Иванов Всеволод ИУ7-21М 04.03.2023

Лабораторная работа 1, вариант 6

"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, тестировать 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).

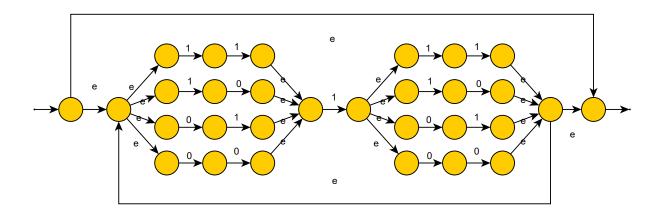
b. Множество цепочек из $\{0,1\}^*$ с четным числом нулей и нечетным числом единиц.

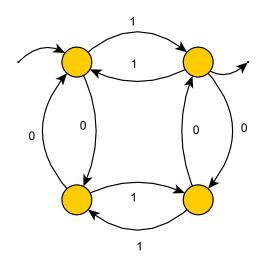
Регулярно.

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

Праволинейная грамматика:

S->A	A->00A A->01A A->10A	B->1C	C->00C C->01C C->10C
	A->11A		C->11C C->e
	A->B		





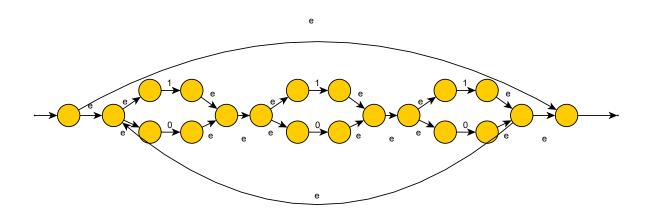
с. Множество цепочек из $\{0,1\}^*$, длины которых делятся на 3.

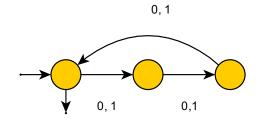
Регулярно.

((0|1)(0|1)(0|1))*

Праволинейная грамматика:

S->A	A->0B	B->0C	C->0A
	A->1B	B->1C	C->1A
	A->e		





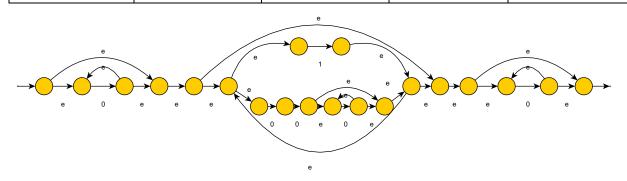
d. Множество цепочек из $\{0,1\}^*$, не содержащих подцепочки 101.

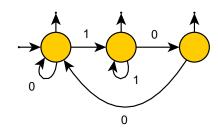
Регулярно.

0*(1|000*)*0*

Праволинейная грамматика:

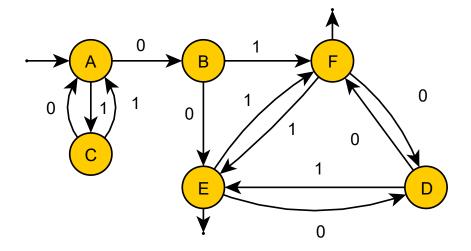
S->A	A->0A	B->1B	C->0C	D->0D
	A . D	B->00C	C->B	D->e
	A->B	B->D		





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	В	С
В	Е	F
C	A	A
D	F	Е
Е	D	F
F	D	Е



По алгоритму минимизации (алгоритм Хопкрофта)

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

Перебираются сплитеры (пара множество состояний - символ).

Сплитер <Р0, 1> разбивает Р0 на множества $\{A,C\}$, $\{B,D\}$

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

Сплитер <P0, 0> разбивает P0 на множества $\{A\}$, $\{C\}$

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

Больше разбиений не происходит. Новый автомат имеет 4 состояния

