

## 编译原理作业 (3)

姓名: \_\_\_\_\_ 学号: \_\_\_\_\_

2024 年 03 月 31 日

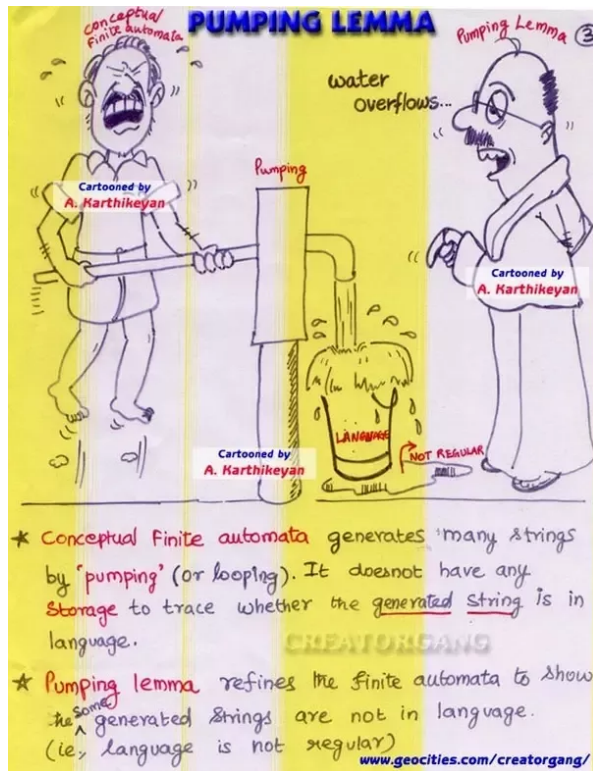
请独立完成作业, 不得抄袭。

若得到他人帮助, 请致谢。

若参考了其它资料, 请给出引用。

鼓励讨论, 但需独立书写解题过程。

允许并鼓励使用 ChatGPT 等工具, 但需明确说明使用方式。



## 1 作业 (必做部分)

### 题目 1

请证明以下上下文无关文法表示语言  $\{x \in \{a, b\}^* \mid x \text{ 中 } a, b \text{ 个数相同}\}$ 。

$$V \rightarrow VV \mid aVb \mid bVa \mid \epsilon$$

解答:

(a) 首先证明该 CFG 可表示的任意字符串  $x$  都满足:  $x$  中  $a, b$  个数相同。

由于任何可用该 CFG 可表示的字符串都对应着至少一颗语法树, 我们可以基于语

法树高度  $h$  进行归纳证明.

奠基:

树高  $h = 1$  时, 能够表达的字符串集合  $S_0$  为  $\{\epsilon\}$ , 满足每个元素的  $a, b$  个数相同.

归纳假设:

树高  $h \leq k - 1$  时, 能够表达的字符串集合  $S_h$ , 满足每个元素的  $a, b$  个数相同.

归纳递推:

树高  $h \leq k$  时, 有三种情况:

1. 匹配  $V \rightarrow V_0 V_1$  规则,  $V_0, V_1$  对应的树高都小于  $k$ , 利用归纳假设可以得到该情况下能够表达的字符串集合  $S_k$  满足条件.
2. 匹配  $V \rightarrow a V_0 b$  规则,  $V_0$  对应的树高小于  $k$ , 利用归纳假设可以得到该情况下能够表达的字符串集合  $S_k$  满足条件.
3. 匹配  $V \rightarrow b V_0 a$  规则,  $V_0$  对应的树高小于  $k$ , 利用归纳假设可以得到该情况下能够表达的字符串集合  $S_k$  满足条件.

综合三种情况, 此时能够表达的字符串集合  $S_k$ , 满足每个元素的  $a, b$  个数相同.

(b) 集合  $\{x \in \{a, b\}^* \mid x \text{ 中 } a, b \text{ 个数相同}\}$  中的元素都可以用该 CFG 表达.

令  $L = \{x \in \{a, b\}^* \mid x \text{ 中 } a, b \text{ 个数相同}\}$ .

对于任意  $s \in L$ , 基于字符串长度  $l = |s|$  进行归纳.

奠基:

字符串长度  $l = 0$  时,  $s$  属于的字符串集合  $S_0$  为  $\{\epsilon\}$ , 满足每个元素都可以用该 CFG 表达.

归纳假设:

字符串长度  $l \leq k - 1$  时,  $s$  属于的字符串集合  $S_l$ , 满足每个元素都可以用该 CFG 表达.

