

Лабораторная работа 1

“Распознавание цепочек регулярного языка”

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.

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).

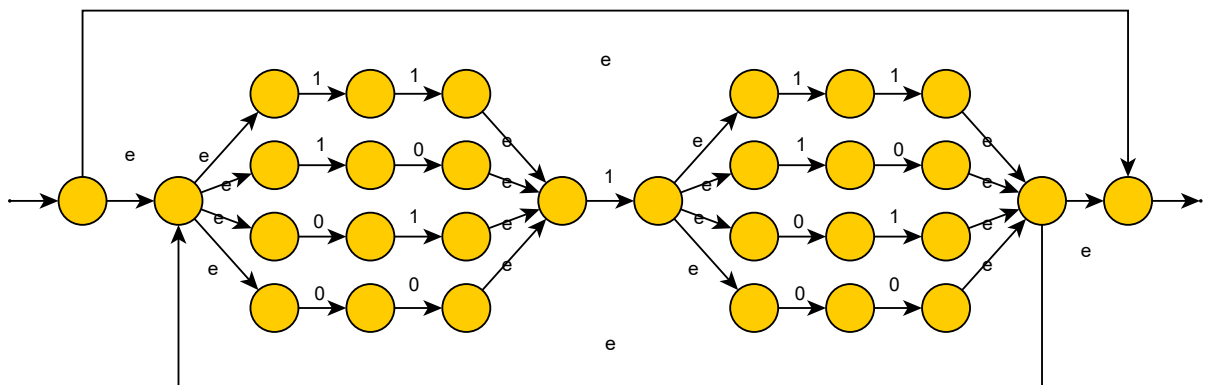
б. Множество цепочек из $\{0, 1\}^*$ с четным числом нулей и нечетным числом единиц.

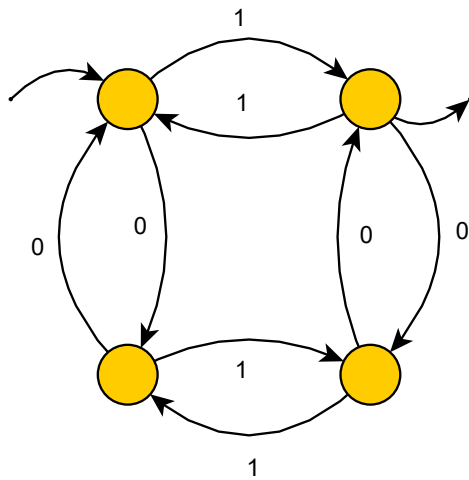
Регулярно.

$(00|01|10|11)^*1(00|01|10|11)^*$

Праволинейная грамматика:

$S \rightarrow A$	$A \rightarrow 00A$ $A \rightarrow 01A$ $A \rightarrow 10A$ $A \rightarrow 11A$ $A \rightarrow B$	$B \rightarrow 1C$	$C \rightarrow 00C$ $C \rightarrow 01C$ $C \rightarrow 10C$ $C \rightarrow 11C$ $C \rightarrow \epsilon$
-------------------	---	--------------------	--





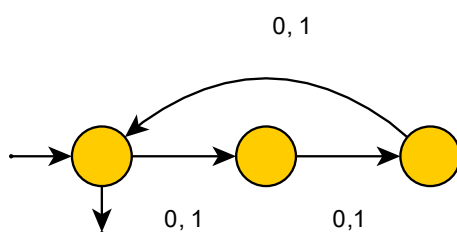
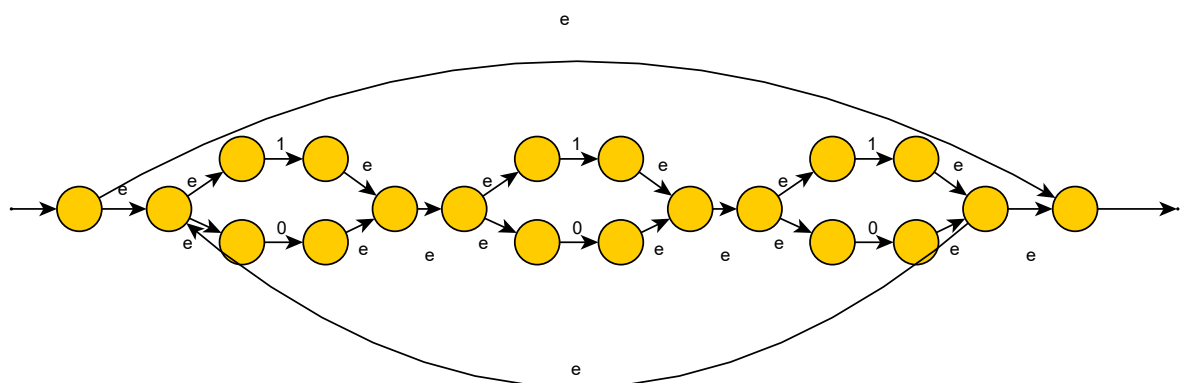
с. Множество цепочек из $\{0, 1\}^*$, длины которых делятся на 3.

