



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 ИУ «Информатика и системы управления»

КАФЕДРА ИУ-7 «Программное обеспечение ЭВМ и информационные технологии»

ЛАБОРАТОРНАЯ РАБОТА № 1

ТЕМА:

«Распознавание цепочек регулярного языка»

ВАРИАНТ №4

Группа: ИУ7-21М

Студент: Дубовицкая Ольга Николаевна

Дисциплина: Конструирование компиляторов

Преподаватель: Ступников Андрей Алексеевич

Москва, 2024 г.

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, конечно-автоматным языком и недетерминированным конечно-автоматным языком.
- 3) Разработать, протестировать и отладить программу распознавания цепочек регулярного языка в соответствии с предложенным вариантом грамматики.

Содержание работы (Вариант 4):

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

- 1) По регулярному выражению строит НКА.
- 2) По НКА строит эквивалентный ему ДКА.
- 3) По ДКА строит эквивалентный ему КА, имеющий наименьшее возможное количество состояний (алгоритм Бржозовского).
- 4) Моделирует минимальный КА для входной цепочки из терминалов исходной грамматики.

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

Недетерминированный конечный автомат (НКА) состоит из:

- 1) Множества состояний S ;
- 2) Множества входных символов Σ (входного алфавита); считаем, что символ, обозначающий пустую строку, не является членом Σ ;
- 3) Функции переходов, которая для каждого состояния и каждого символа из $\Sigma \cup \{\varepsilon\}$ даёт множество последующих состояний;
- 4) Состояния из S , обозначаемого как стартовое (начальное);
- 5) Множества состояний F , являющегося подмножеством S и известное как множество допускающих (конечных состояний)

Детерминированный конечный автомат (ДКА) представляет собой частный случай НКА, в котором:

- 1) Нет переходов для входа ε ;
- 2) Для каждого состояния s и входного символа a имеется ровно 1 дуга, выходящая из s и помеченная a .

Алгоритм Томпсона строит по НКА эквивалентный ему ДКА следующим образом:

Шаг 1. Помещаем в очередь множество Q , состоящее только из стартовой вершины.

Шаг 2. Затем, пока очередь не пуста, выполняем следующие действия:

1. Достаём из очереди множество – назовём его q ;
2. Для всех $s \in \Sigma$ посмотрим, в какое состояние ведёт переход по символу s из каждого состояния в q . Полученное множество состояний кладём в очередь Q , только если оно не лежало там раньше. Каждое такое множество в итоговом ДКА будет отдельной вершиной, в которую будут вести переходы по соответствующим символам;

3. Если в множестве q хотя бы одна из вершин была терминальной в НКА, то соответствующая данному множеству вершина в ДКА также будет терминальной.

Алгоритм Бржозовского для минимизации конечного автомата

Введём следующие обозначения:

- A – конечный автомат;
- $d(A)$ – детерминированный автомат для A ;
- $r(A)$ – обратный автомат для A ;
- $dr(A)$ – результат $d(r(A))$; аналогично для $rdr(A)$ и $drdr(A)$.

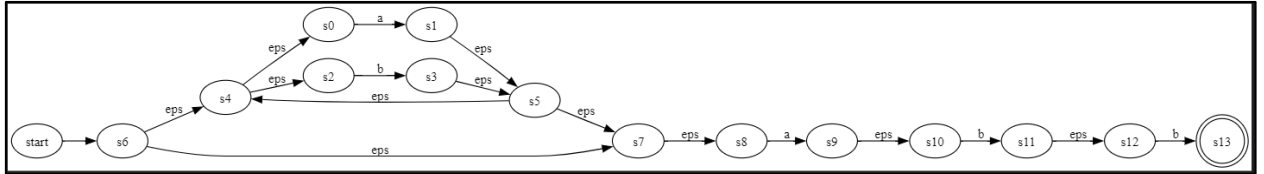
Пусть A – автомат (необязательно детерминированный), который распознаёт язык L . Минимальный детерминированный автомат AL может быть вычислен следующим образом:

