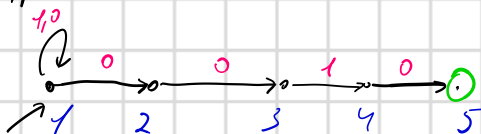
$S_{DKA} = \{S_{HKA}\}$  Row

$\delta(a, c) = \text{next}(a, c)$

$g \in T_{DKA} : g \cap T_{HKA} \neq \emptyset$

алгоритм Томпсона

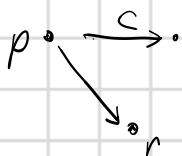
Пр:



$\epsilon$ -переход:

$\epsilon$ -НКА

можно:



можно:



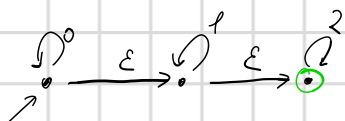
$p \rightsquigarrow q$ , не имеет смысла

НКА  $\xleftrightarrow{!}$   $\epsilon$ -НКА

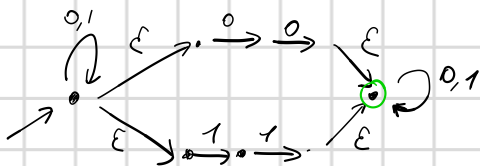
$\epsilon$ -замыкание

Пр:

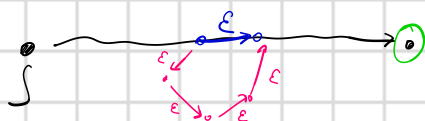
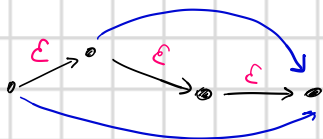
1)  $0^* 1^* 2^*$



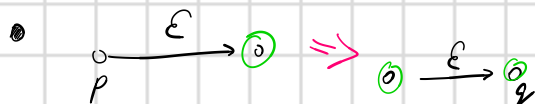
2) либо 00 или 11



• Транзитивно замыкнуть граф  $\epsilon$ -пер.



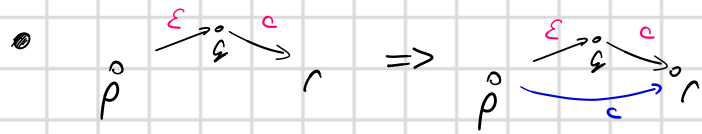
• Если сразу можно замыкнуть, то это можно замыкнуть не сразу 2  $\epsilon$ -пер подряд





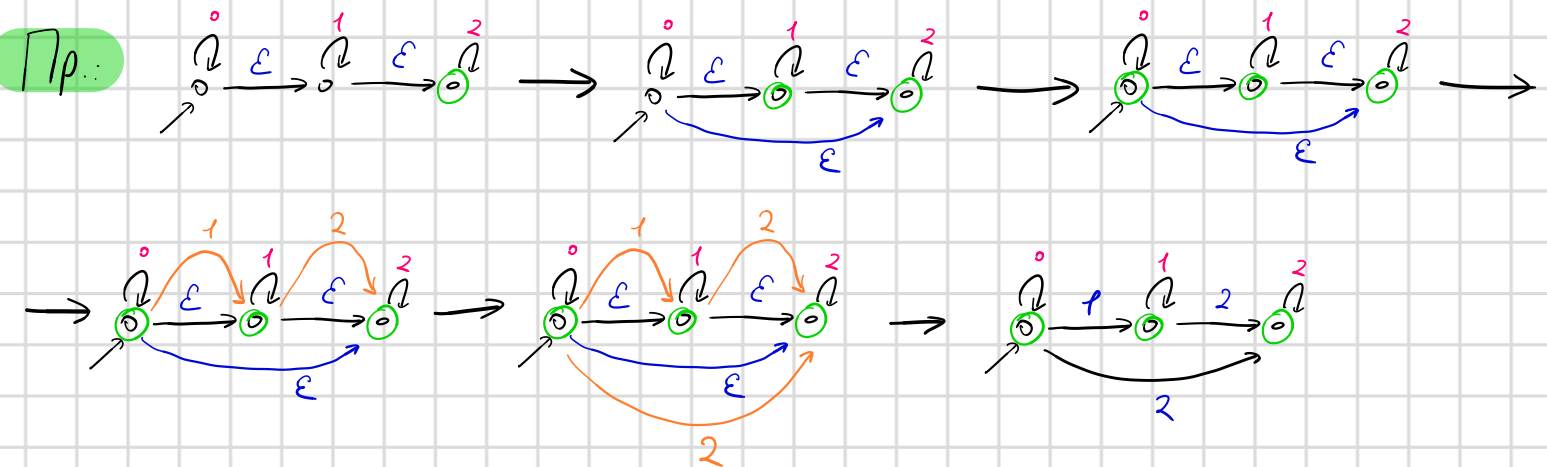
если слово можно пропустить, то его можно пропустить и

