Лексический анализатор

У нас есть кортеж (?) специальных символов, в котором хранятся зарезервированные имена ( program, uses и т.д)

Кортеж для разделителей ( - + и т.д.)

Кортеж для имен и кортеж для чисел.

И список с найденными лексемами

Каждая лексема имеет 3 поля:

- тип (идентификатор, разделитель и тд)

- номер

- значение

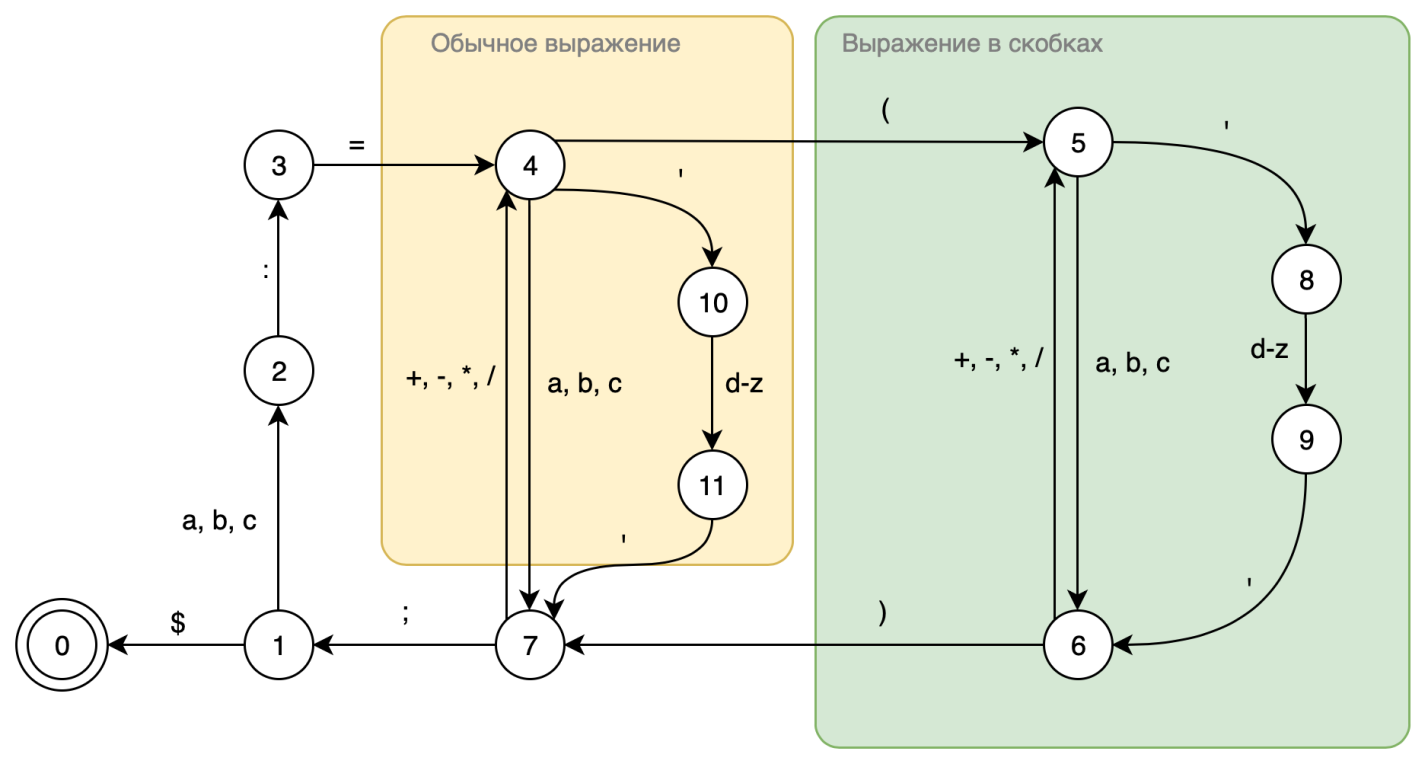
- позиция ( нужна для отображения места ошибки, но пока я не знаю, как этим пользоваться)

Нужен еще буфер для хранения слова. Мы считываем по символу и записываем в буфер. Это нужно для того, чтобы сравнивать с кортежами зарезервированных имен

Смысл:

Принимаем на вход текст проги. Потом посимвольно разбираем и в зависимости от того, что за символ получили, переходим между состояниями автомата.

Пример, для арифметики

Нужно такое же для синтаксиса

Когда мы распознали какое-то слово, то нужно его закинуть в список.

В списке будет “ id:<>, lex: <>, val: <> “.

Разница между лексемой и токеном ?

Лексема – слово в программе

Токен – описание слова “ id:<>, lex: <>, val: <> “