归纳递推:

字符串长度  $\leq k$  时, 有四种情况

1.  $s$  最左端为  $a$ , 最右端为  $b$ , 令  $s = a s_0 b$ , 因为  $|s_0| \leq k - 1$  于是  $s_0$  可以用该 CFG 表达, 进而  $s$  可以用该 CFG 表达
2.  $s$  最左端为  $b$ , 最右端为  $a$ , 与情况 1. 同理
3.  $s$  最左端为  $a$ , 最右端为  $a$ , 令  $s = s_0 s_1$ , 且满足  $s_0$  最左端为  $a$ , 最右端为  $b$ ;  $s_1$  最左端为  $b$ , 最右端为  $a$ . (可以很容易证明必然能够找到满足条件的  $s_0$  和  $s_1$ ) 于是  $s_0$  和  $s_1$  可以用该 CFG 表达, 进而  $s$  可以用该 CFG 表达.
4.  $s$  最左端为  $b$ , 最右端为  $b$ , 与情况 3. 同理

综合四种情况, 此时  $s$  属于的字符串集合  $S_k$ , 满足每个元素都可以用该 CFG 表达.

根据 (a)(b), 可以得到该 CFG 表示语言  $\{x \in \{a, b\}^* \mid x \text{ 中 } a, b \text{ 个数相同}\}$

---

## 题目 2

请证明以下两种 if-else 文法都是无二义性的, 并且实现了“else 与最近的未匹配的 if 匹配”的语义。

(1) 符合 ANTLR 4 中的“最前优先匹配原则”的如下文法:

```
stat : 'if' expr 'then' stat
     | 'if' expr 'then' stat 'else' stat
     | expr
     ;
```

尝试: <https://github.com/courses-at-nju-by-hfwei/2024-compilers-coding/blob/main/src/main/antlr/ifstat/IfStat.g4>

(2) 教材中改写后的文法:

```
stat : matched_stat | open_stat ;

matched_stat : 'if' expr 'then' matched_stat 'else' matched_stat
             | expr
             ;

open_stat: 'if' expr 'then' stat
          | 'if' expr 'then' matched_stat 'else' open_stat
          ;
```

尝试: <https://github.com/courses-at-nju-by-hfwei/2024-compilers-coding/blob/main/src/main/antlr/ifstat/IfStatOpenMatched.g4>

解答:

(1) 无二义性:

antlr4 默认排在前面的语法规则优先级更高, 无二义性.

**满足 else 与最近的未匹配的 if 匹配的语义:**

只需证明在 **stat 规则二** 中的第一个 stat 不会匹配 **stat 规则一** 即可, 采用反证法:

假设在 **stat 规则二** 中的第一个 stat 匹配 **stat 规则一**, 那么根据最前优先匹配规则, else 应该与第一个 stat 中的 if 匹配, 矛盾!

因此 **stat 规则二** 中的第一个 stat 不会匹配 **stat 规则一**, 于是满足 else 与最近的未匹配的 if 匹配的语义.

(2) 无二义性:

若满足 else 与最近的未匹配的 if 匹配的语义, 即可说明该文法不无二义性, 我们直接证明下一步即可.

**满足 else 与最近的未匹配的 if 匹配的语义:**

与 (1) 类似, 只需证明 **matched\_stat 规则一** 和 **open\_stat 规则二** 中的第一个 matched\_stat 不会匹配 **open\_stat 规则一**, 而显然 matched\_stat 不能匹配 **open\_stat** 的任何规则, 因此直接从上下文无关语法层面上保证正确性.

## 2 作业 (选做部分)

### 题目 1

请使用 Pumping Lemma 证明以下语言不是正则语言:

(1)  $L = \{x \in \{a, b\}^* \mid x \text{ 中 } a, b \text{ 个数相同}\}$ . ①

① 提示: 可以考虑  $s = a^p b^p$ .

(2)  $L = \{a^i b^j \mid i > j\}$ . ②

② 提示: 可以考虑  $s = a^{p+1} b^p$ .

### 解答:

(1) 假设  $L$  是正则语言, 那么令  $n$  为  $L$  对应的泵长度.

令  $s = a^n b^n \in L$ .

根据泵引理, 可以令  $s = xyz$ , 其中  $|xy| \leq n, |y| \geq 1$ .

易知  $y$  的只能由  $a$  组成, 于是对于  $\forall k > 1, xy^k z \notin L$ , 矛盾!

因此  $L$  不是正则语言.

(2) 假设  $L$  是正则语言, 那么令  $n$  为  $L$  对应的泵长度.

令  $s = a^{n+1} b^n \in L$ .

根据泵引理, 可以令  $s = xyz$ , 其中  $|xy| \leq n, |y| \geq 1$ .

易知  $y$  的只能由  $a$  组成, 于是对于  $\forall k = 0, xy^0 z = xz \notin L$ , 矛盾!

因此  $L$  不是正则语言.

---

## 3 反馈

请在 Zulip 平台讨论或将反馈发送至 [hfwei@nju.edu.cn](mailto:hfwei@nju.edu.cn)。