Boyer-Moore Algorithms (boyermoores)

จงเขียนโปรแกรมโดยใช้วิธี Boyer-Moore algorithm ในการแก้ปัญหา string pattern matching โดยที่โจทย์จะให้ text t และ pattern p คุณจะต้องเขียนโปรแกรมเพื่อค้นหาว่า มี p อยู่ใน t ที่ตำแหน่งใดเป็นตำแหน่งแรก

เพื่อให้คุณใช้วิธี Boyer-Moore algorithm ในการแก้ปัญหา คุณจะต้องพิมพ์ข้อมูล last ของแต่ละตัวอักขระที่มีอยู่ใน pattern p และต้องพิมพ์จำนวนครั้งที่มีการเปรียบเทียบระหว่าง text กับ pattern ด้วย

ข้อมูลป้อนเข้า

บรรทัดแรก มีจำนวนเต็ม 2 จำนวน คือ n และ m (1<= n <=10,000; 1<= m <=100 และ m <= n) โดยที่ n เป็นความยาวของ text และ m เป็นความยาวของ pattern

บรรทัดที่ 2 มี text t ที่มีความยาว n ตัวอักษร โดยตัวอักษรจะเป็นตัวอักษรภาษาอังกฤษตัวเล็กทั้งหมด บรรทัดที่ 3 มี pattern p ที่มีความยาว m ตัวอักษร โดยตัวอักษรจะเป็นตัวอักษรภาษาอังกฤษตัวเล็กทั้งหมด

ผลลัพธ์

ตามลำดับตัวอักษร

มี s + 2 บรรทัด โดยที่ s เป็นจำนวนตัวอักษรทั้งหมดที่ไม่ซ้ำกันที่อยู่ใน text t ซึ่ง s บรรทัดแรก จะมีตัวอักษร c และค่า last ของแต่ละตัวอักษร c ทั้งหมดที่ไม่ซ้ำกันที่อยู่ใน text t โดยที่คุณจะต้องพิมพ์ค่า last

บรรทัดถัดมาเป็น จำนวนครั้งของการเปรียบเทียบระหว่าง text กับตัวอักษรตัวสุดท้ายของ pattern บรรทัดสุดท้าย มีจำนวนเต็ม i (0 <= i <= n-m หรือ i = -1) ซึ่งจะเป็นตำแหน่งแรกที่ p ปรากฎอยู่ใน t และถ้า p ไม่ปรากฎใน t ให้ ค่า i เป็น -1

ตัวอย่าง 1

ในตัวอย่างนี้ จะมี text t คือ abcddabcacbdacabacac และมี pattern p คือ abac ดังนั้น n=20, m=4 และมีข้อมูล last ดังนี้

ตัวอักษร	а	b	С	d
last	2	1	3	-1

และจะมีการเปรียบเทียบระหว่าง text กับตัวอักษรตัวสุดท้ายของ pattern ซึ่งในที่นี้ก็คือตัว c จำนวน 9 ครั้งตามการเปรียบเทียบใน หน้าที่ 2

นอกจากนั้นจะเห็นว่า pattern abac ปรากฎอยู่ใน t เป็นตำแหน่งแรกที่ตำแหน่งที่ 14 ดังนั้น input และ output จึงเป็นไปตาม ข้อมูลด้านล่าง

input:

20 4

abcddabcacbdacabacac

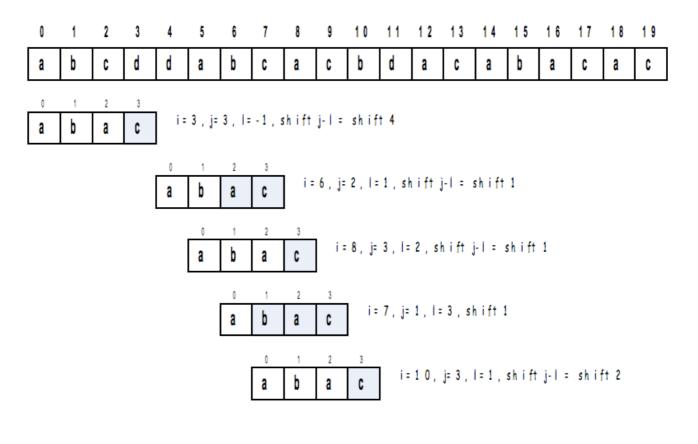
abac

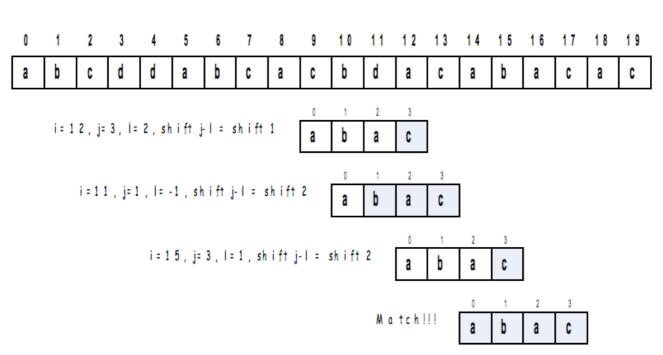
output:

- a 2
- b 1
- c 3
- d -1
- 9
- 14

ตัวอย่างการเปรียบเทียบของวิธี Boyer-Moore algorithm

จะเห็นว่ามีการเปรียบเทียบระหว่าง text กับตัวอักษรตัวสุดท้ายของ pattern ซึ่งในที่นี้ก็คือตัว c จำนวน 9 ครั้ง ซึ่งก็คือจำนวนครั้งที่มี การ shift ตัว pattern นั่นเอง





ตัวอย่าง 2

input:

100 10

vg sobesbmz behulf qb qovwceuld gjv gzvmntbmtim cjtcznsktj gjvt soepvyk biv fpuntsdfrtnnz foheiunlbpt to som the solution of the solution ofutuofzjvrpf ktjgivtsoe

output:

- b -1
- c -1
- d -1
- e 9
- f -1
- g 3
- h -1 i 4
- j 2
- k 0
- 1 -1
- $m \ -1$ n -1
- o 8
- p -1
- q -1 r -1 s 7
- t 6
- u -1
- v 5
- w -1
- y -1
- z -1 8
- 48