

第一次作业

第一题

下列正则表达式定义了什么语言? (用尽可能简短的自然语言描述)

- | | |
|-------------------------------|---|
| 1. $(a b)^*a(a b \epsilon)$ | 1. 在集合 $\{a,b\}$ 中, 所有以 a,aa ,或者 ab 结尾的字符串组成的集合 |
| 2. $(A B ... Z)(a b ... z)^*$ | 2. 以一个大写字母开头而且只含一个大写字母的字符串组成的集合 |
| 3. $(aa b)^*(a bb)^*$ | 3. 任意个 aa 或 b 连接上任意个 a 或 bb 的字符串的集合 |
| 4. $((\epsilon a)b)^*$ | 4. 在集合 $\{a,b\}$ 中, 任意个 b 或者任意个 ab 重复出现的字符串 |
| 5. $b^*ab^*a(b a)^*$ | 5. 在集合 $\{a,b\}$ 中, 至少有两个 a 的字符串 |
| 6. $a(aa)^*bb(bb)^*(cc)^*c$ | 6. 在集合 $\{a,b\}$ 中, 奇数个 a , 大于0的偶数个 b 以及奇数个 c 连接而成的字符串。 |

第二题

按以下自然语言描述，写出正则表达式

1. 在 $\{0, 1\}$ 上不以0开头的、以11结尾的字符串的集合。
2. 最多只含 2 个a 的 $\{a, b\}$ 上的集合。

$11|1(1|0)^*11$

$b^*|b^*ab^*|b^*ab^*ab^*$

第三题

设字母表 $\Sigma=\{a,b\}$ ，用正则表达式（只使用 $a, b, \varepsilon, |, *, +, ?$ ）描述下列语言：

1. 不包含子串 ab 的所有由 a 和 b 组成的字符串 b^*a^*
2. 不包含子串 abb 的所有由 a 和 b 组成的字符串 $b^*(a^*|(ba)^*)^*$
3. 不包含子序列 abb 的所有由 a 和 b 组成的字符串. $b^*a^*ba^*|\varepsilon$

