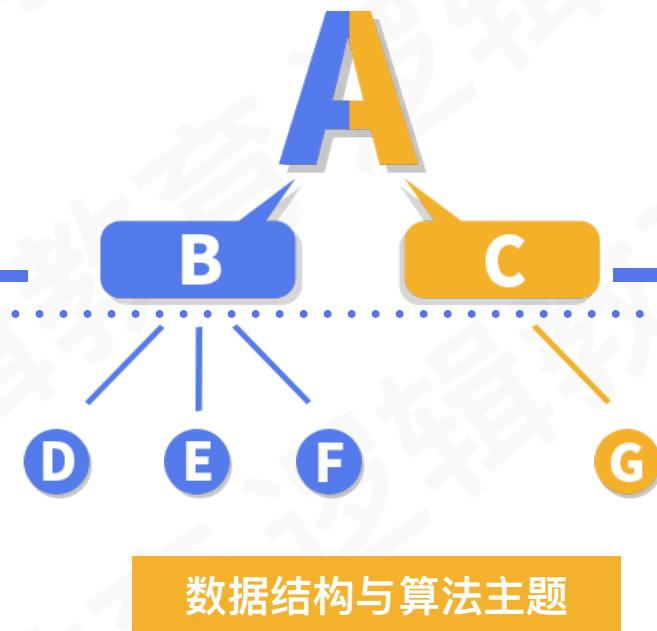




逻辑教育
Logic education

Hello 数据结构与算法

数据结构与算法 一串的理解与应用



@CC老师
全力以赴·非同凡“想”

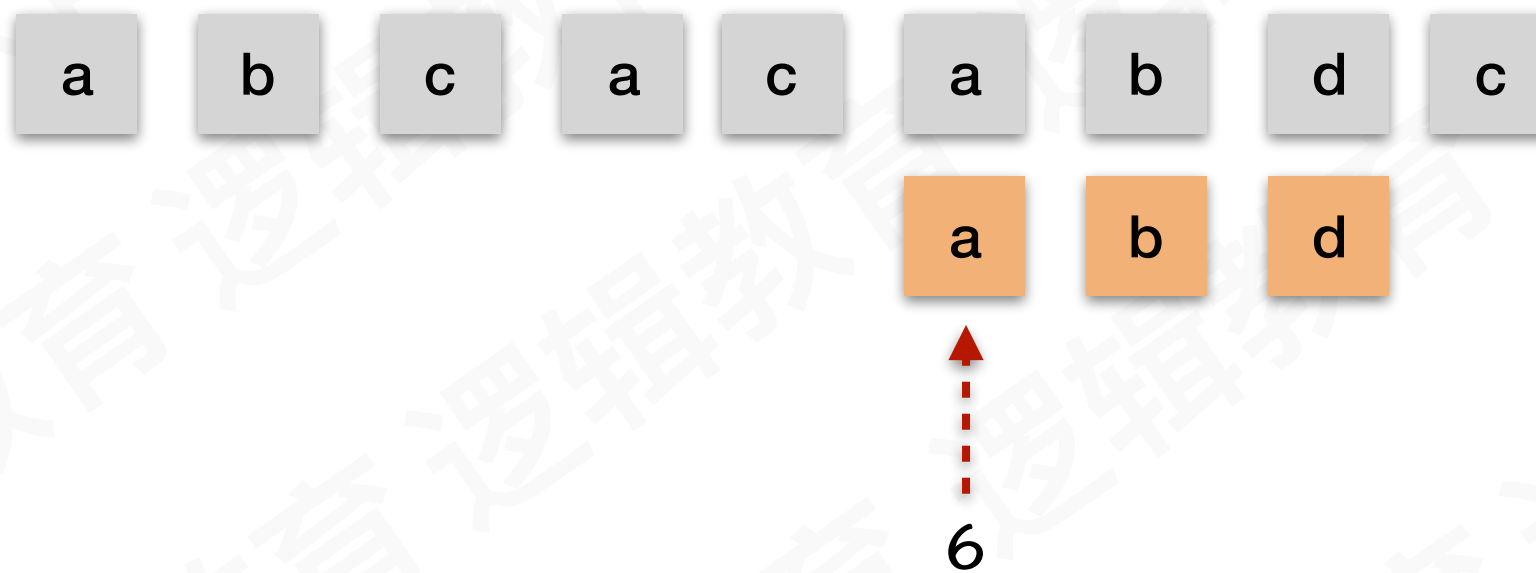
课程研发:CC老师
课程授课:CC老师



字符串匹配

题目: 有一个主串 $S = \{a, b, c, a, c, a, b, d, c\}$, 模式串 $T = \{a, b, d\}$; 请找到模式串在主串中第一次出现的位置;

提示: 不需要考虑字符串大小写问题, 字符均为小写字母;



