Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное учреждение высшего образования

«Новгородский государственный университет имени Ярослава Мудрого»

Кафедра «Информационных технологий и систем»

**Проектирование лексического анализатора**

Лабораторная работа

по дисциплине:

«Теория языков программирования и методы трансляции»

Разработал:

Студент группы 8091

Шляханов Д. А. \_\_\_\_\_\_

«\_\_» \_\_\_\_\_2020г

Проверил:

Макаров В. А. \_\_\_\_\_\_

«\_\_» \_\_\_\_\_2020г

**Великий Новгород**

**2020**

**Постановка задачи**

**Формулировка**

Разработать программу, решающую задачу лексического разбора исходного текста программы. Результат работы лексического анализатора выводятся в файл в виде таблицы распознанных лексем.

**Представление входных данных**

В виде входных данных программе подается текст. Для успешного распознавания он должен удовлетворять правилам грамматики из файла *Grammar.txt* в папке *Grammar\_Diagram*. Описанная в данном файле грамматика выбрана на основе варианта №4(арифметические выражения в языке C).

Алфавит состоит из: [[a-z], [A-Z], ['\_'], [0-9], ['.']], [int, double, char, short, long,], ['=', '\*', '/', '%', '&', '~' , '!' , '+' , '-'], [ ';', '(', ')']

Классы лексем:

1. Идентификаторы;
2. Ключевые слова; [int, double, char, short, long,]
3. Числовые константы;
4. Знаки операций; ['=', '\*', '/', '%', '&', '~' , '!' , '+' , '-']
5. Разделители. [ ';', '(', ')']

**Описание выбранного метода решения задачи ЛА**

Для решения задачи лексического анализатора был выбран конечный автомат, поскольку его использование позволяет обрабатывать сложные константы (например, числа с плавающей точкой со знаком, порядком и мантиссой).

Для хранения состояний были использованы enum, для хранения таблицы ключевых слов и идентификаторов я решил использовать словарь std::map, поскольку от него можно ожидать высокой производительности. Для удобства были реализованы класс Lex, enum Num\_E.

**Описание реализации**

Схема автомата:

