

## 考点 10: 串的模式匹配算法

						• 小题8
2009	2010	2011	2012	2013	2014	2015
			• 小题9			
2016	2017	2018	2019	2020	2021	2022

历年考频: 小题×2

1

## 考点10

串的模式匹配  
算法

王道考研/CSKAOYAN.COM

2

## 串的模式匹配算法

9. 设主串  $T = \text{"abaabaabcabaabc"}$ , 模式串  $S = \text{"abaabc"}$ , 采用 KMP 算法进行模式匹配, 到匹配成功时为止, 在匹配过程中进行的单个字符间的比较次数是
- A. 9                      B. 10                      C. 12                      D. 15

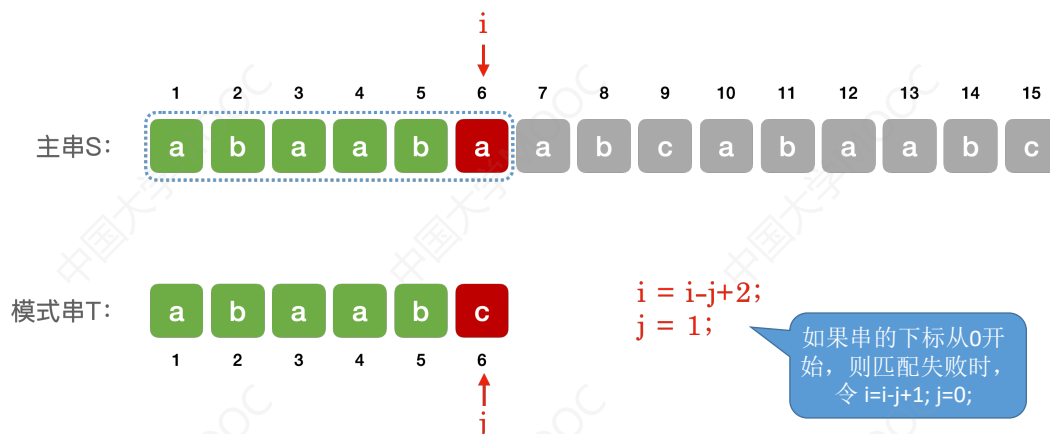
2019真题

8. 已知字符串  $S$  为 "abaabaabacacaabaabcc", 模式串  $t$  为 "abaabc". 采用 KMP 算法进行匹配, 第一次出现 "失配" ( $s[i] \neq t[j]$ ) 时,  $i=j=5$ , 则下次开始匹配时,  $i$  和  $j$  的值分别是\_\_\_\_\_。
- A.  $i=1, j=0$               B.  $i=5, j=0$               C.  $i=5, j=2$               D.  $i=6, j=2$