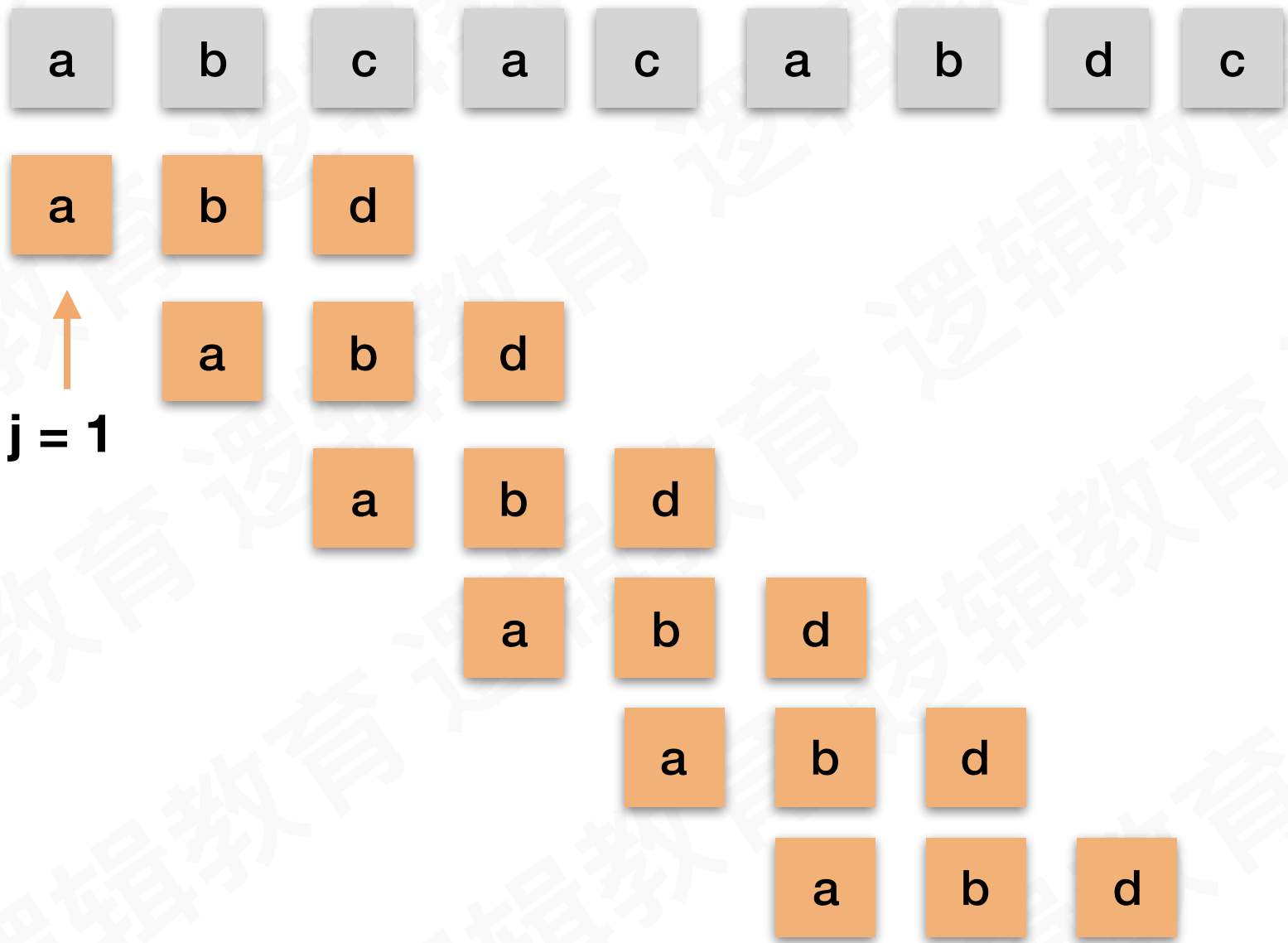


BF 算法

$i = 1$



$j = 1$



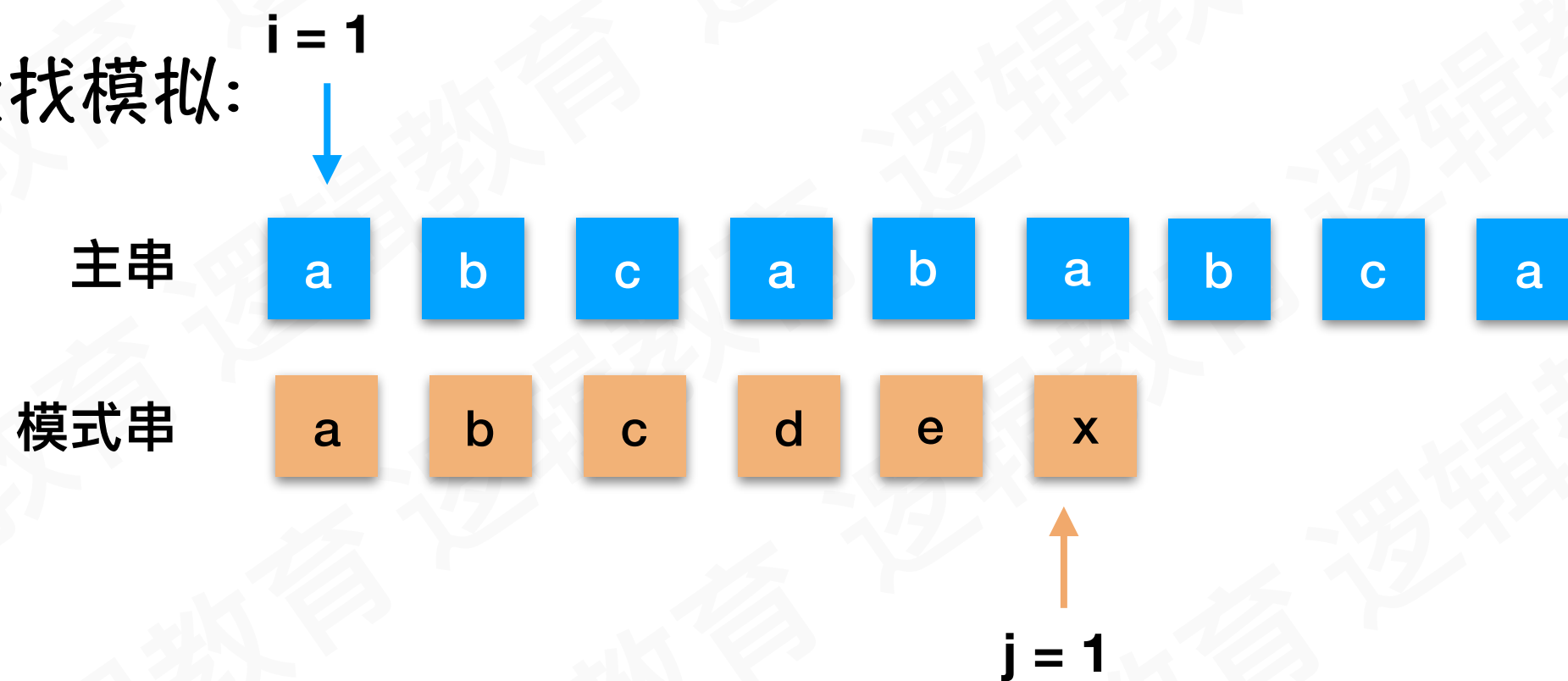
课程研发:CC老师
课程授课:CC老师



BF 算法

假设, 主串 $S = \text{"abcababca"}$; 模式串 $T = \text{"abcde x"}$

BF算法查找模拟:

