

F: Seimei Handan 999.0

原案 : climpet

問題文 : amylase

データセット : climpet

解答 : climpet, fuppy, hos

問題概要

- $1, \dots, L$ からなる整数列 A, B が与えられる。
- 全単射 $f: [1, L] \rightarrow [1, L]$ のうち、次の条件を満たすものは何通りか？
 - 列 $(f(A[0]), f(A[1]), \dots)$ は B を連続部分列として含む。
- 制約
 - $|A|, |B|, L \leq 3 \times 10^5$

定義: Parameterized match (p-match)

- 同じ長さの列 X, Y が次の条件を満たすとき、 X と Y は parameterized match (p-match) すると定義する。
 - ある全単射 f が存在し、任意の位置 i に対し $f(X[i]) = Y[i]$

考察

- A の連続部分列と B が p -match するとき、それに対応する全単射 f は、「ほぼ」一意に定まる。
 - 正確には、 B に一度も出現しない要素についてはどう割り当てても良い。具体的には B に出現する要素が K 種類るとき、ちょうど $(L - K)!$ 通り存在する。
- つまり、次の問題に帰着できる。
 - A の連続部分列のうち、 B と p -match するものは何種類あるか？最後に $(L - K)!$ 倍して答えよ。
- 連続部分列の重複除去は、接尾辞配列やローリングハッシュなどを使えばよい。よって、次の問題が解ければよい。
 - A の連続部分列のうち、 B と p -match するものの出現位置を全列挙せよ。

考察

列 X に対し、整数列 prev_X を次のように定義する。

- $X[i]$ が初出ならば、 $\text{prev}_X[i] = 0$
- $X[i]$ が以前に出現したことがあるならば、その最右位置を j とすると、 $\text{prev}_X[i] = i - j$

	0	1	2	3	4	5	6	7	8	9
X	10	20	30	20	10	40	20	10	30	50
prev_X	0	0	0	2	4	0	3	3	6	0

考察

- 列 X, Y が p -match することと、 $\text{prev}_X = \text{prev}_Y$ であることは同値である。
- したがって、 A の連続部分列 A' のうち、 $\text{prev}_{A'} = \text{prev}_B$ であるものを見つけられればよい。
- ここからは、少なくとも二通りの方針がある。
 - ローリングハッシュを使う方法
 - Knuth-Morris-Pratt や Z-algorithm を基にした方法

方針1: ローリングハッシュ

長さ $|B|$ のウィンドウを一要素ずつずらしながら、 A の連続部分列 A' に対し $\text{hash}(\text{prev}_{A'})$ を計算していく。中間の要素が 0 に変わる場合があることに注意。

	0	1	2	3	4	5	6	7	8	9
A	10	20	30	20	10	40	20	10	30	50

$\text{prev}_{A[1..7]}$

0	0	2	0	0	3	3
---	---	---	---	---	---	---

$\text{prev}_{A[2..8]}$

0	0	0	0	3	3	6
---	---	---	---	---	---	---

方針2: KMP や Z-algorithm など

(Knuth-)Morris-Pratt や Z-algorithm において、文字の比較を行っている部分を、prev を用いたものに置き換える。ただし、prev が大きすぎる場合には 0 とみなす必要がある。

例えば、**この実装** ($y = A, x = B$ に相当) であれば、 $x[i] \neq y[j]$ となっている部分を $\text{prevX}[i] \neq (\text{prevY}[j] \leq i ? \text{prevY}[j] : 0)$ のように置き換える
とよい。 $x[i] \neq x[j]$ となっている部分についても同様。

Z-algorithm についても同様に改修できる。

解法まとめ

1. ローリングハッシュ、Knuth-Morris-Pratt, Z-algorithm などを用いて、 A の連続部分列 A' のうち $\text{prev}_{A'} = \text{prev}_B$ となるものの出現位置を全列挙する。
2. ローリングハッシュや接尾辞配列などを用いて、それらの連続部分列の重複を除去する。
3. 最後に $(L - K)!$ をかける。ただし K は B に出現する要素の種類数。

計算量は全体で $O(N + M + L)$ 時間。

統計情報

- Acceptances
 - 8 teams
- First Acceptance
 - KOMOREBI (163 min)