

天下一プログラマーコンテスト2014本戦

E問題 解説

uwi

問題概要

長さ N の文字列 S と、 Q 個のクエリ $(a[i], b[i])$ が与えられるので、 $S[a[i], b[i]]$ に含まれる繰り返し文字列(Tandem Repeat)の長さの和をそれぞれ求めよ。

$N \leq 10^5$, $Q \leq 10^5$

$N \leq 2 \cdot 10^4$, $Q \leq 10^5$ (部分点2)

$N \leq 1000$, $Q \leq 10^5$ (部分点1)

難しさ

(零点) << (部分点1) <<<<<<<<<<< (部分点2) < (満点)

解法概要

1. 繰り返し文字列をすべて列挙する
2. 各繰り返し文字列を含むクエリ区間に長さを足していく

部分点解法1 ($N \leq 1000$)

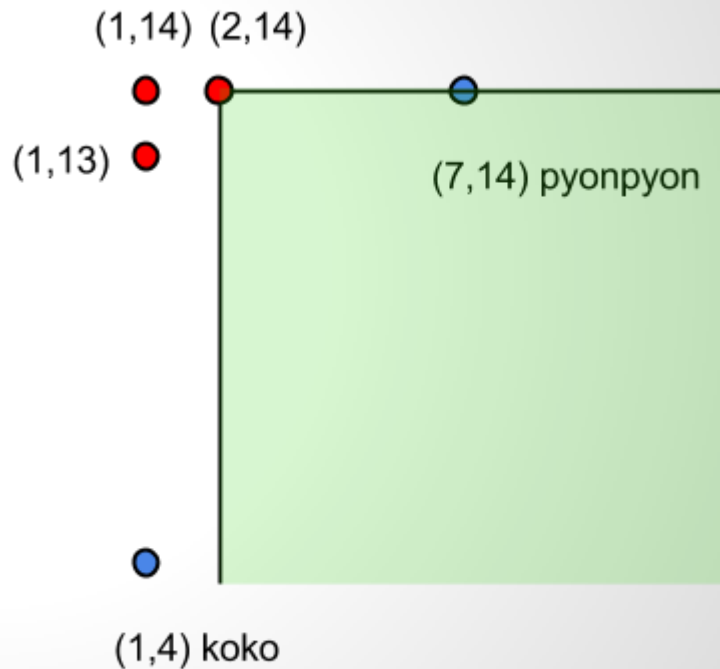
繰り返し文字列の個数はたかだか $N^2/4$ (全部同じ文字の時)なので全列挙できる。

- ・ ローリングハッシュを使う
- ・ Suffix Array+LCP

列挙した繰り返し文字列は”長さ=重み”の区間になっている。
これらとクエリ区間をまぜてBIT等で $O((N^2+Q)\log(N^2+Q))$ で求められる。

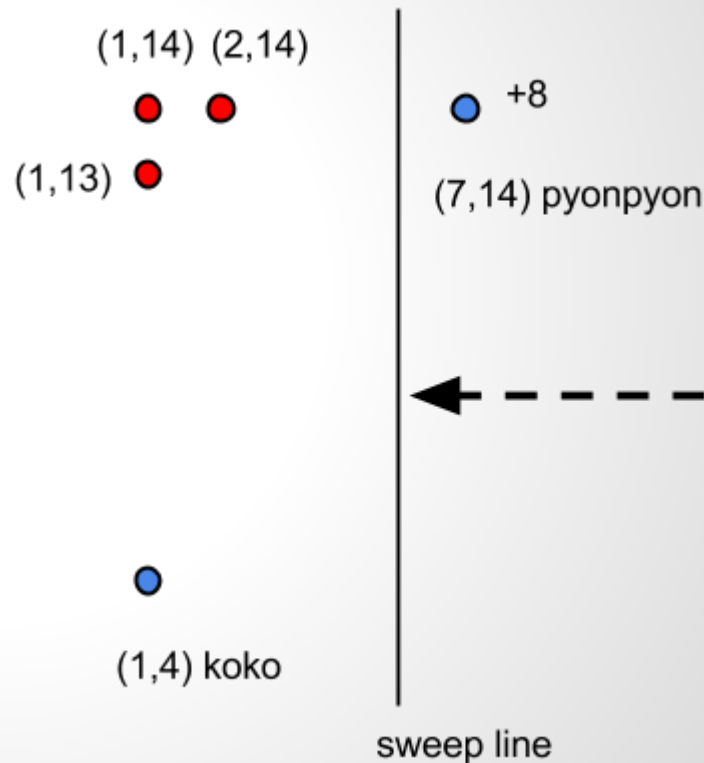
sample 1

繰り返し文字列とクエリを
(始端, 終端)で点で図示。
クエリ点の右下(境界含む)
にある繰り返し文字列の長
さの合計を求めれば良い。



sample 1

- ・ 高さごとに値を足せる
 - ・ ある高さまでの合計を高速に出せる
- データ構造(BITとか)を持ちながら右からline sweep



満点解法&部分点解法2 ($N \leq 10^5$)

