H: Almost Same Substrings

原案: 林
問題文: 寺尾 林
解答: 秋葉 生田 林 矢野
解説: 林

問題の概要

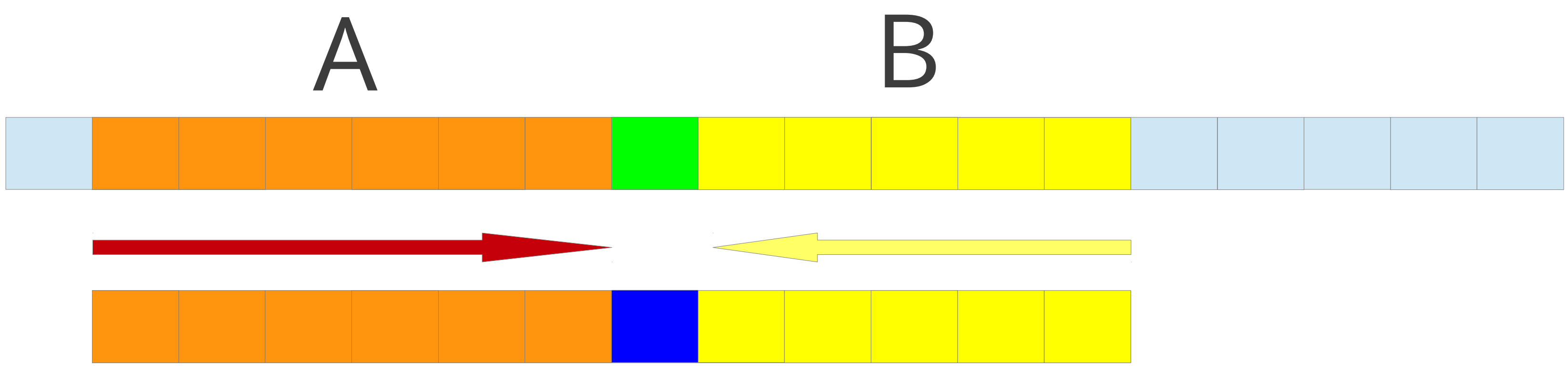
- 文字列S, Tが与えられる。
- Sの部分文字列でTと1文字異なるものを数える。
- S, Tの長さは300,000以下。

- k-mismatch-matchingの特殊ケース
  - k文字異なっている。
  - Suffix TreeでO(k|S|)で可。

解法1(ハッシュ)

- Tの1文字を変えてハッシュ値を計算する。
  - 元のTに対してローリングハッシュで計算して差分を取る
  - 300,000×52（アルファベット数）個のハッシュ値
- Sの部分文字列とTのハッシュ値の一致を見る。
  - 不必要なハッシュ値生成や比較を避けるように実装
  - TLEの可能性があります。
  - ハッシュの衝突に注意。

解法2(Z-algorithm)

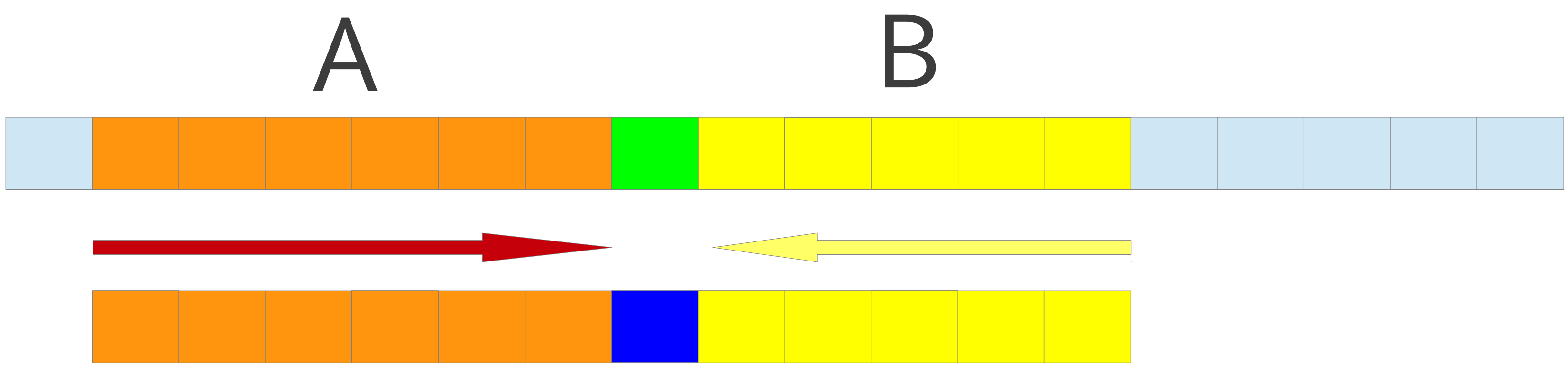
- S[i…i + |T|)がTと1文字違いでマッチする条件は
  - S[i…i+|T|)とTの最長共通接頭辞をA、
  - S[i…i+|T|)とTの最長共通接尾辞をBとしたとき
  - |A|+|B| = |T| - 1となればよい。
- 

解法2(Z-algorithm)

- 各S[i...|S|)とTの最長接頭辞はZ-algorithmで合計O(|S|)で求めることができる。
  - Codeforcesの解説記事がわかりやすい。
    http://codeforces.com/blog/entry/3107
  - 実装が楽で確実な答えが出ます。
- SuffixArrayとLCPテーブルを使っても時間内に計算できますが面倒です。

解法2(Z-algorithm)

- S[i…i+|T|)とTの最長共通接頭辞を各iについて求める
- S[i…i+|T|)とTの最長共通接尾辞をSとTを逆にしたものに対してZ-algorithmを適用して求める。
- 最長共通接頭辞の長さの配列と最長共通接尾辞の配列をなめていくことで線形時間でできる。



解法3

- S[i...|S|)とT[j...|T|)の最長共通接頭辞はローリングハッシュやSuffixArray+LCPArrayで高速に求まる。
- 1 ≦ i ≦ |S|となる各iに対し
  - S[i...|S|),T[j...|T|)の最長共通接頭辞の長さAを求める。
  - S[i+A+1... |S|), T[j+A+1...|T|)の最長共通接頭辞の長さBを求める。
  - A+B=|T|-1となればよい。
- この方法ならk-mismatch-matchingも解けます。

結果

- Accept数　: 12
- 提出数　: 75

- First Accept(全体)　: tomerun (84:26)
- First Accept (Onsite): 0perasan(100:16)