**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)»**

Факультет: Информатика и системы управления

Кафедра: Теоретическая информатика и компьютерные технологии

**Лабораторная работа № 5**

“Лексический распознаватель”

по дисциплине «Конструирование компиляторов»

Вариант 7

Работу выполнил

студент группы ИУ9-62Б

Сербин Денис

Москва, 2022

# **1. Цель работы**

# Целью данной работы является изучение использования детерминированных конечных автоматов с размеченными заключительными состояниями (лексических распознавателей) для решения задачи лексического анализа.

# **2. Исходные данные**

Пусть лексическая структура модельного языка состоит из шести лексических доменов:

1. пробелы непустые последовательности пробельных символов (пробел; горизонтальная табуляция, маркеры конца строки):

2. идентификаторы непустые последовательности латинских букв и десятичных цифр, начинающиеся с буквы:

3. целочисленные литералы непустые последовательности десятичных цифр:

4. ключевые слова (варианты ключевых слов перечислены в таблицах 1, 2 и 3);

5. знаки операций (варианты знаков операций перечислены в таблицах 1, 2 и 3);

6. комментарии или строковые литералы (варианты перечислены в таблицах 1, 2 и 3).

Чтобы не усложнять лексический анализатор, разрешим идентификаторам примыкать справа к целочисленным литералам.

**3. Задание**

Выполнение лабораторной работы состоит из пяти этапов:

1. описание лексических доменов модельного языка в виде регулярных выражений;

2. построение недетерминированного лексического распознавателя для модельного языка;

3. детерминизация построенного лексического распознавателя и факторизация его алфавита;

4. построение массива обобщённых символов, матрицы переходов и массива заключи тельных состояний для полученного детерминированного лексического распознавателя с факторизованным алфавитом;

5. разработка лексического анализатора, работающего на основе интерпретации построен ных структур данных.

Входной поток для лексического анализатора должен загружаться из файла (в ASCII). В результате работы программы в стандартный поток вывода должны выдаваться описания распознанных лексем в формате:

Тег (координаты\_фрагмента): изображение лексемы

При этом лексемы, принадлежащие домену пробелов, должны отбрасываться. Лексический анализатор должен иметь программный интерфейс для взаимодействия с парсером. Рекомендуется реализовывать его как метод nextToken () для императивных языков или функцию, возвращающую список лексем, для функциональных языков.

Входной файл может содержать ошибки, при обнаружении которых лексический анализатор должен выдавать сообщение с указанием координаты, восстанавливаться и продолжать работу.

В лабораторной работе разрешается использовать любой язык программирования, под держивающий массивы с операцией доступа к элементу по индексу, работающей за констант ное время.

