



**Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Московский государственный технический университет  
имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)**

---

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

**Лабораторная работа № 1**

**Дисциплина Конструирование компиляторов**

**Тема Распознавание цепочек регулярного языка**

**Вариант №6**

**Студент Иванов В.А.**

**Группа ИУ7-21М**

**Преподаватель Ступников А.А.**

Москва.  
2023 г.

**Цель работы:** приобретение практических навыков реализации важнейших элементов лексических анализаторов на примере распознавания цепочек регулярного языка.

**Задачи работы:**

- 1) Ознакомиться с основными понятиями и определениями, лежащими в основе построения лексических анализаторов.
- 2) Прояснить связь между регулярным множеством, регулярным выражением, праволинейным языком, конечноавтоматным языком и недетерминированным конечно-автоматным языком.
- 3) Разработать, протестировать и отладить программу распознавания цепочек регулярного или праволинейного языка в соответствии с предложенным вариантом грамматики.

### Вариант 6

Напишите программу, которая в качестве входа принимает произвольное регулярное выражение, и выполняет следующие преобразования:

- 1) Преобразует регулярное выражение непосредственно в ДКА.
- 2) По ДКА строит эквивалентный ему КА, имеющий наименьшее возможное количество состояний (Воспользоваться алгоритмом Хопкрофта)
- 3) Моделирует минимальный КА для входной цепочки из терминалов исходной грамматики.

Текст программы и набор тестов приведены в приложении.

Пример работы программы (в режиме отладки):

```
Regular expression:
(a/b)*abb
Input string:
ababb
|3- "ababb" |2- "babb" |1- "abb" |2- "bb" |1- "b" |0- ""
String valid: True
```

### Выводы

В результате выполнения лабораторной работы были изучены основные понятия построения лексических анализаторов, разработана программа распознавания цепочек регулярного языка.

## Контрольные вопросы

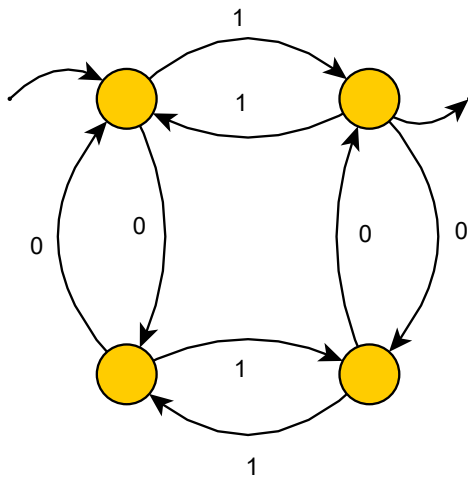
- 1) Какие из следующих множеств регулярны? Для тех, которые регулярны, напишите регулярные выражения.
- 2) Найдите праволинейные грамматики для тех множеств из вопроса 1, которые регулярны.
- 3) Найдите детерминированные и недетерминированные конечные автоматы для тех множеств из вопроса 1, которые регулярны.

а. Множество цепочек с равным числом нулей и единиц.

Нерегулярное. Соответствующий автомат будет иметь бесконечность состояний (состояния соответствуют текущей разнице между количеством нулей и единиц).

б. Множество цепочек из  $\{0, 1\}^*$  с четным числом нулей и нечетным числом единиц.

Регулярно. Составим ДКА,



Если существует ДКА, то из него можно получить соответствующее регулярное выражение:

$(0(11)^*0)^*(1|0(11)^*10)(0(11)^*0|(1|0(11)^*10)(0(11)^*0)^*(1|0(11)^*10))^*$

Праволинейная грамматика:

$S \rightarrow A$	$A \rightarrow 0B$ $B \rightarrow 11B$ $B \rightarrow C$ $C \rightarrow 0A$ $A \rightarrow D$	$D \rightarrow 1E$ $D \rightarrow 0F$ $F \rightarrow 11F$ $F \rightarrow J$ $J \rightarrow 10E$	$E \rightarrow 0H$ $H \rightarrow 11H$ $H \rightarrow I$ $I \rightarrow 0E$ $E \rightarrow e$	$E \rightarrow 1K$ $E \rightarrow 0L$ $L \rightarrow 11L$ $L \rightarrow M$ $M \rightarrow 10K$	$K \rightarrow 0N$ $N \rightarrow 11N$ $N \rightarrow O$ $O \rightarrow 0K$	$K \rightarrow 1E$ $K \rightarrow 0P$ $P \rightarrow 11P$ $P \rightarrow U$ $U \rightarrow 10E$
-------------------	---	---	---	---	--	---

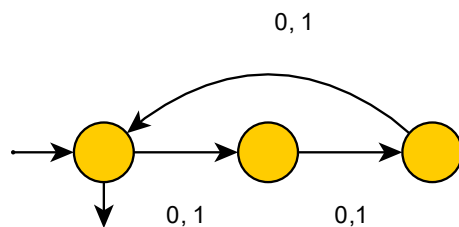
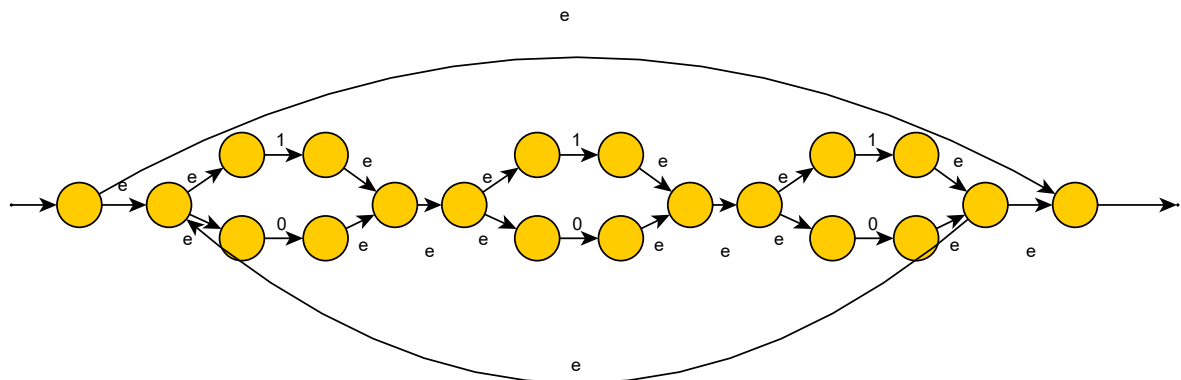
с. Множество цепочек из  $\{0, 1\}^*$ , длины которых делятся на 3.

