清华大学本科生考试试题专用纸

考试课程 《形式语言与自动机》 B 卷 2023 年 1 月 4 日

学号：______________ 姓名：______________ 班级：______

（注：解答请写在自己准备的答题纸上。）

一. (16 分) 判别下列各命题的真假性，回答 true 或者 false: (每小题 2 分)

1．存在一个判定两个正规语言之间存在包含关系（即集合包含$\subseteq$）的算法。

____

2．对 CFG $G = (V, T, P, S)$，开始符号 S 到任意 $w \in T^*$ 的最左推导的唯一性是可判定的。

____

4．不存在可接受语言 $\{ ww \mid w \in \{a, b\}^* \}$ 的可停机的确定的图灵机。

____

4．DFA 中，设状态 r 和 s 通过某个输入符号 a 可分别转移到 p 和 q，即 $\delta(r, a) = p$, $\delta(s, a) = q$，则有：若 r 和 s 是不可区别的，则 p 和 q 也是不可区别的。

____

5．若 L 为上下文无关语言，R 为正规语言，则 $L-R$ 也是上下文无关语言；同时，$R-L$ 也是上下文无关语言。

____

6．固有二义的上下文无关语言不是任何 *DPDA* 的语言。

____

7．对角语言 L_d 可以归约到通用语言 L_u。

____

8．任一 *NP*–完全（*NP-complete*）问题均存在可以接受它的非确定图灵机。

____

二. (12 分) 选择填空（单选） (每小题 2 分)

1．$\{ 0^n1^m \mid n \le m \}$

____

2．$\{ 0^n1^i2^k3^m \mid i+k \ge 100，n+m < 100 \}$

____

3．$\{ 0^n1^m \mid m = n^2 \}$

____

4．$\{ 0^n1^n2^m3^m \mid n\ge1，m\ge1 \} \cup \{ 0^n1^m2^m3^n \mid n\ge1，m\ge1 \}$

____

供以上 1~4 题选择的答案：

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 *PDA* 的语言，但不是任何 *DPDA* 的语言。

F. 不是任何 *PDA* 的语言。

5. 下列问题中，_________ 是可判定的。

供选择的答案：

A. *Post* 对应问题 *PCP*

B. 一个图灵机的语言是否为正规语言

C. 一个图灵机的语言是否为上下文无关语言

D. 可满足性问题 *SAT*

6. 下列语言中，_________ 不是递归可枚举语言。

供选择的答案：

A. 语言 L_u（课程中定义的通用语言）

B. 语言 L_H（课程中图灵机停机问题所定义的语言）

C. 语言 $L_e = \{ M \mid M$ 为图灵机，且 $L(M) = \phi \}$

D. 语言 $L_{ne} = \{ M \mid M$ 为图灵机，且 $L(M) \neq \phi \}$

三. (32 分) 简答题:

1. (4 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'，使得 $L(G') = L(G) - \{\varepsilon\}$. 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''，使得$L(G'') = L(G')$. 给出$P_2 =$ ？

（3）消去 P_2 中的无用符号得到产生式集合P_3. 给出$P_3 =$ ？

2. (4 分) 文法 G（S为开始符号）的产生式集合为：

$$S \to SA \mid AB$$
$$A \to BA \mid a \mid b$$
$$B \to AA \mid b$$

X_{13}		
X_{12}	X_{23}	
X_{11}	X_{22}	X_{33}
a	a	b

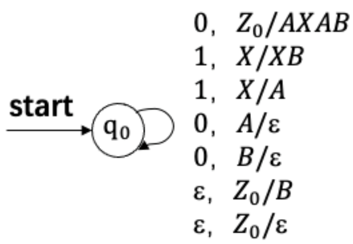
上图表示对于文法 G 和字符串 aab 应用 CYK 算法时所构造的表。

(1) 分别计算图中所有 X_{ij} （$1 \le i, j \le 3$）

(2) 是否有 $aab \in L(G)$?

3. (4 分) 设 A 和 B 是字母表 Σ 上的 DFA，试给出判定 $L(A) \cap L(B)$ 是否为空语言的一个算法。

4. (4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) 对于语言

$$L = \{ a^i b^j c^k \mid i + j = 2k \}$$

可以利用 Pumping 引理证明 L 不是正规语言，以下是一个证明概要：

对于任意的 $n \ge 1$，取 $w =$ __①__ $\in \{a,b,c\}^*$，则有 $w \in L$。
对任意满足条件 $w = xyz \wedge y \neq \varepsilon \wedge |xy| \le n$ 的 x, y, z，取 $k =$ __②__，有 $xy^kz \notin L$。

试在其中 ① 和 ② 处填写适当的内容。如需分类讨论可自行添加填空。

6. (4 分) 定义如下同态 $h: \{0,1,2\} \to \{a,b\}$，$h(0) = aba, h(1) = ab, h(2) = b$.