**4. Индивидуальный вариант**

# case, break, \*, (, ), комментарии ограничены знаками (\*, \*) могут пересекать границы строк текста.

# **5. Результат выполнения**

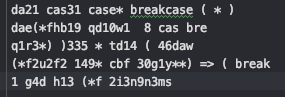


Рисунок 1. Текст входного потока

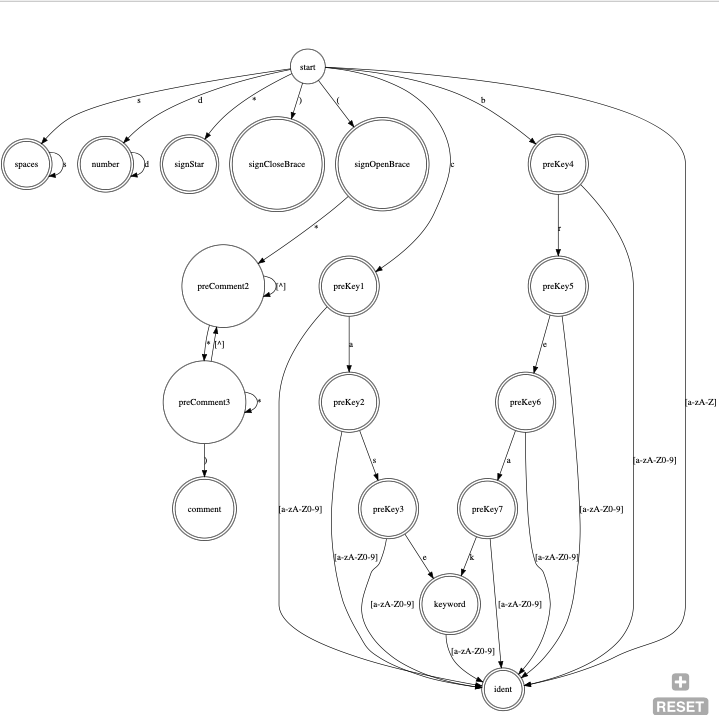


Рисунок 2. Детерминированный лексический распознаватель

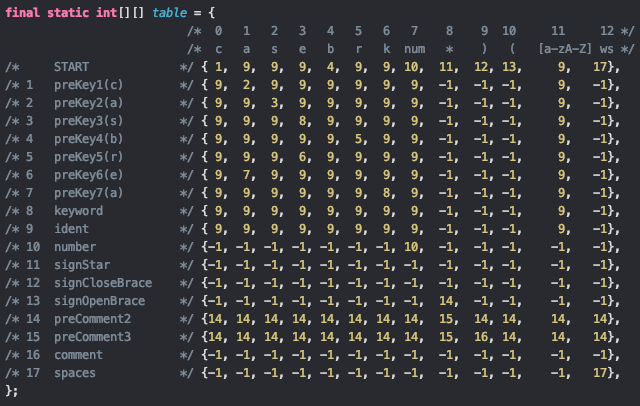


Рисунок 3. Матрица переходов

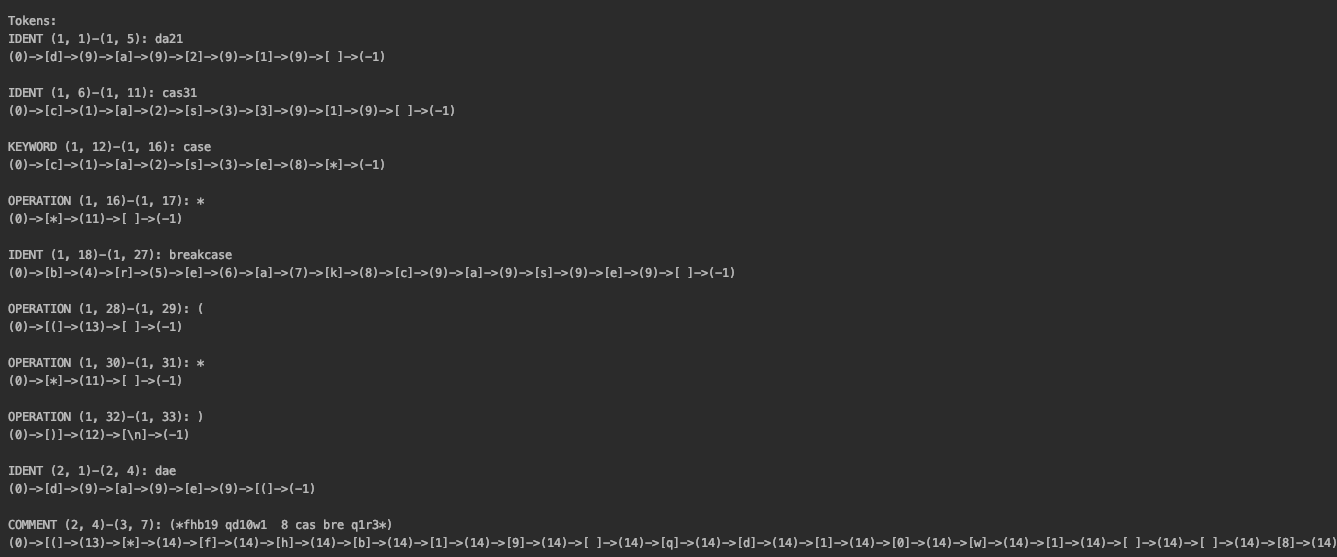


Рисунок 4. Результат

# **6. Вывод**

# В рамках данной лабораторной работы было изучено использование детерминированных конечных автоматов с размеченными заключительными состояниями (лексических распознавателей) для решения задачи лексического анализа.