

一. 填空:

1. 编译程序的工作过程一般划分为 6 个阶段：

- 词法分析
- 语法分析
- 语义分析
- 中间代码生成
- 代码优化
- 目标代码生成

其中，词法分析器用于识别 **单词**。

2. **符号表管理** 和 **出错处理** 是编译程序各阶段都涉及到的工作。

3. 程序的语义错误可分为 **静态语义错误** 和 **动态语义错误**。

4. 推导的过程可以用一棵树来表示，被称为 **分析树**（或称语法树或称推导树）。

5. Lex 和 Yacc 是用于生成 **词法分析器** 和 **语法分析器** 的工具。

6. 语义规则的两种表示方式是 **语法制导定义** 和 **翻译方案**。

7. 语法的概念：由语言基本符号组成程序中各个语法成分（包括程序）的一组规则叫做**语法**。

8. 词法分析器分析的单词通常可以分为：**关键词、标识符、运算符、常数和界符** 几种。

9. 在编译器设计中，在生成源代码之前，通常在内部采用一种不依赖目标机的结构的代码标识原代码，这种代码被称为 **中间代码**。

10. 运行编译程序的计算机为 **宿主机**，运行编译程序所产生的目标代码的计算机被称为 **目标机**。

11. 一般高级语言的翻译程序有 **编译程序** 和 **解释程序** 两种。

12. 字母表 Σ ，用 Σ^* 表示 Σ 上所有有穷长的串的集合， Σ^* 称为 Σ 的闭包， Σ^+ 称为 Σ 的正闭包，

Σ^* 与 Σ^+ 的区别是 Σ^* 集合中包含 ϵ ，而 Σ^+ 中不含有 ϵ 。

13. 写出描述含 010、0|1 串的正规式： $(0|1)^*010(0|1)^*$

14. 令 $\Sigma = \{a,b\}$ ，则 Σ 上所有以 b 为首的字符串构成的正规集的正规式为： $b(a|b)^*$

15. $\Sigma = \{a,b\}$ 上的正规式 ab 的正规集是 $\{ab\}$ 。

16. 设 $\Sigma = \{a,b\}$ ， Σ 上的正规式 $(a|b)(a|b)$ 相应的正规集是 $\{aa,ab,ba,bb\}$ 。

17. 不确定有穷自动机中的有穷是指 **状态的数量** 是有限的。

18. 识别 3 型文法描述语言的自动机是 **有穷自动机**。

19. 有穷自动机接受的语言是 **正规语言**。

20. NFA 的映像 f 是从“状态 X”映射到“状态 Y”， f 为 **多** 值函数。

21. 在自上而下的语法分析中，需要消除左递归以 **避免产生死循环**，需要提取左因子以 **避免虚假匹配**。

22. 在自上而下的语法分析中，应该先消除文法的 **直接左递归**，再消除文法的 **间接左递归**。

23. 规范规约是最 **左** 规约。

24. LR 分析是按规范句型的 **句柄**（或最左素短语）为可规约串。

25. 为了将非 LL(1) 变换为与之等价的 LL(1) 文法，通常采用 **消除左递归** 和 **提取左公共因子** 对文法进行等价变换。
26. 文法产生二义性的原因是缺少对文法符号 **优先级** 和 **结合性** 的规定。
27. 当文法的 **相同左部非终结符所形成产生式** 的 SELECT 集两两不相交时，该文法是 LL(1) 文法，对应的句子的分析不含有 **回溯性**。
28. 源程序是用高级语言编写的，目标程序是 **机器语言程序或汇编程序**，则其翻译程序称为 **编译程序**。
29. 若源程序是用高级语言编写的，目标程序是 **机器语言程序或汇编程序**，则其翻译程序称为 **编译程序**。
30. 编译方式与解释方式的根本区别在于 **是否生成目标代码**。
31. **翻译程序** 是这样一种程序，它能够将 **用甲语言书写的程序** 转换成与其等价的 **用乙语言书写的程序**。
32. 如果编译程序生成的目标程序是机器代码程序，则源程序的执行分为两大阶段：**编译阶段** 和 **运行阶段**。如果编译程序生成的目标程序是汇编语言程序，则源程序的执行分为三个阶段：**编译阶段，汇编阶段 和 运行阶段**。
33. 若一个文法存在某个句子对应两棵不同的语法树，则称这个文法是 **二义的** 或者，若一个文法存在某个句子有两个不同的最左（右）推导，则称这个文法是 **二义的**
34. **自上而下分析法：**
从文法的开始符号出发，反复使用文法的产生式，寻找与 **输入符号串** 匹配的 **推导**，或者说，为输入串寻找一个最左推导。
35. **自下而上分析法：**
从输入符号串开始，逐步进行 **归约**，直至归约到文法的 **开始符号**。
36. **自上而下方法** 是从文法符号开始，将它做为语法树的根，向下逐步建立语法树，使语法树的结果正好是 **输入符号串**。
37. **自下而上方法** 则是从 **输入符号串** 开始，以它做为语法树的结果，自底向上的构造 **语法树**。
38. 句型分析过程中需要解决的有关问题：在自上而下的分析方法中需要解决的问题是 **如何选择使用哪个产生式进行推导**。在自下而上的分析方法中需要解决的问题是 **如何识别可归约的串**。
39. 一个上下文无关文法所含四个组成部分是 **一组终结符号、一组非终结符号、一个开始符号、一组产生式**。
40. 产生式是用于定义 **语法成分** 的一种书写规则。
41. 扫描器的任务是从源程序中识别出一个个 **单词符号**。
42. 语法分析最常用的两类方法是 **自上而下** 和 **自下而上** 分析法。
43. **语法分析** 的任务是识别给定的终结符串是否为给定文法的句子。
44. 递归下降法不允许任一非终结符是 **直接左递归的**。
45. 自顶向下的语法分析方法的关键是 **如何选择候选式** 的问题。
46. 递归下降分析法是 **自顶向上** 分析方法。
47. **简单优先方法** 每次归约当前句型的 **句柄**，算符优先方法每次归约当前句型的 **最左素短语**，二者都是不断移进输入符号，直到符号栈顶出现 **可归约串** 的尾，再向前找到 **可归约串** 的头，然后归约。
48. 在 LR(0) 分析法的名称中，L 的含义是 **自左向右** 的扫描输入串，R 的含义是 **最左归约**，0 的含义是 **向貌似句柄的符号串后查看 0 个输入符号**。
49. 在 SLR(1) 分析法的名称中，S 的含义是 **简单的**。

二. 填空题

1、语法分析是基于_____文法进行的，即识别的是该类文法的句子。语法分析的有效工具是_____。

2、扫描器的任务是从_____中识别出一个个_____。

3、语法分析最常用的两类方法是_____和_____分析法。

4、自顶向下的语法分析方法的基本思想是：从文法的_____开始，根据给定的输入串并按照文法的产生式一步一步的向下进行_____，试图推导出文法的_____，使之与给定的输入串_____。

5、对编译程序而言，输入数据是_____，输出结果是_____。

6、自底向上的语法分析方法的基本思想是：从输入串入手，利用文法的产生式一步一步地向上进行_____，力求_____到文法的_____。

答案：

1. 上下文无关 语法树
2. 源程序、 单词符号
3. 自上而下、 自下而上
4. 开始符号 直接推导 句子 匹配
5. 源程序、 目标程序
6. 直接规约、 规约、 开始符号