注意：关于子串（substring）和子序列（subsequence）的区别可以参考课本的内容。

第四题

把下列的正则表达式直接转换成DFA，对生成的DFA请用初始状态，终止状态集合，转换表格的方式进行文本描述。

$a(a|b)^*a$

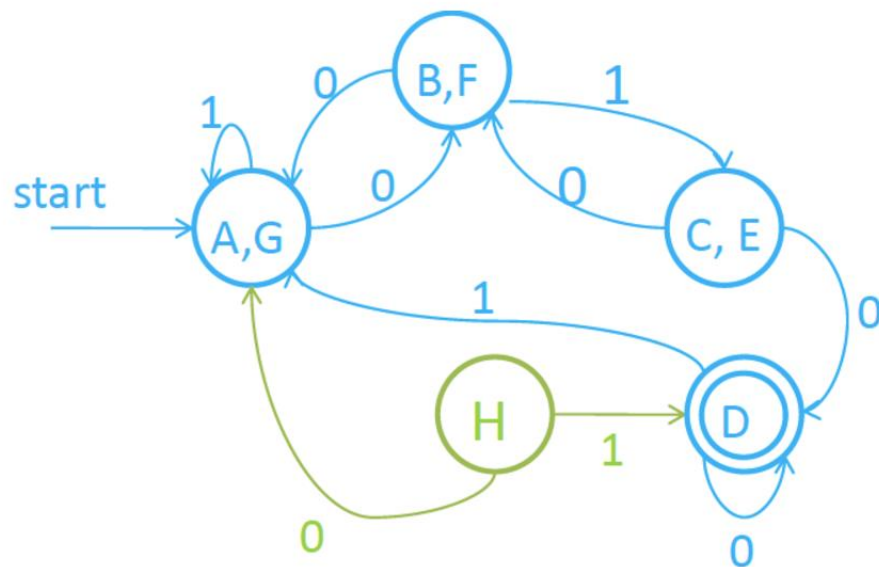
$((\epsilon|a)b^*)^*$

$a^*ba^*ba^*ba^*$

第五题

考虑以下DFA 的状态迁移表，其中0，1 为输入符号，A~H 代表状态

	0	1
A	B	A
B	A	C
C	D	B
D	D	A
E	D	F
F	G	E
G	F	G
H	G	D

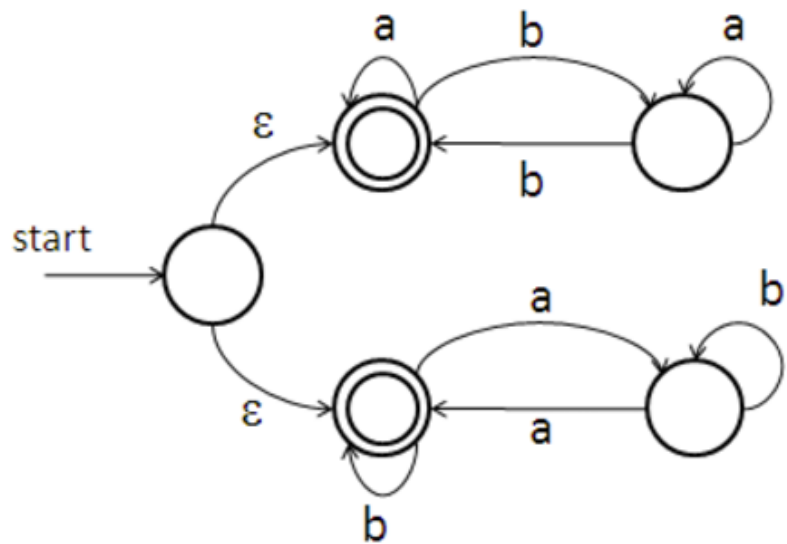


其中A 为初始状态，D 为接受状态，请画出与此DFA 等价的最小DFA，并在新的DFA 状态中标明它对应的原DFA 状态的子集。

对最小化后的DFA，请用起始状态，终止状态，状态迁移表的方式进行文本描述。

第六题

考虑以下的DFA

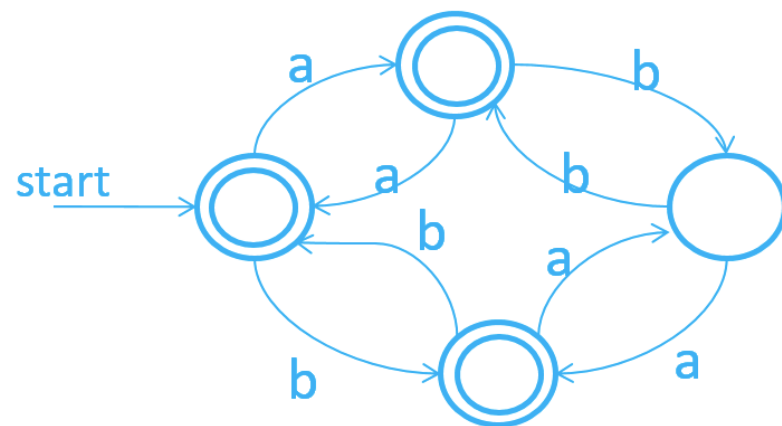


1. 这一 NFA 接受什么语言（用自然语言描述）？
2. 构造接受同一语言的 DFA .

