会津合宿2015 Day3
F: みこみー文字列

原案・解説・問題文：井上
解答：井上・鈴木・田中

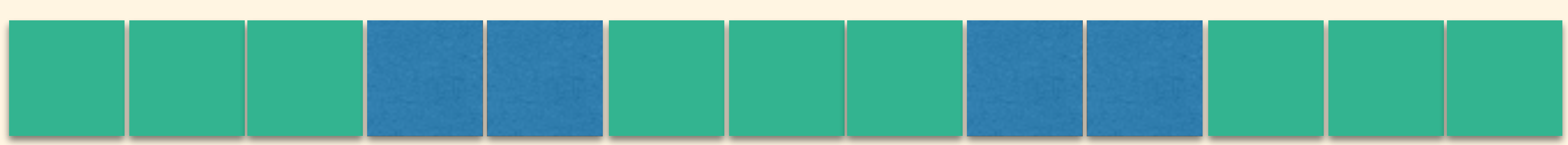
問題概要

- 文字列Sが与えられる
- S = ABABAとなるような非空文字列A，Bがあるか判定せよ
- あるなら|AB|が最小になるものを出力せよ

- 制約: $1 \leq S \leq 10^6$

想定 TLE 解法：全探索

- Aの長さを決め打ちする（1 ≤ |A| ≤ |S|/3)
- するとBの長さは一意（B = (|S| - 3|A|) / 2)
- あとは以下が一致するか調べればよい
  - S[0:A) と S[A+B:2A+B) と S[2A+2B:3A+B)
  - S[A:A+B) と S[2A+B:2A+2B)
- O(N)通りについて、長さO(N)の文字列比較を行うのでO(N^2)
  → N = 10^6なのでTLE



想定解法: ローリングハッシュ

- 部分文字列 s[l:r) に以下のようなハッシュ値h(l,r)を割り当てる
  - $h(l,r) = \Sigma_{l \leq i \leq r} (\mathrm{int}) s_i * p^{(r-i)} \bmod M$
  - p, Mは互いに素（基本p<Mで、素数とか）
- このハッシュ値が一致 ⇒ 文字列が一致

想定解法：ローリングハッシュ

- $h(l,r) = \Sigma_{l \le i \le r} (int)s_i * p^{(r-i)} \bmod M$
- このハッシュ値が一致 ⇒ 文字列が一致

- つまり……以下を調べればよい
  - h(0,A) = h(A+B,2A+B) = h(2A+2B,3A+2B)
  - h(A,A+B) = h(2A+B,2A+2B)
- 数値なのでO(1)で判定できる
- ただし、ハッシュ一致 ⇐ 文字列一致は言えない
  - まともな p, M を使えば確率的にほとんど起こらない

想定解法：ローリングハッシュ

- $h(l,r) = \Sigma_{l \le i \le r}$ (int)s_i * $p^{(r-i)}$ mod M
- けど h を計算するのに O(N) かかるのでは？
  - $h(l,r) = (h(0,r) - h(0,l) * p^{(r-l)})$ mod M
  - $h(0,i) = h(0,i-1) * p +$ (int)s_i と計算できるので、あらかじめ h(0,i) を O(N)で計算しておけば h(l,r) の計算は O(1)
- 全体で O(N)

別解: Suffix Array + LCP + RMQ

- 接尾辞配列 (SA) を作り、SAで隣との共通部分接頭辞 (LCP)の長さを記録した配列に対して区間最小値クエリ (RMQ) を投げる
- $S[l_1,l_1+k)$ と $S[l_2, l_2+k)$ の文字列比較をするときは、l_1とl_2に該当するSAのインデックスを区間としてRMQすると、答えがl_1,l_2の共通接頭辞の長さになる
- これがkより長ければ$S[l_1,l_1+k) = S[l_2, l_2+k)$
- 計算量: O(SA構築 + NlogN)
  - 蟻本のSA構築は $O(N\log^2 N)$ なのでTLE的に厳しい

余談

- ぶっちゃけナイーブ$O(N^2)$が速すぎたので$|S| \leq 10^6$になった
- のでSA+LCP+RMQは厳しくなった
- 個人的にはこっちもすんなり通したかった

ジャッジ解

- 井上 (C++) 54行 1057B
- 鈴木 (C++) 38行 1035B
- 田中 (C++) 43行 1331B

回答状況

- Accept / Submit
  - 11 / 30 (36.7%)
- First Acceptance
  - onsite: syumi_plus (01:53)
  - online: natsugiri (00:24)