二. 名词解释：

50. **文法定义：**文法 G 定义为四元组 (V_N, V_T, P, S) 其中

V_N ：非终结符号集；

V_T ：终结符号集；

P：规则的集合；

V_N, V_T 和 P 是 非空有穷集。

S：称作识别符号或开始符号，是一个非终结符，它至少要在一条产生式中作为左部出现。

V_N 和 V_T 不含公共的元素，即 $V_N \cap V_T = \emptyset$

用 V 表示 $V_N \cup V_T$ ，称为文法 G 的字母表或字汇表

P 中的规则：称产生式或生成式，是形如 $\alpha \rightarrow \beta$ 或 $\alpha ::= \beta$ 的 (α, β) 有序对，其中 α 是字母表 V 的正闭包 V^* 中的一个符号串， β 是 V 中的一个符号串。 α 称为产生式的左部， β 称作产生式的右部。

51. **直接推导“ \Rightarrow ”：** $\alpha \rightarrow \beta$ 是文法 G 的产生式，若有 v, w 满足： $v = \gamma \alpha \delta, w = \gamma \beta \delta$,

其中 $\gamma \in V^*, \delta \in V^*$ ，则称 v 直接推导产生 w ，记作 $v \Rightarrow w$ ，也称 w 直接归约到 v

若存在 $v \Rightarrow w_0 \Rightarrow w_1 \Rightarrow \dots \Rightarrow w_n = w$, ($n \geq 1$), 则记为 $v \Rightarrow^+ w$, 称作 v 推导出 w , 或 w 归约到 v 。

若有 $v \Rightarrow^+ w$ 或 $v = w$, 则记为 $v \Rightarrow^* w$, ($n \geq 0$)

52. **句型:** 有文法 G , 若 $S \Rightarrow^* x$, 则称 x 是文法 G 的句型。

53. **句子:** 有文法 G , 若 $S \Rightarrow^* x$, 且 $x \in V_T^*$, 则称 x 是文法 G 的句子。

54. **由文法 G 生成的语言** 记为 $L(G)$, 它是文法 G 的一切句子的集合:

$$L(G) = \{x | S \Rightarrow^* x, \text{ 其中 } S \text{ 为文法的开始符号, 且 } x \in V_T^*\}$$

55. **文法的等价:** 若 $L(G_1) = L(G_2)$, 则称文法 G_1 和 G_2 是等价的。

56. **文法的类型:** 通过对产生式施加不同的限制, Chomsky 将文法分为四种类型:

其中 1 型 (也称上下文有关) 文法, 2 型 (也称上下文无关) 文法, 3 型 (也称正规) 文法。

0型文法: 对任一产生式 $\alpha \rightarrow \beta$, 都有

$$\alpha \in (V_N \cup V_T)^+, \quad \beta \in (V_N \cup V_T)^*$$

1型文法: 对任一产生式 $\alpha \rightarrow \beta$, 都有 $|\beta| \geq |\alpha|$,
仅仅 $S \rightarrow \epsilon$ 除外

2型文法: 对任一产生式 $\alpha \rightarrow \beta$, 都有 $\alpha \in V_N$

3型文法: 任一产生式 $\alpha \rightarrow \beta$ 的形式都为 $A \rightarrow aB$ 或
 $A \rightarrow a$, 其中 $A \in V_N$, $B \in V_N$, $a \in V_T$

3型文法产生的语言是有穷自动机 (FA) 所接受的集合。

定理 设 $G = (V_N, V_T, P, S)$ 是 3 型文法, 则存在一个有穷自动机 $M = (K, \Sigma, f, A, Z)$, 使得 $L(M) = L(G)$

57. **最左推导:** 在推导的任何一步 $\alpha \Rightarrow \beta$, 其中 α 、 β 是句型, 都是对 α 中的最左非终结符进行替换。

58. **最右推导:** 在推导的任何一步 $\alpha \Rightarrow \beta$, 其中 α 、 β 是句型, 都是对 α 中的最右非终结符进行替换。

59.

名称	概念
短语	定义: 若 S 为文法 G 的开始符号, $\alpha\beta\delta$ 是该文法的一个句型, 即 $S \Rightarrow^* \alpha\beta\delta$, 且有 $A \Rightarrow^+ \beta$, 则称 β 是句型 $\alpha\beta\delta$ 相对于非终结符 A 的短语。 语法树: 在语法树中表示所有分支结点对应子树, 短语即子树叶子对应的符号。 注: 子树包括语法树本身, 及句型本身也可以称为短语。
直接短语	定义: 若 $S \Rightarrow^* \alpha\beta\delta$, 且文法中包含产生式 $A \rightarrow \beta$, 则称 β 是句型 $\alpha\beta\delta$ 相对于非终结符 A 的直接短语。 语法树: 在语法树中表示为该短语只有上下相邻父子两代
句柄	“可规约串”, 句柄对应某个产生式的右部, 是某个, 但不是任意一个。作为一种规约对象, 句柄表示最左直接短语 。 语法树: 在语法树上, 则表示为最左边的只包含相邻父子节点的短语 (最左直接短语)
素短语	定义: 是指一个短语至少包含一个终结符, 并且除它自身之外不再包含其他素短语
最左素短语	定义: 最左素短语就是句型最左边的素短语, 是算符优先分析法的规约对象。 语法树: 通过语法树分析时, 要注意先判断是否为素短语, 再找相对最左端的素短语。

60. 自顶向下的语法分析方法的基本思想是：从文法的 **开始符号** 开始，根据给定的输入串并按照文法的产生式一步一步的向下进行**直接推导**，试图推导出文法的 **句子**，使之与给定的**输入串匹配**。
61. 自底向上的语法分析方法的基本思想是：从给定的**终结符串**开始，根据文法的规则一步一步的向上进行**直接归约**，试图归约到文法的 **开始符号**。
62. 什么是 NFA？

不确定的有穷自动机NFA

定义

NFA $M = \{K, \Sigma, f, S, Z\}$, 其中 K 为状态的有穷非空集, Σ 为有穷输入字母表, f 为 $K \times \Sigma^*$ 到 K 的子集 (2^K) 的一种映射, $S \subseteq K$ 是初始状态集, $Z \subseteq K$ 为终止状态集.

63. 什么是 DFA？

确定的有穷自动机DFA

DFA 定义：

一个确定的有穷自动机 (DFA) M 是一个五元组: $M = (K, \Sigma, f, S, Z)$ 其中

1. K 是一个有穷集, 它的每个元素称为一个状态;
2. Σ 是一个有穷字母表, 它的每个元素称为一个输入符号, 所以也称 Σ 为输入符号表;

3. f 是转换函数, 是在 $K \times \Sigma \rightarrow K$ 上的映射, 即, 如 $f(k_i, a) = k_j$, ($k_i \in K, k_j \in K$) 就意味着, 当前状态为 k_i , 输入符为 a 时, 将转换为下一个状态 k_j , 我们把 k_j 称作 k_i 的一个后继状态;
4. $S \in K$ 是唯一的一个初态;
5. $Z \subseteq K$ 是一个终态集, 终态也称可接受状态或结束状态。

64. 什么是编译程序?

编译程序: 如果源语言为高级语言, 目标语言为某台计算机上的汇编语言或机器语言, 则此翻译程序称为编译程序。

65. 什么是文法的二义性?

若对于一个文法的某一句子存在两棵不同的语法树, 则该文法是二义性文法。

66. 一个典型的编译程序通常由哪些部分组成?

一个典型的编译程序通常包含 8 个组成部分, 它们是词法分析程序、语法分析程序、语义分析程序、中间代码生成程序、中间代码优化程序、目标代码生成程序、表格管理程序和错误处理程序。

67. 什么是正规文法?