$N^2/4$ が大きすぎて直接列挙はとても無理！
runを列挙する。

runとは

- ・ 周期的な連続部分文字列で、2周期以上の長さを持つもののこと
- ・ (長さ-2*周期+1)個の繰り返し文字列を持つ。
- ・ S内に極大なrun(それ以上周期を保ったまま左右に伸ばせない)はO(N)個存在する。

例: 周期3のrun

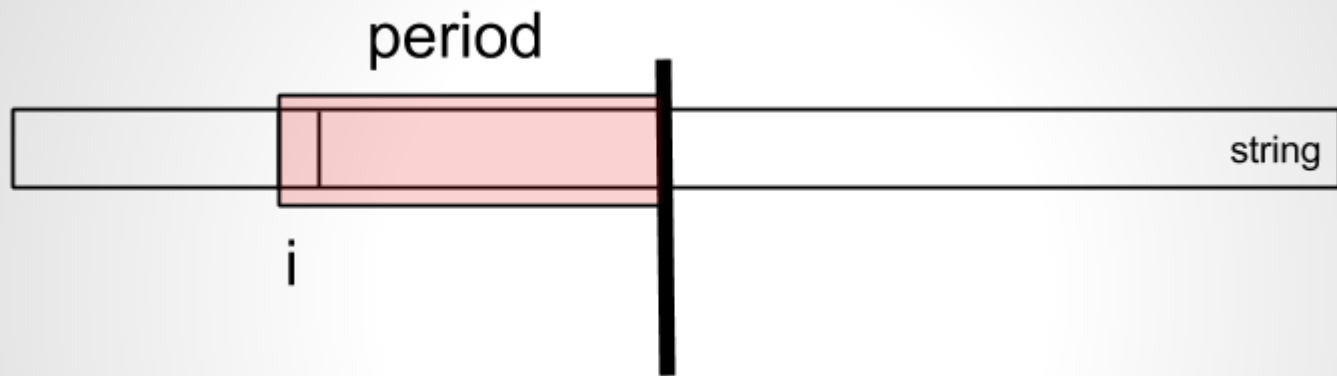
ABABBABBAABAB

ABABBABBAABAB

runの列挙

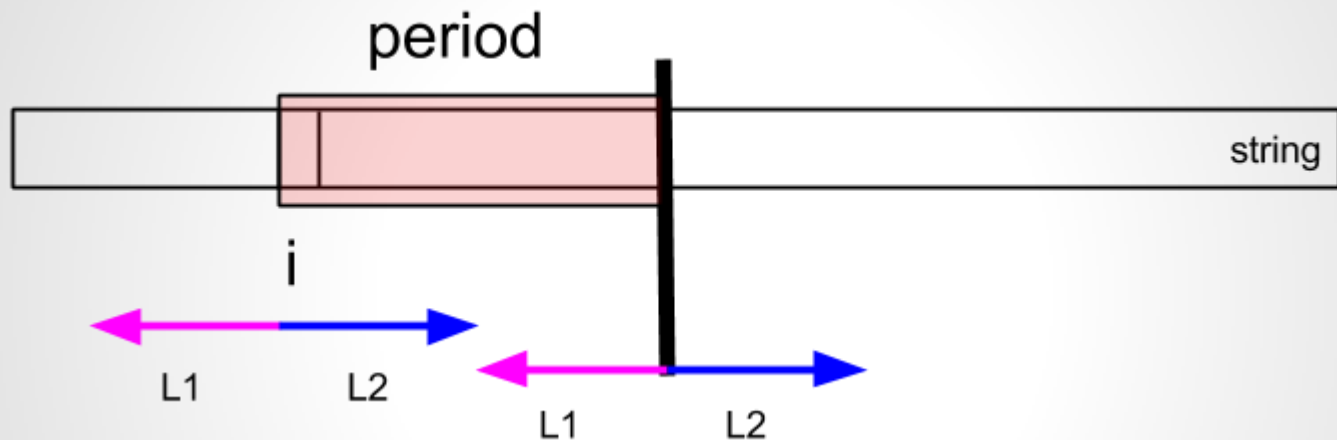
- **Main-Lorentzのアルゴリズム** ($O(N \log N)$)
分割統治+Z algorithm
- **Gusfieldのアルゴリズム** ($O(N)$)
Suffix Tree -> LZ-factorization -> Z algorithmで極大runを列挙
- **Crochemoreのアルゴリズム** ($O(N)$)
Suffix Array+LCP -> LZ-factorization -> Z algorithmで極大runを列挙
重複を除かないといけないのと、ソートその他も $O(N)$ でやらなければ
いけないので面倒
重複とか関係ない、最大長最大繰り返しを求める的な問題なら強い。

