

NTU DSA HW4

How's Problem

B06902093 王彦仁

題目連結

<https://tioj.ck.tp.edu.tw/pmisc/ntudsa/hw4.html>

Background (完全不是重點!!!)

浩浩是個喜歡讀書的天才小兒童，天底下的所有問題都難不倒他，因此有一個廣為人知的稱號——「一眼秒題皓大爺」。今天喜歡看書的他在看一本名為「DSA」的書，其中有一題讓他看了兩秒還想不出答案，因此他便很高興的拿著這一題和他的好朋友——裴裴討論，可惜的是兩個臭皮匠勝不過一個諸葛亮(因為要三個才夠(X))，對這一題依然沒有半點頭緒，請問你能寫個程式來幫助浩浩解決這個難題嗎？

Problem Statement

給你一個初始的字串 S ，及一個正整數 Q 。

接下來有 Q 個問題，每種問題有三種形式，分別如下：

1. 「1 c 」其中 c 是一個字元，代表要加一個字元 c 在字串 S 的前方。
2. 「2 c 」其中 c 是一個字元，代表加一個字元 c 在字串 S 的後方。
3. 「3 T_i 」其中 T_i 是一個字串，如果是這種形式的問題，要輸出字串 T_i 在 S 中出現幾次。

Example 1

如果一開始 $S = \text{"cd"}$, $Q = 5$, 並且依序的4個問題如下:

1 'b' → 此時 $S = \text{"bcd"}$

1 'a' → 此時 $S = \text{"abcd"}$

2 'e' → 此時 $S = \text{"abcde"}$

3 "cd" → 此時 $S = \text{"abcde"}$, "cd"在S中出現一次, 因此要輸出1

3 "aa" → 此時 $S = \text{"abcde"}$, "aa"在S中出現零次, 因此要輸出0

Example 2

$S = \text{"ababa"}$

$Q = 1$

問題為3 "aba" → 答案應該要是2

String matching與hash的關係 - 1 (很重要!!!)

如果 a 是一個字串, $a[i]$ 代表字串的第 i 項,

考慮一個多項式 $f(x) = a[0] * x^{(n-1)} + a[1] * x^{(n-2)} + \dots + a[n-2] * x + a[n-1]$

如果 b 是一個字串, $b[i]$ 代表字串的第 i 項,

考慮一個多項式 $g(x) = b[0] * x^{(n-1)} + b[1] * x^{(n-2)} + \dots + b[n-2] * x + b[n-1]$

兩個字串相等, 代表 $a = b$, 同時也會有 $f(x) = g(x)$

String matching與hash的關係 - 2

我們考慮隨便找個數字 x 代入多項式。代入後如果 $f(x)$ 的數值與 $g(x)$ 的數值相同，我們就可以視為兩字串是一樣的!!!

但是代入後可能會有 $f(x)$ 及 $g(x)$ 數值overflow的問題，因此我們可以再挑個數字 M ，換成觀察 $f(x) \bmod M$ 是否與 $g(x) \bmod M$ 的結果是否相同，如果相同，那麼我們就可以說有「高機率」是一樣的!!!

這種string matching的方式叫做Rolling hash，其中 M 的選定最好是一個質數

https://en.wikipedia.org/wiki/Rolling_hash

字串的hash具體來說該怎麼做呢？

對於一個給定的字串 $S[1...N]$ ，我們選定兩個數字 x (29)和 M ($10^{**}9 + 7$)

預處理的部分如下(某種prefix sum的感覺):

1. $hash[0] = 0$
2. $hash[i] = (hash[i - 1] * x + (S[i] - 'a' + 1)) \% M$, 其中 $i \geq 1$

在預處理之後，我們可以快速知道哪些事情？

我們可以快速知道 $S[l \dots r]$ 的hash value!!!

$$\text{hash_value}(l, r) = (\text{hash}[r] - \text{hash}[l - 1] * x^{(r - l + 1)} \bmod M + M) \bmod M$$

因為我們有

$$\text{hash}[l - 1] = (S[1] * x^{(l - 2)} + S[2] * x^{(l - 3)} + \dots + S[l - 2] * x + S[l - 1]) \bmod M$$

$$\text{hash}[r] = (S[1] * x^{(r - 1)} + S[2] * x^{(r - 2)} + \dots + S[l - 1] * x^{(r - l + 1)} + \dots + S[r]) \bmod M$$

還不知道怎麼快速的算出 x^k ？沒關係，這也可以預處理!!!

Hash的碰撞機率

最直觀的模型應該是「a, b 是兩相異物, 所以他們的 hash 值有 $M * M$ 種挑法, 碰撞的有 M 種挑法, 所以一次比較碰撞的機率是 $M / (M * M) = 1 / M$ 」(by 集貴)

延伸閱讀:

Hash Collision Probabilities

(<http://preshing.com/20110504/hash-collision-probabilities/>)

Birthday attack (https://en.wikipedia.org/wiki/Birthday_attack)

測資範圍

所有字元皆為英文小寫字母

$1 \leq \text{字串} S \text{ 的初始長度} \leq 10^{**}5$

$0 \leq \text{所有字串} T_i \text{ 的長度總和} \leq 10^{**}5$

實作時請注意演算法的時間複雜度!!!

Subtask 1 (5 pts)

$1 \leq \text{字串} S \text{ 的初始長度} \leq 10^{**5}$

$Q = 1$

所有問題皆為第三種問題

可以用來測試string matching algorithm有沒有寫壞

Subtask 2 (5 pts)

$1 \leq \text{字串} S \text{ 的初始長度} \leq 1000$

$1 \leq Q \leq 1000$

$0 \leq \text{所有字串 } T_i \text{ 的長度總和} \leq 10^{**4}$

可以知道加東西在字串的前後有沒有寫壞

Subtask 3 (30 pts)

$1 \leq \text{字串} S \text{ 的初始長度} \leq 10^{**5}$

$1 \leq Q \leq 10^{**5}$

所有字串 T_i 的長度 $\geq 10^{**4}$

$0 \leq \text{所有字串 } T_i \text{ 的長度總和} \leq 10^{**5}$

想一想，這一筆subtask有什麼特殊性？

(可以仔細盯著綠色字的地方看)

Subtask 4 (30 pts)

$1 \leq \text{字串} S \text{ 的初始長度} \leq 10^5$

$1 \leq Q \leq 10^5$

$1 \leq \text{所有字串 } T_i \text{ 的長度} \leq 10$

$0 \leq \text{所有字串 } T_i \text{ 的長度總和} \leq 10^5$

Subtask 5 (30 pts)

原題設條件

解題關鍵 (小提示)

1. 字串的hash
2. 分類討論(平方分割): 對於不同的測資使用不同的方法
(combine subtask 3 and 4 to solve subtask 5)
3. 高中數學 - 算幾不等式

小花絮1 - 一眼秒題皓大爺 的 第一次上傳結果

00: TLE

01: AC (5 pts)

02: WA + RE + TLE

03: WA + RE

04: WA + RE

05: WA + TLE

06: WA + TLE

07: WA + TLE

08: WA + TLE

09: WA + TLE

Total: 5pts !!!

小花絮2 - 裴裴 的 第一次上傳結果

00: AC (5 pts)

01: AC (5 pts)

02: AC (15 pts)

03: AC (15 pts)

04: TLE

05: TLE

06: TLE

07: TLE

08: TLE

09: TLE

Total: 40 pts!!!

希望大家會喜歡這一題~

謝謝大家^^