# 清华大学 2024—2025 学年秋季学期期末试题

# 形式语言与自动机

2025.1

**一、(16 分)判别下列各命题的真假性,回答 true 或者 false**

1. 不使用星号运算符 $*$ 的正则表达式能表达的语言都是有限的正则语言.

   ________

2. 若 $L$ 是正则语言,$F$ 是有限的,那么 $L \cup F$ 也是正则语言.

   ________

3. 设问题 $A$ 和 $B$ 都是 NP 问题,且 $A$ 可多项式时间归约到 $B$. 若 $B$ 是 NP—完全问题,则 $A$ 也是 NP—完全问题.

   ________

4. 任何一个多带图灵机都可以被某个单带图灵机模拟.

   ________

5. 两个上下文无关文法的等价性是可判定的.

   ________

6. 上下文无关文法的二义性是可判定的.

   ________

7. 若 $L$ 是正则语言,那么对于它的某个子语言 $L' \subset L$,$L'$ 也一定是正则语言.

   ________

8. 如果某个上下文无关文法对一个字符串有 2 种不同的最左推导,那么该文法是二义的.

   ________

**二、(12 分)单项选择题**

1. $\{0^m 1 0^n \mid m \geqslant n \geqslant 1\}$
2. $\{ww^R \mid w \in \{0,1\}^+, w^R$ 为 $w$ 的反向$\}$
3. $\{w \mid w$ 是图灵机 $M$ 的编码,且 $w \in L(M)\}$

   供以上 1～3 题选择的答案 A～I:

   A. 是某个有限自动机的语言,也是某个空栈接受方式的 DPDA 的语言
   B. 是某个有限自动机的语言,但不是任何空栈接受方式的 DPDA 的语言
   C. 既是某个终态接受方式的 DPDA 的语言,又是某个空栈接受方式的 DPDA 的语言,但不是任何有限自动机的语言
   D. 是某个终态接受方式的 DPDA 的语言,但不是任何空栈接受方式的 DPDA 的语言,也不是任何有限自动机的语言
   E. 是某个无二义上下文无关文法的语言,但不是任何 DPDA 的语言
   F. 是某个 PDA 的语言,但不是任何无二义上下文无关文法的语言
   G. 是递归语言,但不是任何 PDA 的语言
   H. 是递归可枚举语言,但不是递归语言
   I. 不是递归可枚举语言

4. 下列各种自动机中,与基本图灵机的表达能力不等价的是

   A. 具有一个计数器的计数器机　　B. 具有两个栈的确定性下推自动机
   C. 具有一条无穷带的非确定性图灵机　　D. 具有两条半无穷带的图灵机

5. 下列各种语言中，不能够由一个空栈接受的 PDA 所表达的是

A. 任意两个上下文无关语言的交　　B. 任意两个上下文无关语言的并

C. 任意一个上下文无关语言的反向　　D. 任意一个上下文无关语言的闭包

6. 下列问题中不可判定的是

A. 给定一个图灵机 $M$，是否存在一个长度不超过 2025 的串 $w$，满足 $M$ 接受 $w$

B. 给定一个图灵机 $M$，是否存在一个串 $w$，使 $M$ 在 2025 步以内接受 $w$

C. 给定两个正规表达式，它们是否描述同一语言

D. 给定一个布尔表达式，它是否为可满足的

**三、(28 分)简答题**

1. (6 分)设 CFG $G=(\{S,A,B,C,D\},\{a,b,c\},P,S)$，其中 $P$ 由下列产生式构成：

$$
\begin{aligned}
&S \to ABb\\
&A \to DBa \mid B\\
&B \to Bb \mid \varepsilon\\
&C \to ABc \mid \varepsilon\\
&D \to c \mid \varepsilon
\end{aligned}
$$

(1)消去 $P$ 中的 ε－产生式得到产生式集合 $P_1$，构成 CFG $G_1$，使得 $L(G_1)=L(G)-\{\varepsilon\}$. 给出 $P_1=$？

