

**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)**

**ОТЧЕТ**

По лабораторной работе №1

По курсу «Конструирование компиляторов»

Тема: «Распознавание цепочек регулярного языка»

Вариант 5

Студент:

Ковалев М.С.

Группа:

ИУ7-22М

Преподаватель:

Ступников А.А.

Москва 2023

**Цель и задачи работы**

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.

**Теоретическая часть**

Распознавание цепочек регулярного языка является центральной задачей лексического анализа, который образует первый этап процесса компиляции. На этом этапе символы, составляющие исходную программу, считываются и группируются в отдельные лексические элементы, называемые лексемами. Большую часть того, что происходит в течение лексического анализа можно моделировать с помощью конечных преобразователей, работающих последовательно или параллельно. Например, лексический анализатор может состоять из ряда последовательно соединенных конечных преобразователей. Первый преобразователь в этой цепи может устранять из исходной программы все несущественные пробелы, второй ликвидирует комментарии, третий ищет константы и т. д. Другая возможность - завести набор конечных преобразователей, каждый из которых ищет определенную лексическую конструкцию. Лексические анализаторы бывают по существу двух видов - прямые и непрямые. Можно показать как по регулярным выражениям, описывающим соответствующие лексемы, строятся анализаторы обоих видов. При непрямом лексическом анализе требуется, прочитав цепочку знаков, определить, появилась ли подцепочка, образующая некоторую конкретную лексему. Если множество возможных цепочек, которые могут образовывать эту лексему, обозначается, как это обычно бывает, регулярным выражением, то проблему построения непрямого лексического анализатора для данной лексемы можно представить себе как проблему реализации конечного преобразователя. Конечный преобразователь - это почти конечный автомат (распознаватель) в том смысле, что он читает вход, не производя выхода, пока не обнаружит присутствие лексемы данного типа (т. е. достигнет заключительного состояния). Тогда он сигнализирует о том, что эта лексема появилась, и выдает на выходе цепочку символов, образующих эту лексему. Проблема непрямого лексического анализа является, таким образом, по существу проблемой построения детерминированного конечного автомата (ДКА) по заданному регулярному выражению и его программной реализации. В простейшем случае сначала по регулярному выражению строят недетерминированный конечный автомат (НКА). Затем этот НКА превращают в ДКА, либо моделируют его работу, прослеживая параллельно всевозможные последовательности тактов. Пусть множество лексем данного типа обозначается регулярным выражением или задается в виде праволинейной грамматики. Во втором случае требуется преобразование праволинейной грамматики в регулярное выражение, которое выполняется за два шага: Шаг 1. Построение стандартной системы уравнений с регулярными коэффициентами по праволинейной грамматике. Шаг 2. Решение стандартной системы уравнений с регулярными коэффициентами. Построение НКА по регулярному выражению может выполняться либо с помощью алгоритма «Конструктор Томпсона», либо с помощью алгоритма построения НКА по расширенному регулярному выражению (см. Алгоритм 3.2. АУ1). Наконец, осуществляют детерминированное моделирование НКА для заданной входной цепочки знаков. При моделировании конечного автомата необходимо учитывать следующие обстоятельства. Работа автомата завершается, если обнаруживается хотя бы одна допускающая конфигурация, достижимая из начальной конфигурации автомата. Если входная цепочка построена синтаксически неправильно, то придется рассмотреть все возможные последовательности тактов автомата. Если исчерпаны все возможные последовательности тактов, а допустимая конфигурация не обнаружена, то надо выдать сообщение об ошибке. Внимание! При моделировании НКА или ДКА необходимо показать все конфигурации алгоритма

**Вариант**

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

1) По регулярному выражению строит ДКА.

2) По ДКА строит эквивалентный ему КА, имеющий наименьшее возможное количество состояний. Указание. Воспользоваться алгоритмом, приведенным по адресу:

[http://neerc.ifmo.ru/wiki/index.php?title=Минимизация\_ДКА,\_алгоритм\_за\_O(n%5E2)\_с\_построением\_пар\_разл ичимых\_состояний](http://neerc.ifmo.ru/wiki/index.php?title=Минимизация_ДКА,_алгоритм_за_O(n%5E2)_с_построением_пар_разл%20ичимых_состояний%20)

3) Моделирует минимальный КА для входной цепочки из терминалов исходной грамматики.

**Текст программы**

<https://github.com/martel42/compilers/tree/master/Lab1/DKA>

**Работа программы**

В начале вводится регулярное выражение, после по нему строится и выводится ДКА и минимальный КА. Далее вводится входная цепочка, которую проверяют проходит ли она КА, после чего выводится результат. Работа программы показана на рисунке 1.

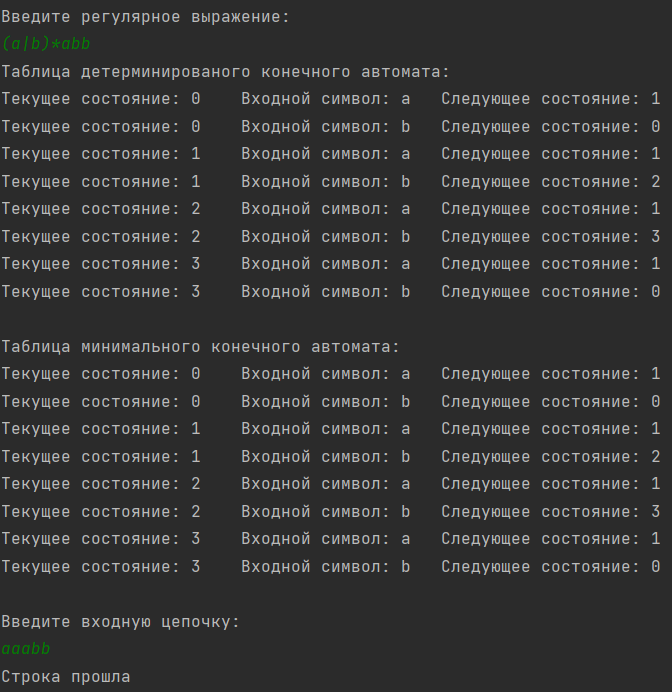
**Вывод**

Рисунок 1

В результате выполнения лабораторной работы были приобретены практических навыков реализации важнейших элементов лексических анализаторов на примере распознавания цепочек регулярного языка. Кроме этого была реализована программа, строящая ДКА по регулярному выражению, минимизирующая его, а также выполняющая функцию проверки входной цепочки на соответствие ДКА.

**Список литературы**

1. БЕЛОУСОВ А.И., ТКАЧЕВ С.Б. Дискретная математика: Учеб. Для вузов / Под ред. В.С. Зарубина, А.П. Крищенко. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2001. 2. АХО А., УЛЬМАН Дж. Теория синтаксического анализа, перевода и компиляции: В 2-х томах. Т.1.: Синтаксичечкий анализ. - М.: Мир, 1978. 3. АХО А.В, ЛАМ М.С., СЕТИ Р., УЛЬМАН Дж.Д. Компиляторы: принципы, технологии и инструменты. – М.: Вильямс, 2008. 4. Notes on lexical analysis.pdf. 5. БУНИНА Е.И., ГОЛУБКОВ А.Ю. Формальные языки и грамматики. Учебное пособие. – М.: Изд-во МГТУ им. Н.Э.Баумана, Москва, 2006. URL: http://iu9.bmstu.ru/data/book/fl.pdf