Main-Lorentzのアルゴリズム



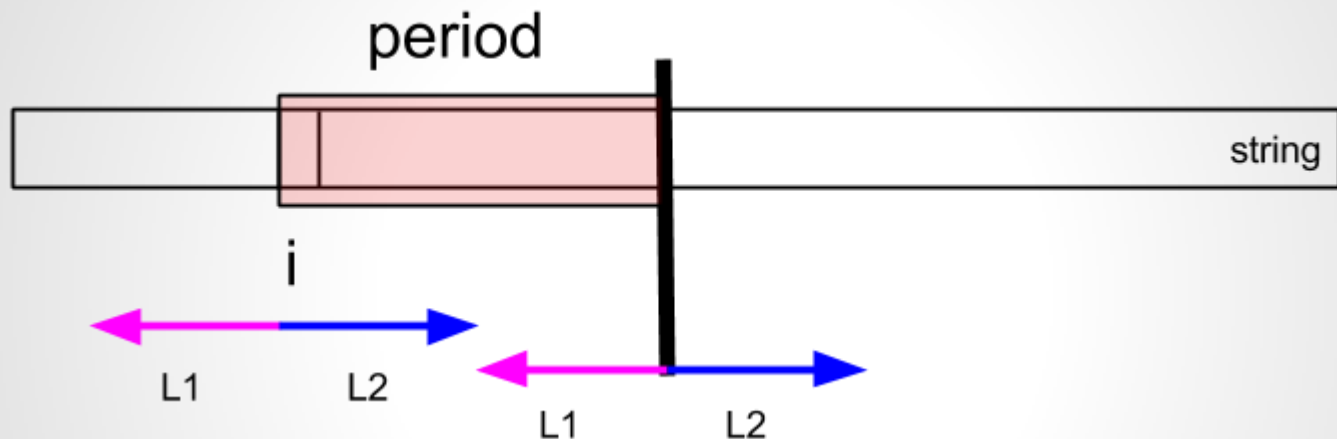
文字列を分割。分割線にまたがる繰り返し文字列を考える。
分割線の左側に i を固定。 i から分割線までを周期と考える
(右側に周期がある場合は文字列を逆転して同様)。

Main-Lorentzのアルゴリズム



i , 分割線から左に $L1$ 文字ずつ一致, 右に $L2$ 文字ずつ一致しているならば、 $[i-L1, \text{分割線の直後}+L2-1]$ が繰り返し文字列になる。 $L1+L2=\text{period}$ であり、 $L1, L2$ それぞれ上限がある。
 $[i-L1 \text{ 上限}, \text{分割線の直後}+L2 \text{ 上限}-1]$ が周期 period のrunになる。

Main-Lorentzのアルゴリズム



i を動かした時の各L1の上限は分割線から逆行した文字列のZ-algorithmで列挙できる。

各L2の上限は、分割の後半+分割の前半 に対するZ-algorithmで列挙できる。

Main-Lorentzのアルゴリズム

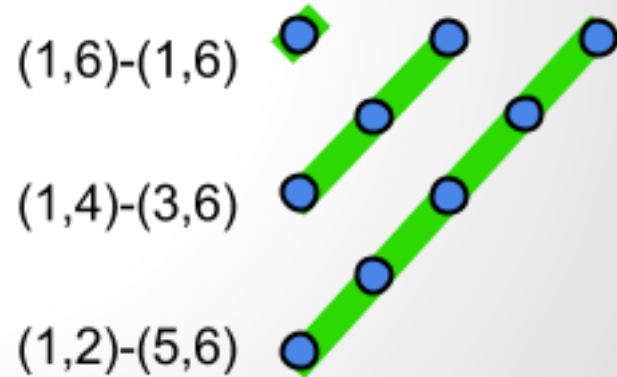
以上を再帰的に行うと、すべてのrunを重複なく列挙可能。
時間計算量 $O(N \log N)$, 空間計算量 $O(N)$.
 $O(N \log N)$ 個のrunが列挙される。