Регулярно.

$((0|1)(0|1)(0|1))^*$

Праволинейная грамматика:

$S \rightarrow A$	$A \rightarrow 0B$ $A \rightarrow 1B$ $A \rightarrow e$	$B \rightarrow 0C$ $B \rightarrow 1C$	$C \rightarrow 0A$ $C \rightarrow 1A$
-------------------	---	--	--



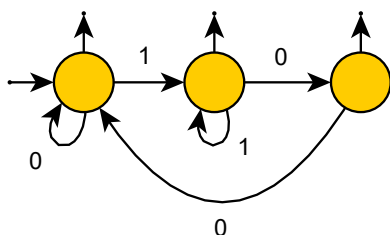
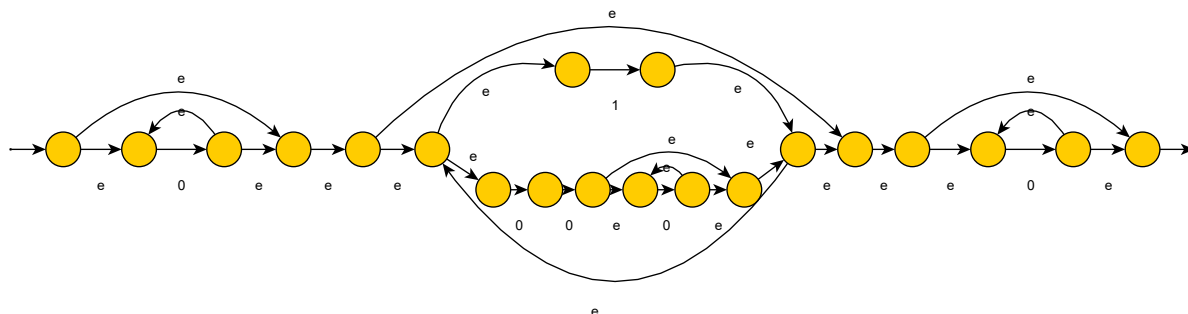
d. Множество цепочек из  $\{0, 1\}^*$ , не содержащих подцепочки 101.

Регулярно.

$0^*(1|000)^*0^*$

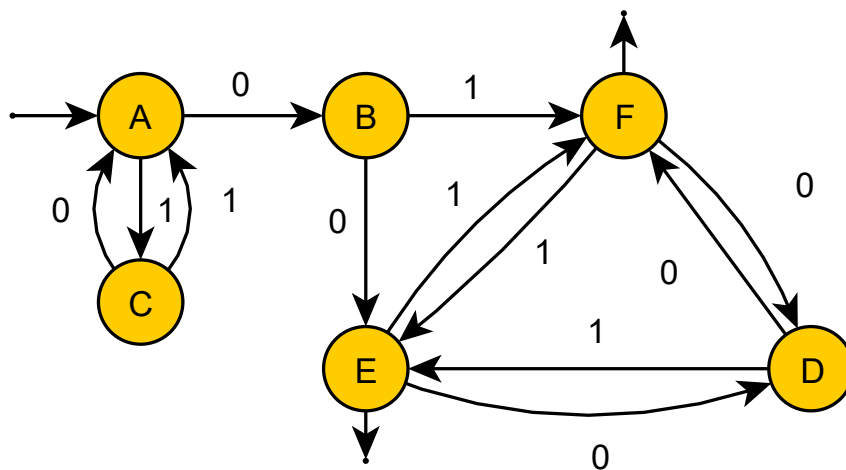
Праволинейная грамматика:

$S \rightarrow A$	$A \rightarrow 0A$ $A \rightarrow B$	$B \rightarrow 1B$ $B \rightarrow 00C$ $B \rightarrow D$	$C \rightarrow 0C$ $C \rightarrow B$	$D \rightarrow 0D$ $D \rightarrow e$
-------------------	---	--	---	---



4) Найдите конечный автомат с минимальным числом состояний для языка, определяемого автоматом  $M = (\{A, B, C, D, E\}, \{0, 1\}, d, A, \{E, F\})$ , где функция  $d$  задается таблицей

Состояние	Вход	
	0	1
A	B	C
B	E	F
C	A	A
D	F	E
E	D	F
F	D	E



По алгоритму минимизации (алгоритм Хопкрофта)

$P = \{\{A, B, C, D\}, \{E, F\}\}$  – начальное разбиение

Перебираются сплитеры (пара множество состояний - символ).

Сплитер  $\langle P_0, 1 \rangle$  разбивает  $P_0$  на множества  $\{A, C\}, \{B, D\}$

$P = \{\{A, C\}, \{B, D\}, \{E, F\}\}$

Сплитер  $\langle P_0, 0 \rangle$  разбивает  $P_0$  на множества  $\{A\}, \{C\}$

$P = \{\{A\}, \{C\}, \{B, D\}, \{E, F\}\}$

Больше разбиений не происходит. Новый автомат имеет 4 состояния