после пер. на  $\epsilon$

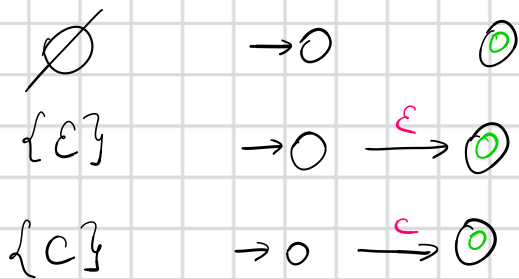


если слово можно пропустить, то его можно пропустить без  $\epsilon$ -пер.

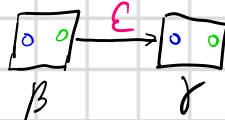
$\Rightarrow \epsilon\text{-НКА} \rightarrow \text{НКА}$



$\boxed{0 \ 0}$  ← автомат  $\boxed{1}$  - терм,  $\boxed{2}$  - вх.с

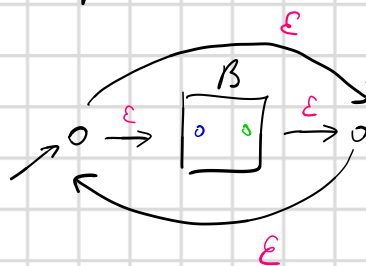
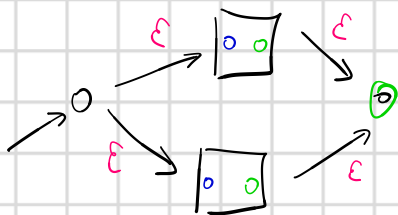


$L = \beta \gamma$   $w = ab$



$L = \beta \gamma$

$L = \beta^*$



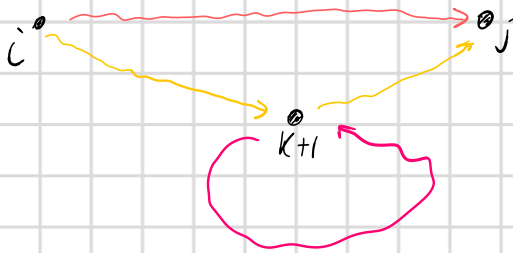
$\Leftarrow$  Aut  $\subset$  Reg

DKA  $\rightarrow$  per. Bop.

$\mathcal{G}_{ijk} - \text{per. Bop. jne cob.} \Leftrightarrow i \xrightarrow[\leq k]{\sim} j$

$\mathcal{G}_{i10} = \mathcal{E} | \underbrace{\mathcal{C}_1 | \mathcal{C}_2}_{\substack{\text{символы} \\ \text{на нечетк}}} \dots$   
 $\mathcal{C}_1, \mathcal{C}_2$

$$\mathcal{G}_{ijk+1} = \mathcal{G}_{ijk} / \mathcal{G}_{i,k+1,k} \left( \mathcal{G}_{k+1,k+1,k} \right)^* \mathcal{G}_{k+1,j,k}$$



орун поз:  $\mathcal{G}_{i,k+1,k}$   $\mathcal{G}_{k+1,i,k}$

два поз:  $\mathcal{G}_{i,k+1,k}$   $\mathcal{G}_{k+1,k+1,k}$   $\mathcal{G}_{k+1,j,k}$

Ответ:  $\mathcal{G}_{st_1n} / \mathcal{G}_{st_2n} / \mathcal{G}_{st_3n} \dots, \epsilon_i \in T \blacktriangleleft$