设 $G = (V_N, V_T, P, S)$, 若 P 中的每一个产生式的形式都是 $A \rightarrow aB$ 或 $A \rightarrow a$, 其中 A 和 B 都是非终结符, a 属于 V_T 的闭包, 则 G 是 3 型文法或正规文法。

- 68.

三. 综合题:

例题 1: 求短语, 直接短语, 句柄

给定句型: $T^*P \uparrow(T^*F)$

给定文法 $G[T]$:

$T \rightarrow T^*F|F$

$F \rightarrow F \uparrow P|P$

$P \rightarrow (T)|i$

解析最右推导步骤为:

$T \Rightarrow T^*F$

$\Rightarrow T^*F \uparrow P$

$\Rightarrow T^*F \uparrow(T)$

$\Rightarrow T^*F \uparrow(T^*F)$

$\Rightarrow T^*P \uparrow(T^*F)$

或 (非最左非最右推导:)

$T \Rightarrow T^*F$

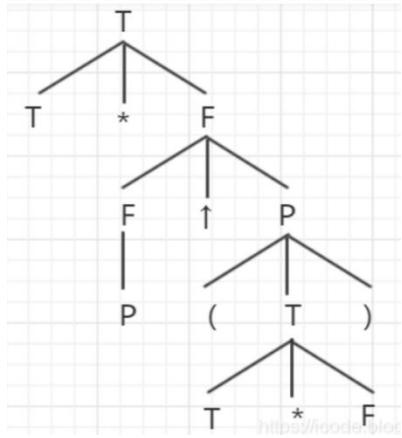
$\Rightarrow T^*F \uparrow P$

$\Rightarrow T^*P \uparrow P$

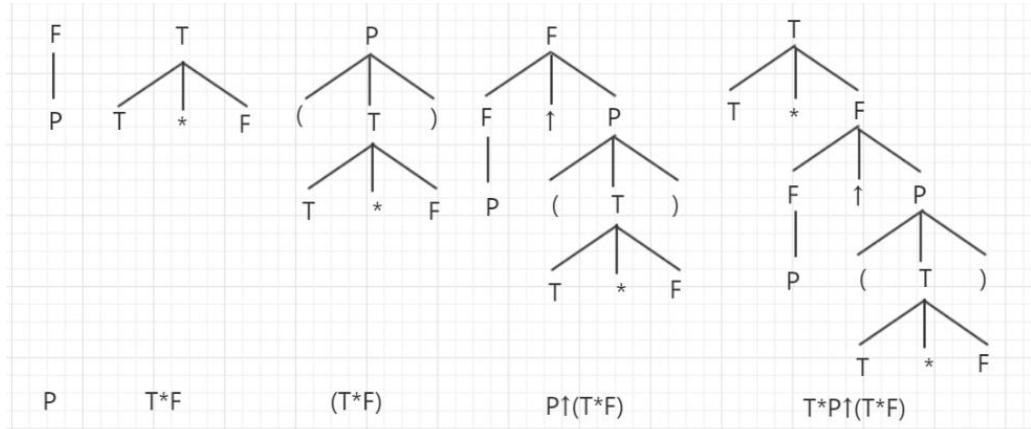
$\Rightarrow T^*P \uparrow(T)$

$\Rightarrow T^*P \uparrow(T^*F)$

画出语法树为:



该语法树的 5 个子树及 5 个短语为：



求直接短语方法：该句型的语法树有两颗直接子树（最左边的两颗子树），由这两颗直接子树的叶子结点组成的符号串（或者说只包含两层的子树叶子结点对应的），就是句型的两个直接短语，直接短语 P 和 T*F。

求句柄：因为 P 相对 T*F，在语法树上的左侧，所以句柄是 P

最终结果：

类型	内容
短语 5 个	P, T*F, (T*F), P↑(T*F), T*P↑(T*F)
直接短语 2 个	P, T*F
句柄	P

例题 2：求素短语，最左素短语

给定句型：FF↑*a*

给定文法 G[T]：

$E \rightarrow ET + | T$

$T \rightarrow TF^* | F$

$F \rightarrow F \uparrow | a$

解析最左推导步骤为：

$E \Rightarrow T$

$T \Rightarrow TF^*$

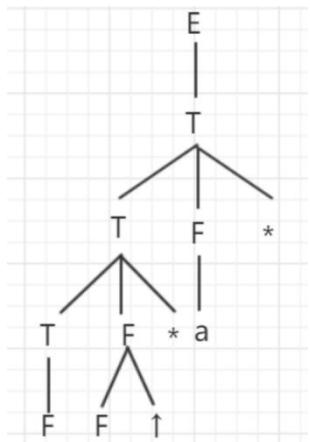
$\Rightarrow TF^*F^*$

$\Rightarrow FF^*F^*$

$\Rightarrow FF\uparrow^*F^*$

$\Rightarrow FF\uparrow^*a^*$

画出语法树：



最终结果：

根据定义可以找出素短语有：

类型	内容
素短语 2 个	$F\uparrow, a$
最左素短语	$F\uparrow$

例题 3. 设文法 $G[S]$ 的产生式集如下：

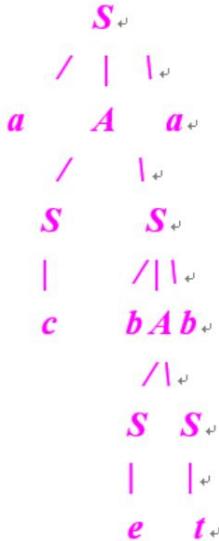
$S \rightarrow aAa | bAb | e | c | t$

$A \rightarrow SS$

(1) 试给出句子 $a c b e t b a$ 的最右推导过程以及语法树；

(2)指出此句型的所有短语、直接短语、句柄和素短语和最左素短语。

解: (1)



$S \Rightarrow aAa \Rightarrow aSSa \Rightarrow aSbAba \Rightarrow aSbSSba \Rightarrow aSbStba \Rightarrow aSbetba \Rightarrow acbetba$

(2)

短语: e 、 c 、 t 、 et、 betb 、 cbetb、 acbetba

直接短语: e、 c 、 t

句柄: c

素短语: c、 e、 t

最左素短语: c

例题 4. 已知文法 $G[S]$ 为:

$$S \rightarrow dAB$$

$$A \rightarrow aA|a$$

$$B \rightarrow Bb|\epsilon$$

① $G[S]$ 产生的语言是什么?

② $G[S]$ 能否改写为等价的正规文法?

解: ① $G[S]$ 产生的语言是 $L(G[S]) = \{da^n b^m \mid n \geq 1, m \geq 0\}$ 。

② $G[S]$ 能改写为等价的正规文法, 其改写后的等价的正规文法 $G[S']$ 为:

$$S' \rightarrow dA$$

$$A \rightarrow aA|aB|a$$

$$B \rightarrow bB|b$$

习题及答案:

1. 已知文法 $G = (\{A, B, S\}, \{a, b, c\}, P, S)$ 其中 P 为:

$$S \rightarrow Ac|aB$$

$$A \rightarrow ab$$

$$B \rightarrow bc$$

写出 $L(G[S])$ 的全部元素。

答案: $L(G[S])=\{abc\}$

2. 已知文法 $G[N]$ 为:

$$N \rightarrow D|ND$$

$$D \rightarrow 0|1|2|3|4|5|6|7|8|9$$

$G[N]$ 的语言是什么?

答案:

$G[N]$ 的语言是 V^+ 。 $V=\{0,1,2,3,4,5,6,7,8,9\}$

$N \Rightarrow ND \Rightarrow NDD \dots \Rightarrow NDDDD \dots D \Rightarrow D \dots \dots D$

或者: 允许 0 开头的非负整数?