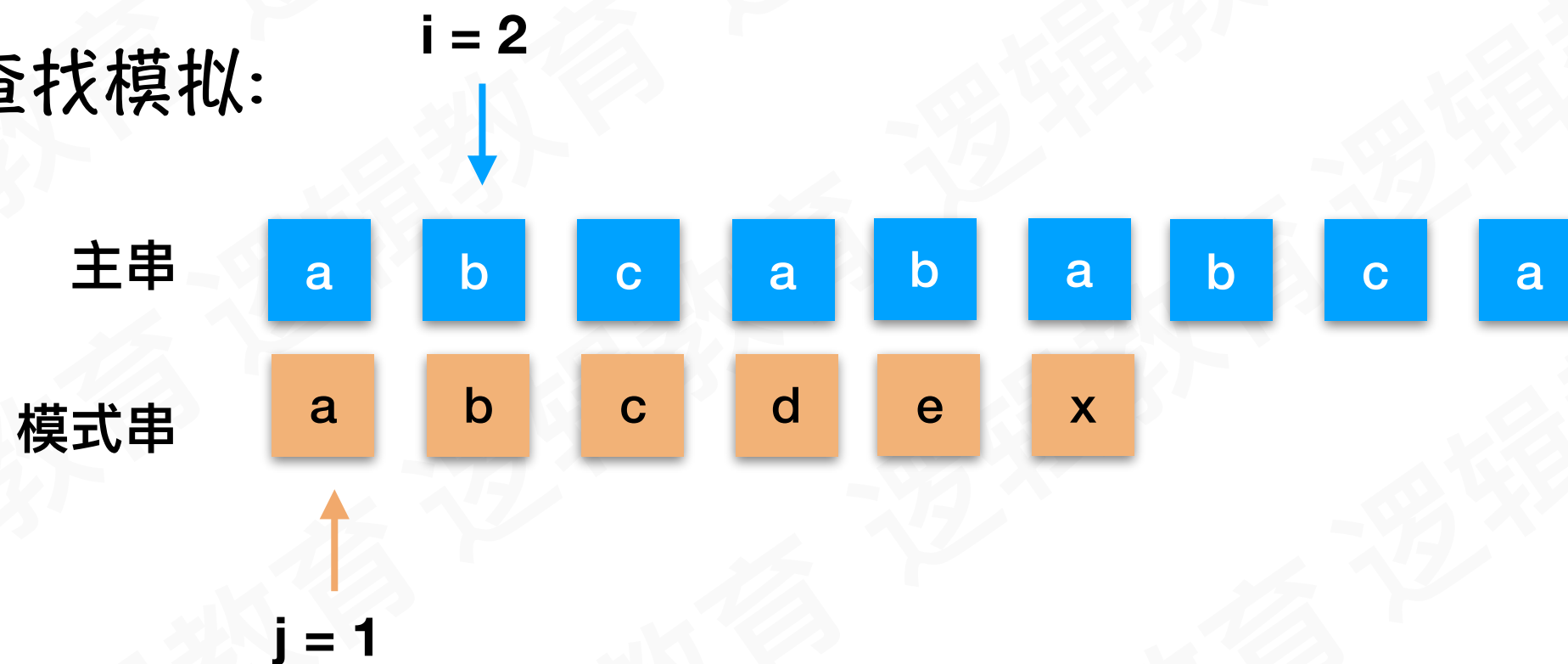




BF 算法

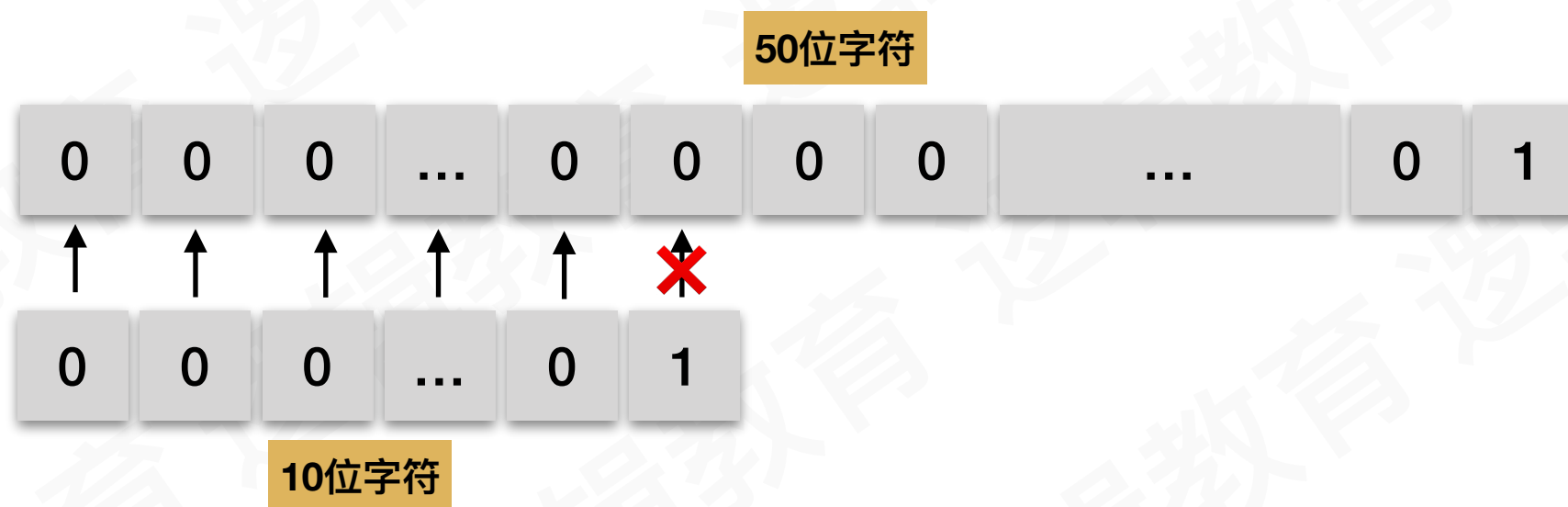
假设, 主串 $S = \text{"abcababca"}$; 模式串 $T = \text{"abcdex"}$

BF算法查找模拟:





BF 算法效率分析



T在第一个位置判断了10次发现字符串不匹配



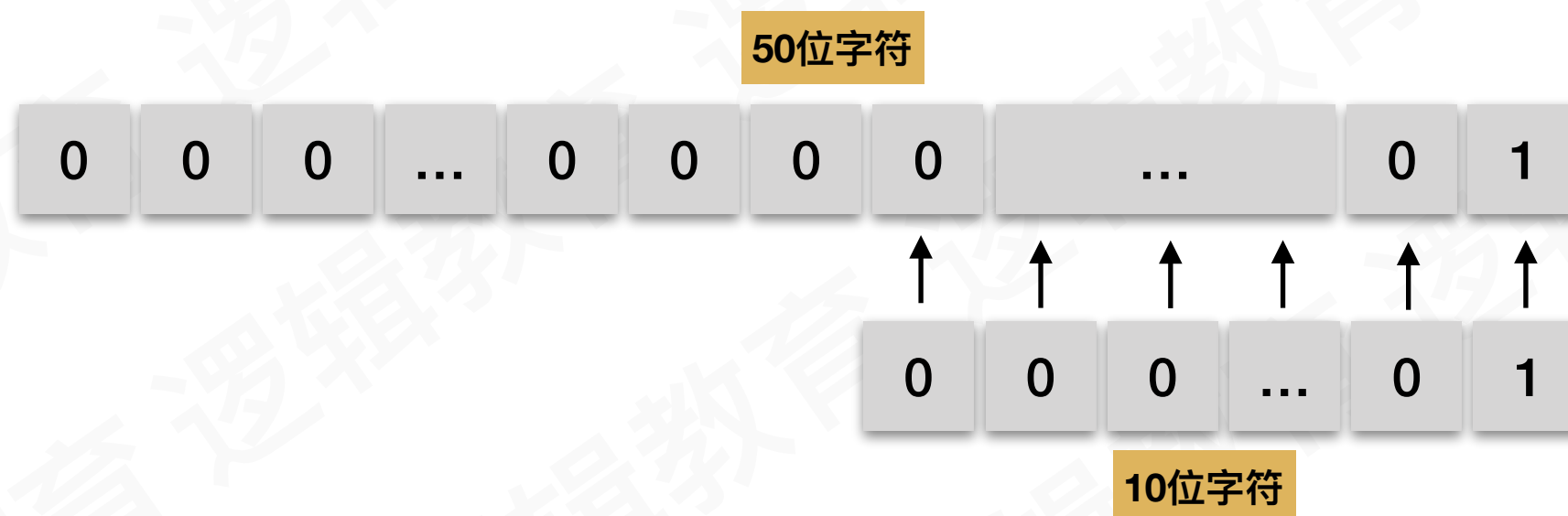
BF 算法效率分析



T在第二个位置判断了10次发现字符串不匹配



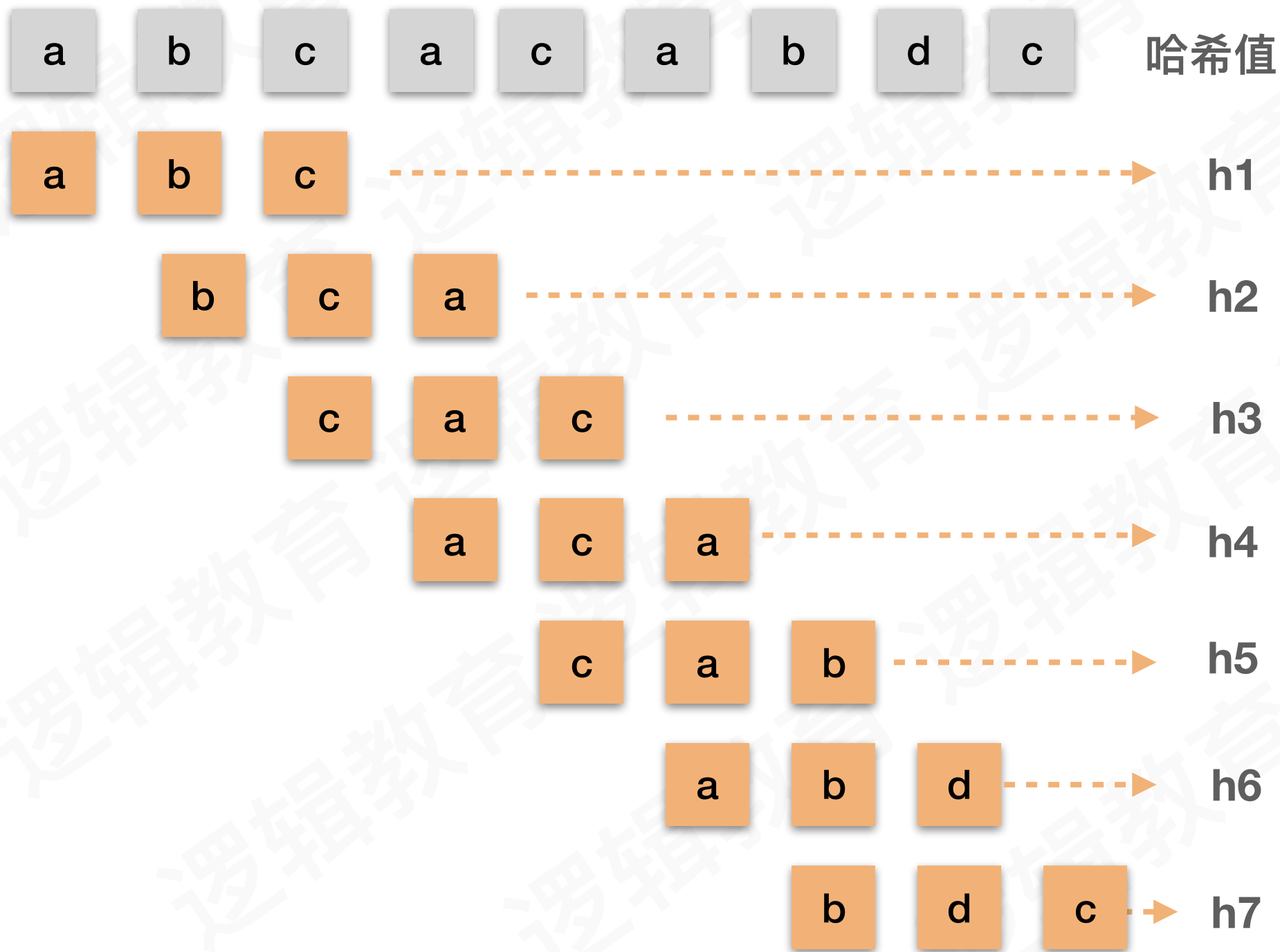
BF 算法效率分析



T在第41个位置判断了10次发现字符串终于匹配成功
期间进行了 $(50-10+1)*10$ 次判断操作



RK 算法核心思想



课程研发:CC老师
课程授课:CC老师



逻辑教育
Logic education

RK 算法核心思想

如何将模式串或者主串拆分后的子串换算成一个哈希值？

课程研发:CC老师
课程授课:CC老师



Hash (哈希算法)

Hash (哈希). 一般中文也翻译做“散列”; 也可以直接音译“哈希”; 散列在开发中是常见手段! 比如大家常用的MD5 算法就是哈希算法; 哈希算法在安全方面应用是非常多,一般体现在如下这几个方面:

1. 文件校验
2. 数字签名
3. 鉴权协议



Hash (哈希算法)

将不同的字符组合能够通过某种公式的计算映射成不同的数字!

例如

比较“abc”与“cde”; 比较 123 与 456; 是一样的吗?



逻辑教育
Logic education

RK 算法核心思想

$$657 = 6 * 10 * 10 + 5 * 10 + 7 * 1$$

$$657 = 6 \times 10^2 + 5 \times 10^1 + 7 \times 10^0$$

课程研发:CC老师
课程授课:CC老师



逻辑教育
Logic education

RK 算法核心思想

字母换算成哈希值

课程研发:CC老师
课程授课:CC老师



RK 算法核心思想

将'当前字母' - 'a' = 数字

例如

$$a - a = 0;$$

$$b - a = 1;$$

$$c - a = 2;$$

$$d - a = 3;$$

$$e - a = 4;$$

...

课程研发:CC老师

课程授课:CC老师



逻辑教育
Logic education

RK 算法核心思想

小写字母之间存在的进制

课程研发:CC老师
课程授课:CC老师



RK 算法核心思想

$$\begin{aligned}\text{"cba"} &= 'c' * 26 * 26 + 'b' * 26 + 'a' * 1 \\ &= 2 * 26 * 26 + 1 * 26 + 0 * 1 \\ &= 1378\end{aligned}$$

$$\begin{aligned}\text{"cba"} &= c \times 26^2 + b \times 26^1 + a \times 26^0 \\ &= 2 \times 26^2 + 1 \times 26^1 + 0 \times 26^0 \\ &= 1352 + 26 + 0 \\ &= 1378\end{aligned}$$

