

输出函数: i	output(i)
3	the
6	she
9	che
11	this
13	think
15	third
19	hers

1.output 出多个结果的情况, 是一个接收状态的 fail 指针, 指向了另外一个接收状态, 此时, 原来的那个接收状态需要多打印一个结果。比如 she 中, e 的 fail 指针指向了 he 中的 e, 那么 she 的接收状态 i 的 output 函数就需要添加一个 he。

② 求 shift 表与 HASH 表: i: 0 1 2 3 4 5 6 7 8 9 10, 11, 12, 13, 14, 15, 16

因: mod 28, 则 shift: shift(i): 5 5 4 5 4 3 5 5 2 0 5 5 5 5 5 0 5

表项共对 HASH 2 3

首先看 i=0 时, 无块, 则 5 (取长度)

i=2 时, 有 sh, 取其右侧字母 h 距最右侧的距离 4, 若多次出现, 取最小值。

再求 HASH 表, 其记录 i 是哪时模式串的最右块 \Rightarrow next 0, next 1, ch=2, h

2.这个地方有误, 没有块对应时, 计算方法应该为 $length - B + 1$, 其中 B 是块长, 此处是 2, length 是模式串长, 此处是 6, 求得找不到时右移结果为 $6 - 2 + 1 = 5$ 。

3.对于 BM 算法的复杂度, 最小时为 $O(n/m)$, 即每次都不匹配; 最大时为 $O(n*m)$, 即每次都匹配。注意到即使 BM 算法使用了好后缀规则, 相同的部分也需要再次比较。