(2)消去 $P_1$ 中的 Unit 产生式得到产生式集合 $P_2$，构成 CFG $G_2$，使得 $L(G_2)=L(G_1)$. 给出 $P_2=$？

(3)消去 $P_2$ 中的无用符号得到产生式集合 $P_3$，构成 CFG $G_3$，使得 $L(G_3)=L(G_2)$. 给出 $P_3=$？

2. (4 分)文法 $G$($S$ 为开始符号)的产生式集合为：

$$
\begin{aligned}
&S \to AB\\
&A \to AB \mid AC \mid a\\
&B \to SS \mid SA \mid b\\
&C \to BC \mid BS \mid c
\end{aligned}
$$

| | | | |
|---|---|---|---|
| $X_{14}$ | | | |
| $X_{13}$ | $X_{24}$ | | |
| $X_{12}$ | $X_{23}$ | $X_{34}$ | |
| $X_{11}$ | $X_{22}$ | $X_{33}$ | $X_{14}$ |
| $a$ | $b$ | $c$ | $b$ |

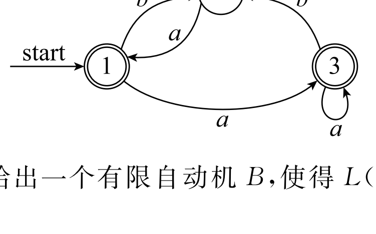
上图右侧表示对于文法 $G$ 和字符串 $abcb$ 应用 CYK 算法时所构造的表.

(1)分别计算图中所有 $X_{ij}$ $(1\leqslant i,j\leqslant 4)$；

(2)是否有 $abcb\in L(G)$？

3. (6 分)下图表示一个有限自动机 $A$：

(注意：本题中多处出现有限自动机的描述，可以是也可以不是 DFA)

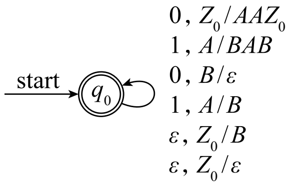



(1)试采用课程中的方法，给出一个有限自动机 $B$，使得 $L(B)=L(A)^R$；($L(A)^R$ 为 $L(A)$ 的反向)

(2)试采用课程中的方法，给出一个有限自动机 $C$，使得 $L(C)=\{a,b\}^*-L(A)$；

(3)设映射 $h:\{0,1\}\to\{a,b\}^*$ 定义为 $h(0)=aa$，$h(1)=ab$，试构造一个有限自动机 $D$，使得 $L(D)=h^{-1}(L(A))$.

4. (3 分)下图刻画了 PDA $P=(\{q_0\},\{0,1\},\{Z_0,X,A,B\},\delta,q_0,Z_0)$ 的转移规则，试严格利用课程中介绍的从空栈接受的 PDA 到 CFG 的转换算法，定义一个与该 PDA 等价的 CFG，开始符号设为 $S$.




5. (4 分)一台全路径 NFA(all－NFA)$M$ 是一个 5 元组$(Q,\Sigma,\delta,q_0,F)$. 如果 $M$ 对 $x\in\Sigma^*$ 的每一个可能的计算都结束在 $F$ 中的状态，即 $\hat{\delta}(q_0,x)\subseteq F$，则 $M$ 接受 $x$.（相反地，普通 NFA 只需有一个计算结束在接受状态，就接受这个字符串.）实际上，全路径 NFA 识别正则语言类. 请给出一个将全路径 NFA 转化为 DFA 的算法，无需证明算法的正确性.

6.(5 分)对于语言：

$$L=\{a^icb^j \mid j\leqslant i^2\}$$

可以利用 Pumping 引理证明 $L$ 不是上下文无关语言，以下是一个证明概要：

用反证法假设 $L$ 是上下文无关语言，设 $n\geqslant 1$ 为上下文无关语言 Pumping 引理给出的整数.

取 $z=$ ___①___，则 $z\in L$.

对任意满足条件 $z=uvwxy\wedge vx\neq\varepsilon\wedge|vwx|\leqslant n$ 的 $u,v,w,x,y$，

如果 $c$ 在 $vx$ 中，则 $k>1$ 时 $uv^kwx^ky\notin L$；在 $c$ 不在 $vx$ 中的情况下，

若___②___，取 $k=$ ___③___，则 $uv^kwx^ky\notin L$；

若___④___，取 $k=$ ___⑤___，则 $uv^kwx^ky\notin L$.