课程研发:CC老师
课程授课:CC老师



RK 算法核心思想

为了让大家知道接下来推演过程,以数字为例,会更容易让大家理解,它的全集是 $\{0,1,2,3,4,5,6,7,8,9\}$. $d = 10$; 模式串 $p = 123$, 主串 $s = 65127451234$

123

65127451234

651 512 127 274 745 451 512 123 234

求解 子串A的哈希值 $= 1 \times 10^2 + 2 \times 10^1 + 3 \times 10^0$. 代码实现!



逻辑教育
Logic education

RK 算法核心思想

主串拆解的子串与模式串的哈希值比较？

课程研发:CC老师
课程授课:CC老师

转载需注明出处,不得用于商业用途.已申请版权保护



RK 算法核心思想

主串: 

主串: 

模式串: 



RK 算法核心思想

主串: d b c e d b

$$= 3 \times 26^2 + 1 \times 26^1 + 2 \times 26^0$$

主串: d b c e d b

$$= 1 \times 26^2 + 2 \times 26^1 + 4 \times 26^0$$



RK 算法核心思想

子串哈希值求解规律:

相邻的2个子串 $s[i]$ 与 $s[i+1]$ (i 表示子串从主串中的起始位置,子串的长度都为 m). 对应的哈希值计算公式有交集. 也就说我们可以使用 $s[i-1]$ 计算出 $s[i]$ 的哈希值;



RK 算法核心思想

以数字为例, 它的全集是 $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. $d = 10$;
模式串 $p = 123$, 主串 $s = 65127451234$

123

65127451234								
651	512	127	274	745	451	512	123	234

$$s[i] = 1 \times 10^2 + 2 \times 10^1 + 7 \times 10^0$$

274

$$s[i+1] = 2 \times 10^2 + 7 \times 10^1 + 4 \times 10^0$$



RK 算法核心思想

$$s[i] = 1 \times 10^2 + 2 \times 10^1 + 7 \times 10^0$$

$$s[i+1] = 2 \times 10^2 + 7 \times 10^1 + 4 \times 10^0$$

$$s[i+1] = 10 \times (127 - 1 \times 10^2) + 4$$

$$s[i+1] = 10 \times (s[i] - 1 \times 10^2) + 4$$

$s[i+1]$ 实现上是上一个 $s[i]$ 去掉最高位数据,其余的 $m-1$ 为字符乘以 d 进制. 再加上最后一个为字符得到;