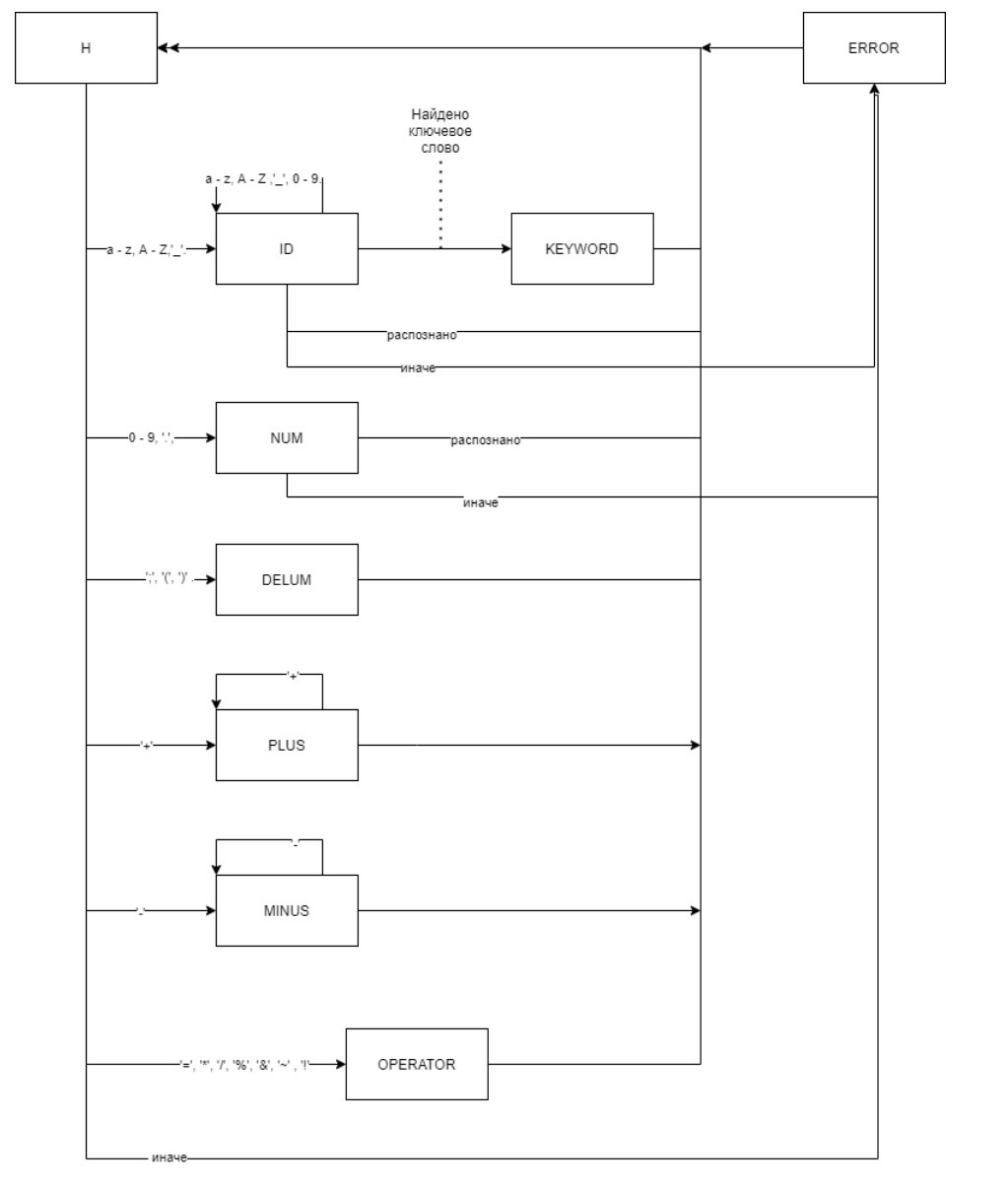
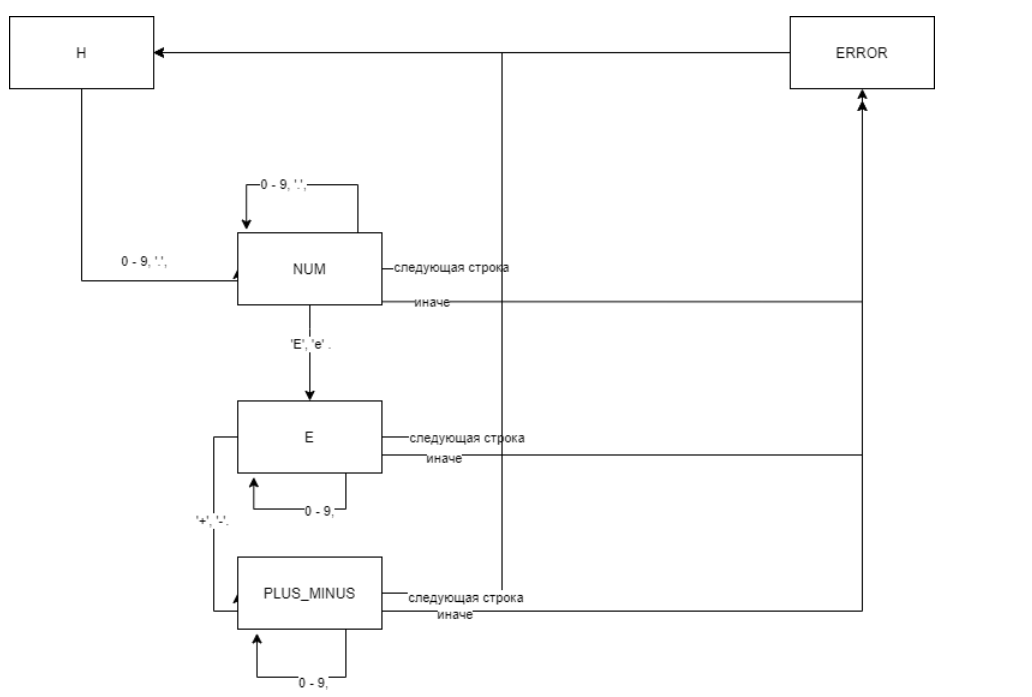


Схема автомата чисел:



**Обработка ошибок**

Лексема помечается «ошибочной», если следующий рассматриваемый символ будет недопустимым, а также если в начальном состоянии придет символ, необрабатываемый программой.

**Методика тестирования**

При тестировании было использовано два файла *test.txt* и *testNumber.txt*

**Вывод**

В итоге проделанной работы был разработан лексический анализатор на основе рассматриваемой грамматики, позволяющий распознать простейшие лексемы, чтобы в будущем использовать их для компиляции программы.