極大のrunは $O(N)$ 個存在するので、上記をマージするだけで $O(N)$ 個に落とせる。部分点2はマージしないで後半を行った場合を想定。

満点解法&部分点解法2 ($N \leq 10^5$)

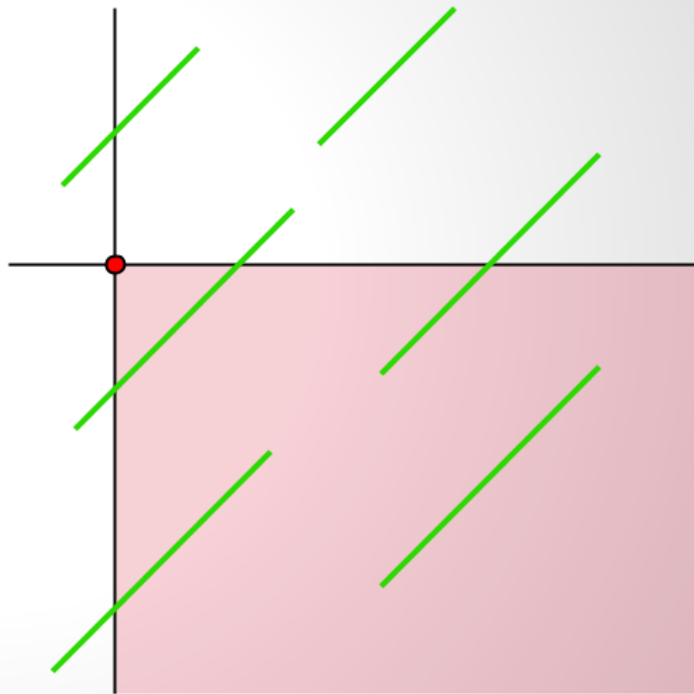
runが全部求まったが、これらをクエリに足し込まなければならない。

“aaaaaa”



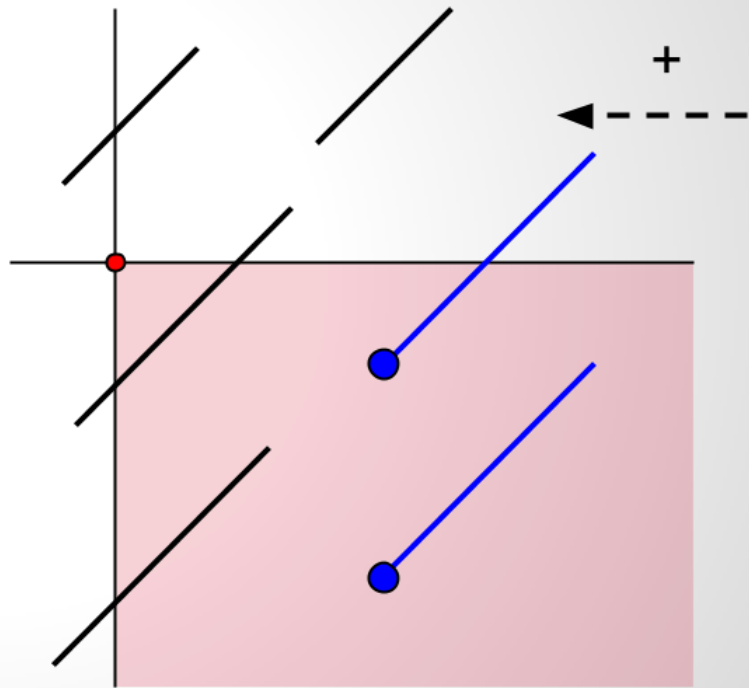
満点解法&部分点解法2 ($N \leq 10^5$)