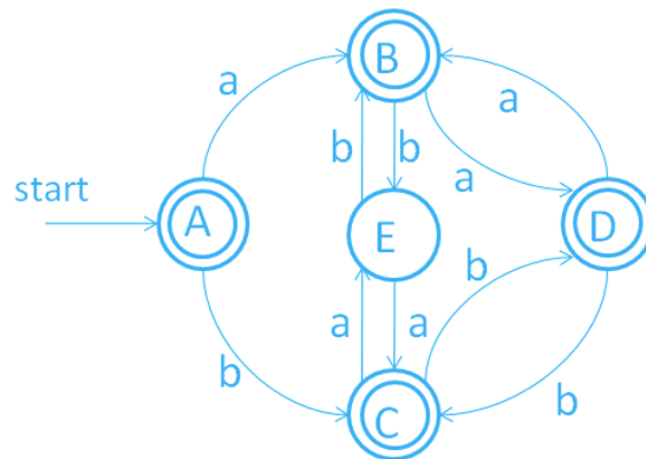
$$a^*(ba^*b)^*a^* \mid b^*(ab^*a)^*b^*$$

所有只含有字母a和b，并且a出现偶数次或b出现偶数次的字符串.

直接构造通常得：

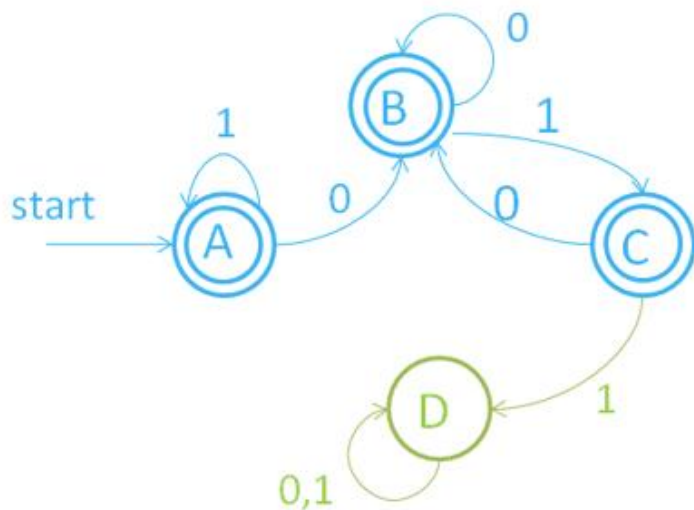


由 NFA 构造 DFA 得这一答案：



第七题

1. 画出一个 DFA，该DFA 恰好识别所有不含011 子串的所有二进制串.
2. 假如要求你再画一个 DFA，该DFA 恰好识别所有不含101 子串的所有二进制串。
你从中总结出解决这一类问题的普遍规律是什么？
3. 再证明：对任一正则表达式R，一定存在另一正则表达式R'，使得 $L(R')$ 是 $L(R)$ 的补集.



证明：根据正则表达式与 DFA 的等价性，一定存在识别语言 $L(R)$ 的 DFA. 设这一 DFA 为 M , 则将 M 的所有接受状态改为非接受状态, 所有非接受状态改为接受状态, 得到新的 DFA M' . 易知 M' 识别语言 $L(R)$ 的补集. 再由正则表达式与 DFA 的等价性知必存在正则表达式 R' , 使得 $L(R')$ 是 $L(R)$ 的补集.

第八题

设有一门小小语言仅含 z 、 o 、 $/$ （斜杠）3 个符号，该语言中的一个注释由 $/ o$ 开始、以 $o /$ 结束，并且注释禁止嵌套.

1. 请给出单个正则表达式，它仅与一个完整的注释匹配，除此之外不匹配任何其他串。书写正则表达式时，要求仅使用最基本的正则表达式算子（ ϵ ， $|$ ， $*$ ， $+$ ， $?$ ）。
2. 给出识别上述正则表达式所定义语言的确有限自动机（DFA）。你可根据问题直接构造DFA，不必运用机械的算法从上一小题的正则表达式转换得到DFA。

$/o(o^*z|/)^*o+ /$

思路：基本思路是除了最后一个 $o /$ ，在注释中不能出现 o 后面紧跟着 $/$ 的情况；还有需要考虑的是最后一个 $o /$ 之前也可以出现若干个 o 。

