

1. 答：在编译系统所使用的的文法中定义的字母表的所有元素都被称作字符；记号是编译系统所使用的文法所能推导出来的一个短语；单词是由字符组成的，字符无实际语义，而单词是具有独立意义的最小语法单位；如在C语言中，所有ASCII码都是有效字符，但是只有while、for等预定义或用户自定义的单词具有实际语义。

6. 答：词法分析器以源程序作为字符流输入，依据程序功能等价的要求，将字节流划分组合成一个个单词，并按照先后顺序输出单词序列。

7. 答：单词是由字母所组成的，而字母往往是有限的，难以表达更加复杂和高级的含义，所以由字母依照一定的文法所组成的单词就承担了语义表达的作用，而在单词之上则是句子，能够表达更加复杂的语义，所以相较之下，单词就是具有独立意义的最小语法单位了。

15. 答：双缓冲区是用来解决缓冲区用完时等待读入以及保证单词识别完整性的方案，用于高效和便捷地进行超前查找。

23. (1) $(0|1)^*$

(2) $(0|1)^+$

(3) $1^*(01^+)^*(01|0|1)$

(4) $\varepsilon|1^*(01^+)^*(01|0|1)$

(5) $(0|1)^*10110(0|1)^*$

(6) $(0|1)^*10110(0|1)^*$

(7) $(0|1)^91(0|1)^*$

(8) $0(0|1)^*1$

(9) $(0|1)^*1(0|1)^*1(0|1)^*$

(10) \emptyset

(11) ε

25. 解： $letter(letter|digit)^*(\varepsilon|(-|.)(letter|digit)^+)$

其中： $letter$ 为字母字符集， $digit$ 为数字字符集

28. (1) $(+|-)((0|1|2|3|4)^+|(\varepsilon|. (0|1|2|3|4)^+))$

(2) $(a|b)^+(-|.)^+(c|d|e|f)^*c|(a|b)^*b(-|.)^+$

(3) $(0|1|2)^*(3|4)(0|1|2|3|4)^+$

31. 画出有穷状态自动机的状态转换图, 分别识别下列文法生成的语言。

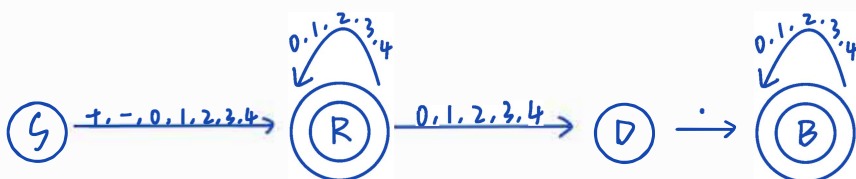
(1) G_1 :

$S \rightarrow +R \mid -R \mid 0R \mid 1R \mid 2R \mid 3R \mid 4R$

$R \rightarrow 0R \mid 1R \mid 2R \mid 3R \mid 4R \mid 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 0D \mid 1D \mid 2D \mid 3D \mid 4D$

$D \rightarrow .B$

$B \rightarrow 0B \mid 1B \mid 2B \mid 3B \mid 4B \mid 0 \mid 1 \mid 2 \mid 3 \mid 4$



(2) G_2 :

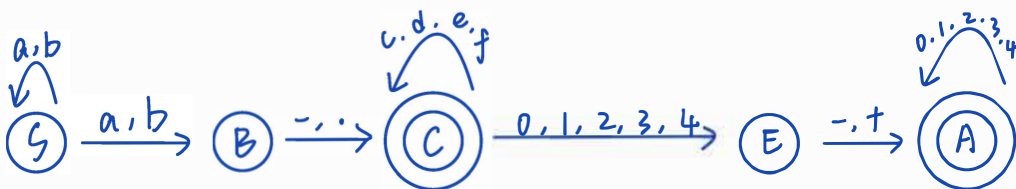
$S \rightarrow .aS \mid aB \mid bB \mid bS$

$B \rightarrow .C \mid .C$

$C \rightarrow .cC \mid dC \mid eC \mid fC \mid c \mid 0E \mid 1E \mid 2E \mid 3E \mid 4E$

$E \rightarrow .A \mid +A$

$A \rightarrow 0A \mid 1A \mid 2A \mid 3A \mid 4A \mid 0 \mid 1 \mid 2 \mid 3 \mid 4$



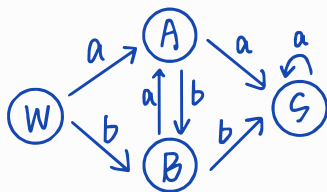
(3) G_3 :

$S \rightarrow .Sa \mid Aa \mid Bb$

$A \rightarrow Ba \mid a$

$B \rightarrow Ab \mid b$

设 $G_3 = G(W)$



批改: 这里的A与B都是两个圈

(4) G_4 :

$S \rightarrow .aA \mid a$

$A \rightarrow aA \mid bB \mid a$

$B \rightarrow bB \mid b$

