

Формальные языки

домашнее задание до 23:59 23.03

1. Написать лексер для языка L (спецификация ниже), используя любимый генератор лексеров из семейства Lex (или любой другой понравившийся инструмент, только не парсер-комбинаторы, и не пишите лексер вручную). (8 баллов за полностью выполненное задание)
 - Структуры данных для лексем должны однозначно их идентифицировать, а также содержать привязку к коду (в какой строке исходного кода и с какого по какой символ располагается лексема).
 - Составить набор тестов, демонстрирующий правильность работы полученного лексера (качество тестового покрытия важно!).
 - Сделать консольное приложение, принимающее на вход путь к файлу, содержащему программу на языке L, производящее лексический анализ и печатающее полученный поток лексем.
 - Результатом работы лексического анализатора должен быть поток лексем, который печатается при помощи отдельной процедуры печати.
 - Код должен быть размещен на гитхабе, собираться одним скриптом, содержать инструкцию по сборке и запуску собранного приложения, собираться на чистой Ubuntu 16.04 или Windows 10. Все зависимости, в случае их отсутствия в системе, должны либо доставляться скриптом, либо перечисляться в явном виде.

Лексическая спецификация языка L

Программы на языке L записываются символами ASCII. Ограничителями строк являются ASCII-символы CR, LF или 2 подряд идущих символа: CR LF. Пробельными символами являются символы-ограничители строк и символы пробела (SP), табуляции (HT) и перевода страницы (FF). Пробельные символы не имеют значения, для них не должно генерироваться лексем.

Комментарии только однострочные. Комментарием считается суффикс строки до символа-ограничителя строк, начинающийся с "//".

Идентификаторы в языке — произвольная последовательность символов (до первого пробельного символа), начинающаяся либо с буквы, либо символа подчеркивания (), которая не является ключевым словом.

Ключевые слова языка: if, then, else, while, do, read, write, begin, end

Литералы языка: числа с плавающей точкой, true, false.

Операторы языка: +, -, *, /, %, ==, !=, >, >=, <, <=, &&, ||

Разделители языка: (,), ;

Пример программы:

```
read x; if y + 1 == x then write y else write x
```

Результат лексического анализа должен быть в духе:

```
KW_Read(0, 0, 3); Ident("x", 0, 5, 5); Colon(0, 6, 6); KW_If(0, 8, 9); Ident("y", 0, 11, 11);  
Op(Plus, 0, 13, 13); Num(1, 0, 15, 15); Op(Eq, 0, 17, 18); Ident("x", 0, 20, 20); KW_Then(0, 22, 25);  
KW_Write(0, 27, 31); Ident("y", 0, 33, 33); KW_Else(0, 35, 38); KW_Write(0, 40, 44); Ident("x", 0, 46, 47)
```