试在其中①、②、③、④和⑤处填写适当的内容.（若需更多分支，可自行添加）

**四、(25 分)设计题(必要时解释设计思路)**

1. (5 分)试构造接受下列语言的 DFA，要求用状态转移图的方式给出答案，且总状态数不超过 6 个：

$$L=\{w\in\{0,1\}^* \mid w \text{ 不包含子串 010 且不包含子串 101}\}.$$

2. (5 分)试构造接受下列语言的一个正规表达式：

$$L=\{w\in\{0,1\}^* \mid w \text{ 以 1 开头和结尾，且 } w \text{ 中 1 的个数为偶数}\}.$$

3. (5 分)试构造接受下列语言的一个上下文无关文法：

$$L=\{w\in\{0,1\}^* \mid |w| \text{ 为 3 的倍数，且 } w \text{ 的前三分之一必须包含 1}\}.$$

4. (5 分)试构造以终态接受方式接受下列语言的一个 DPDA，用状态转移图描述你的设计(必要时给出设计思路)：

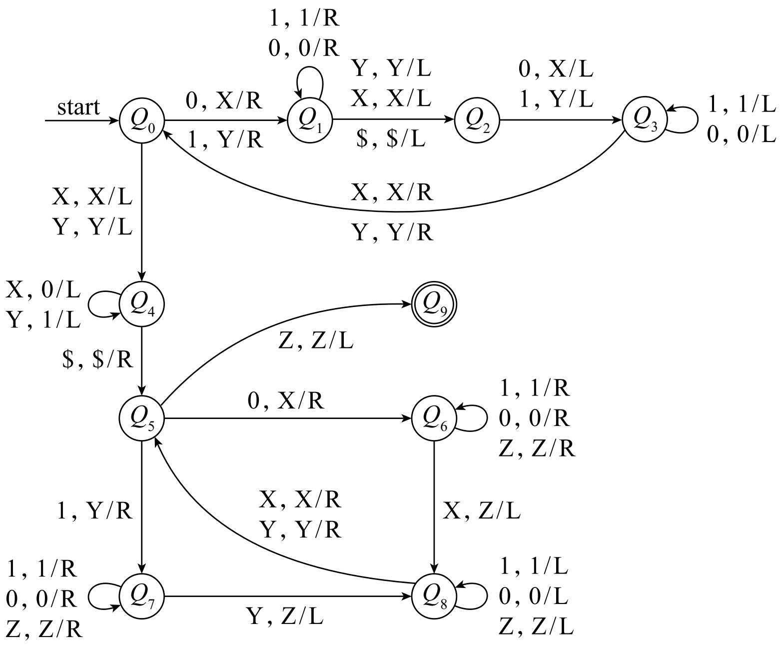
$$L=\{w\in\{a,b\}^* \mid w \text{ 中串 } aa \text{ 的出现次数不少于串 } aba \text{ 的出现次数}\}.$$

(例如串 $abab aaab$ 中 $aa$ 出现了 2 次，$aba$ 也出现了 2 次，故 $ababaaab\in L$)

5. (5 分)以下图片描述了一个单带读写图灵机. 该图灵机的输入为由 0,1 构成的字符串,\$ 表示空白符. 转移规则描述为:读入一个字符,将其替换,然后进行一步移动. 以节点 $Q_0$ 至 $Q_1$ 的一条转移规则"0,X/R"为例,该转移规则表示读入 0 后,将当前字符替换为 X,然后向右移动一步.

(1)当输入是 1010 时,详细写出每一步的转移与对应的纸带上的字符串. 判断该输入是否被接受;

(2)描述该图灵机所接受的语言.