RK 算法核心思想

$m = 3; n = 8$

模式串:

1 2 3

循环结束位置 ← END



主串S:

1 2 7 4 5 1 2 3



注意: d 指的是进制!
 d 在数字里指的是十进制

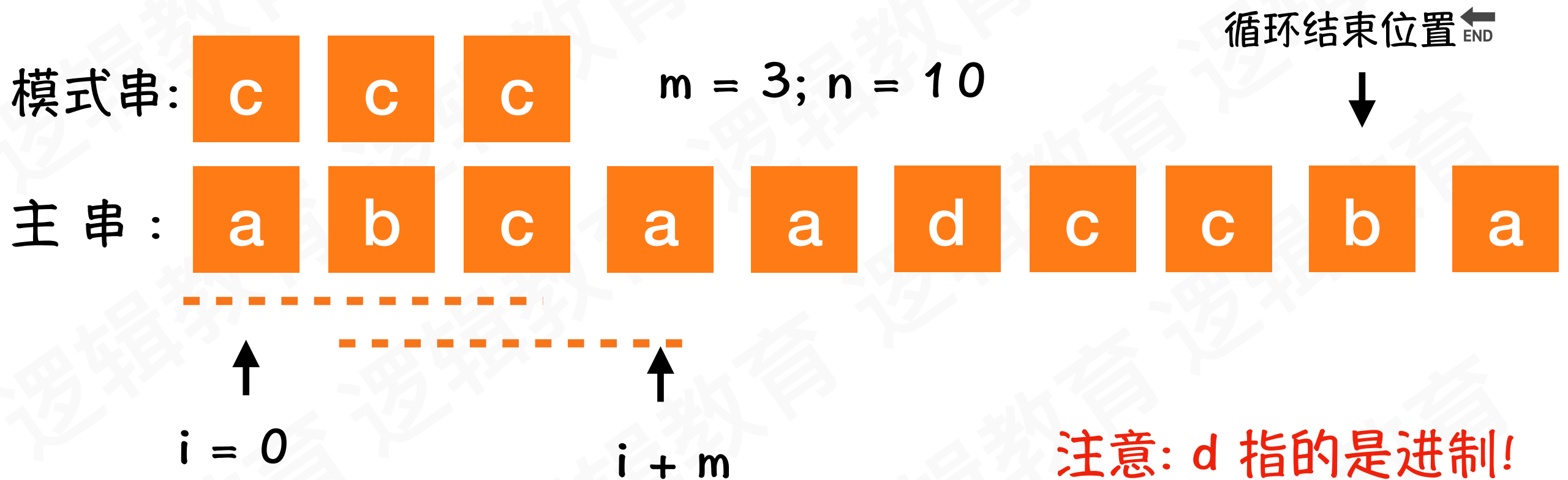
$$274 = (127 - 1 \times d^2) \times d + 4$$

$$\text{主串分割后的哈希值: } St[i+1] = (st[i] - d^2 \times s[i]) \times d + s[i+m]$$

课程研发:CC老师
课程授课:CC老师



RK 算法核心思想



$$bca = (abc - a \times d^2) \times d + a$$

$$St[i] = (st[i-1] - d^2 \times (s[i] - 'a') \times d + (s[i+m] - 'a'))$$



逻辑教育
Logic education

RK 算法核心思想

请用代码实现 刚刚关于主串与子串哈希值计算与比较的代码!

课程研发:CC老师
课程授课:CC老师



逻辑教育
Logic education

RK 算法核心思想

哈希冲突解决方案!

课程研发:CC老师
课程授课:CC老师



逻辑教育
Logic education

RK 算法核心思想

哈希冲突解决方案!

1. 设计更复杂的哈希公式; (后面课程会讲解关于哈希值的相关课程)