(1) 试写出正规表达式 R 使 $L(R) = h(L(0 + 12^*))$

(2) 试写出正规表达式 R 使 $L(R) = h^{-1}(L((ab)^*))$

7. (8 分) 对于语言

$$L = \{ a^n b^n c^i \mid i \le n \}$$

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：

对于任意的 $n \geq 1$，取 $z =$ ___①___ $\in \{a,b,c\}^*$，则有 $z \in L$。
对任意满足条件 $z = uvwxy \wedge vx \neq \varepsilon \wedge |vwx| \leq n$ 的 u,v,w,x,y，取 $k =$ ___②___，有 $uv^kwx^ky \notin L$。

试在其中 ① 和 ② 处填写适当的内容。如需分类讨论可自行添加填空。

此外，试给出两个上下文无关语言 L_1, L_2（不须证明）使得 $L_1 \cap L_2 = L$，由此说明交运算 ∩ 对于上下文无关语言不封闭。

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一个有限自动机（*DFA*，*NFA*，$\varepsilon - NFA$均可），要求用**状态转移图**的方式给出答案：

$L = \{ w \mid w \in \{0, 1, 2, 3\}^*$，$w$ 的长度是 3 的倍数，且 w 的后三位包含"22"，w 的头三位包含"23" $\}$

2. (5 分) 给出下列正规语言 L 的一个正规表达式：

$L = \{ w \mid w \in \{a, b\}^*$，且 w 中不包含子串 $abab \}$

3. (5 分) 试给出下列语言的一个上下文无关文法：

$L = \{ a^n b^m \mid 2 \leqslant n+m \leqslant 2+3n \}$

4. (5 分) 试构造接受下列语言的一个 *PDA*（终态接受和空栈接受均可，必要时给出设计思路），要求该 *PDA* 的堆栈符号数不超过 3，且用状态转移图描述你的设计，并简述设计思路：

$L = \{ w \mid w = a^i b^j c^k$，其中 $i=2j$ 或 $j \neq 2k \}$

5. (5 分) 试设计一个可停机图灵机 $M = (Q, \{0,1\}, \{0,1 \ldots, B\}, \delta, q_0, B, \{q_f\})$ 可以将串 $w \in \{0, 1\}^*$ 作为输入，当到达终态 q_f 时，带上的内容为将w循环右移1位的结果。例：如输入串为101010，则到达终态时，带上的内容应为010101。如输入串为10101，则到达终态时，带上的内容应为11010。如输入串长度$n \leq 1$，到达终态时带上内容不变。该图灵机的状态数不超过 7。到达q_f 时，对读写头在何处不作要求。要求该图灵机 M 从开始执行至停机的移动步数与输入串长度 n 成线性关系，即时间复杂度为O(n)。用状态转移图描述你所设计的图灵机，并简述设计思路。

五. (15 分) 证明题：（要求证明过程严谨，步骤明确。）

1. (5 分) 定义 $A/B = \{ w \mid wx \in A$ 对于某个$x \in B \}$。试证明若 A 为正规语言，B 为任意语言，则 A/B 为正规语言。

2. (5 分) 对于课堂上讲授的空栈接受的 *PDA*：$\mathrm{P} = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$，其栈的长度是无限的：即，在 *PDA* 上的转移过程中，我们允许向其栈中压入的元素个数不设上限；也即，*PDA* 上的推导过程中，任一时刻的当前格局（instantaneous descriptions, ID）的三元组 (q, w, γ) 中，当前栈中内容 γ 的长度是无限的。本题考虑一种新的情况：当空栈接受的 *PDA* 的栈的长度最大为N，即，对于这一 *PDA* 的任意可能的 ID = (q, w, γ)，总要求 γ 的长度 $|\gamma| \leqslant N < +\infty$时，该 *PDA* 的性质将如何变化。

具体地，对于任意一个 栈的长度无限的 空栈接受的 *PDA* P，任意 $N \in \mathbb{Z}^+$，将 P 的栈的长度限制为小于等于 N，得到 *PDA* P'_N，命题

$$L(P'_N) \subseteq \{ w \mid w \in L(P), 且 |w| \leq 2N - 1 \}$$

是否总是成立？若是，则给出证明过程；若不是，则说明理由。

提示：默认初始 ID 中 $\gamma=Z_0$，即开始时 *PDA P* 的栈中存在一个栈底元素 Z_0。

3. (5 分) 设 Σ 和 T 为字母表，$h: \Sigma\to T^*$。若 S 为 Σ 上的正规语言，则 S 的反同态 $h^{-1}(S)$ 也是正规语言。

附加题（5分，直接加入总评成绩）

(注意：附加题只有能呈现出核心思路才有可能得到部分分数，建议大家在前面题目已做完且进行充分检查之后，再看是否有时间考虑下列题目之一)

若字符串 y 可由字符串 x 重排序后生成，则我们称 y 为 x 的重排列。例如，011为101的一个重排列。定义 $\text{shuffle}(L)$ 为 L 中的字符串的所有重排列所构成的集合。如 $L=\{0^n1^n \mid n\geq 0\}$，则 $\text{shuffle}(L)$ 为所有 0、1 个数相等的字符串的集合。试证明：若 L 为大小为 2 的字母表 Σ 上的一个正规语言，则 $\text{shuffle}(L)$ 一定是一个上下文无关语言。并举例说明当 $|\Sigma|>2$ 时如上结论不成立。