3. 写一文法, 使其语言是偶正整数的集合。要求:

- (1) 允许 0 打头;
- (2) 不允许 0 打头。

答案:

(1) 允许 0 开头的偶正整数集合的文法

$$E \rightarrow NT|D$$

$$T \rightarrow NT|D$$

$$N \rightarrow D|1|3|5|7|9$$

$$D \rightarrow 0|2|4|6|8$$

(2) 不允许 0 开头的偶正整数集合的文法

$$E \rightarrow NT|D$$

$$T \rightarrow FT|G$$

$$F \rightarrow N|0$$

$$G \rightarrow D|0$$

4. 给出下述文法所对应的正规式:

$$S \rightarrow 0A|1B$$

$$A \rightarrow 1S|1$$

$$B \rightarrow 0S|0$$

答案:

$$R = (01 \mid 10)(01 \mid 10)^*$$

5. 已知文法:

$$S \rightarrow a|\wedge|(T)$$

$$T \rightarrow T, S|S$$

写出对 $(a,(a,a))$ 和 $((a,a),\wedge,a),a)$ 的最左推导。 (6 分)

答案:

(1) 对 $(a,(a,a))$ 的最左推导为:

$$S \Rightarrow (T) \Rightarrow (T, S) \Rightarrow (S, S) \Rightarrow (a, S) \Rightarrow (a, (T)) \Rightarrow (a, (T, S)) \Rightarrow (a, (S, S)) \Rightarrow (a, (a, S)) \Rightarrow (a, (a, a))$$

(2)

$S \Rightarrow (T)$
 $\Rightarrow (T, S)$
 $\Rightarrow (S, S)$
 $\Rightarrow ((T), S)$
 $\Rightarrow ((T, S), S)$
 $\Rightarrow ((T, S, S), S)$
 $\Rightarrow ((S, S, S), S)$
 $\Rightarrow (((T), S, S), S)$
 $\Rightarrow (((T, S), S, S), S)$
 $\Rightarrow (((S, S), S, S), S)$
 $\Rightarrow (((a, S), S, S), S)$
 $\Rightarrow (((a, a), S, S), S)$
 $\Rightarrow (((a, a), \wedge, S), S)$

6. 已知文法 $G[E]$:

$$E \rightarrow ET + | T$$

$$T \rightarrow TF^* | F$$

$$F \rightarrow F^\wedge | a$$

试证: $FF^{\wedge\wedge*}$ 是文法的句型, 指出该句型的短语、简单短语和句柄.

答案:

该句型对应的语法树如下:

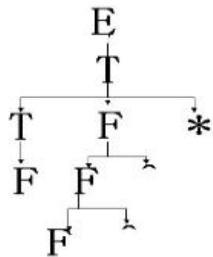
该句型相对于 E 的短语有 $FF^{\wedge\wedge*}$

相对于 T 的短语有 $FF^{\wedge\wedge*}, F$

相对于 F 的短语有 $F^\wedge; F^{\wedge\wedge}$

简单短语有 $F; F^\wedge$

句柄为 F .



7. 已知文法 $G[S]$:

$$S \rightarrow M H | a$$

$$H \rightarrow L S o | \epsilon$$

$$K \rightarrow d M L | \epsilon$$

$$L \rightarrow e H f$$

$$M \rightarrow K l b L M$$

(1) 判断 G 是否为 LL(1) 文法? (2) 如果是 LL(1) 文法, 请构造 LL(1) 预测分析表.

答案:

文法展开为：

- 0) $S \rightarrow M H$
- 1) $S \rightarrow a$
- 2) $H \rightarrow L S o$
- 3) $H \rightarrow \epsilon$
- 4) $K \rightarrow d M L$
- 5) $K \rightarrow \epsilon$
- 6) $L \rightarrow e H f$
- 7) $M \rightarrow K$
- 8) $M \rightarrow b L M$

非终结符	FIRST 集	FOLLOW 集
S	{a,d,b, ϵ ,e}	{#,o}
M	{d, ϵ ,b}	{e,#,o}
H	{ ϵ ,e}	{#,f,o}
L	{e}	{a,d,b,e,o,#}
K	{d, ϵ }	{e,#,o}

对相同左部的产生式可知：

$$\text{SELECT}(S \rightarrow M H) \cap \text{SELECT}(S \rightarrow a) = \{ d, b, e, \#, o \} \cap \{ a \} = \emptyset$$

$$\text{SELECT}(H \rightarrow L S o) \cap \text{SELECT}(H \rightarrow \epsilon) = \{ e \} \cap \{ \#, f, o \} = \emptyset$$

$$\text{SELECT}(K \rightarrow d M L) \cap \text{SELECT}(K \rightarrow \epsilon) = \{ d \} \cap \{ e, \#, o \} = \emptyset$$

$$\text{SELECT}(M \rightarrow K) \cap \text{SELECT}(M \rightarrow b L M) = \{ d, e, \#, o \} \cap \{ b \} = \emptyset$$

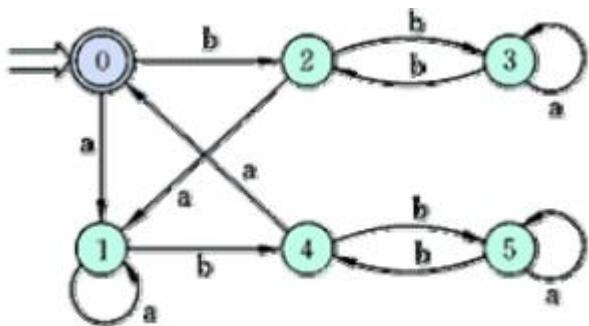
所以文法是 LL(1) 的。

预测分析表：

	a	o	d	e	f	b	#
S	$\rightarrow a$	$\rightarrow MH$	$\rightarrow MH$	$\rightarrow MH$		$\rightarrow MH$	$\rightarrow MH$
M		$\rightarrow K$	$\rightarrow K$	$\rightarrow K$		$\rightarrow bLM$	$\rightarrow K$
H		$\rightarrow \epsilon$		$\rightarrow LSo$	$\rightarrow \epsilon$		$\rightarrow \epsilon$
L				$\rightarrow eHf$			
K		$\rightarrow \epsilon$	$\rightarrow dML$	$\rightarrow \epsilon$			$\rightarrow \epsilon$

由预测分析表中无多重入口也可判定文法是 LL(1) 的。

8. 请将如下 DFA 最小化。



答案:

初始分划得

Π_0 : 终态组{0}，非终态组{1,2,3,4,5}

对非终态组进行审查：

$\{1,2,3,4,5\}a \subset \{0,1,3,5\}$

而{0,1,3,5}既不属于{0}，也不属于{1,2,3,4,5}

$\because \{4\} a \subset \{0\}$ ，所以得到新分划

Π_1 : {0}, {4}, {1,2,3,5}

对{1,2,3,5}进行审查：

$\because \{1,5\} b \subset \{4\}$

$\{2,3\} b \subset \{1,2,3,5\}$ ，故得到新分划

Π_2 : {0}, {4}, {1,5}, {2,3}

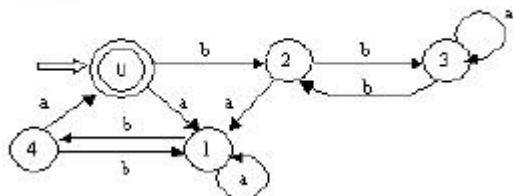
$\{1,5\} a \subset \{1,5\}$

$\{2,3\} a \subset \{1,3\}$ ，故状态2和状态3不等价，得到新分划

Π_3 : {0}, {2}, {3}, {4}, {1,5}

这是最后分划了

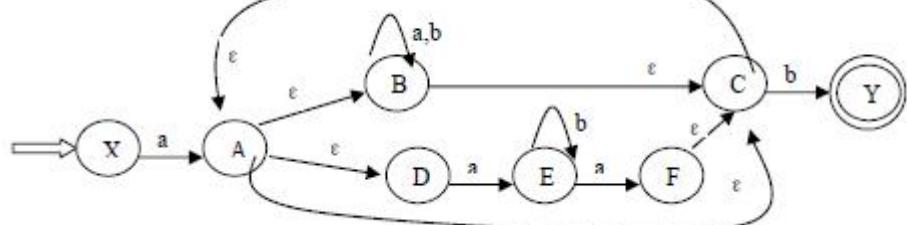
最小 DFA:



9. 构造正规式所对应的 DFA。

$a((alb)^*|ab^*a^*)b$

先构造 NFA:

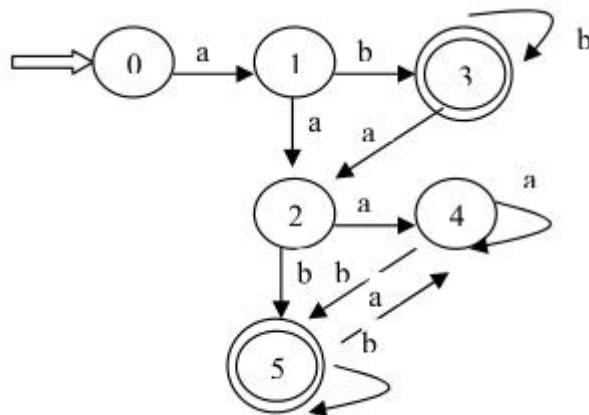


用子集法将 NFA 确定化

	ϵ	a	b
X	X		
$T_0=X$		A	
A	ABCD		
$T_1=ABCD$		BE	BY
BE	ABCDE		
BY	ABCDY		
$T_2=ABCDE$		BEF	BEY
BEF	ABCDEF		
BEY	ABCDEY		
$T_3=ABCDY$		BE	BY
$T_4=ABCDEF$		BEF	BEY
$T_5=ABCDEY$		BEF	BEY

将 T_0 、 T_1 、 T_2 、 T_3 、 T_4 、 T_5 重新命名，分别用 0、1、2、3、4、5 表示。因为 3、5 中含有 Y，所以它们都为终态。

	a	b
0	1	
1	2	3
2	4	5
3	2	3
4	4	5
5	4	5



10. 对于下列文法 $G[E]$:

$$E \rightarrow E+T|T$$

$$T \rightarrow T^*F|F$$

$$F \rightarrow (E)|a$$

(1) 消除文法中的左递归，改写文法形成文法 $G[E']$ 。

- (2) 判断 G 是否为 LL(1) 文法? (3) 如果是 LL(1) 文法, 请构造 LL(1) 预测分析表。
 (4) 进行输入符号串 #a+a# 的递归下降分析, 分析其是否为文法所能识别的符号串。

消除文法中左递归规则

消除直接左递归

形如: $P \rightarrow P \alpha \mid \beta$
 $\alpha \neq \epsilon, \alpha, \beta$ 不以 P 打头

改写为: $P \rightarrow \beta Q$
 $Q \rightarrow \alpha Q \mid \epsilon$
 其中 Q 为新增加的非终结符

答案: (1) 消除左递归:

$$G[E']:$$

(1) $E \rightarrow TE'$	(2) $E' \rightarrow +TE'$	
(3) $E' \rightarrow \epsilon$	(4) $T \rightarrow FT'$	
(5) $T' \rightarrow *FT'$	(6) $T' \rightarrow \epsilon$	
(7) $F \rightarrow (E)$	(8) $F \rightarrow a$	

(2) 判断 G 是否为 LL(1) 文法?

各非终结符的 FIRST 集合如下:

$$\begin{aligned} \text{FIRST}(E) &= \{ (, a \} \\ \text{FIRST}(E') &= \{ +, \epsilon \} \\ \text{FIRST}(T) &= \{ (, a \} \\ \text{FIRST}(T') &= \{ *, \epsilon \} \\ \text{FIRST}(F) &= \{ (, a \} \end{aligned}$$

各非终结符的 FOLLOW 集合为:

$$\begin{aligned} \text{FOLLOW}(E) &= \{ \), \# \} \\ \text{FOLLOW}(E') &= \{ \), \# \} \\ \text{FOLLOW}(T) &= \{ +,), \# \} \\ \text{FOLLOW}(T') &= \{ +,), \# \} \\ \text{FOLLOW}(F) &= \{ *, +,), \# \} \end{aligned}$$

$$\begin{aligned} E' \rightarrow +TE' \mid \epsilon &\quad \text{FIRST}(+TE') = \{ + \} \\ &\quad \text{FOLLOW}(E') = \{ \), \# \} \\ T' \rightarrow *FT' \mid \epsilon &\quad \text{FIRST}(*FT') = \{ * \} \\ &\quad \text{FOLLOW}(T') = \{ +,) \}, \# \} \\ F \rightarrow (E) \mid a &\quad \text{FIRST}((E)) = \{ (\} \\ &\quad \text{FIRST}(a) = \{ a \} \end{aligned}$$

所以 $G[E]$ 是 LL(1) 的

(3) 如果是 LL(1)文法, 请构造 LL (1) 预测分析表。

预测分析表						
	a	+	*	()	#
E	(1)			(1)		
E'		(2)			(3)	(3)
T	(4)			(4)		
T'		(6)	(5)		(6)	(6)
F	(8)			(7)		

(4) 进行输入符号串#a+a#的递归下降分析

分析输入串**#a+a#**的步骤

栈内容	栈顶符号	当前输入	余留串	M[X,b]
1 #E	E	a	+a#	E → TE'
2 #E'T	T	a	+a#	T → FT'
3 #E'T'F	F	a	+a#	F → a
4 #E'T'a	a	a	+a#	
5 # E'T'	T'	+	a#	T' → ε
6 #E'	E'	+	a#	E' → +TE'
7 #E'T+	+	+	a#	
8 # E'T	T	a	#	T → FT'
9 #E'T'F	F	a	#	F → a
10 #E'T'a	a	a	#	
11 #E'T'	T'	#		T' → ε
12 #E'	E'	#		E' → ε
13	#	#		acc 34

11.G[S]为:

S → a A c B e

A→b

A→Ab

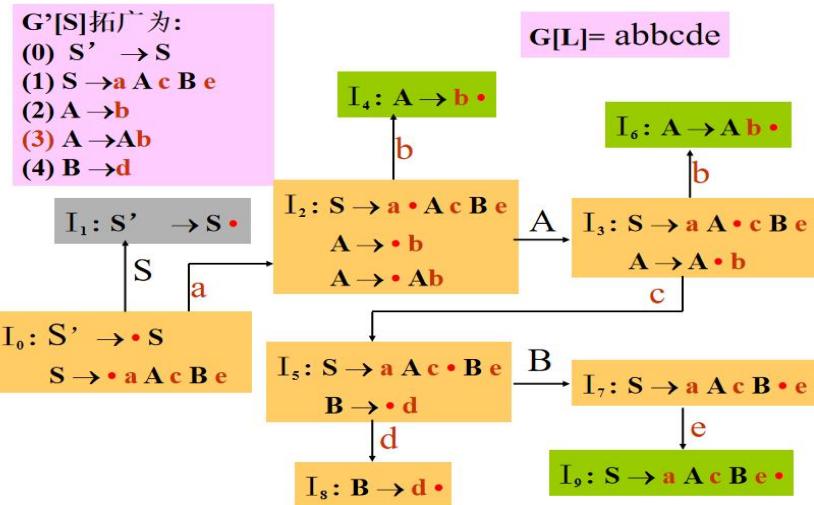
B→d

1)构造识别活前缀的 DFA

2)构造它的 LR(0)分析表。

3)分别给出对输入符号串 abbcde 和 Abbbce 的 LR(0)分析步骤。

答案: 1)构造识别活前缀的 DFA



(2)