2015真题

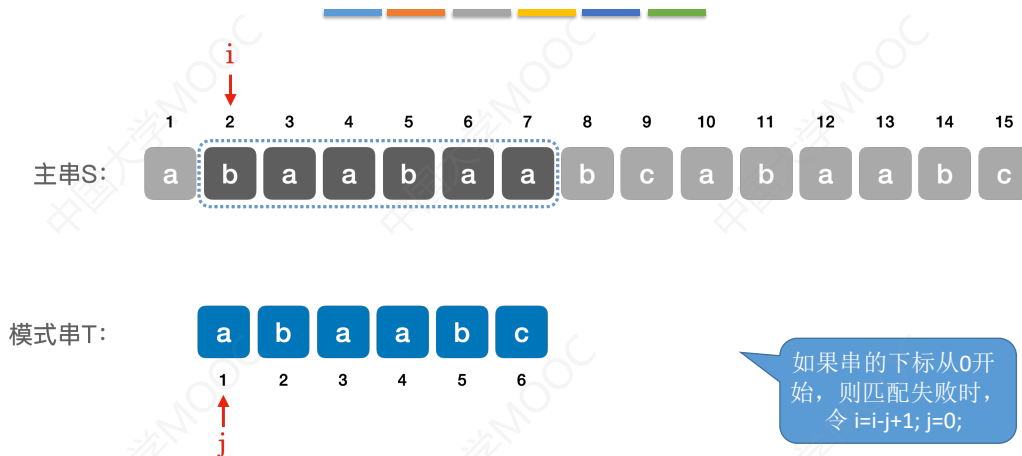
3

## 朴素模式匹配算法



4

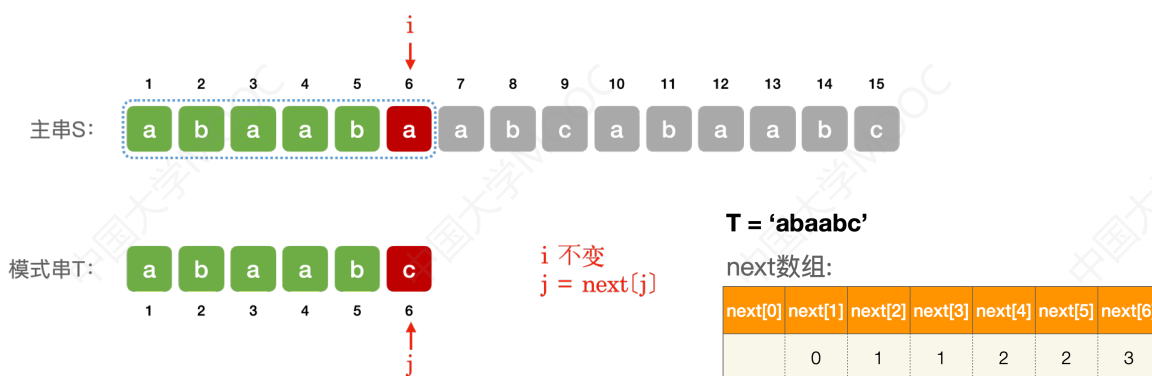
## 朴素模式匹配算法



若模式串长度为 $m$ ，主串长度为 $n$ ，则直到匹配成功/匹配失败最多需要  $(n-m+1)*m$  次比较  
最坏时间复杂度:  $O(nm)$

5

## KMP算法



KMP算法: 比较  $S_i$  和  $T_j$ , 若相等则指针各增 1, 否则  $j$  再退回到  $j=\text{next}[j]$  的位置, 再继续比较  $S_i$  和  $T_j$ 。当子串和模式串不匹配时, 主串指针  $i$  不回溯。

算法平均时间复杂度:  $O(n+m)$

6

### KMP算法

123456789101112131415

a b a a b a a b c a b a a b c

123456

a b a a b c

主串S:

模式串T:

T = 'abaabc'

next数组:

next[0]	next[1]	next[2]	next[3]	next[4]	next[5]	next[6]
	0	1	1	2	2	3

KMP算法: 比较 S<sub>i</sub> 和 T<sub>j</sub>, 若相等则指针各增 1, 否则 j 再退回到下一个 j=next[j] 的位置, 再继续比较 S<sub>i</sub> 和 T<sub>j</sub>。当子串和模式串不匹配时, 主串指针 i 不回溯, 模式串指针 j=next[j]

算法平均时间复杂度:  $O(n+m)$

7

### Next数组

next[j] 的含义: 当主串 S<sub>i</sub> 与子串 T<sub>j</sub> 比较发生失配时, 则子串退回到 next[j] 的位置与主串当前位置 (S<sub>i</sub>) 继续进行比较。

模式串 T = ababaa

ababaa

序号j	1	2	3	4	5	6
模式串	a	b	a	b	a	a
next[j]	0	1				

next[1]都无脑写 0  
next[2]都无脑写 1

其他 next: 在不匹配的位置前, 划一根美丽的分界线  
模式串一步一步往后退, 直到分界线之前“能对上”, 或模式串完全跨过分界线为止。此时 j 指向哪儿, next数组值就是多少

8

Next数组

考场上，你手里有：①纸质试卷；②纸质答题卡；③草稿纸。答题卡上不能乱写东西

计算 ababaa 的next数组：在稿纸上写上模式串的各种前缀，试卷空白处写上完整的模式串并标好下标

9

Next数组

next[j]的含义：当主串  $S_i$  与子串  $T_j$  比较发生失配时，则子串退回到 next[j]的位置与主串当前位置 ( $S_i$ ) 继续进行比较。

模式串：        a b a a b c

next数组：     0 1 1 2 2 3

大多数教材默认串的下标从1开始，因此next[1]无脑写0，next[2]无脑写1

如果下标从0开始，则所有next值都要减1。next[0]无脑写-1，next[1]无脑写0

编号	0	1	2	3	4	5
t	a	b	a	a	b	c
next	-1	0	0	1	1	2

10

## Next数组

4. 已知串 S='aaab', 其 next 数组值为 ( )。
- A. 0123                      B. 0112  
C. 0231                      D. 1211
5. 串 'ababaaababaa' 的 next 数组值为 ( )。
- A. 01234567899              B. 012121111212  
C. 011234223456              D. 0123012322345
6. 串 'ababaaababaa' 的 next 数组为 ( )。
- A. -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 8, 8  
B. -1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1  
C. -1, 0, 0, 1, 2, 3, 1, 1, 2, 3, 4, 5  
D. -1, 0, 1, 2, -1, 0, 1, 2, 1, 1, 2, 3

Tips: 根据Next数组的第一个值判断下标是从0还是从1开始

next 数组第一个值为0→下标从1开始  
next 数组第一个值为-1→下标从0开始

11

## 考点 10: 串的模式匹配算法

						• 小题8
2009	2010	2011	2012	2013	2014	2015
			• 小题9			
2016	2017	2018	2019	2020	2021	2022

历年考频: 小题×2

12