$$AL = drdr(A)$$

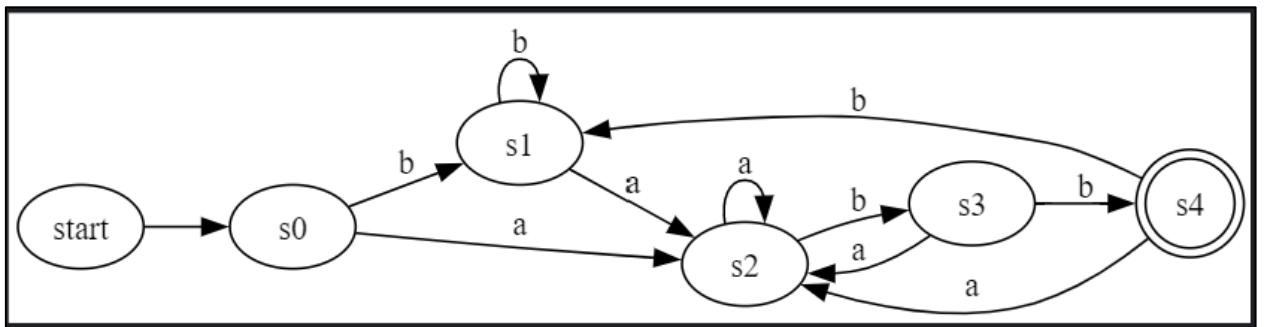
РЕЗУЛЬТАТЫ РАБОТЫ

Введённое регулярное выражение: $(a|b)^*abb$

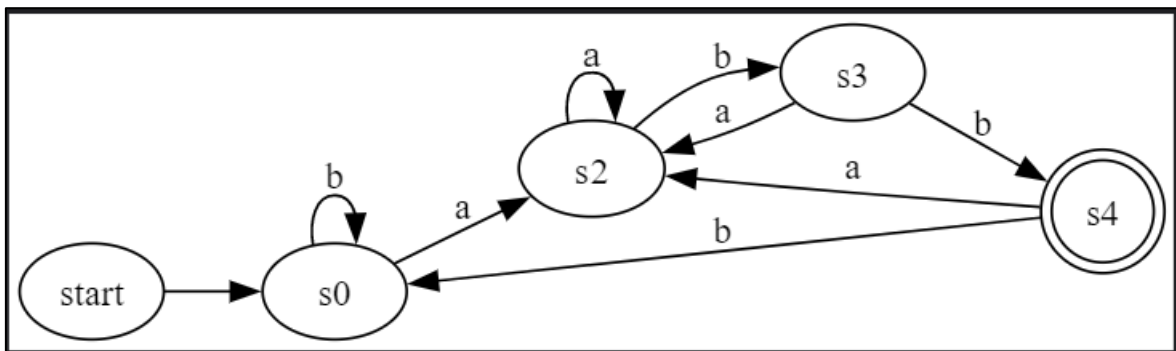
Построенный по нему НКА:



ДКА, эквивалентный построенному КА:



КА, эквивалентный построенному ДКА и имеющий наименьшее возможное количество состояний (алгоритм Бржозовского):



Входная цепочка	Результат
abb	True
babb	True
bbb	False

ЗАКЛЮЧЕНИЕ

В ходе выполнения лабораторной работы была реализована программа на языке Python, позволяющая распознавать цепочки регулярного языка.

В программе было реализовано построение НКА по произвольно введённому регулярному выражению, по НКА был построен эквивалентный ему ДКА, по ДКА был построен эквивалентный ему КА, имеющий наименьшее возможное количество состояний (в соответствии с алгоритмом Бржозовского), а также был смоделирован минимальный КА для входной цепочки из терминалов исходной грамматики.

Таким образом, в результате выполнения лабораторной работы были приобретены практические навыки реализации важнейших элементов лексических анализаторов на примере распознавания цепочек регулярного языка.