

ЛЕКСИЧЕСКИЙ АНАЛИЗАТОР

к.т.н., доц. С.Ю. Гавриленко, к.т.н., проф. Ю.Н. Колыбин, студент А.В. Квасенко, Национальный технический университет "ХПИ", г. Харьков

Лексический анализатор (сканер) – это часть компилятора, которая читает исходную программу и выделяет в её тексте структурные единицы языка – лексемы.

Применение лексического анализатора упрощает работу с текстом исходной программы на этапе синтаксического разбора и сокращает объём обрабатываемой информации.

В основном лексические анализаторы выполняют исключение из текста исходной программы комментариев, незначащих пробелов, символов табуляции и перевода строки, а также выделение лексем следующих типов: идентификаторов, числовых, строковых и символьных констант, знаков операций и разделителей, ключевых слов входного языка. Выделенные лексем сохраняются в таблице лексем. Если же лексема является идентификатором, то она дополнительно заносится в таблицу идентификаторов. В случае обнаружения недопустимых лексем, сканер выдает сообщение об ошибке.

Для описания лексических анализаторов, как правило, используют либо формализм регулярных выражений, либо формализм контекстно-свободных грамматик.

По описанию лексического анализатора в виде регулярного выражения или автоматной грамматики строится конечный автомат, распознающий соответствующий язык. Конечный автомат для каждой входной цепочки языка дает ответ на вопрос о том, принадлежит или нет цепочка языку, заданному автоматом.

Выделение границ лексем представляет определённую проблему, поскольку во входном тексте программы лексем не ограничены никакими специальными символами. В некоторых случаях требуется параллельная работа сканера, синтаксического разбора и, возможно, семантического анализа.