

## **Лабораторная работа 2. Лексический анализ.**

Освоение работы с существующими лексическими анализаторами (по желанию). Разработка лексического анализатора подмножества языка программирования, определенного в лабораторной работе 1. Определяются лексические правила. Выполняется перевод потока символов в поток лексем (токенов).

Лексический анализатор читает поток символов, составляющих исходную программу, и группирует эти символы в значащие последовательности, называемые лексемами.

На вход программы подается текстовый файл (напр. с именем INPUT.TXT), содержащий строки символов анализируемой программы. Например, строка может задавать переменной значения арифметического выражения в виде

*ПЕРЕМЕННАЯ = ВЫРАЖЕНИЕ.*

Выражение может включать:

Знаки сложения и умножения («+» и «\*»);

Круглые скобки («(» и «)»);

Константы (например, 5; 3.8; 1e+18, 8.41E-10);

### ***Имена переменных.***

Имя переменной – это последовательность букв и цифр, начинающаяся с буквы.

Разбор выражения  $COST = (PRICE + TAX) * 0.98$ .

### ***Проанализируем выражение:***

COST, PRICE и TAX – лексемы-идентификаторы;

0.98 – лексема-константа; =, +, \* – просто лексемы.

Пусть все константы и идентификаторы можно отображать в лексемы типа <идентификатор> (<ИД>). Тогда выходом лексического анализатора будет последовательность лексем  $\langle ИД1 \rangle = (\langle ИД2 \rangle + \langle ИД3 \rangle) * \langle ИД4 \rangle$ .

Вторая часть компоненты лексемы (указатель, т.е. номер лексемы в таблице имен) – показана в виде индексов. Символы «=», «+» и «\*» трактуются как лексемы, тип которых представляется ими самими. Они не имеют связанных с ними данных и, следовательно, не имеют указателей.

## **РАБОТА С ТАБЛИЦЕЙ ИМЕН**

После того, как в результате лексического анализа лексемы распознаны, информация о некоторых из них собирается и записывается в таблицу имен.

Для нашего примера COST, PRICE и TAX – переменные с плавающей точкой. Рассмотрим вариант такой таблицы. В ней перечислены все идентификаторы вместе с относящейся к ним информацией (таблица 1).

Таблица 1 – Таблица имен

Номер элемента	Идентификатор	Информация
1	COST	Переменная с плавающей точкой
2	PRICE	Переменная с плавающей точкой
3	TAX	Переменная с плавающей точкой
4	0.98	Константа с плавающей точкой

Если позднее во входной цепочке попадаете идентификатор, надо справиться в этой таблице, не появлялся ли он ранее. Если да, то лексема, соответствующая новому вхождению этого идентификатора, будет той же, что и у предыдущего вхождения.

Необходимо создание таблиц констант, ключевых слов, разделителей, логических/математических операторов.

При определении неверной последовательности символов, необходимо обнаружить эту ошибку и выдать сообщение о ней. Всего необходимо показать скриншоты нахождения 4-х лексических ошибок.