例 $G[S]$ 的LR(0)分析表

	ACTION						GOTO		
	a	c	e	b	d	#	S	A	B
0	S_2						1		
1						acc			
2				S_4				3	
3		S_5		S_6					
4	r_2	r_2	r_2	r_2	r_2	r_2			
5					S_8				7
6	r_3	r_3	r_3	r_3	r_3	r_3			
7			S_9						
8	r_4	r_4	r_4	r_4	r_4	r_4			
9	r_1	r_1	r_1	r_1	r_1	r_1			

(3)

对输入串 $abbcde\#$ 的分析过程

Step	states.	Syms.	The rest of input	action goto
1	0	#	abbcde#	s2
2	02	#a	bbcde#	s4
3	024	#ab	bcde#	r2 3
4	023	#aA	bcde#	s6
5	0236	#aAb	cde#	r3 3
6	023	#aA	cde#	s5
7	0235	#aAc	de#	s8
8	02358	#aAcd	e#	r4 7
9	02357	#aAcB	e#	s9
10	023579	#aAcBe	#	r1 1
11	01	#S	#	acc

对输入串abbce#的分析过程

<u>Step</u>	<u>states.</u>	<u>Syms.</u>	<u>The rest of input</u>	<u>action goto</u>
1	0	#	abbce#	s2
2	02	#a	bbce#	s4
3	024	#ab	bce#	r2 3
4	023	#aA	bce#	s6
5	0236	#aAb	ce#	r3 3
6	023	#aA	ce#	s5
7	0235	#aAc	e#	出错

说明abbce#不是 文法 G[S]的句子

12. 文法 G[S]为:

- (1) $S \rightarrow rD$ (2) $D \rightarrow D, i$ (3) $D \rightarrow i$

判断该文法可否用 LR(0)分析法分析, 如果不可以, 判断该文法可否用 SLR (1) 分析法分析, 如果可以用 SLR (1) 分析法分析, 请构造 SLR (1)预测分析表。

答案:

- (1) 构造该文法的拓广文法如下:

$$\begin{array}{ll} (0) S' \rightarrow S & (1) S \rightarrow rD \\ (2) D \rightarrow D, i & (3) D \rightarrow i \end{array}$$

- (2) 构造 LR(0)项目如下:

LR (0) 项目

$$\begin{array}{lll} 1) S' \rightarrow . S & 2) S' \rightarrow S. & 3) S \rightarrow . rD \\ 4) S \rightarrow r. D & 5) S \rightarrow rD. & 6) D \rightarrow . D, i \\ 7) D \rightarrow D. , i & 8) D \rightarrow D, . i & 9) D \rightarrow D, i. \\ 10) D \rightarrow . i & 11) D \rightarrow i. & \end{array}$$

- (3) 构造 LR (0) 项目集规范族:

LR (0) 项目集规范族

$I_0: \quad S' \rightarrow . S$ $\quad \quad \quad S \rightarrow . r D$	$I_3: \quad S \rightarrow r D.$ $\quad \quad \quad D \rightarrow D. , i$
$I_1: \quad S' \rightarrow S.$	$I_4: \quad D \rightarrow i.$
$I_2: \quad S \rightarrow r. D$ $\quad \quad \quad D \rightarrow . D, i$ $\quad \quad \quad D \rightarrow . i$	$I_5: \quad D \rightarrow D, . i$ $I_6: \quad D \rightarrow D, i.$

因为状态集 I_3 中含有移进/归约冲突, 所以该文法不是 LR(0)文法, 不可以采用 LR(0)分析法分析输入符号串。

(4) 因为 I_3 中规约项目 $S \rightarrow rD$. 产生式左部的非终结符 S 的后随符号集合与输入符号集合 (I_3 中的移进项目 $D \rightarrow D, i$ 可以看出输入符号集为 {, }) 相交为空集, 即 $\text{FOLLOW}(S) \cap \{, \} = \emptyset$

所以该文法是 SLR (1) 文法, 可以用 SLR (1) 分析法分析。

构造的 SLR (1) 预测分析表如下:

状态	ACTION			GOTO		
	r	,	i	#	S	D.
0	S ₂					1
1				acc		
2			S ₄			3
3		S ₅		r ₁		
4		r ₃		r ₃		
5			S ₆			
6		r ₂		r ₂		

13. 已知文法 G(S): (1) $S \rightarrow BB$ (2) $B \rightarrow aB$ (3) $B \rightarrow b$

(1) 请构造该文法的 LR(1)项目集规范族及转换函数。

(2) 构造该文法的 LR(1)分析表。

答案:

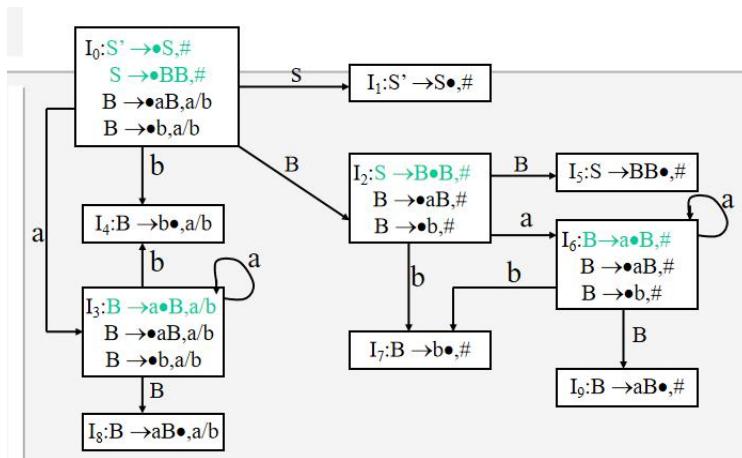
(1) 构造该文法的 LR(1)项目集规范族及转换函数:

构造该文法 G(S) 的拓广文法 G(S') 如下:

(0) $S' \rightarrow S$ (1) $S \rightarrow BB$ (2) $B \rightarrow aB$ (3) $B \rightarrow b$

构造 LR (1) 项目集规范族:

$I_0: S' \rightarrow .S \quad S \rightarrow .BB \quad B \rightarrow .aB \quad B \rightarrow .b$
 $I_1: S' \rightarrow S.$
 $I_2: S \rightarrow B.B \quad B \rightarrow .aB \quad B \rightarrow .b$
 $I_3: B \rightarrow a.B \quad B \rightarrow .aB \quad B \rightarrow .b$
 $I_4: B \rightarrow aB.$
 $I_5: B \rightarrow b.$
 $I_6: S \rightarrow BB.$



(2) 构造该文法的 LR(1)分析表:

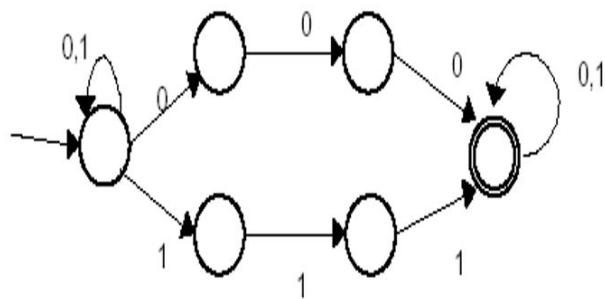
LR(1)分析表

状态	ACTION			GOTO	
	a	b	#	S	B
♦ 0	S3	S4		1	2
♦ 1			acc		
♦ 2	S6	S7			5
♦ 3	S3	S4			8
♦ 4	r3	r3			
♦ 5			r1		
♦ 6	S6	S7			9
♦ 7			r3		
♦ 8	r2	r2			
♦ 9			r2		

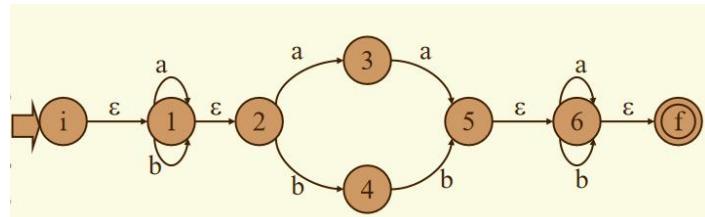
14. 写出正规式所对应的有穷自动机:

$$(0|1)^*(000|111)(0|1)^*$$

答案:



15. NFA 的确定化和最小化, NFA 如下:

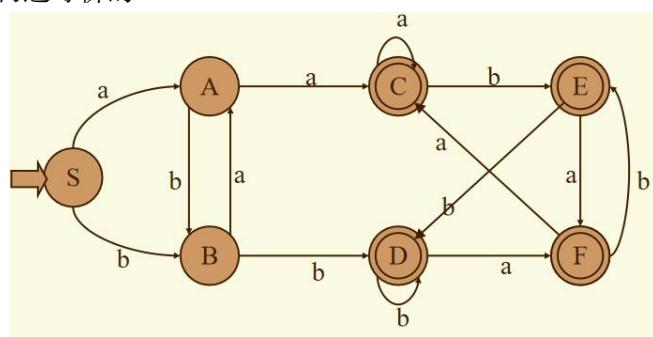


答案: (1) 构造等价的 DFA:

①构造 NFA 的状态子集:

	Ia	Ib
{i,1,2}	S	{1,2,3} A
{1,2,3}	A	{1,2,3,5,6,f} C
{1,2,4}	B	{1,2,3} A
{1,2,3,5,6,f}	C	{1,2,3,5,6,f} C
{1,2,4,5,6,f}	D	{1,2,3,6,f} F
{1,2,4,6,f}	E	{1,2,3,6,f} F
{1,2,3,6,f}	F	{1,2,3,5,6,f} C
		{1,2,4,6,f} E

②构造等价的 DFA:



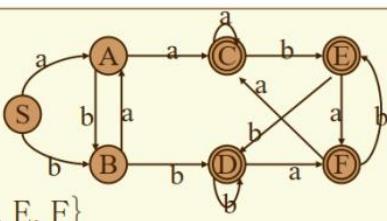
(2) DFA 的最小化

$\Pi_0: \{S, A, B\}$

$\{C, D, E, F\}$

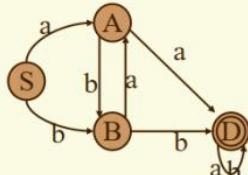
$\Pi_1: \{S, A, B\}$

$\frac{a}{\{A\}} \quad \frac{b}{\{S, B\}}$ $\{C, D, E, F\}$



$\Pi_2:$

$\frac{a}{\{S\}} \frac{b}{\{B\}}$



四. 选择题

注意：请选择题答案填写到以下选择题答题卡中

题号 1 2 3 4 5 6 7

答案 D B D C B A

1. 编译程序绝大多数时间花在 D 上。

- A. 词法分析 B. 语法分析

C. 语义分析 D. 表格管理

2. 汇编程序是将_____B_____翻译成_____；编译程序是将_____翻译成_____。

①高级语言 ②汇编语言 ③机器语言 ④高级语言或汇编语言 ⑤汇编语言或机器语言

- A. ①③①⑤ B. ②③①⑤ C. ④③①③ D. ②③①③

3. 给定文法 $G[A]: A \rightarrow bA|cc$

下面____D______是该文法的句子。

- A.bcbc B.bcbcc
C.bccbcc D.bbbccc

4. 已知语言 $\{a^n b^n c^l \mid n \geq 1, l \geq 0\}$, 则下述文法中, ____C____可以产生该语言。

- A. $S \rightarrow AB, A \rightarrow aAb|ab, B \rightarrow cB|c$ B. $S \rightarrow aAb, A \rightarrow aBb, B \rightarrow cB|c$
C. $S \rightarrow AB, A \rightarrow aAb|ab, B \rightarrow cB|\epsilon$ D. $S \rightarrow aSb|A, A \rightarrow bAc|c$

5. 在通常的语法分析方法中, ____D____特别适用于表达式的分析。

- A. LR(1)分析法 B. LL(1)分析法
C. 递归下降分析法 D. 算符优先分析法

6. 在 DFA 的最小化过程中, 初始应该把集合根据____B____划分子集。

- A. 初态和非初态 B. 终态和非终态
C. 是否含有 ϵ D. 是否初态

7. 若文法 G 定义的语言是无限集, 则文法必然是____A____。

- A. 递归的 B. 二义性的
C. 前后文无关的 D. 无二义性的

8. 一个句型中最左的 (D) 称为该句型的句柄。

- A、短语
B、非终结符号
C、终结符号
D、直接短语

9. 最左简单子树的末端结点构成的符号串称为 (B)

- A、简单短语
B、句柄
C、最左素短语
D、素短语

10. 设 G 是一个给定的文法, S 是文法的开始符号, 如果 $S \Rightarrow^* x$ (其中 $x \in V^*$) , 则称 x 是文法 G 的一个 (D) 。

- A、产生式
B、单词
C、候选式

D、句型

11. 若一个文法是递归的，则它产生的句子个数是 (B)
- A、有限个
 - B、无穷个
 - C、可能有限个
 - D、以上均不对
12. 乔姆斯基 (Chomsky) 把文法分为四种类型，即 0 型、1 型、2 型、3 型。其中 2 型文法是 (B)。
- A、正则文法
 - B、上下文无关文法
 - C、上下文有关文法
 - D、短语文法
13. 文法 $G[E]: E \rightarrow T \mid E + T, T \rightarrow F \mid T * F, F \rightarrow a \mid (E)$ 该文法句型 $E + F * (E + T)$ 的简单短语是下列符号串中的。① $(E + T)$ ② $E + T$ ③ F ④ $F * (E + T)$ 可选项有 (C)
- A、②和③
 - B、③
 - C、③和④
 - D、①和③
14. 若 a 为终结符，则 $A \rightarrow \alpha \cdot a \beta$ 为 (C) 项目。
- A、待约
 - B、接受
 - C、移进
 - D、归约
15. 下面哪种不是自底向上的语法分析文法？(C)
- A、LR (1)
 - B、SLR (1)
 - C、LL (K)
 - D、算符优先法
16. 一个上下文无关文法消除了左递归，提取了左公共因子后是满足 LL (1) 文法的 (C)。
- A、无关系
 - B、充分必要条件
 - C、必要条件
 - D、充分条件
17. 一个上下文无关文法 G 包括四个组成部分，它们是：一组非终结符号，一组终结符号，一个开始符号，以及一组 (B)。

