

第三章

1. 描述下列正规式（正则）定义的语言（请用汉语直白地描述该语言）

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

以 0 开头，以 0 结尾，长度至少为 2 的二进制串。

(b) $((\varepsilon|0)1^*)^*$

0 不连续出现的二进制串。

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

倒数第 3 位是 0，且长度至少为 3 的二进制串。

(d) $0^*10^*10^*10^*$

1 的个数正好为 3 个的二进制串。

(e) $(00|11)^*((01|10)(00|11)^*(01|10)(00|11)^*)^*$

整个二进制串的长度必须为偶数，之后相邻两个一组，其中 01 和 10 的个数也为偶数的二进制串。

2. 为下列语言写出正规定义（或正规式、正规文法、有限自动机）

(a) 包含五个元音，且按顺序排列的所有字母串（RE）

为了方便起见，这里认为“按顺序排列”指所有字母按顺序排列，而不只是元音按顺序排列。且这里认为只有大写字母，没有小写字母。那么 RE 为：

$AB^*C^*D^*EF^*G^*H^*IJ^*K^*L^*M^*N^*OP^*Q^*R^*S^*T^*UV^*W^*X^*Y^*Z^*$

(b) 字母按字典升序排列的所有字母串 (*RE*)
$$A^*B^*C^*D^*E^*F^*G^*H^*I^*J^*K^*L^*M^*N^*O^*P^*Q^*R^*S^*T^*U^*V^*W^*X^*Y^*Z^*$$
(c) 以/* 开始, */结束的注释, 中间不能包含 */, 除非包含在"和"中 (*RE*)

这里用 Σ 表示所有的字符, 用 $!(*/)$ 表示不是 $*/$ 的字符串, 用 $!"$ 表示不是 $"$ 的字符。用 $(/*)$ 和 $(*/)$ 表示 $/*$ 和 $*/$, 暂时不考虑 $*$ 的转义问题。

$$(/*)((!("!")*)*(!(*/))*)*(*/)$$
(d) 不包含重复数字的数字串 (正规定义)

继续用 $!$ 表示“不匹配”, 用 $\backslash 1$ 表示之前匹配过的带括号的捕获组。

$$\text{digit} \rightarrow 0|1|2|3|4|5|6|7|8|9$$

$$\text{repeat_digit} \rightarrow (\text{digit}) \text{digit}^* \backslash 1$$

$$\text{not_repeat_digit_string} \rightarrow (!\text{repeat_digit})^*$$
(e) 包含至多一个重复数字的数字串 (正规定义)

可能不包含重复数字, 也可能恰好包含一个重复数字。看上去只比上一题多了一个要求, 但变得非常复杂, 所以直接换种方式吧。

$$\text{repeat_max_1_digit} \rightarrow \{x; x \text{ 是 } (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) \cup [0-9] \text{ 的一个全排列的一个子序列}\}$$
(f) 包含偶数个 0 和奇数个 1 的 0、1 串 (正规定义)

参考: https://blog.csdn.net/weixin_44001521/article/details/112852266

https://blog.csdn.net/weixin_42787270/article/details/88959203

https://blog.csdn.net/qq_42689684/article/details/82763075

$$\text{even_0_even_1} \rightarrow (00|11|(01|10)(00|11)^*(01|10))^* \quad (\text{这个好像和 1 (e) 是等价的})$$

$$\text{even_0_odd_1} \rightarrow 1 \text{ even_0_even_1} | 0(00|11)^*(01|10) \text{ even_0_even_1}$$