Регулярно.

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

Праволинейная грамматика:

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



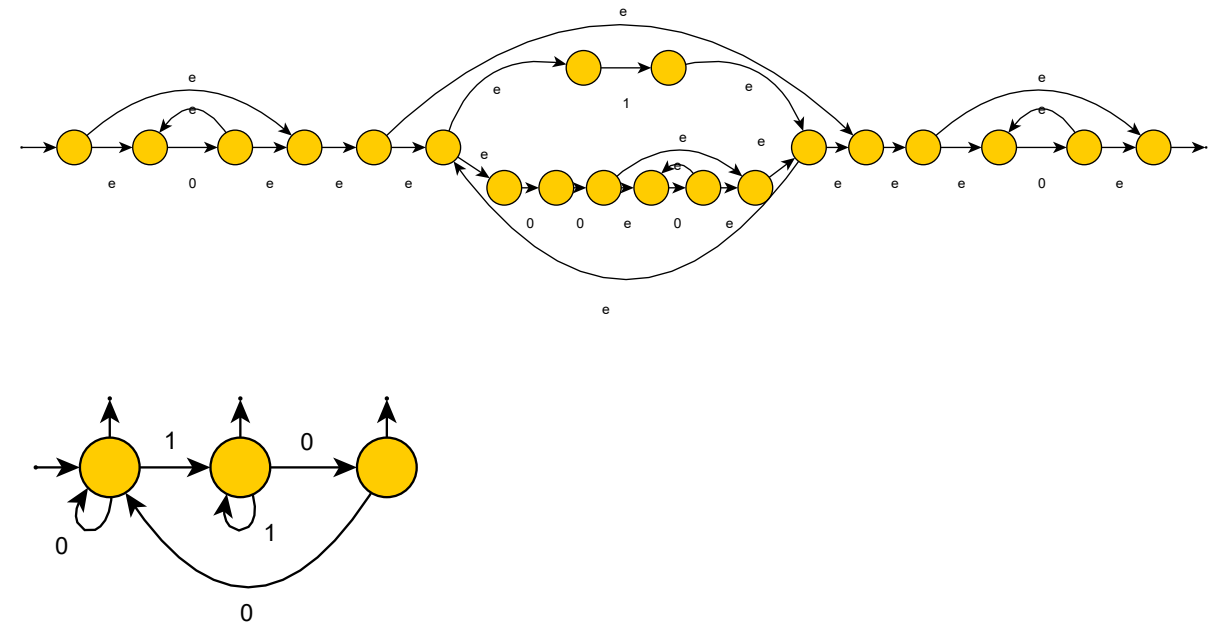
d. Множество цепочек из $\{0, 1\}^*$, не содержащих подцепочки 101.

Регулярно.

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

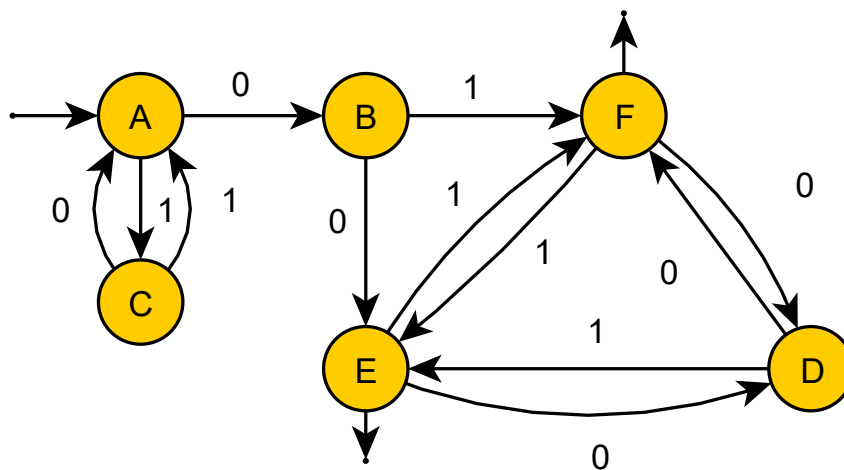
Праволинейная грамматика:

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



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