**五、(19 分)证明题**

1. (4 分)已知语言 $L_{01}=\{0^k1^k \mid k\geqslant 0\}$不是正规语言,试利用该结论以及正规语言的封闭运算,证明如下语言 $L$ 不是正规语言:

$$L=\{w \mid w=a^mb^n,\text{其中 } m,n\geqslant 0,\text{且 } 202m\neq 5n\}$$

2. (5 分)对于一个含有 $n$ 个状态的有限自动机,它的转移关系 $\delta$ 可以用一个 $n\times n$ 的矩阵 $G$ 表示出来. $G$ 的第 $u$ 行、第 $v$ 列的元素为从状态 $u$ 转移到状态 $v$ 的字符集合:

$$G_{uv}=\{a\in\Sigma \mid \delta(u,a)=v\}.$$

考虑所有在状态集合 $Q$,字母表 $\Sigma$ 上的这些 $n\times n$ 方阵,我们还可以定义这些矩阵之间的运算:

$$(A+B)_{uv}\overset{\text{def}}{=}A_{uv}\cup B_{uv},\quad (AB)_{uv}\overset{\text{def}}{=}\bigcup_{w\in Q}A_{uw}B_{wv}.$$

注意集合之间的运算 $A_{uw}B_{wv}$ 表示连接(concatenation). 此外,定义单位矩阵 $I$:

$$I_{uv}\overset{\text{def}}{=}\begin{cases}\{\varepsilon\} & \text{if } u=v,\\ \varnothing & \text{otherwise.}\end{cases}$$

矩阵的幂:

$$A^0\overset{\text{def}}{=}I,\quad A^{n+1}\overset{\text{def}}{=}A^nA,\quad (A^*)_{uv}\overset{\text{def}}{=}\bigcup_{n\geqslant 0}(A^n)_{uv}.$$

假设 $s$ 是自动机的初始状态,$F$ 是接受状态的集合,试证明:$L(M)=\bigcup_{t\in F}(A^*)_{st}$.

3. (5 分)在$\{0,1\}^*$中，如果两个串 $x,y$ 长度相同$|x|=|y|$，定义 $x$ 和 $y$ 的 **Hamming 距离** $H(x,y)$为 $x\oplus y$ 中 1 的个数(其中$\oplus$表示按位异或)，也就是使 $x$ 和 $y$ 对应处的字符不同的位置数. 假设 $L$ 是字母表$\Sigma=\{0,1\}$上的正规语言，试证明如下语言 $L_1$ 也是正规语言：

$$L_1=\{x\in\{0,1\}^* \mid \text{存在 } y\in L \text{ 使} |x|=|y| \text{且 } H(x,y)\leqslant 1\}$$

提示：考虑 $L$ 对应的 DFA $A=(Q,\Sigma,\delta,q_0,F)$，基于 $A$ 构造新的有限自动机并证明其恰好识别 $L_1$.

4. (5 分)考虑以下判定问题：给定一个上下文无关语言 $L$ 的语法(以上下文无关文法的形式描述)，问 $L$ 是否生成无限多个字符串. 该问题是否是可判定的？若否，给出证明；若是，给出一个判定算法(无复杂度要求)，并证明算法的正确性.

**附加题(5 分，直接加入总评成绩)**

考虑字母表$\Sigma$上的语言$L$. 设 $x$ 和 $y$ 是两个$\Sigma^*$中的字符串，如果存在字符串 $z$，使得 $xz$ 和 $yz$ 中恰好有一个是 $L$ 的成员，则称 $x$ 和 $y$ 是用$L$ 可区分的；否则，对每一个字符串 $z$，$xz$ 和 $yz$ 要么都是、要么都不是 $L$ 的成员，则称 $x$ 和 $y$ 是用$L$ 不可区分的，记作 $x\equiv_L y$. 易证$\equiv_L$是一个等价关系，因此它可以把$\Sigma^*$划分成一些等价类，定义 $L$ 的**指数**为这些等价类的数量，即$|\Sigma^*/\equiv_L|$.

**Myhill－Nerode 定理**声称一个语言是正则的当且仅当它有有穷的指数. 为了表明这一命题成立，请试证明：

(1)如果 $L$ 被一台有 $k$ 个状态的 DFA 识别，则 $L$ 的指数不超过$k$；

(2)如果 $L$ 的指数是一个自然数$k$，则可以构造一台有 $k$ 个状态的 DFA 识别 $L$.