run(傾き1の線分)のうちクエリ点の右下領域との、
(交わりの長さ)*(runの周期)の合計がほしい。



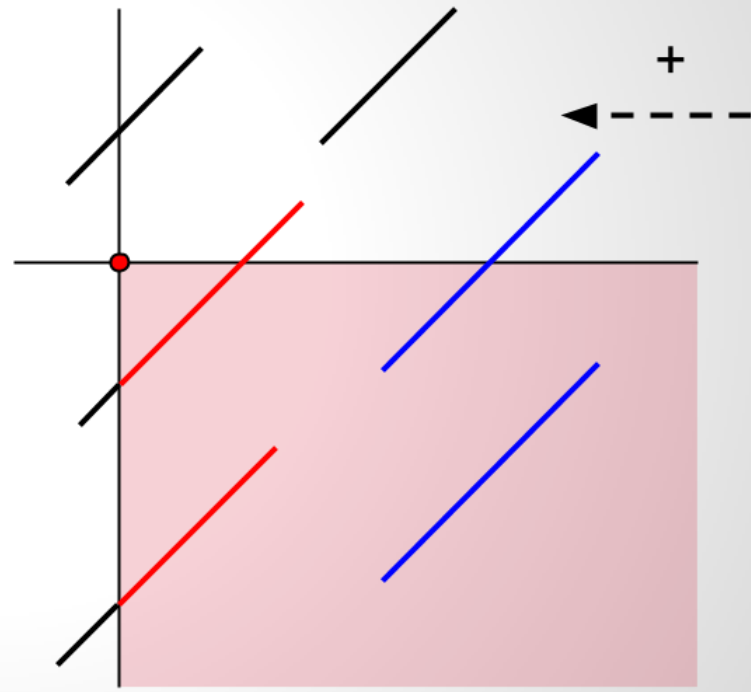
これを

右から左へ走査。
runの左下端にrunの
重みすべてを乗せる。
点なら、クエリ点の
右下にあるものの合
計は楽勝。



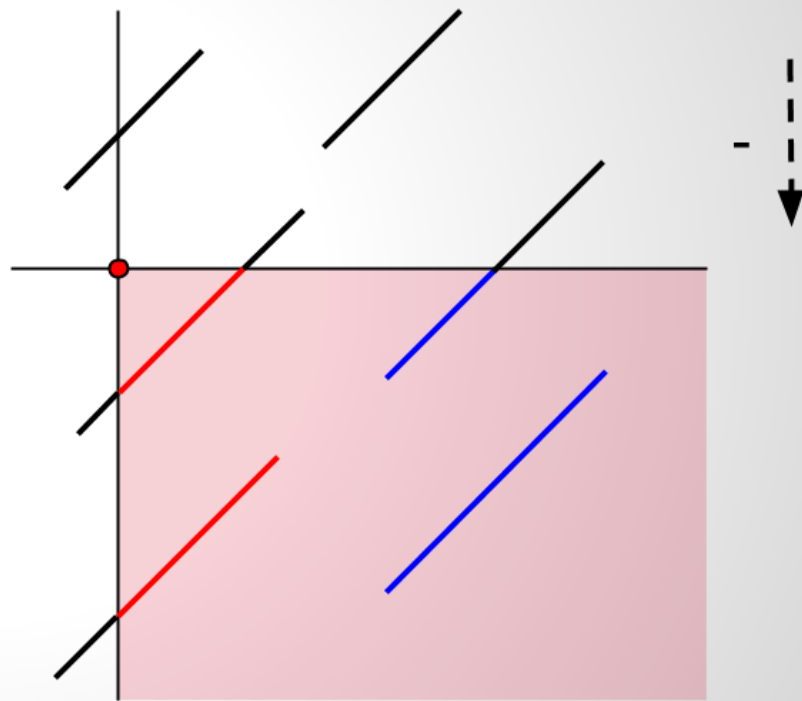
こうして

右から左へ走査。
(クエリ点の下方を通過するrunの右上端x座標の合計) - (クエリ点のx座標) * (クエリ点の下方を通過するrunの個数)



こうじゃ

上から下へ走査。
(クエリ点の右方を通過するrunの右上端y座標の合計) - (クエリ点のy座標) * (クエリ点の右方を通過するrunの個数)



満点解法&部分点解法2 ($N \leq 10^5$)

以上の3回のsweepはどれも $O((N+Q)\log(N+Q))$ で可能。

全体で時間計算量 $O(N\log^2 N + (N+Q)\log(N+Q))$ でできる。

背景

回文の問題はよく出てくるのに、繰り返し文字列の問題はほとんどなかったので、教育的な意味も込めて出題してみました。

既出だったらごめんなさい