

22.22.1 Охарактеризувати лексичний аналізатор компілятора Асемблера з врахуванням обробки макросів.

В результаті лексичного аналізу формується таблиця лексем чергового рядка програми. Лексема – один або більше символів в рядку, які Асемблером розглядаються як єдиний об'єкт. Лексеми – “неподільні атоми” мови програмування. В зв'язку з цим часто використовується термін «термінальний символ» мови програмування. В мові Асемблера розрізняють наступні лексеми:

- Ідентифікатори – послідовності букв та цифр, які розпочинаються з букви. До букв відносяться малі та великі букви відповідних алфавітів. В MASM та TASM таким алфавітом є латинський алфавіт, а також символи `_`, `@`, `?`, `$`. Загальноприйнято не розрізняти малі та великі букви. В той же час малі та великі букви в ASCII мають різні коди. Тому при обробці ідентифікаторів компілятор перетворює всі малі букви в великі. Довжина ідентифікатора практично не обмежується, але значущими є перші 32 символи.
- Числові константи - послідовності цифр та деяких букв, які розпочинаються з цифри. Розрізняють двійкові, восьмеричні, десяткові та шістнадцятирічні константи. В якості допустимих букв числової константи є великі та малі (вони як і у ідентифікаторах не розрізняються) букви `a,b,c,d,e,f` та букви основи системи числення – `b,o,q,d,h`. Двійкові константи складаються із цифр 0 та 1 і обов'язково закінчуються буквою `b`. Восьмеричні константи складаються із цифр від 0 до 7 і обов'язково закінчуються буквою `o` або `q`. Десяткові константи складаються із цифр від 0 до 9 та можуть закінчуватись буквою `d`. Особливістю десятичних констант є те, що в них допускається відсутність букви `d`. Ознакою закінчення десятикової константи в цьому випадку є поява символу, який не входить в перелік цифр від 0 до 9. Шістнадцятирічні константи складаються із цифр від 0 до 9 і букв `a,b,c,d,e,f` та обов'язково закінчуються буквою `h`.
- Текстові константи - послідовність довільних символів, які можна ввести з клавіатури і які розпочинаються та закінчуються символом `'` або `"`. Символи в текстових константах ніяким перетворенням не підлягають.
- Розподільники лексем - символи пропуску, табуляції та точка з комою
- Одно символні лексеми – всі інші символи, які містяться в алфавіті мови Асемблера. Розподільник між двома лексемами є обов'язковим, якщо конкатенація (при стикуванні) цих лексем є також лексемою.

Таблиця лексем може мати, наприклад, наступну структуру

№ п/п	Лексема	Довжина лексеми в символах	Тип лексеми

В результаті формування таблиці лексем з подальшого розгляду видаляються всі розподільники та коментарі.