课程研发:CC老师
课程授课:CC老师



逻辑教育
Logic education

RK 算法核心思想

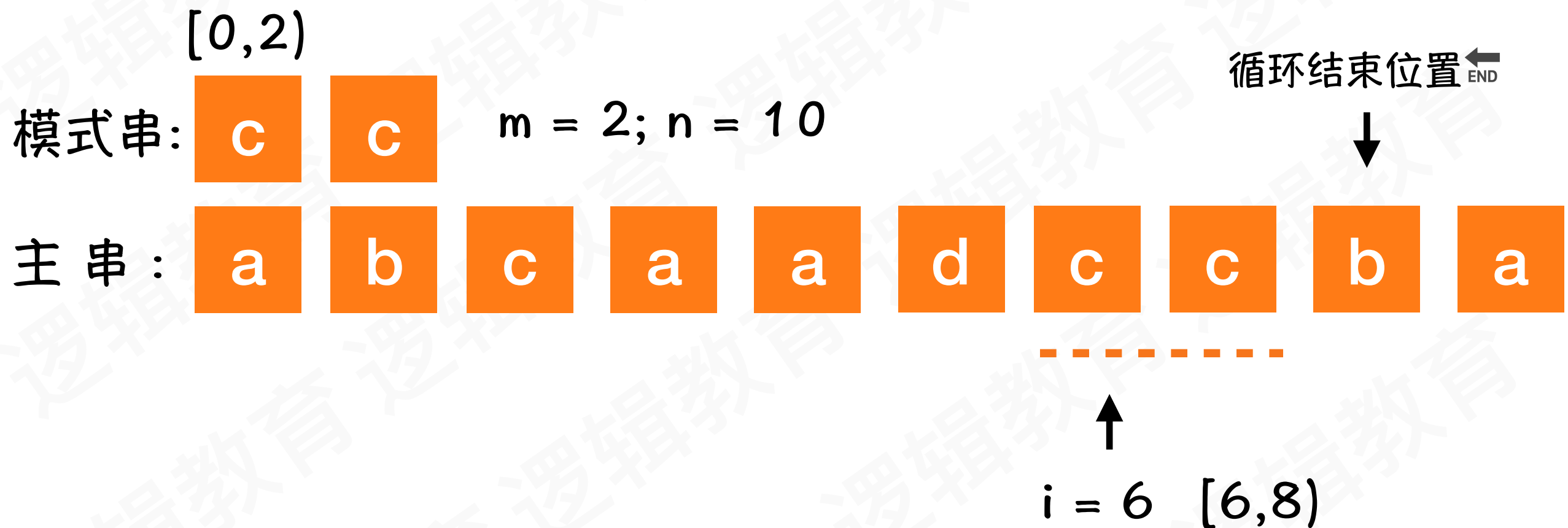
哈希冲突解决方案!

2. 如果相等时,不要直接返回结果. 而是重新核实!

课程研发:CC老师
课程授课:CC老师



RK 算法核心思想



请用代码实现 刚刚关于主串 $[i, i+m)$ 的哈希值与子串 $(0, m)$ 哈希值相等时,二次核实函数!



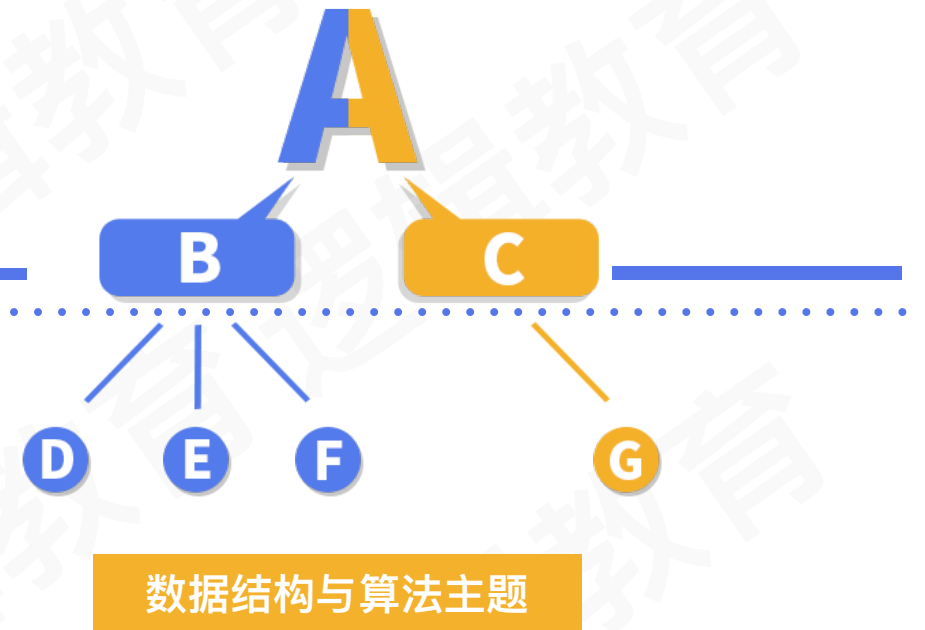
RK 算法核心思想

1. 如果不做哈希冲突二次核查 比较次数是 $n-m+1$ 次; 那么时间复杂度为 $O(n)$;
2. 但是要想解决冲突存在可能性.就需要添加二次核查! 那么就需要 m 次比对; 那么时间复杂度为 $O(n*m)$;



逻辑教育
Logic education

*Class Ending !
thanks, see you next time*



@CC老师

全力以赴·非同凡“想”

课程研发:CC老师
课程授课:CC老师

转载需注明出处,不得用于商业用途.已申请版权保护



逻辑教育
Logic education

课程研发:CC老师
课程授课:CC老师

转载需注明出处,不得用于商业用途.已申请版权保护