- A、单词
 - B、产生式
 - C、句型
 - D、句子
18. 下面哪个不是单词的描述工具? (D)
- A、正规式
 - B、正规文法
 - C、有穷自动机
 - D、下推自动机
19. 正规式 M1 和 M2 等价是指 (D)。
- A、M1 和 M2 的有向弧条数相等
 - B、M1 和 M2 的状态数相等
 - C、M1 和 M2 状态数和有向弧条数相等
 - D、M1 和 M2 所识别的语言集相等
20. 编译程序中语法分析器接收以 (C) 为单位的输入。
- A、句子
 - B、表达式
 - C、单词
 - D、产生式
21. 一个句型中的可归前缀为 (C)
- A、短语
 - B、句柄
 - C、规范前缀，且句柄位于该规范前缀的后端
 - D、简单短语
22. 字母表 $\{a,b\}$ 上以 aa 打头任何符号串的集合，可用正规式表示为 (C)
- A、 $aa(a^*|b^*)$
 - B、 $(a|b)^*aa$
 - C、 $aa(a|b)^*$
 - D、以上都不对
23. 下面不是翻译程序的是 (B)
- A、汇编程序
 - B、源程序
 - C、编译程序
 - D、解释程序

24. LR (k) 方法是 (B)
- A、从左到右分析，每次向前预测 k 步的一种编译方法。
 - B、从左到右分析，是否归约句柄要向后看 k 个输入符号的一种编译方法。
 - C、从左到右分析，共经过 k 步的一种编译方法。
 - D、从左到右分析，每次走 k 步的一种编译方法。
25. 算符优先分析是一种自底向上的分析方法，它是以什么作为每一步归约的对象？ (C)
- A、句柄
 - B、最右直接短语
 - C、最左素短语
 - D、最右素短语
26. 语法分析器则可以发现源程序中 (D) 。
- A、语法和语义错误
 - B、错误并校正
 - C、语义错误
 - D、语法错误
27. 文法 G 所描述的语言是 (B) 的集合。
- A、文法 G 的字母表 V 的闭包 V^* 中的所有符号串
 - B、由文法的开始符号推出的所有终结字符串
 - C、由文法的开始符号推出的所有符号串
 - D、文法 G 的字母表 V 中所有符号组成的符号串
28. 在通常的语法分析方法中， (C) 特别适用于表达式的分析.
- A、简单优先法
 - B、LR 分析法
 - C、算符优先法
 - D、LL (1) 分析法
29. 高级语言编译程序常用的语法分析方法中，递归分析法属于哪种分析方法？ (D)
- A、自右至左
 - B、自左至右
 - C、自底向上
 - D、自顶向下
30. 文法 G 产生的 (D) 的全体叫语言。
- A、句型
 - B、终结符集
 - C、非终结符集

D、句子

31. 若文法 G 定义的语言是无限的，则文法必然是 (A)
A、递归的
B、上下文有关的
C、二义的
D、无二义的
32. 编译程序中词法分析器所完成的任务是从源程序识别出一个一个具有独立意义的 (D)
A、表达式
B、语句
C、过程
D、单词符号
33. 一个确定有穷自动机是一个 (A)
A、五元组 (K, Σ, f, S, Z)
B、四元组 $(V N, V T, P, S)$
C、四元组 (K, Σ, f, S)
D、三元组 $(V N, V T, P)$
34. 文法 $G[N] = (\{b\}, \{N, B\}, N, \{N \rightarrow b \mid bB, B \rightarrow bN\})$ ，该文法所描述的语言是 (C)
A、 $L(G[N]) = \{b^i \mid i \geq 0\}$
B、 $L(G[N]) = \{b^{2i} \mid i \geq 0\}$
C、 $L(G[N]) = \{b^{2i+1} \mid i \geq 0\}$
D、 $L(G[N]) = \{b^{2i+1} \mid i \geq 1\}$
35. 若项目集 I_k 含有 $A \rightarrow \alpha \cdot$ ，则在状态 k 时，仅当面临的输入符号 $a \in FOLLOW(A)$ 时，才采取“ $A \rightarrow \alpha \cdot$ ”规约动作的一定是 (D)。
A、LALR 文法
B、LR(0)文法
C、LR(1)文法
D、SLR(1)文法
36. 在语法分析处理中，FIRST 集合、FOLLOW 集合、SELECT 集合均是 (B)。
A、非终结符集
B、终结符集
C、字母表
D、状态集
37. (B) 和代码优化部分不是每个编译程序都必需的。
A、语法分析

- B、中间代码生成
C、词法分析
D、目标代码生成
38. 文法 $G: S \rightarrow xSx|y$ 所识别的语言是 (C)。
- A、 xyx
B、 $(xyx)^*$
C、 $x^n y x^n (n \geq 0)$
D、 $x^* y x^*$
39. 如果文法 G 是无二义的，则它的任何句子 α (A)。
- A、最左推导和最右推导对应的语法树必定相同
B、最左推导和最右推导对应的语法树可能不同
C、最左推导和最右推导必定相同
D、可能存在两个不同的最左推导，但它们对应的语法树相同
40. 给定文法 $G: A \rightarrow bA|cc$ ，在下面的符号串中，为该文法句子的是(D)。
- A、 $bcbc$
B、 $bcbcc$
C、 $bccbcc$
D、 $bbbcc$
41. 设 $\Sigma = \{x, y\}$ ，则对其上的正规式 ϵ 所对应的正规集是(C)。
- A、 ϵ
B、 $\{\epsilon, x\}$
C、 $\{\epsilon\}$
D、 $\{\epsilon, x, y\}$
44. 一个 LR 分析器实质上是一个带有先进后出存储栈的(A)。
- A、DFA
B、NFA
C、ACTION 表
D、LR 分析表
45. 自下而上语法分析的主要分析动作是(C)。
- A、移进
B、推导
C、归约
D、匹配
46. 规范归约称为(B)。

- A、最右归约
 - B、最左归约
 - C、最左推导的逆过程
 - D、直接归约
47. 中间代码生成时所依据的是(D)。
- A. 语法规则 B. 词法规则 C. 语义规则 D. 等价变换规则
48. 动态存储分配可采用的分配方案有(C)。
- A、对式存储分配
 - B、线性存储分配
 - C、栈式存储分配
 - D、链式存储分配
49. 下列(C)不属于局部优化。
- A、合并已知量
 - B、利用公共子表达式
 - C、强度削弱
 - D、删除无用赋值
50. 给定文法 $G: V \rightarrow aaV|bc$, 在下面符号串中, 为该文法句子的是(B)。
- A、aa
 - B、aabbc
 - C、abc
 - D、以上都不对
51. 设定义在字母表{a,b,c,x,y,z}上的正规式 $r=(a|b|c)(x|y|z)$, 则 $L(r)$ 中的元素有 (A) 个。
- A、9
 - B、6
 - C、18
 - D、27
52. 规范归约的关键问题是寻找(A)。
- A、句柄
 - B、短语
 - C、最左素短语
 - D、直接短语
53. LR 分析法属于(D)。
- A、自顶向下分析法
 - B、LALR 分析法

C、SLR 分析法

D、自底向上分析法

54. 算符优先分析法实际上是一种 (C)。

A、规范规约

B、规范推导

C、自下而上规约

D、自上而下推导

55. 巴科斯-诺尔范式 (即产生式) 是一种广泛采用的 (C) 的工具。

A、描述规则

B、描述语言

C、描述文法

D、描述句子

56. 一个句型中的最左 (B) 称为该句型的句柄。

A、短语

B、直接短语

C、素短语

D、终结符号

57. 在规范规约中，任何可规约串的出现都在 (C)。

A、栈的内部

B、剩余的输入串中

C、栈顶

D、在栈顶和输入串中